KRYGER • ROTH • DEMENT

Principles and Practice of SLEEP MEDICINE

SIXTH EDITION

Basner • Hirshkowitz • Lavigne • Malow • Scammell • Turek

ELSEVIER

Principles and Practice of
SLEEP MEDICINE

SIXTH EDITION

Meir Kryger, MD, FRCPC
Professor
Pulmonary, Critical Care, and Sleep Medicine
Yale University School of Medicine
New Haven, Connecticut

Thomas Roth, PhD
Division Head
Sleep Disorders and Research Center
Henry Ford Hospital
Detroit, Michigan

William C. Dement, MD, PhD
Lowell W. and Josephine Q. Berry Professor
of Psychiatry and Behavioral Sciences
Stanford University School of Medicine
Sleep Sciences and Medicine
Stanford, California

ELSEVIER

ELSEVIER

1600 John F. Kennedy Blvd.
Ste 1800
Philadelphia, PA 19103-2899

PRINCIPLES AND PRACTICE OF SLEEP MEDICINE, SIXTH EDITION.
ISBN: 978-0-323-24288-2

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

With respect to any drug or pharmaceutical products identified, readers are advised to check the most current information provided (i) on procedures featured or (ii) by the manufacturer of each product to be administered, to verify the recommended dose or formula, the method and duration of administration, and contraindications. It is the responsibility of practitioners, relying on their own experience and knowledge of their patients, to make diagnoses, to determine dosages and the best treatment for each individual patient, and to take all appropriate safety precautions.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Names: Kryger, Meir H., editor. | Roth, T. (Tom), editor.
Title: Principles and practice of sleep medicine / [edited by] Meir Kryger, Thomas Roth.
Description: Sixth edition. | Philadelphia, PA : Elsevier, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2015044903 | ISBN 9780323242882 (hardcover : alk. paper)
Subjects: | MESH: Sleep Disorders. | Sleep–physiology.
Classification: LCC RC547 | NLM WL 108 | DDC 616.8/498–dc23 LC record available at http://lccn.loc.gov/2015044903

Executive Content Strategist: Kellie Heap
Content Development Specialist: Laura Schmidt
Publishing Services Manager: Patricia Tannian
Project Manager: Amanda Mincher
Design Direction: Amy Buxton

Printed in China

Last digit is the print number: 9 8 7 6 5 4 3 2 1

We dedicate this volume to

Barbara Kryger, Jay and Shelley Gold, Emily and Michael Kryger, Steven Kryger

Toni Roth, Daniel and Jeanne Roth, Adam and Carol Roth, Jonathan and Cheyna Roth, Andrea and Justin Leibow

Catherine Dement Roos and Gary Roos, Elizabeth (Liz) Anne Dement, John Nicholas (Nick) Dement, and Stacy Seibert; and in loving memory of Pat Dement

From the Arts

Every Tuesday, Queen Elizabeth II of the United Kingdom (played by Dame Helen Mirren) had a private audience with her Prime Minister in the Private Audience Room on the first floor of Buckingham Palace. This is dramatized in Peter Morgan's play, The Audience. In this scene Elizabeth is meeting with Prime Minister Gordon Brown.

Elizabeth So, back to your weekend, and all this industriousness. Were you up very early?

Brown Four thirty.

Elizabeth Oh, dear.

Brown It's all right. I never sleep much.

Elizabeth Since when?

Brown Since always.

Elizabeth Harold Wilson always used to say, "The main requirement of a Prime Minister is a good night's sleep … and a sense of history." Mrs Thatcher taught herself to need very little towards the end. But I'm not sure how reassured I am by that. I like the idea of any person with the power to start nuclear war being rested. (*A beat.*) Besides, lack of sleep can have a knock-on effect in other areas.

Brown Such as?

Elizabeth One's general sense of health.

A silence.

And happiness.

A silence.

And equilibrium.

Brown looks up. A silence.

I gather there's been some concern …

Brown About what?

Elizabeth Your happiness. Don't worry. You wouldn't be the first in your position to feel overwhelmed. Despondent.

She searches for the right word.

Depressed.

From Morgan, Peter. THE AUDIENCE, Faber and Faber, 2013.
Used with permission of Mr. Peter Morgan.

Blessings on him who first invented sleep.—It covers a man all over, thoughts and all, like a cloak.—It is meat for the hungry, drink for the thirsty, heat for the cold, and cold for the hot.—It makes the shepherd equal to the monarch, and the fool to the wise.—There is but one evil in it, and that is that it resembles death, since between a dead man and a sleeping man there is but little difference.

From DON QUIXOTE
By Saavedra M. de Cervantes

"To sleep! To forget!" he said to himself with the serene confidence of a healthy man that if he is tired and sleepy, he will go to sleep at once. And the same instant his head did begin to feel drowsy and he began to drop off into forgetfulness. The waves of the sea of unconsciousness had begun to meet over his head, when all at once—it was as though a violent shock of electricity had passed over him. He started so that he leapt up on the springs of the sofa, and leaning on his arms got in a panic on to his knees. His eyes were wide open as though he had never been asleep. The heaviness in his head and the weariness in his limbs that he had felt a minute before had suddenly gone.

From ANNA KARENINA, Part IV, Chapter XVIII
By Leo Tolstoy

But the tigers come at night,
With their voices soft as thunder,
As they tear your hope apart,
As they turn your dream to shame.

From I Dreamed a Dream, LES MISÉRABLES,
with permission, Cameron Mackintosh, producer

Contributors

Sabra M. Abbott, MD, PhD
Assistant Professor
Ken and Ruth Davee Department of Neurology
Northwestern University Feinberg School of Medicine
Chicago, Illinois
Circadian Dysregulation in Mental and Physical Health
Circadian Disorders of the Sleep–Wake Cycle

Peter Achermann, PhD
Professor
Institute of Pharmacology and Toxicology
Professor
Zurich Center for Interdisciplinary Sleep Research
Professor
Zurich Center for Integrative Human Physiology
University of Zurich
Professor
Neuroscience Center
University and ETH Zurich
Zurich, Switzerland
Sleep Homeostasis and Models of Sleep Regulation

Philip N. Ainslie, PhD
Professor
Health and Exercise Sciences
University of British Columbia
Kelowna, British Columbia, Canada
Respiratory Physiology: Sleep at High Altitudes

Torbjörn Åkerstedt, PhD
Professor
Stress Research Institute
Stockholm University
Professor
Clinical Neuroscience
Karolinska Institute
Stockholm, Sweeden
Introduction: Occupational Sleep Medicine
Sleep, Occupational Stress, and Burnout

Ravi Allada, MD
Professor and Chair
Department of Neurobiology
Weinberg College of Arts and Sciences
Northwestern University
Evanston, Illinois
Introduction: Genetics and Genomics of Sleep
Genetics and Genomic Basis of Sleep in Simple Model Organisms

Richard P. Allen, PhD
Associate Professor of Neurology
Johns Hopkins University
Baltimore, Maryland
Restless Legs Syndrome and Periodic Limb Movements During Sleep

Fernanda R. Almeida, DDS, MSc, PhD
Associate Professor
Division of Orthodontics
Oral Health Sciences
University of British Columbia
Vancouver, British Columbia, Canada
Role of Dentistry and Otolaryngology in Sleep Medicine
Oral Appliances for the Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome and for Concomitant Sleep Bruxism

Amy W. Amara, MD, PhD
Assistant Professor
Neurology
University of Alabama at Birmingham
Birmingham, Alabama
Epidemiology of Sleep Medicine

Sonia Ancoli-Israel, PhD
Professor Emeritus of Psychiatry and Medicine
Professor of Research
University of California, San Diego
La Jolla, California
Sleep and Fatigue in Cancer Patients
Insomnia in Older Adults
Circadian Rhythms in Older Adults
Actigraphy

Chelsea Angel, BA
Research Specialist II
Departments of Anesthesiology and Psychology
University of Tennessee
Knoxville, Tennessee
Opiate Action on Sleep and Breathing

Taro Arima, DDS, PhD
Lecturer
Division of International Affair
Graduate School of Dental Medicine
Hokkaido University
Sapporo, Japan
Sleep Bruxism: Definition, Prevalence, Classification, Etiology, and Consequences

J. Todd Arnedt, PhD
Associate Professor
Director, Behavioral Sleep Medicine Program
Departments of Psychiatry and Neurology
University of Michigan Medical School
Ann Arbor, Michigan
Insomnia Diagnosis, Assessment, and Evaluation

Isabelle Arnulf, MD, PhD
Sleep Disorders Unit
Pitie-Salpetriere University Hospital
Sorbonne University
Pierre and Marie Curie University
Paris, France
Parkinsonism
Kleine-Levin Syndrome
Nightmares and Dream Disturbances

Alon Y. Avidan, MD, MPH
Director, University of California, Los Angeles Sleep Disorders Center
Director, University of California, Los Angeles Neurology Clinic
Professor of Neurology
Department of Neurology
David Geffen School of Medicine at University of California, Los Angeles
Los Angeles, California
Physical Examination in Sleep Medicine
Non–Rapid Eye Movement Parasomnias: Clinical Spectrum, Diagnostic Features, and Management

John Axelsson, MSc, PhD
Associate Professor
Department of Clinical Neuroscience
Karolinska Institute
Affiliated Researcher
Stress Research Institute
Stockholm University
Stockholm, Sweden
Optimizing Shift Scheduling

M. Safwan Badr, MD
Professor and Chief
Division of Pulmonary, Critical Care, and Sleep Medicine
Wayne State University School of Medicine
Detroit, Michigan
Anatomy and Physiology of Upper Airway Obstruction

Helen A. Baghdoyan, PhD
Beaman Professor
Departments of Anesthesiology and Psychology
University of Tennessee
Knoxville, Tennessee
Opiate Action on Sleep and Breathing

Fiona C. Baker, PhD
Senior Program Director, Human Sleep Research
Center for Health Sciences
SRI International
Menlo Park, California
Honorary Senior Research Fellow
Brain Function Research Group, School of Physiology
University of the Witwatersrand
Johannesburg, South Africa
Sex Differences and Menstrual-Related Changes in Sleep and Circadian Rhythms
Sleep and Menopause

Thomas J. Balkin, PhD
Behavioral Biology Branch
Walter Reed Army Institute of Research
Silver Spring, Maryland
Performance Deficits During Sleep Loss and Their Operational Consequences
Sleep and Performance Prediction Modeling

Bilgay Izci Balserak, PhD
Assistant Professor
Department of Women, Children, and Family Health Science
Center for Narcolepsy, Sleep, and Health Research
University of Illinois, College of Nursing
Chicago, Illinois
Sleep and Sleep Disorders Associated with Pregnancy
Sleep-Disordered Breathing in Pregnancy

Siobhan Banks, PhD
Centre for Sleep Research
University of South Australia
Adelaide, Australia
Sleep Deprivation

Steven R. Barczi, MD
Professor of Medicine
University of Wisconsin School of Medicine and Public Health
Associate Director
Madison VA Geriatric Research, Education and Clinical Center
William S. Middleton Veterans Affairs Hospital
Madison, Wisconsin
Psychiatric and Medical Comorbidities and Effects of Medications in Older Adults

Mathias Basner, MD, PhD, MSc
Unit for Experimental Psychiatry
Division of Sleep and Chronobiology
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania
Sleep Deprivation

Claudio L. Bassetti, MD
Chairman and Head
Neurology Department
Inselspital, University Hospital
Bern, Switzerland
Idiopathic Hypersomnia
Sleep and Stroke

Christian R. Baumann, MD
Department of Neurology
University Hospital Zurich
University of Zurich
Zurich, Switzerland
Pathophysiology of Sleep-Wake Disturbances After Traumatic Brain Injury
Sleep Disorders After Traumatic Brain Injury

Mihaela Bazalakova, MD, PhD
Assistant Professor
Department of Neurology
Center for Sleep Medicine and Sleep Research
University of Wisconsin–Madison
Madison, Wisconsin
Wake-Promoting Medications: Efficacy and Adverse Effects

Simon Beaulieu-Bonneau, PhD
Research Associate
École de Psychologie
Université Laval, Québec
Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale
Quebec, Canada
Cognitive Behavior Therapies for Insomnia I: Approaches and Efficacy

Gregory Belenky, MD
Research Professor
Sleep and Performance Research Center
Washington State University
Spokane, Washington
Introduction: Occupational Sleep Medicine
Fatigue Risk Management Systems

Ruth M. Benca, MD, PhD
Professor
Department of Psychiatry
Director
Center for Sleep Medicine and Sleep Research
University of Wisconsin-Madison
Madison, Wisconsin
Wake-Promoting Medications: Efficacy and Adverse Effects
Unipolar Major Depression

Kathleen L. Benson, PhD
Research Associate
Brain Imaging Center
McLean Hospital
Belmont, Massachusetts
Research Associate
Department of Psychiatry
Harvard Medical School
Boston, Massachusetts
Schizophrenia

Mark B. Berger, MD
Chief Medical Officer
Precision Pulmonary Diagnostics, LLC
Houston, Texas
Obstructive Sleep Apnea in the Workplace

Richard B. Berry, MD
Professor of Medicine
Division of Pulmonary, Critical Care, and Sleep Medicine
University of Florida
Gainesville, Florida
Sleep Related Breathing Disorders: Classification

Donald L. Bliwise, PhD
Professor of Neurology
Emory University School of Medicine
Atlanta, Georgia
Normal Aging

Bradley F. Boeve, MD
Professor of Neurology
Center for Sleep Medicine and Department of Neurology
Mayo Clinic College of Medicine
Rochester, Minnesota
Alzheimer Disease and Other Dementias
Rapid Eye Movement Sleep Parasomnias

Alexander A. Borbély, MD
Institute of Pharmacology and Toxicology
University of Zurich
Professor Emeritus
Zurich Center for Interdisciplinary Sleep Research
University of Zurich
Zurich, Switzerland
Sleep Homeostasis and Models of Sleep Regulation

Daniel B. Brown, BA, JD
Taylor English Duma LLP
Atlanta, Georgia
Legal Obligations of Persons Who Have Sleep Disorders or Who Treat or Hire Them
Legal Aspects of Fatigue- and Safety-Sensitive Professions
Sleep Medicine Clinical Practice and Compliance—United States

Luis Buenaver, PhD
Assistant Professor
Psychiatry and Behavioral Sciences
The Johns Hopkins University and Hospital School of Medicine
Baltimore, Maryland
Medical and Device Treatment for Obstructive Sleep Apnea: Alternative, Adjunctive, and Complementary Therapies
Pharmacotherapy, Complementary, and Alternative Medicine for Sleep Bruxism

Keith R. Burgess, MBBS, MSc, PhD, FRACP, FRCPC
Clinical Associate Professor
Department of Medicine
University of Sydney
Medical Director
Peninsula Sleep Clinic
Senior Staff Specialist
Critical Care
Manly Hospital
Director
Peninsula Respiratory Group
Sydney, New South Wales, Australia
Respiratory Physiology: Sleep at High Altitudes

Jane E. Butler, BSc(Hons), PhD
Principal Research Fellow
Neuroscience Research Australia
Senior Research Fellow
National Health and Medical Research Council of Australia
Associate Professor
School of Medical Sciences
University of New South Wales
Sydney, Australia
Respiratory Physiology: Understanding the Control of Ventilation

Orfeu M. Buxton, PhD
Associate Professor
Biobehavioral Health
Pennsylvania State University
University Park, Pennsylvania
Lecturer on Medicine
Division of Sleep Medicine
Harvard Medical School
Associate Neuroscientist
Department of Medicine
Brigham and Women's Hospital
Adjunct Associate Professor
Social and Behavioral Sciences
Harvard School of Public Health
Boston, Massachusetts
Human Circadian Timing System and Sleep–Wake Regulation

Daniel J. Buysse, MD
Professor of Psychiatry and Clinical and Translational Science
Department of Psychiatry
University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania
Clinical Pharmacology of Other Drugs Used as Hypnotics
Insomnia: Recent Developments and Future Directions
Bipolar Disorder

Enda M. Byrne, PhD
Research Fellow
Queensland Brain Institute
Brisbane, Australia
Visiting Scholar
Center for Sleep and Circadian Neurobiology
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania
Genetics and Genomic Basis of Sleep Disorders in Humans

Michelle T. Cao, DO
Clinical Assistant Professor
Psychiatry and Behavioral Sciences
Sleep Medicine Division
Stanford University School of Medicine
Stanford, California
Narcolepsy: Diagnosis and Management
Sleep and Neuromuscular Diseases

Colleen E. Carney, PhD
Associate Professor
Department of Psychology
Director, Sleep and Depression Laboratory
Ryerson University, Toronto
Toronto, Canada
Psychological and Behavioral Treatments for Insomnia II: Implementation and Specific Populations

Michelle Carr, BSc
PhD Candidate
Biomedical Science
Université de Montréal
Researcher
Dream and Nightmare Laboratory
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Nightmares and Nightmare Function

Maria Clotilde Carra, DMD, PhD
Assistant Professor
Department of Periodontology
Rothschild Hospital, Paris
Faculty of Odontology
Paris Diderot University
Paris, France
Oral Appliances for the Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome and for Concomitant Sleep Bruxism

Santiago J. Carrizo, MD
Senior Consultant
Respiratory Service
Hospital Universitario Miguel Servet
Zaragoza, Spain
Overlap Syndromes of Sleep and Breathing Disorders

Mary A. Carskadon, PhD
Professor, Psychiatry and Human Behavior
Alpert Medical School of Brown University
Providence, Rhode Island
Director, Sleep and Chronobiology Laboratory
EP Bradley Hospital
East Providence, Rhode Island
Professor, Psychology, Social Work, and Social Policy
University of South Australia
Adelaide, South Australia
Director, Centre for Sleep Research
University of South Australia
Adelaide, South Australia
Normal Human Sleep: An Overview
Daytime Sleepiness and Alertness

Eduardo Castrillon, DDS, MSc, PhD
Associate Professor
Section of Orofacial Pain and Jaw Function
School of Dentistry, Aarhus University
Aarhus, Denmark
Sleep Bruxism: Definition, Prevalence, Classification, Etiology, and Consequences

Etienne Challet, PhD
Institute of Cellular and Integrative Neurosciences
University of Strasbourg
Strasbourg, France
Central and Peripheral Circadian Clocks

Ronald D. Chervin, MD, MS
Professor of Neurology
Michael S. Aldrich Collegiate Professor of Sleep Medicine
Director, Sleep Disorders Center
University of Michigan Health System
Ann Arbor, Michigan
Use of Clinical Tools and Tests in Sleep Medicine

Peter A. Cistulli, MBBS, PhD, MBA, FRACP
ResMed Chair in Sleep Medicine
Sydney Medical School
University of Sydney
Sydney, Australia
Director
Centre for Sleep Health and Research
Royal North Shore Hospital
St. Leonards, Australia
Oral Appliances for the Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome and for Concomitant Sleep Bruxism

Samuele Cortese, MD, PhD
Clinical Associate Professor/Honorary Consultant
University of Southampton
Southampton, United Kingdom
Adjunct Associate Professor
New York University
New York, New York
Sleep Disturbances in Attention-Deficit/Hyperactivity Disorder

Anita P. Courcoulas, MD, MPH, FACS
Professor of Surgery
Minimally Invasive Bariatric and General Surgery
University of Pittsburgh Medical Center
Pittsburgh, Pennsylvania
Obstructive Sleep Apnea, Obesity, and Bariatric Surgery

Robert Craft, MD
Professor and Chair
Department of Anesthesiology
University of Tennessee Graduate School of Medicine
Knoxville, Tenneesee
Opiate Action on Sleep and Breathing

Michel A. Cramer-Bornemann, MD
Lead Investigator
Sleep Forensics Associates
Minneapolis/Saint Paul, Minnesota
Director
Sleep Medicine
Care Services
Olmsted Medical Center
Rochester, Minnesota
Sleep Forensics: Criminal Culpability for Sleep-Related Violence

Antonio Culebras, MD
Professor of Neurology
SUNY Upstate Medical University
Syracuse, New York
Other Neurologic Disorders

Charles A. Czeisler, PhD, MD, FRCP, FAPS
Frank Baldino, Jr., PhD Professor of Sleep Medicine
Professor of Medicine
Director, Division of Sleep Medicine
Harvard Medical School
Chief, Division of Sleep and Circadian Disorders
Departments of Medicine and Neurology
Brigham & Women's Hospital
Boston, Massachusetts
Human Circadian Timing System and Sleep-Wake Regulation

Michael Czisch, PhD
Max Planck Institute of Psychiatry
Munich, Germany
Lucid Dreaming

Yves Dauvilliers, MD, PhD
National Reference Network for Orphan Diseases
(Narcolepsy, Hypersomnia, Kleine-Levin Syndrome)
Sleep Unit, Department of Neurology
Gui de Chauliac Hospital
Montpellier, France
Idiopathic Hypersomnia

Judith R. Davidson, PhD
Associate Professor (Adjunct)
Departments of Psychology and Oncology
Queen's University
Psychologist
Kingston Family Health Team
Kingston, Ontario, Canada
Cognitive Behavior Therapies for Insomnia I: Approaches and Efficacy

O'Neill F. D'Cruz, MD, MBA
Chief Medical Officer
Cyberonics
Houston, Texas
Cardinal Manifestations of Sleep Disorders

Tom Deboer, PhD
Associate Professor
Molecular Cell Biology
Leiden University Medical Center
Leiden, Netherlands
Thermoregulation in Sleep and Hibernation

Luigi De Gennaro, PhD
Associate Professor of Physiological Psychology
Department of Psychology
University of Rome Sapienza
Rome, Italy
Brain Correlates of Successful Dream Recall

William C. Dement, MD, PhD
Lowell W. and Josephine Q. Berry Professor of Psychiatry and Behavioral Sciences
Stanford University School of Medicine
Sleep Sciences and Medicine
Stanford, California
History of Sleep Physiology and Medicine
Normal Human Sleep: An Overview
Daytime Sleepiness and Alertness

Jerome A. Dempsey, PhD
John Robert Sutton Professor Emeritus of Population Health Sciences
Director, John Rankin Laboratory of Pulmonary Medicine
University of Wisconsin–Madison
Madison, Wisconsin
Sleep and Breathing at High Altitude

Derk-Jan Dijk, PhD
Professor
Surrey Sleep Research Centre
University of Surrey
Guildford, United Kingdom
Genetics and Genomic Basis of Sleep in Healthy Humans

David F. Dinges, PhD
Unit for Experimental Psychiatry
Division of Sleep and Chronobiology
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania
Sleep Deprivation

G. William Domhoff, PhD
Distinguished Professor Emeritus and Research Professor, Department of Psychology
University of California
Santa Cruz, California
Dream Content: Quantitative Findings

Jill Dorrian, PhD
Centre for Sleep Research
University of South Australia
Adelaide, Australia
Sleep Deprivation

Anthony G. Doufas, MD, PhD
Associate Professor of Anesthesiology
Perioperative and Pain Medicine
Stanford University School of Medicine
Stanford, California
Pain and Sleep

Luciano F. Drager, MD, PhD
Hypertension Unit
Heart Institute, University of Sao Paulo
Sao Paulo, Brazil
Sleep and Cardiovascular Disease: Present and Future

Christopher L. Drake, PhD
Director of Sleep Research
Sleep Disorders and Research Center
Henry Ford Hospital
Associate Professor
Psychiatry and Behavioral Neuroscience
Wayne State University School of Medicine
Detroit, Michigan
Shift Work, Shift Work Disorder, and Jet Lag

Martin Dresler, PhD
Max Planck Institute of Psychiatry
Munich, Germany
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands
Lucid Dreaming

Peter R. Eastwood, PhD
Winthrop Professor and Director
Centre for Sleep Science
School for Anatomy, Physiology, and Human Biology
University of Western Australia
Senior Scientist
West Australian Sleep Disorders Research Institute
Department of Pulmonary Physiology and Sleep Medicine
Sir Charles Gairdner Hospital
Perth, Australia
Anesthesia in Upper Airway Surgery for Obstructive Sleep Apnea

Danny J. Eckert, PhD
Principal Research Fellow
Neuroscience Research Australia
R.D. Wright Fellow
National Health and Medical Research Council of Australia
Associate Professor
School of Medical Sciences
University of New South Wales
Sydney, Australia
Respiratory Physiology: Understanding the Control of Ventilation

Jack D. Edinger, PhD
Professor of Medicine
National Jewish Health
Denver, Colorado
Adjunct Professor
Psychiatry and Behavioral Sciences
Duke University Medical Center
Durham, North Carolina
Psychological and Behavioral Treatments for Insomnia II: Implementation and Specific Populations

Jason Gordon Ellis, PhD
Professor of Sleep Science
Northumbria Centre for Sleep Research
Northumbria University
Newcastle, United Kingdom
Etiology and Pathophysiology of Insomnia

E. Wesley Ely, MD, MPH
Professor of Medicine
Department of Allergy, Pulmonary, and Criticial Care Medicine
Vanderbilt University Medical Center
Nashville, Tennessee
Sleep in the Critically Ill Patient

Daniel Erlacher, PhD
Institute of Sport Science
University of Bern
Bern, Switzerland
Lucid Dreaming

Gregory K. Essick, DDS, PhD
Professor
Department of Prosthodontics and Center for Pain Research and Innovation
University of North Carolina School of Dentistry
Chapel Hill, North Carolina
Orofacial Pain and Temporomandibular Disorders in Relation to Sleep-Disordered Breathing and Sleep Bruxism

Francesca Facco, MD
Assistant Professor
Department of Obstetrics, Gynecology, and Reproductive Sciences
University of Pittsburgh School of Medicine
Magee-Womens Hospital of UPMC
Pittsburgh, Pennsylvania
Sleep-Disordered Breathing in Pregnancy

Siavash Farshidpanah, MD
Sleep Medicine Fellow
Neurology, Division of Sleep Medicine
Vanderbilt University Medical Center
Nashville, Tennessee
Sleep in the Critically Ill Patient

Irwin Feinberg, MD
Professor Emeritus
Department of Psychiatry and Behavioral Sciences
University of California, Davis
Davis, California
Schizophrenia

Luigi Ferini-Strambi, MD
Professor of Neurology
Deptartment of Clinical Neuroscience
Università Vita-Salute San Raffaele
Milano, Italy
Restless Legs Syndrome and Periodic Limb Movements During Sleep

Julio Fernandez-Mendoza, PhD, CBSM
Assistant Professor of Psychiatry
Sleep Research and Treatment Center
Penn State College of Medicine
Penn State Milton S. Hershey Medical Center
Hershey, Pennsylvania
Insomnia and Health

Michele Ferrara, PhD
Department of Life, Health, and Environmental Sciences
University of L'Aquila
L'Aquila, Italy
Brain Correlates of Successful Dream Recall

Raffaele Ferri, MD
Sleep Research Centre, Department of Neurology I.C.
Oasi Research Institute (IRCCS)
Troina, Italy
Recording and Scoring Sleep-Related Movements

Stuart Fogel, PhD
Research Scientist
The Brain and Mind Institute
Western University
Adjunct Professor
Department of Psychology
Western University
London, Ontario, Canada
Memory Processing in Relation to Sleep

Paul Franken, PhD
Associate Professor
Centre Intégrative de Génomique
Bâtiment Le Génopode
Université de Lausanne
Lausanne, Switzerland
Genetics and Genomic Basis of Sleep in Rodents

Karl A. Franklin, MD, PhD
Associate Professor
Surgical and Perioperative Science, Surgery
Umeå University
Umeå, Sweden
Coronary Artery Disease and Obstructive Sleep Apnea

Neil Freedman, MD
Division of Pulmonary and Critical Care Medicine
Department of Medicine
NorthShore University Healthsystem
Evanston, Illinois
Positive Airway Pressure Treatment for Obstructive Sleep Apnea

Stephany Fulda, PhD
Sleep and Epilepsy Center
Neurocenter of Southern Switzerland/Civic Hospital (EOC) of Lugano
Lugano, Switzerland
Recording and Scoring Sleep-Related Movements

Rylie J. Gabehart, BS
Postbaccalaureate Research Assistant
Sleep and Performance Research Center
Washington State University
Spokane, Washington
Circadian Rhythms in Sleepiness, Alertness, and Performance

Charlene E. Gamaldo, MD
Associate Professor
Department of Neurology
Johns Hopkins Medicine
Baltimore, Maryland
Sleep-Related Movement Disorders and Their Unique Motor Manifestations

Philippa H. Gander, PhD
Professor
Sleep/Wake Research Centre
Massey University
Wellington, New Zealand
Fatigue Risk Management Systems

Philip R. Gehrman, PhD, CBSM
Sleep and Traumatic Stress Program
Department of Psychiatry
University of Pennsylvania
Philadelphia, Pennsylvania
Genetics and Genomic Basis of Sleep Disorders in Humans
Insomnia Diagnosis, Assessment, and Evaluation

Avram R. Gold, MD
Associate Professor of Clinical Medicine
Pulmonary, Critical Care, and Sleep Medicine
Stony Brook University School of Medicine
Stony Brook, New York
Staff Physician
Pulmonary Section, Medical Service
DVA Medical Center
Northport, New York
Snoring and Pathologic Upper Airway Resistance Syndromes

Cathy A. Goldstein, MD, MS
Assistant Professor of Neurology
Sleep Disorders Center
University of Michigan Health System
Ann Arbor, Michigan
Use of Clinical Tools and Tests in Sleep Medicine

Joshua J. Gooley, PhD
Program in Neuroscience and Behavioral Disorders
Duke-NUS Graduate Medical School
Singapore City, Singapore
Anatomy of the Mammalian Circadian System

Nadia Gosselin, PhD
Assistant Professor
Department of Psychology
Université de Montréal
Researcher
Center for Advanced Research in Sleep Medicine
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Pathophysiology of Sleep-Wake Disturbances After Traumatic Brain Injury

Harly Greenberg, MD
Professor of Medicine
Pulmonary, Critical Care, and Sleep Medicine
Hofstra North Shore LIJ School of Medicine
New Hyde Park, New York
Obstructive Sleep Apnea: Clinical Features, Evaluation, and Principles of Management

Edith Grosbellet, PhD
Institute of Cellular and Integrative Neurosciences
University of Strasbourg
Strasbourg, France
Central and Peripheral Circadian Clocks

Ludger Grote, MD, PhD
Assistant Professor
Sleep Disorders Center
Department of Pulmonary Medicine
Sahlgrenska University Hospital
Gothenburg, Sweden
Pulse Wave Analysis During Sleep

Christian Guilleminault, MD
Professor
Psychiatry and Behavioral Sciences
Sleep Medicine Division
Stanford University School of Medicine
Stanford, California
Narcolepsy: Diagnosis and Management
Sleep and Neuromuscular Diseases

Seema Gulyani, PhD, CRNP
Senior Research Fellow
Laboratory of Neurosciences
NIH National Institute on Aging
Baltimore, Maryland
Sleep-Related Movement Disorders and Their Unique Motor Manifestations

Martica H. Hall, PhD
Professor of Psychiatry, Psychology, and Clinical and Translational Science
University of Pittsburgh
Pittsburgh, Pennsylvania
Insomnia and Health

Ronald M. Harper, PhD
Distinguished Professor of Neurobiology
David Geffen School of Medicine
Member, Brain Research Institute
University of California Los Angeles
Los Angeles, California
Cardiovascular Physiology and Coupling with Respiration: Central and Autonomic Regulation

Allison G. Harvey, PhD
Professor of Psychology
University of California, Berkeley
Berkeley, California
Insomnia: Recent Developments and Future Directions
Bipolar Disorder

Jan Hedner, MD, PhD
Professor
Department of Sleep Medicine
Respiratory Medicine and Allergology
Sahlgrenska University Hospital
Gothenburg, Sweden
Coronary Artery Disease and Obstructive Sleep Apnea

Raphael Heinzer, MD, MPH
Director
Center for Investigation and Research in Sleep
University Hospital of Lausanne
Senior Lecturer
University of Lausanne
Lausanne, Switzerland
Physiology of Upper and Lower Airways

John H. Herman, PhD, FAASM
Adjunct Professor
Departments of Psychiatry and Psychology
University of Texas Southwestern Medical Center
Dallas, Texas
Chronobiologic Monitoring Techniques

David R. Hillman, MBBS, FANZCA
West Australian Sleep Disorders Research Institute
Department of Pulmonary Physiology and Sleep Medicine
Sir Charles Gairdner Hospital
Perth, Australia
Anesthesia in Upper Airway Surgery for Obstructive Sleep Apnea

Max Hirshkowitz, PhD
Consulting Professor
Division of Public Mental Health and Population Sciences
Stanford University School of Medicine
Stanford, California
Professor (Emeritus)
Department of Medicine
Baylor College of Medicine
Houston, Texas
Polysomnography and Beyond
Sleep Stage Scoring
Monitoring Techniques for Evaluating Suspected Sleep-Related Breathing Disorders
Evaluating Sleepiness

Laura Hoeg, BA
Research Assistant
Sleep and Performance Research Center
Washington State University
Spokane, Washington
Fatigue Risk Management Systems

Aarnoud Hoekema, MD, DMD, PhD
Associate Professor
Academic Centre for Dentistry Amsterdam
Amsterdam, Netherlands
Doctor
Department of Oral and Maxillofacial Surgery
University Medical Center Groningen
Groningen, Netherlands
Staff Surgeon
Department of Oral and Maxillofacial Surgery
Tjongerschans Hospital
Heerenveen, Netherlands
Upper Airway Surgery to Treat Obstructive Sleep-Disordered Breathing

Birgit Högl, MD
Professor of Neurology
Innsbruck Medical University
Innsbruck, Austria
Restless Legs Syndrome and Periodic Limb Movements During Sleep

Hyun Hor, MD, PhD
Department of Clinical Neurosciences
Lausanne University Hospital
Lausanne, Switzerland
Genetics of Normal Human Sleep

Richard L. Horner, PhD
Professor
Departments of Medicine and Physiology
University of Toronto Faculty of Medicine
Canada Research Chair in Sleep and Respiratory Neurobiology,
Toronto, Ontario, Canada
Respiratory Physiology: Central Neural Control of Respiratory Neurons and Motoneurons During Sleep

Steven R. Hursh, PhD
President
Institutes for Behavior Resources, Inc.
Adjunct Professor
Department of Psychiatry and Behavioral Biology
The Johns Hopkins University School of Medicine
Baltimore, Maryland
Performance Deficits During Sleep Loss and Their Operational Consequences
Sleep and Performance Prediction Modeling

Nelly Huynh, PhD
Assistant Research Professor
Faculty of Dental Medicine
Université de Montréal
Montreal, Quebec, Canada
Role of Dentistry and Otolaryngology in Sleep Medicine
Oropharyngeal Growth and Skeletal Malformations

Adriana G. Ioachimescu, MD, PhD, FACE
Associate Professor of Medicine
Co-director Emory Pituitary Center
Emory University
Atlanta, Georgia
Endocrine Disorders

Octavian C. Ioachimescu, MD, PhD
Section Chief and Medical Director
Sleep Medicine Center
Atlanta Veterans Affairs Medical Center
Decatur, Georgia
Associate Professor of Medicine
Department of Medicine
Division of Pulmonary, Critical Care, and Sleep Medicine
Emory University School of Medicine
Atlanta, Georgia
Endocrine Disorders

Mary Sau-Man Ip, MD
Mok Hing Yiu Endowed Professor and Chair
Department of Medicine
Li Ka Shing Faculty of Medicine
University of Hong Kong
Obstructive Sleep Apnea and Metabolic Disorders

Alex Iranzo, MD, PhD
Neurologist
Hospital Clinic Barcelona
Barcelona, Spain
Other Parasomnias

Shahrokh Javaheri, MD
Medical Director
SleepCare Diagnostics, Inc.
Cincinnati, Ohio
Sleep and Breathing at High Altitude
Sleep and Cardiovascular Disease: Present and Future
Cardiovascular Effects of Sleep-Related Breathing Disorders
Systemic and Pulmonary Hypertension in Obstructive Sleep Apnea
Heart Failure

Peng Jiang, PhD
Postdoctoral Fellow
Center for Sleep and Circadian Biology
Northwestern University
Evanston, Illinois
Genetics and Genomics of Circadian Clocks
Genetics and Genomic Basis of Sleep in Rodents

Hadine Joffe, MD, MSc
Associate Professor of Psychiatry
Harvard Medical School
Vice Chair for Research
Director, Women's Hormone and Aging Research Program
Brigham and Women's Hospital
Director of Psycho-Oncology Research
Department of Psychosocial Oncology and Palliative Care
Dana Farber Cancer Institute
Boston, Massachusetts
Sleep and Menopause

Mark E. Josephson, MD
Herman Dana Professor of Medicine
Harvard Medical School
Beth Israel Deaconess Medical Center
Boston, Massachusetts
Cardiac Arrhythmogenesis During Sleep: Mechanisms, Diagnosis, and Therapy

Stefanos N. Kales, MD, MPH
Associate Professor of Medicine
Harvard Medical School
Associate Professor and Program Director
Occupational Medicine Residency
Harvard School of Public Health
Boston, Massachusetts
Division Chief
Occupational Medicine
Cambridge Health Alliance
Cambridge, Massachusetts
Obstructive Sleep Apnea in the Workplace

Eliot S. Katz, MD
Assistant Professor of Pediatrics
Division of Respiratory Diseases
Boston Children's Hospital
Harvard Medical School
Boston, Massachusettes
Central Sleep Apnea: Definitions, Pathophysiology, Genetics, and Epidemiology

Göran Kecklund, PhD
Deputy Director
Stress Research Institute
Stockholm University
Stockholm, Sweeden
International Research Fellow
Behavioral Science Institute
University of Nijmegen
Nijmegen, Netherlands
Sleep, Occupational Stress, and Burnout
Optimizing Shift Scheduling

Brendan T. Keenan
Biostatistician
Center for Sleep and Circadian Neurobiology
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania
Genetics and Genomic Basis of Sleep Disorders in Humans

Sharon Keenan, PhD
Director
The School of Sleep Medicine, Inc.
Palo Alto, California
Sleep Stage Scoring

John C. Keifer, MD
Associate Professor
Department of Anesthesiology
Duke University Medical Center
Durham, North Carolina
Opiate Action on Sleep and Breathing

Thomas S. Kilduff, PhD
Center Director
Center for Neuroscience
Biosciences Division
SRI International
Menlo Park, California
Hypnotic Medications: Mechanisms of Action and Pharmacologic Effects

Douglas Kirsch, MD, FAAN, FAASM
Medical Director, Sleep Medicine
Carolinas HealthCare System
Clinical Associate Professor
University of North Carolina School of Medicine
Charlotte, North Carolina
Fibromyalgia and Chronic Fatigue Syndromes

Christopher E. Kline, PhD
Assistant Professor of Health and Physical Activity
University of Pittsburgh
Pittsburgh, Pennsylvania
Insomnia and Health

Jacqueline DeMichele Kloss, PhD
Associate Professor of Psychology
Drexel University
Philadelphia, Pennsylvania
Etiology and Pathophysiology of Insomnia

Melissa Pauline Knauert, MD, PhD
Assistant Professor
Section of Pulmonary, Critical Care, and Sleep Medicine
Department of Internal Medicine
Yale University School of Medicine
New Haven, Connecticut
Sleep-Disordered Breathing in Pregnancy

Sanjeev V. Kothare, MD
Director, Pediatric Sleep Program
Professor of Neurology
New York University Medical Center
New York, New York
Epilepsy, Sleep, and Sleep Disorders

Kiyoshi Koyano, DDS, PhD
Professor
Implant and Rehabilitative Dentistry
Faculty of Dental Science
Kyushu University
Fukuoka, Japan
Sleep Bruxism: Diagnostic Considerations

Kurt Kräuchi, NO
Thermophysiological Chronobiology
Centre for Chronobiology
Psychiatric University Clinics
Basel, Switzerland
Thermoregulation in Sleep and Hibernation

James M. Krueger, PhD, MDHC
Regents Professor
Medical Sciences
Washington State University
Spokane, Washington
Sleep and Host Defense

Meir Kryger, MD, FRCPC
Professor
Pulmonary, Critical Care, and Sleep Medicine
Yale University School of Medicine
New Haven, Connecticut
Relevance of Sleep Physiology for Sleep Medicine Clinicians
Physical Examination in Sleep Medicine
Monitoring Techniques for Evaluating Suspected Sleep-Related Breathing Disorders

Andrew D. Krystal, MD, MS
Professor of Psychiatry and Behavioral Sciences
Duke University School of Medicine
Durham, North Carolina
Pharmacologic Treatment of Insomnia: Other Medications
Anxiety Disorders and Posttraumatic Stress Disorder
Unipolar Major Depression

Scott J. Kutscher, MD
Assistant Professor
Department of Neurology
Vanderbilt University
Nashville, Tennessee
Sleep and Athletic Performance

Anthony B. Kwan, MD Cand
College of Medicine
State University of New York
Downstate Medical Center
Brooklyn, New York
Sleep-Related Movement Disorders and Their Unique Motor Manifestations

Viera Lakticova, MD
Assistant Professor of Medicine
Hofstra North Shore LIJ School of Medicine
New Hyde Part, New York
Obstructive Sleep Apnea: Clinical Features, Evaluation, and Principles of Management

Amanda Lamp, BS
PhD Candidate
Sleep and Performance Research Center
Washington State University
Spokane, Washington
Fatigue Risk Management Systems

Hans-Peter Landolt, PhD
Professor
Institute of Pharmacology and Toxicology
Clinical Research Priority Program "Sleep & Health"
Zürich Center for Interdisciplinary Sleep Research
University of Zürich
Zürich, Switzerland
Genetics and Genomic Basis of Sleep in Healthy Humans

Paola A. Lanfranchi, MD, MSc
Center for Sleep Studies
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders

Gilles Lavigne, DMD, FRCDC, PhD
Professor of Oral Medicine
Canada Research Chair in Pain, Sleep, and Trauma
Faculty of Dental Medicine
Université de Montreal
Montreal, Quebec, Canada
Relevance of Sleep Physiology for Sleep Medicine Clinicians
Role of Dentistry and Otolaryngology in Sleep Medicine
Sleep Bruxism: Definition, Prevalence, Classification, Etiology, and Consequences
Orofacial Pain and Temporomandibular Disorders in Relation to Sleep-Disordered Breathing and Sleep Bruxism

Michel Lecendreux, MD
Senior Consultant
Hospital Robert Debré
Paris, France
Sleep Disturbances in Attention-Deficit/Hyperactivity Disorder

Kathryn Aldrich Lee, PhD
Professor
Family Health Care Nursing
University of California, San Francisco
San Francisco, California
Sleep and Sleep Disorders Associated with Pregnancy
Sleep and Menopause

Melanie K. Leggett, PhD
Staff Psychologist
VA Medical Center
Associate Professor
Department of Psychiatry and Behavioral Sciences
Duke University Medical Center
Durham, North Carolina
Psychological and Behavioral Treatments for Insomnia II: Implementation and Specific Populations

Christopher J. Lettieri, MD
Professor of Medicine
Uniformed Services University
Program Director, Sleep Medicine
Pulmonary, Critical Care, and Sleep Medicine
Walter Reed National Military Medical Center
Bethesda, Maryland
Oral Appliances for the Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome and for Concomitant Sleep Bruxism

Kenneth L. Lichstein, PhD
Professor
Department of Psychology
University of Alabama
Tuscaloosa, Alabama
Insomnia: Epidemiology and Risk Factors

Frank Lobbezoo, DDS, PhD
Professor and Chair
Department of Oral Health Sciences
Academic Centre for Dentistry Amsterdam
Amsterdam, Netherlands
Sleep Bruxism: Diagnostic Considerations

Geraldo Lorenzi-Filho, MD, PhD
Associate Professor
Cardiopulmonology
Heart Institute, University of Sao Paulo
Sao Paulo, Brazil
Sleep and Cardiovascular Disease: Present and Future

Judette Louis, MD, MPH
Assistant Professor
Department of Obstetrics and Gynecology
College of Medicine
Assistant Professor
Department of Community and Family Health
College of Public Health
University of South Florida
Tampa, Florida
Sleep-Disordered Breathing in Pregnancy

Ralph Lydic, PhD
Robert H. Cole Professor of Neuroscience
Departments of Anesthesiology and Psychology
University of Tennessee
Knoxville, Tennessee
Opiate Action on Sleep and Breathing

Madalina Macrea, MD, MPH, PhD
Associate Professor of Medicine
Salem Veterans Affairs Medical Center
Salem, Virginia
Associate Professor of Medicine
University of Virginia
Charlottesville, Virginia
Central Sleep Apnea: Definitions, Pathophysiology, Genetics, and Epidemiology

Mary Halsey Maddox, MD
Assistant Professor
Department of Pediatrics
Division of Pulmonary and Sleep Medicine
University of Alabama at Birmingham
Birmingham, Alabama
Epidemiology of Sleep Medicine

Mark W. Mahowald, MD
Professor of Neurology (ret.)
University of Minnesota Medical School
Minneapolis, Minnesota
Adjunct Clinical Professor
Psychiatry and Behavioral Sciences
Stanford University
Stanford, California
Sleep Forensics: Criminal Culpability for Sleep-Related Violence

Atul Malhotra, MD
Professor of Medicine
Division Chief, Pulmonary and Critical Care Medicine
Director of Sleep Medicine
Kenneth M. Moser Professor
Department of Medicine
University of California, San Diego
San Diego, California
Obstructive Sleep Apnea in the Workplace
Central Sleep Apnea: Definitions, Pathophysiology, Genetics, and Epidemiology

Roneil G. Malkani, MD
Assistant Professor
Department of Neurology
Northwestern University Feinberg School of Medicine
Chicago, Illinois
Circadian Dysregulation in Mental and Physical Health

Beth A. Malow, MD, MS
Professor and Director, Sleep Disorders Division
Neurology and Pediatrics
Vanderbilt University
Nashville, Tennessee
Approach to the Patient with Disordered Sleep
Neurologic Monitoring Techniques

Rachel Manber, PhD
Professor
Department of Psychiatry
Stanford University School of Medicine
Stanford, California
Psychological and Behavioral Treatments for Insomnia II: Implementation and Specific Populations

Daniele Manfredini, DDS, PhD
Associate Professor
Department of Maxillofacial Surgery
University of Padova
Padova, Italy
Sleep Bruxism: Diagnostic Considerations

Pierre Maquet, MD, PhD
Cyclotron Research Center
University of Liege
Department of Neurology
Liege University Hospital
Liege, Belgium
What Brain Imaging Reveals About Sleep Generation and Maintenance

Jose M. Marin, MD
Head, Respiratory Sleep Disorders Unit
Hospital Universitario Miguel Servet
Asociated Professor of Medicine
Department of Medicine
University of Zaragoza
Zaragoza, Spain
Overlap Syndromes of Sleep and Breathing Disorders

Jeffrey Masor, JD, CAMS
Contract Attorney
The Daniel Brown Law Group
Dunwoody, Georgia
Legal Aspects of Fatigue- and Safety-Sensitive Professions

Christina S. McCrae, PhD
Professor
Department of Health Psychology
University of Missouri–Columbia
Columbia, Missouri
Insomnia: Epidemiology and Risk Factors

Dennis McGinty, PhD
Department of Psychology
University of California Los Angeles
Research Service
VA Greater Los Angeles Healthcare System
Los Angeles, California
Neural Control of Sleep in Mammals

Reena Mehra, MD, MS
Associate Professor of Medicine
Sleep Center, Neurologic Institute
Cleveland Clinic Lerner College of Medicine
Case Western Reserve University
Cleveland, Ohio
Sleep Breathing Disorders: Clinical Overview

Thomas A. Mellman, MD
Director, Clinical and Translational Research and Stress and Sleep Studies Programs
Professor of Psychiatry
Howard University College of Medicine
Washington, D.C.
Dreams and Nightmares in Posttraumatic Stress Disorder

Wallace B. Mendelson, MD
Professor of Psychiatry and Clinical Pharmacology (Ret.)
The University of Chicago
Chicago, Illinois
Medical Director
San Benito County Behavioral Health
Hollister, California
Hypnotic Medications: Mechanisms of Action and Pharmacologic Effects

Emmanuel Mignot, MD, PhD
Director
Center for Sleep Sciences and Medicine
Stanford University
Stanford, California
Wake-Promoting Medications: Basic Mechanisms and Pharmacology
Narcolepsy: Genetics, Immunology, and Pathophysiology

Jared D. Minkel, PhD
Psychiatry and Behavioral Sciences
Duke University Medical Center
Durham, North Carolina
Unipolar Major Depression

Murray A. Mittleman, MD, DrPH
Professor of Epidemiology
Harvard T.H. Chan School of Public Health
Associate Professor of Medicine
Harvard Medical School
Boston, Massachusetts
Sleep-Related Cardiac Risk

Vahid Mohsenin, MD
Professor of Medicine
Yale University
New Haven, Connecticut
Sleep and Breathing at High Altitude

Babak Mokhlesi, MD, MSc
Director, Sleep Disorders Center and Sleep Medicine Fellowship Program
Department of Medicine
Section of Pulmonary and Critical Care
University of Chicago
Chicago, Illinois
Obesity-Hypoventilation Syndrome

Jacques Montplaisir, PhD
Professor and Director of the Canadian Research Chair in Sleep Medicine
Departement of Psychiatry
Université de Montréal
Center for Advanced Research on Sleep Medicine
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Restless Legs Syndrome and Periodic Limb Movements During Sleep
Alzheimer Disease and Other Dementias

Charles M. Morin, PhD
Professor
École de Psychologie
Université Laval, Québec
Researcher
Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec
Quebec, Canada
Cognitive Behavior Therapies for Insomnia I: Approaches and Efficacy

Mary J. Morrell, PhD
Professor of Sleep and Respiratory Physiology
National Heart and Lung Institute
Imperial College
London, United Kingdom
Obstructive Sleep Apnea and the Central Nervous System: Neural Adaptive Processes, Cognition, and Performance

Douglas E. Moul, MD, MPH
Sleep Psychiatrist, Staff Physician
Sleep Disorders Center, Neurological Institute
Cleveland Clinic
Cleveland, Ohio
Sleep Breathing Disorders: Clinical Overview

Tore Nielsen, PhD
Professor of Psychiatry
Université de Montréal
Director, Dream and Nightmare Laboratory
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Nightmares and Nightmare Function

F. Javier Nieto, MD, MPH, PhD
Professor and Chair of Population Health Sciences
School of Medicine and Public Health
University of Wisconsin–Madison
Madison, Wisconsin
Systemic and Pulmonary Hypertension in Obstructive Sleep Apnea

Seiji Nishino, MD, PhD
Professor of Psychiatry and Behavioral Sciences
Stanford University School of Medicine
Director
Stanford Sleep and Circadian Neurobiology Laboratory
Stanford, California
Wake-Promoting Medications: Basic Mechanisms and Pharmacology

Eric A. Nofzinger, MD
Founder, Director, and Chief Medical Officer
Cerêve, Inc.
Oakmont, Pennsylvania
What Brain Imaging Reveals About Sleep Generation and Maintenance

Louise M. O'Brien, PhD, MS
Associate Professor
Sleep Disorders Center
Associate Professor
Obstetrics and Gynecology
Associate Research Scientist
Oral and Maxillofacial Surgery
University of Michigan
Ann Arbor, Michigan
Sex Differences and Menstrual-Related Changes in Sleep and Circadian Rhythms

Bruce F. O'Hara, PhD
Professor of Biology
University of Kentucky
Lexington, Kentucky
Genetics and Genomic Basis of Sleep in Rodents

Eric J. Olson, MD
Associate Professor of Medicine
Mayo Clinic College of Medicine
Division of Pulmonary and Critical Care Medicine
Co-Director, Center for Sleep Medicine
Mayo Clinic
Rochester, Minnesota
Obstructive Sleep Apnea, Obesity, and Bariatric Surgery

Jason C. Ong, PhD, CBSM
Associate Professor
Department of Behavioral Sciences
Director, Behavioral Sleep Medicine Training Program
Rush University Medical Center
Chicago, Illinois
Insomnia Diagnosis, Assessment, and Evaluation

Mark R. Opp, PhD
Professor and Vice Chair for Basic Research
Anesthesiology and Pain Medicine
University of Washington
Seattle, Washington
Sleep and Host Defense

Edward F. Pace-Schott, PhD
Assistant Professor of Psychiatry
Harvard Medical School
Massachusetts General Hospital
Charlestown, Massachusetts
Neurobiology of Dreaming

Allan I. Pack, MBChB, PhD
John Miclot Professor of Medicine
Director, Center for Sleep and Circadian Neurobiology
Chief, Division of Sleep Medicine
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania
Genetics and Genomic Basis of Sleep Disorders in Humans

Daniel A. Paesani, DDS
Profesor of Stomathognatic Phisiology
School of Dentistry
University of Salvador/AOA
Buenos Aires, Argentina
Sleep Bruxism: Diagnostic Considerations

John G. Park, MD
Assistant Professor of Medicine
Division of Pulmonary and Critical Care Medicine
Mayo Clinic
Rochester, Minnesota
Sleep and Chronic Kidney Disease

Liborio Parrino, MD
Professor of Neurology
Department of Neuroscience
University of Parma
Parma, Italy
Central Nervous System Arousals and Cyclic Alternating Patterns

Susheel P. Patil, MD, PhD
Assistant Professor of Medicine
The Johns Hopkins University and Hospital School of Medicine
Baltimore, Maryland
Medical and Device Treatment for Obstructive Sleep Apnea: Alternative, Adjunctive, and Complementary Therapies
Pharmacotherapy, Complementary, and Alternative Medicine for Sleep Bruxism

Milena K. Pavlova, MD
Medical Director—Faulkner Sleep Testing Center
Neurology
Brigham and Women's Hospital
Assistant Professor of Neurology
Harvard Medical School
Boston, Massachusetts
Epilepsy, Sleep, and Sleep Disorders

John H. Peever, PhD
Professor
Laboratory for Sleep Research
Department of Cell and Systems Biology, and Physiology
University of Toronto
Toronto, Ontario
Canada
Novel Techniques for Identifying Sleep Mechanisms and Disorders
Sensory and Motor Processing During Sleep and Wakefulness

Philippe Peigneux, PhD
Full Professor
Faculty of Psychological Sciences
Université Libre de Bruxelles
Director
Neuropsychology and Functional Neuroimaging Research Unit
Centre for Research in Cognition and Neurosciences
Université Libre de Bruxelles Neurosciences Institute
Brussels, Belgium
Memory Processing in Relation to Sleep

Yüksel Peker, MD, PhD
Professor
Department of Pulmonary Medicine
Marmara University
Istanbul, Turkey
Department of Molecular and Clinical Medicine/Cardiology
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden
Coronary Artery Disease and Obstructive Sleep Apnea

Rafael Pelayo, MD
Clinical Professor
Sleep Medicine Center
Stanford University School of Medicine
Stanford, California
History of Sleep Physiology and Medicine

Thomas Penzel, PhD
Professor
Department of Cardiology
Interdisciplinary Sleep Medicine Center
Charité–Universitätsmedizin Berlin
Berlin, Germany
Sleep Medicine Clinical Practice and Compliance—Europe
Home Sleep Testing

Jean-Louis Pépin, MD, PhD
Université Grenoble Alpes
Laboratoire HP2
Inserm, U1042
CHU de Grenoble
Laboratoire EFCR
Pôle Thorax et Vaisseux
Grenoble, France
Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders

Paul E. Peppard, MS, PhD
Associate Professor
Population Health Sciences
University of Wisconsin–Madison
Madison, Wisconsin
Systemic and Pulmonary Hypertension in Obstructive Sleep Apnea

Michael Lloyd Perlis, PhD
Associate Professor
Departments of Psychiatry and Nursing
University of Pennsylvania
Philadelphia, Pennsylvania
Etiology and Pathophysiology of Insomnia

Lampros Perogamvros, MD
Department of Psychiatry
University Hospitals of Geneva
University of Geneva
Geneva, Switzerland
Emotion, Motivation, and Reward in Relation to Dreaming

Aleksander Perski, PhD
Associate Professor
Stress Research Institute
Stockholm, Sweeden
Sleep, Occupational Stress, and Burnout

Dominique Petit, PhD
Center for Advanced Research in Sleep Medicine
Hôpital du Sacré-Coeur de Montréal
Montreal, Quebec, Canada
Alzheimer Disease and Other Dementias

Megan E. Petrov, PhD
Assistant Professor
College of Nursing and Health Innovation
Arizona State University
Phoenix, Arizona
Insomnia: Epidemiology and Risk Factors

Pierre Philip, MD, PhD
Sleep, Attention, and Neuropsychiatry
University of Bordeaux
University Hospital Pellegrin
Bordeaux, France
Drowsiness in Transportation Workers

Barbara A. Phillips, MD, MSPH, FCCP
Professor
Division of Pulmonary, Critical Care, and Sleep Medicine
University of Kentucky College of Medicine
Lexington, Kentucky
Obstructive Sleep Apnea in the Elderly

Dante Picchioni, PhD
Scientist
Advanced MRI Section
National Institute of Neurological Disorders and Stroke
Scientist
Section on Neuroadaptation and Protein Metabolism
National Institute of Mental Health
Bethesda, Maryland
Neurobiology of Dreaming

Wilfred R. Pigeon, PhD
Research Director
Center of Excellence for Suicide Prevention
Canandaigua VA Medical Center
Canandaigua, New York
Director, Sleep and Neurophysiology Research Lab
Psychiatry
University of Rochester Medical Center
Rochester, New York
Dreams and Nightmares in Posttraumatic Stress Disorder

Margaret A. Pisani, MD, MPH
Associate Professor
Pulmonary, Critical Care, and Sleep Medicine
Yale University School of Medicine
New Haven, Connecticut
Sleep in the Critically Ill Patient

Benjamin T. Pliska, DDS, MSc, FRCD(C)
Assistant Professor of Orthodontics
Department of Oral Health Sciences
University of British Columbia
Vancouver, British Columbia, Canada
Oropharyngeal Growth and Skeletal Malformations

Ronald Postuma, MD, MSc
Associate Professor
Neurology
Montreal General Hospital
Montreal, Quebec, Canada
Parkinsonism

Stacey Dagmar Quo, DDS, MS
Clinical Professor of Orofacial Sciences
University of California, San Francisco
San Francisco, California
Adjunct Assistant Clinical Professor
Psychiatry
Stanford School of Medicine
Palo Alto, California
Oropharyngeal Growth and Skeletal Malformations

Kannan Ramar, MD
Associate Professor of Medicine
Division of Pulmonary and Critical Care Medicine
Mayo Clinic
Rochester, Minnesota
Sleep and Chronic Kidney Disease

Angela C. Randazzo, PhD
Clinical and Research Psychologist
Sleep Medicine and Research Center
St. Luke's Hospital
Chesterfield, Missouri
Drugs that Disturb Sleep and Wakefulness

Karen G. Raphael, PhD
Professor of Oral and Maxillofacial Pathology, Radiology, and Medicine
New York University College of Dentistry
Professor of Psychiatry
New York University School of Medicine
New York, New York
Orofacial Pain and Temporomandibular Disorders in Relation to Sleep-Disordered Breathing and Sleep Bruxism

Susan Redline, MD, MPH
Farrell Professor of Sleep Medicine
Harvard Medical School
Brigham and Women's Hospital
Beth Israel Deaconess Medical Center
Boston, Masachusettes
Obstructive Sleep Apnea: Phenotypes and Genetics

Kathryn J. Reid, PhD
Research Associate Professor
Ken and Ruth Davee Department of Neurology
Northwestern University Feinberg School of Medicine
Chicago, Illinois
Circadian Disorders of the Sleep-Wake Cycle

Albert Rielly, MD, MPH
Physician
Department of Medicine
Cambridge Health Alliance
Cambridge, Massachusetts
Clinical Instructor
Harvard Medical School
Boston, Massachusetts
Obstructive Sleep Apnea in the Workplace

Dieter Wilhelm Riemann, PhD
Professor of Clinical Psychology and Psychophysiology
Center for Mental Disorders/University Medical Center
Freiburg, Germany
Etiology and Pathophysiology of Insomnia

Timothy Roehrs, PhD
Senior Bioscientist
Sleep Disorders and Research Center
Henry Ford Health System
Detroit, Michigan
Daytime Sleepiness and Alertness
Medication and Substance Abuse

Alan M. Rosenwasser, PhD
Professor
Department of Psychology
Cooperating Professor
School of Biology and Ecology
University of Maine
Orono, Maine
Physiology of the Mammalian Circadian System

Ivana Rosenzweig, MD, PhD, MRCPsych
Wellcome Research Fellow and Consultant Neuropsychiatrist
Sleep and Brain Plasticity Centre
Department of Neuroimaging
King's College London
Sleep Disorders Centre
Guy's and St. Thomas' Hospital
London, United Kingdom
Obstructive Sleep Apnea and the Central Nervous System: Neural Adaptive Processes, Cognition, and Performance

Thomas Roth, PhD
Division Head
Sleep Disorders and Research Center
Henry Ford Hospital
Detroit, Michigan
Daytime Sleepiness and Alertness
Effects of Hypnotic Drugs on Driving Performance
Pharmacologic Treatment of Insomnia: Benzodiazepine Receptor Agonists
Medication and Substance Abuse

James A. Rowley, MD
Professor of Medicine
Division of Pulmonary, Critical Care, and Sleep Medicine
Wayne State University School of Medicine
Detroit, Michigan
Anatomy and Physiology of Upper Airway Obstruction

Patricia Sagaspe, PhD
Sleep, Attention, and Neuropsychiatry
University of Bordeaux
University Hospital Pellegrin
Bordeaux, France
Drowsiness in Transportation Workers

Rachel E. Salas, MD
Associate Professor
Department of Neurology
Johns Hopkins Medicine
Baltimore, Maryland
Sleep-Related Movement Disorders and Their Unique Motor Manifestations

Mikael Sallinen, PsyD
Team Leader
Finnish Institute of Occupational Health
Helsinki, Finland
Research Professor
University of Jyväskylä
Jyväskylä, Finland
Optimizing Shift Scheduling

Charles Samuels, MD
Clinical Assistant Professor
Family Medicine
Adjunct Professor
Faculty of Kinesiology
University of Calgary
Calgary, Alberta, Canada
Sleep Problems in First Responders and in Deployed Military Personnel

Anne E. Sanders, MS, PhD
Associate Professor
Department of Dental Ecology
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina
Orofacial Pain and Temporomandibular Disorders in Relation to Sleep-Disordered Breathing and Sleep Bruxism

Clifford B. Saper, MD, PhD
Professor and Chairman
Neurology
Beth Israel Deaconess Medical Center
Harvard Medical School
Boston, Massachusettes
Anatomy of the Mammalian Circadian System

Michael J. Sateia, MD
Professor of Psychiatry (Sleep Medicine), Emeritus
Geisel School of Medicine at Dartmouth
Lebanon, New Hampshire
Classification of Sleep Disorders

Josée Savard, PhD
School of Psychology
Université Laval
CHU de Québec-Université Laval Research Center
Université Laval Cancer Research Centre
Quebec, Canada
Sleep and Fatigue in Cancer Patients

Marie-Hélène Savard, PhD
CHU de Québec-Université Laval Research Center
Université Laval Cancer Research Centre
Quebec, Canada
Sleep and Fatigue in Cancer Patients

Steven M. Scharf, MD, PhD
Professor of Medicine
University of Maryland
Baltimore, Maryland
Obstructive Sleep Apnea: Clinical Features, Evaluation, and Principles of Management

Michael Schredl, PhD
Head of Research
Sleep Laboratory
Central Institute of Mental Health
Medical Faculty Mannheim/Heidelberg University
Mannheim, Germany
Incorporation of Waking Experiences into Dreams

Sophie Schwartz, PhD
Professor of Neuroscience
Department of Neuroscience
University of Geneva
Geneva, Switzerland
Emotion, Motivation, and Reward in Relation to Dreaming

Paula K. Schweitzer, PhD
Sleep Medicine and Research Center
St. Luke's Hospital
Chesterfield, Missouri
Drugs that Disturb Sleep and Wakefulness

Michael K. Scullin, PhD
Assistant Professor
Psychology and Neuroscience
Baylor University
Waco, Texas
Normal Aging

Frédéric Sériès, MD
Centre de Recherche
Institut Universitaire de Cardiologie et de Pneumologie de l'Université Laval
Quebec City, Quebec, Canada
Physiology of Upper and Lower Airways

Barry J. Sessle, MDS, PhD
Professor of Dentistry and Medicine
University of Toronto
Toronto, Ontario, Canada
Sensory and Motor Processing During Sleep and Wakefulness

Amir Sharafkhaneh, MD, PhD
Professor
Department of Medicine
Section of Pulmonary, Critical Care, and Sleep Medicine
Baylor College of Medicine
Houston, Texas
Evaluating Sleepiness

Katherine M. Sharkey, MD, PhD
Assistant Professor of Medicine and Psychiatry and Human Behavior
Brown University Alpert Medical School
Staff, Division of Pulmonology, Critical Care, and Sleep Medicine
Rhode Island Hospital
Providence, Rhode Island
Postpartum Period and Early Motherhood

Priyattam J. Shiromani, PhD
Professor
Department of Psychiatry
Ralph H. Johnson VA and Medical University of South Carolina
Charleston, South Carolina
Novel Techniques for Identifying Sleep Mechanisms and Disorders

Tamar Shochat, DSc
Associate Professor
Department of Nursing
University of Haifa
Haifa, Israel
Insomnia in Older Adults

Jerome M. Siegel, PhD
Professor
Department of Psychiatry and Biobehavioral Sciences
University of California Los Angeles
Chief, Neurobiology Research
Veterans Affairs Greater Los Angeles Healthcare System
Los Angeles, California
Rapid Eye Movement Sleep
Sleep in Animals: A State of Adaptive Inactivity

Michael H. Silber, MB, ChB
Professor of Neurology
Center for Sleep Medicine and Department of Neurology
Mayo Clinic College of Medicine
Rochester, Minnesota
Rapid Eye Movement Sleep Parasomnias

Michael Simmons, DMD
Lecturer
Department of Orofacial Pain and Oral Medicine
University of California, Los Angeles School of Dentistry
Clinical Assistant Professor
Division of Diagnostic Sciences
Herman Ostrow School of Dentistry of USC
Los Angeles, California
Role of Dentistry and Otolaryngology in Sleep Medicine

Carlyle Smith, PhD
Psychology Department
Trent University
Peterborough, Ontario, Canada
Neuroscience Department
Queens University
Kingston, Ontario, Canada
Memory Processing in Relation to Sleep

Michael T. Smith, PhD
Professor
Psychiatry and Behavioral Sciences
The Johns Hopkins University and Hospital School of Medicine
Baltimore, Maryland
Medical and Device Treatment for Obstructive Sleep Apnea: Alternative, Adjunctive, and Complementary Therapies
Pharmacotherapy, Complementary, and Alternative Medicine for Sleep Bruxism

Adriane M. Soehner, PhD
University of Pittsburgh
Pittsburgh, Pennsylvania
Bipolar Disorder

Virend K. Somers, MD, PhD
Professor of Medicine
Department of Internal Medicine
Division of Cardiovascular Diseases
Mayo Medical School/Mayo Clinic
Rochester, Minnesota
Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders
Cardiovascular Effects of Sleep-Related Breathing Disorders

Victor I. Spoormaker, PhD
Max Planck Institute of Psychiatry
Munich, Germany
Lucid Dreaming

Erik K. St. Louis, MD, MS
Associate Professor of Neurology
Center for Sleep Medicine and Department of Neurology
Mayo Clinic College of Medicine
Rochester, Minnesota
Alzheimer Disease and Other Dementias
Rapid Eye Movement Sleep Parasomnias

Murray B. Stein, MD, MPH
Professor
Psychiatry and Family and Preventive Medicine
University of California, San Diego
La Jolla, California
Staff Psychiatrist
Psychiatry Service
VA San Diego Healthcare System
San Diego, California
Anxiety Disorders and Posttraumatic Stress Disorder

Robert Stickgold, PhD
Associate Professor
Department of Psychiatry
Beth Israel Deaconess Medical Center
Department of Psychiatry
Harvard Medical School
Boston, Massachusettes
Introduction: Psychobiology and Dreaming
Why We Dream

Katie L. Stone, MA, PhD
Senior Scientist
Research Institute
California Pacific Medical Center
San Francisco, California
Circadian Rhythms in Older Adults
Actigraphy

Riccardo Stoohs, MD
Director
Sleep Disorders Clinic
Somnolab
Doermund, Germany
Snoring and Pathologic Upper Airway Resistance Syndromes

Robyn Stremler, RN, PhD
Associate Professor
Lawrence S. Bloomberg Faculty of Nursing
University of Toronto
Adjunct Scientist
The Hospital for Sick Children
Toronto, Ontario, Canada
Postpartum Period and Early Motherhood

Kingman P. Strohl, MD
Professor of Medicine and Anatomy
University Hospitals of Cleveland
Cleveland Veterans Affairs Medical Center
Case Western Reserve University
Cleveland, Ohio
Sleep Breathing Disorders: Clinical Overview

Peter Svensson, DDS, PhD, Dr.Odont
Professor and Head
Section of Orofacial Pain and Jaw Function
School of Dentistry, Aarhus University
Aarhus, Denmark
Sleep Bruxism: Definition, Prevalence, Classification, Etiology, and Consequences

Steven T. Szabo, MD, PhD
Assistant Professor
Psychiatry and Behavioral Sciences
Duke University Medical Center
Attending Psychiatrist
Mental Health Service Line
Durham Veterans Affairs Medical Center
Durham, North Carolina
Anxiety Disorders and Posttraumatic Stress Disorder

Ronald Szymusiak, PhD
Professor
Department of Medicine
David Geffen School of Medicine
University of California, Los Angeles
Research Service
VA Greater Los Angeles Healthcare System
Los Angeles, California
Neural Control of Sleep in Mammals

Mehdi Tafti, PhD
Center for Integrative Genomics
University of Lausanne
Center for Investigation and Research in Sleep
Lausanne University Hospital
Lausanne, Switzerland
Genetics of Normal Human Sleep

Jacques Taillard, PhD
Sleep, Attention, and Neuropsychiatry
CNRS
University of Bordeaux
University Hospital Pellegrin
Bordeaux, France
Drowsiness in Transportation Workers

Esra Tasali, MD
Assistant Professor of Medicine
Sleep, Health, and Metabolism Center
University of Chicago
Chicago, Illinois
Endocrine Physiology in Relation to Sleep and Sleep Disturbances

Daniel J. Taylor, PhD, CBSM, DABSM
Associate Professor
Department of Psychology
University of North Texas
Denton, Texas
Insomnia: Epidemiology and Risk Factors

Mihai C. Teodorescu, MD
Associate Professor of Medicine
Division of Geriatrics and Gerontology
University of Wisconsin School of Medicine and Public Health
Wm. S. Middleton Veterans Administration Hospital
Madison, Wisconsin
Psychiatric and Medical Comorbidities and Effects of Medications in Older Adults

Mario Giovanni Terzano, MD
Professor of Neurology
Department of Neuroscience
University of Parma
Parma, Italy
Central Nervous System Arousals and Cyclic Alternating Patterns

Robert Joseph Thomas, MD, MMSc
Associate Professor of Medicine
Pulmonary, Critical Care, and Sleep Division
Beth Israel Deaconess Medical Center
Boston, Massachusetts
Central Sleep Apnea: Diagnosis and Management
Cardiopulmonary Coupling Sleep Spectrograms

Michael J. Thorpy, MD
Professor of Clinical Neurology
The Saul R. Korey Department of Neurology
Albert Einstein College of Medicine at Yeshiva University
Director
Sleep-Wake Disorders Center
Montefiore Medical Center
Bronx, New York
Classification of Sleep Disorders

Gregory J. Tranah, PhD
Professor
Research Institute
California Pacific Medical Center
San Francisco, California
Circadian Rhythms in Older Adults

Claudia Trenkwalder, Prof., Dr.
Professor of Neurology
Department of Neurosurgery
University Medical Center
Goettingen, Germany
Paracelsus-Elena Hospital
Kassel, Germany
Parkinsonism

Fred W. Turek, PhD
Charles E. and Emma H. Morrison Professor of Biology
Department of Neurobiology
Weinberg College of Arts and Sciences
Director, Center for Sleep and Circadian Biology
Northwestern University
Evanston, Illinois
Introduction: Genetics and Genomics of Sleep
Genetics and Genomics of Circadian Clocks
Genetics and Genomic Basis of Sleep in Rodents
Introduction: Master Circadian Clock and Master Circadian Rhythm
Physiology of the Mammalian Circadian System

Shachi Tyagi, MD, MS
Assistant Professor of Medicine
Department of Medicine
University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania
Clinical Pharmacology of Other Drugs Used as Hypnotics

Raghu Pishka Upender, MD
Assistant Professor
Neurology
Vanderbilt University
Nashville, Tennessee
Sleep Medicine, Public Policy, and Public Health

Philipp O. Valko, MD
Department of Neurology
University Hospital Zurich
University of Zurich
Zurich, Switzerland
Sleep Disorders After Traumatic Brain Injury

Eve Van Cauter, PhD
Frederick H. Rawson Professor in Medicine
Sleep, Health, and Metabolism Center
University of Chicago
Chicago, Illinois
Endocrine Physiology in Relation to Sleep and Sleep Disturbances

Aurora J.A.E. van de Loo, MSc
PhD Candidate
Division of Pharmacology
Utrecht University
Utrecht, Netherlands
Effects of Hypnotic Drugs on Driving Performance

Margo van den Berg, BA
Junior Research Officer
Sleep/Wake Research Centre
Massey University
Auckland, New Zealand
Fatigue Risk Management Systems

Olivier M. Vanderveken, MD, PhD
Consultant ENT, Head and Neck Surgeon
Antwerp University Hospital
Professor
Faculty of Medicine and Health Sciences
University of Antwerp
Antwerp, Belgium
Role of Dentistry and Otolaryngology in Sleep Medicine
Anesthesia in Upper Airway Surgery for Obstructive Sleep Apnea
Upper Airway Surgery to Treat Obstructive Sleep-Disordered Breathing

Hans P.A. Van Dongen, MS, PhD
Research Professor and Director
Sleep and Performance Research Center
Washington State University
Spokane, Washington
Circadian Rhythms in Sleepiness, Alertness, and Performance
Performance Deficits During Sleep Loss and Their Operational Consequences
Sleep and Performance Prediction Modeling

Bradley V. Vaughn, MD
Professor of Neurology
University of North Carolina School of Medicine
Chapel Hill, North Carolina
Cardinal Manifestations of Sleep Disorders
Parasomnias: Overview and Approach

Richard L. Verrier, PhD
Associate Professor of Medicine
Harvard Medical School
Beth Israel Deaconess Medical Center
Boston, Massachusetts
Cardiovascular Physiology and Coupling with Respiration: Central and Autonomic Regulation
Sleep-Related Cardiac Risk
Cardiac Arryhthmogenesis During Sleep: Mechanisms, Diagnosis, and Therapy

Joris C. Verster, PhD
Doctor of Pharmacology
Utrecht University
Utrecht, Netherlands
Centre for Human Psychopharmacology
Swinburne University
Melbourne, Australia
Effects of Hypnotic Drugs on Driving Performance

Alexandros N. Vgontzas, MD
Professor of Psychiatry
Research Director
Sleep Research and Treatment Center
Penn State College of Medicine
Penn State Milton S. Hershey Medical Center
Hershey, Pennsylvania
Insomnia and Health

Bryan Vila, PhD
Professor
Sleep and Performance Research Center
Washington State University–Spokane
Spokane, Washington
Sleep Problems in First Responders and in Deployed Military Personnel

Martha Hotz Vitaterna, PhD
Research Associate Professor
Center for Sleep and Circadian Biology
Northwestern University
Evanston, Illinois
Genetics and Genomics of Circadian Clocks

James K. Walsh, PhD
Executive Director and Senior Scientist
Sleep Medicine and Research Center
St. Luke's Hospital
St. Louis, Missouri
Pharmacologic Treatment of Insomnia: Benzodiazepine Receptor Agonists

Arthur Scott Walters, MD
Professor of Neurology
Associate Director of Sleep Medicine
Vanderbilt University School of Medicine
Nashville, Tenneessee
Restless Legs Syndrome and Periodic Limb Movements During Sleep

Erin J. Wamsley, PhD
Assistant Professor
Psychology
Furman University
Greenville, South Carolina
Why We Dream

Paula L. Watson, MD
Assistant Professor
Pulmonary, Critical Care, and Sleep Medicine
Vanderbilt University Medical Center
Nashville, Tennessee
Sleep in the Critically Ill Patient

Edward M. Weaver, MD, MPH
Professor
Otolaryngology/Head and Neck Surgery
Co-Director
Sleep Center
University of Washington
Staff Surgeon
Surgery Service
VA Puget Sound Healthcare System
Seattle, Washington
Upper Airway Surgery to Treat Obstructive Sleep-Disordered Breathing

Terri E. Weaver, PhD, RN, FAAN
Professor and Dean
University of Illinois at Chicago
College of Nursing
Chicago, Illinois
Obstructive Sleep Apnea and the Central Nervous System: Neural Adaptive Processes, Cognition, and Performance

Nancy J. Wesensten, PhD
Air Traffic Organization Safety and Technical Training
Safety Services (AJI-15)
Federal Aviation Administration
Washington, D.C.
Introduction: Occupational Sleep Medicine
Sleep Problems in First Responders and in Deployed Military Personnel

Ephraim Winocur, DMD
Senior Lecturer in Orofacial Pain
Oral Rehabilitation
Tel Aviv University
Tel Aviv, Israel
Medical and Device Treatment for Obstructive Sleep Apnea: Alternative, Adjunctive, and Complementary Therapies
Pharmacotherapy, Complementary, and Alternative Medicine for Sleep Bruxism

Amy R. Wolfson, PhD
Professor of Psychology
Vice President for Academic Affairs
Loyola University Maryland
Baltimore, Maryland
Postpartum Period and Early Motherhood

Christine Won, MD, MS
Assistant Professor
Department of Medicine (Pulmonary)
Director, Women's Sleep Health Program
Director, Yale Sleep Center
Yale University School of Medicine
New Haven, Connecticut
Fibromyalgia and Chronic Fatigue Syndromes

Kenneth P. Wright, Jr., PhD
Associate Professor
Integrative Physiology
University of Colorado Boulder
Boulder, Colorado
Shift Work, Shift Work Disorder, and Jet Lag

Lora J. Wu, PhD
Research Officer
Sleep/Wake Research Centre
Massey University
Wellington, New Zealand
Fatigue Risk Management Systems

Mark Wu, MD, PhD
Associate Professor of Neurology, Medicine, and Neuroscience
The Johns Hopkins University School of Medicine
Attending Physician
Sleep Disorders Center
The Johns Hopkins Hospital
Baltimore, Maryland
Genetics and Genomic Basis of Sleep in Simple Model Organisms

Terry Young, PhD
Professor of Population Health Sciences
School of Medicine and Public Health
University of Wisconsin–Madison
Madison, Wisconsin
Systemic and Pulmonary Hypertension in Obstructive Sleep Apnea

Antonio Zadra, PhD
Department of Psychology
Université de Montréal
Montreal, Quebec, Canada
Dream Content: Quantitative Findings

Phyllis C. Zee, MD, PhD
Professor of Neurology, Neurobiology, and Physiology
Ken and Ruth Davee Department of Neurology
Northwestern University Feinberg School of Medicine
Chicago, Illinois
Introduction: Master Circadian Clock and Master Circadian Rhythm
Circadian Dysregulation in Mental and Physical Health
Circadian Disorders of the Sleep-Wake Cycle

Chunbai Zhang, MD, MPH
University of Washington
Valley Medical Center
Renton, Washington
Obstructive Sleep Apnea in the Workplace

Andrey V. Zinchuk, MD
Fellow
Pulmonary, Critical Care, and Sleep Medicine
Yale University School of Medicine
New Haven, Connecticut
Central Sleep Apnea: Diagnosis and Management

Ding Zou, MD, PhD
Center for Sleep and Vigilance Disorders
Department of Internal Medicine and Clinical Nutrition
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden
Pulse Wave Analysis During Sleep

Foreword

Don't Blink!

Perhaps my favorite phrase to parents welcoming home a newborn is: "Don't blink!" In what seems like only a moment in time, parents are suddenly reflecting on how quickly their child has grown: talking, walking, in school, driving, perhaps college, relationships, jobs, their own kids? As this sixth edition of *Principles and Practice of Sleep Medicine* is published, my question to Drs. Kryger, Roth, and Dement is: "Did you blink?"

Who could have imagined that *Principles and Practice of Sleep Medicine* would grow to 21 distinct sections that include 171 chapters? The breadth, depth, and quality represented by the scientific and clinical knowledge in this text are quite amazing. Peruse the range of topics covered across those 21 sections: normal sleep, sleep mechanisms, and phylogeny to why we dream to occupational sleep medicine to the classics of sleep medicine (insomnia, sleep disordered breathing, parasomnias, and narcolepsy). Then consider the depth of knowledge represented: 16 chapters on sleep-disordered breathing, 12 each on physiology in sleep and instrumentation and methodology, and even the "newest" areas, such as genetics and genomic basis of sleep (6 chapters) and legal topics in sleep medicine (5 chapters) have enough content for multiple chapters.

Actually, if anyone could have imagined this textbook growing so quickly and so broadly, it would include Drs. Meir Kryger, Tom Roth, and Bill Dement. *Principles and Practice of Sleep Medicine* has expanded as a reflection of the field, mirroring the incredible advancements in sleep, circadian, and sleep medicine knowledge and practice that have occurred over the past half century. However, the text is more than an invaluable resource and repository of current knowledge; it provides a vision to the future as well.

Sleep medicine, sleep, and chronobiology touch every human at our most basic cellular level (genetics and genomics) and are critical at every level of our society (e.g., occupational, legal). There is an emerging acknowledgment that our safety, health, performance, and mood are fundamentally linked to our sleep, circadian rhythms, and sleep health. While still nascent, this societal recognition grows daily due to the ever-increasing knowledge generated by the sleep medicine community and sleep and circadian scientists. The application of this knowledge and the practice of sleep medicine are creating the foundation for changing societal attitudes and behaviors about sleep, sleep disorders, and circadian factors.

So don't blink; keep your eyes wide open as sleep medicine continues to grow, evolve, and become fully integrated into the safety and health of our society. Just imagine the tenth edition of *Principles and Practice of Sleep Medicine* …

Mark R. Rosekind, PhD
Washington, D.C.

Exciting times

These are exciting times for the field of sleep medicine! The success of any field of medicine is often directly proportional to the scope and comprehensiveness of the knowledge base available to physicians, scientists, trainees, and the general public. For sleep medicine, we are fortunate in that there continues to be dramatic growth in this knowledge, derived from both patient care and clinical/basic research. When one takes a step back and reflects on the rapid development of this field, one reveals how this knowledge base has developed in so short a time. It has been less than 65 years since the discovery of rapid eye movement (REM) sleep, which initiated the organized, scientific study of sleep, and barely 30 years since the invention of continuous positive airway pressure (CPAP), which was the first effective treatment for obstructive sleep apnea. In this short time, the sleep field has expanded to the point where we have over 11,000 accredited member American sleep centers and individual members, including physicians, scientists, and other health care professionals, of the American Academy of Sleep Medicine. Our field has blossomed to the point that it is truly interdisciplinary, comprising specialists from the areas of pulmonary medicine, neurology, psychiatry, internal and family medicine, pediatrics, psychology, otolaryngology, and others. Exciting breakthroughs in sleep research have affected other disciplines of science and research as well, and it is not unusual for sleep medicine specialists to collaborate with other diverse fields of medicine, such as cardiology, endocrinology, genetics, and immunology. Additionally, sleep medicine is practiced worldwide, and a new world sleep organization will be formed in 2017 after the merger of two of our large international sleep organizations (the World Sleep Federation and the World Association of Sleep Medicine).

Despite our amazing growth, there are still many questions yet to be answered, including the holy grail of our field: the function of sleep. To explore these questions, funding from the government, industry, and foundations; support from institutions; and strong mentorship by experienced investigators are important cornerstones. As members of the field, we must collectively strive to ensure that funding, support, and mentorship continue in order to safeguard continued success, even in times of economic downturns and increased competition from other fields. For without breakthroughs in research, there won't be new diagnostic methods, medications, or treatments to help us manage the nearly 90 different sleep disorders that are currently identified.

The growth of our field and the exploration of critical research areas cannot exist without adequate education and training of our young clinicians and investigators to ensure that bright, talented, and dedicated individuals are provided the necessary tools to establish a successful independent clinical and research career. We are indeed privileged that we have excellent resources available that enable trainees to learn more about sleep and sleep medicine. For countless students, *Principles and Practice of Sleep Medicine* has served as the primary textbook, the study guide for the sleep medicine board

certification examination, and/or the basic resource for any sleep-related condition or question about sleep. Often fondly referred to as simply "P&P," it continues to rise in prominence and demand. I've had the great pleasure to learn from and collaborate with Drs. Kryger, Roth, and Dement, and not only are they among the top clinicians and scientists within our field, but they have continued to produce a sleep medicine reference that has remained the gold standard over the span of almost 30 years. Our field is deeply indebted to their dedication, hard work, and diligence.

Clete A. Kushida, MD, PhD, RPSGT
President, World Sleep Federation
Professor, Stanford University Medical Center
Medical Director, Stanford Sleep Medicine Center
Director, Stanford Center for Human Sleep Research
Stanford University, California
clete@stanford.edu

Sixth Edition Preface

It has been about 30 years since we started to work together on the first edition of *Principles and Practice of Sleep Medicine.* The field at the time was in an embryonic stage. We have witnessed the growth of the science and the practice of sleep medicine through its birth, childhood, and adolescence. Sleep is an accepted part of scientific inquiry and the practice of medicine. Almost everyone knows someone who is being treated for a sleep disorder.

This edition continues the overall organization of the very first edition: the first part reviews the principles of sleep medicine, the second part the practice of sleep medicine. If one compares the first to this, the sixth, edition, there have been dramatic improvements that have always been a result of what readers wanted and needed to know: how best to understand the science and to treat their patients. New content areas have been added in subsequent editions. They include genetics, circadian disorders, geriatrics, women's health, cardiovascular diseases, occupational sleep medicine, legal aspects of sleep medicine, and dental sleep medicine. The latter two sections were added in this edition. The volume has gone from being a 722-page book to a volume more than twice that size, with an enormous amount of digital content that is viewable on virtually all connected devices. The spirit that drove the conception of the first edition (see the preface of that edition next) is still in our hearts.

Probably about a thousand authors have contributed to all the editions of this book. As a group, they are brilliant, have a pioneering spirit, and generously shared their knowledge. We cannot thank the section editors enough for all of their magnificent and hard work. The editors have their own unique style and methods of ensuring scientific accuracy and readability. It was an absolute pleasure working with them. The editors were given the authority to make decisions for their sections, and they had the last word on what went into their sections.

In the more than quarter century that this book has existed many authors and section editors retired, and sadly some have died. Many of the contributors started early in their careers. They established themselves and are continuing to lead the field of sleep medicine into the next generation. Some of the authors of this edition were not even born when *Principles and Practice of Sleep Medicine* was first conceived. They will lead the field into the future.

Meir Kryger
Tom Roth
Bill Dement

First Edition Preface

Medical disorders related to sleep are obviously not new. Yet the discipline of sleep disorders medicine is in its infancy. There is a large body of knowledge on which to base the discipline of sleep disorder medicine. We hope that this textbook will play a role in the evolution of this field.

Douglas Hofstadter reviewed how ideas and concepts evolve and are transmitted.[1] In 1965, Roger Sperry[2] wrote the following: "Ideas cause ideas and help evolve new ideas. They interact with each other and with other mental forces in the same brain, in neighboring brains, and thanks to global communication, in far distant, foreign brains. And they also interact with the external surroundings to produce *in toto* a burstwise advance in evolution that is far beyond anything to hit the evolutionary scene yet, including the emergence of the living cell." Jacques Monod[3] wrote the following in *Chance and Necessity:* "For a biologist it is tempting to draw a parallel between the evolution of ideas and that of the biosphere. For while the abstract kingdom stands at a yet greater distance above the biosphere than the latter does above the non-living universe, ideas have retained some of the properties of organisms. Like them they tend to perpetuate their structure and to breed; they too can fuse, recombine, segregate their content; indeed they too can evolve, and in this evolution selection must surely play an important role." Hofstadter has called this universe of ideas the ideosphere analogous to the biosphere. The ideosphere's counterpart to the biosphere gene has been called meme by Richard Dawkins.[4] He wrote "just as genes propagate themselves in a gene pool by leaping from body to body via sperm or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain. … If a scientist hears or reads about a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on it can be said to propagate itself spreading from brain to brain … memes should be regarded as living structures, not just metaphorically but technically."

Thus, this textbook represents an attempt to summarize the body of science and ideas that up to now has been transmitted verbally, in articles, and in a few more specialized books. The memes in this volume are drawn from a variety of disciplines, including psychology, psychiatry, neurology, pharmacology, internal medicine, pediatrics, and basic biological sciences. That a field evolves from multidisciplinary roots certainly has precedents in medicine. The field of infectious diseases has its in microbiology, and its practitioners are expected to know relevant aspects of internal medicine, surgery, gynecology, and pediatrics. Similarly, oncology has its roots in surgery, hematology, and internal medicine, and its practitioners today must also know virology and molecular biology. Patients with sleep problems have in the past 'fallen through the cracks.' It is not uncommon to see a patient with classic narcolepsy who has seen five to ten specialists before a diagnosis is finally made. There is a clinical need for physicians to know about sleep and its disorders.

[1]Hofstadter DR. Chapter 3. In: *Metamagical Themas: Questing for the Essence of Mind and Pattern*. Toronto: Bantam Books; 1986.
[2]Sperry R. Mind, brain, and humanist values. In: Platt JR, editor. *New Views of the Nature of Man*. Chicago: The university of Chicago Press; 1965.
[3]Monod J. *Chance and Necessity*. New York: Vintage Books; 1972.
[4]Dawkins R. *The Selfish Gene*. Oxford: Oxford University Press; 1976. p. 206.

Acknowledgments

We have been working on *Principles and Practice of Sleep Medicine* for over a quarter of a century. Thousands of people have been involved in the production of the six editions. As much as we would like to thank each person, there is no way that we can thank them all. Some have retired, some have died, and some made important contributions in the production of the various editions but are unknown to us. This group includes secretaries, copyeditors, artists, designers, people who dealt with the page proofs, internet programmers, and those who physically produced the books.

We would like to acknowledge all the extraordinary Elsevier editors who gave birth to each previous edition of the book. These include Bill Lamsback, Judy Fletcher, Richard Zorab, Cathy Carroll, Todd Hummell, and Dolores Meloni. They fueled the dream that helped establish a new field of medicine.

Many people helped in the preparation of the content of this volume, the sixth edition, including those listed below.

The staff members at Elsevier who helped this book in its sixth journey were Helene Caprari, Laura Kuehl-Schmidt, Amanda Mincher, and many others involved in production and design for both the printed volume and the online content.

We also must acknowledge the family members of all the people involved in the book because they indirectly helped produce a work that we believe may have had important positive impact on the lives of thousands, perhaps millions, of people.

Finally, we wish to thank the many hundreds of authors and the magnificent work of the section editors and their deputy editors. All their contributions were so great that they cannot be measured.

Section and Deputy Editors

1E 1989

Mary Carskadon
Michael Chase
Richard Ferber
Christian Guilleminault
Ernest Hartmann
Meir Kryger
Timothy Monk
Anthony Nicholson
Allan Rechtschaffen
Gerald Vogel
Frank Zorick

2E 1994

Michael Aldrich
Mary Carskadon
Michael Chase
J. Christian Gillin
Christian Guilleminault
Ernest Hartmann
Meir Kryger
Anthony Nicholson
Allan Rechtschaffen
Gary Richardson
Thomas Roth
Frank Zorick

3E 2000

Michael Aldrich
Michael Chase
J. Christian Gillin
Christian Guilleminault
Max Hirshkowitz
Mark W. Mahowald
Wallace B. Mendelson
R.T. Pivik
Leon Rosenthal
Mark Sanders
Fred Turek
Frank Zorick

4E 2005

Michael Aldrich
Ruth Benca
J. Christian Gillin
Max Hirshkowitz
Shahrokh Javaheri
Meir Kryger
Mark W. Mahowald
Wallace B. Mendelson
Jacques Montplaiser
John Orem
Timothy Roehrs
Mark Sanders
Robert Stickgold
Fred Turek

5E 2011

Sonia Ancoli-Israel
Gregory Belenky
Ruth Benca
Daniel Buysse
Michael Cramer-Bornemann
Charles George
Max Hirshkowitz
Meir Kryger
Gilles Lavigne
Kathryn Aldrich Lee
Beth A. Malow
Mark W. Mahowald
Wallace B. Mendelson
Jacques Montplaisir
Tore Nielsen
Mark Sanders
Jerome Siegel
Fred Turek

6E 2017

Sonia Ancoli-Israel
Robert Basner
Gregory Belenky
Dan Brown
Daniel Buysse
Jennifer DeWolfe
Max Hirshkowitz
Shahrokh Javaheri
Andrew Krystal
Gilles Lavigne
Kathryn Aldrich Lee
Beth A. Malow
Timothy Roehrs
Thomas Roth
Thomas Scammell
Jerome Siegel
Robert Stickgold
Katie L. Stone
Fred Turek
Bradley V. Vaughn
Erin J. Wamsley
Christine Won

Contents

Video Contents

Abbreviations

AASM: American Academy of Sleep Medicine
ACC: anterior cingulate cortex
Ach: acetylcholine
ACTH: adrenocorticotropic hormone
AD-ACL: Activation-Deactivation Adjective Check List
ADHD: attention-deficit/hyperactivity disorder
AHI: apnea-hypopnea index
AIM: ancestry informative marker
AMPA: α-amino-3hydroxy-5-methylisozazole-4-propionic acid
AMPK: adenosine-monophosphate-activated protein kinase
AMS: acute mountain sickness
ANS: autonomic nervous system
ApoE: apoliprotein E; ApoE-ε4
ASPS: advanced sleep phase syndrome
ASPT: advanced sleep phase type
AVAPS: average volume assured pressure support
AW: active wakefulness
BA: Brodman area
BAC: blood alcohol content
BCOPS: Buffalo Cardio-Metabolic Occupational Police Stress
BD: bipolar disorder
BF: basal forebrain
BMAL1: brain and muscle ARNT-like
BMI: body mass index
BNST: bed nucleus of the stria terminalis
BPD: biliopancreatic diversion
BPDDS: biliopancreatic diversion with duodenal switch
BzRA: benzodiazepine receptor agonist
CAD: coronary artery disease
CAPS: cyclic alternating pattern sequence(s)
CBT: cognitive behavior therapy
CBT-I: cognitive behavior therapy for insomnia
CHF: congestive heart failure
CI: confidence interval
CPS/HHPRI: Calgary Police Service Health and Human Performance Research Initiative
COMT: catechol-*O*-methyltransferase
COPD: chronic obstructive pulmonary disease
CPAP: continuous positive airway pressure
CRP: C-reactive protein
CRY: cryptochrome
CSN: cold-sensitive neuron
CYP: cytochrome P-450
DA: dopamine
DAT: dopamine transporter
DBP: D-element binding protein
DD: constant dark
DIM: digital integration mode
DLMO: dim-light melatonin onset
DLPFC: dorsolateral prefrontal cortex
DMD: Duchenne's muscular dystrophy
DSISD: Duke Structured Interview for Sleep Disorders
DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition
DSPS: delayed sleep phase syndrome
DSPT: delayed sleep phase type
DTs: delirium tremens
DU: duodenal ulcer
ECG: electrocardiogram, electrocardiographic
EDS: excessive daytime sleepiness
EEG: electroencephalogram, electroencephalographic
EMG: electromyogram
ENS: enteric nervous system
EOG: electrooculogram
EPS: extrapyramidal side effects
EPSP: excitatory postsynaptic potential
ERP: event-related potential
ESS: Epworth Sleepiness Scale
FAID: Fatigue Audit InterDyne
^{18}FDG: 2-deoxy-2-[^{18}F]fluoro-d-glucose
F-DOPA: 6-[^{18}F]fluoro-l-dopa
FEV_1: forced expiratory volume in 1 second
FFT: fast Fourier transform
FIRST: Ford Insomnia Response to Stress Test
fMRI: functional magnetic resonance imaging
FOQA: flight operations quality assurance
FOSQ: Functional Outcomes of Sleep Questionnaire
FRA: Federal Railroad Administration
FRC: functional residual capacity
FSIVGTT: frequently sampled intravenous glucose tolerance test
GABA: gamma-aminobutyric acid
GAD: generalized anxiety disorder
GAHMS: genioglossus advancement, hyoid myotomy, and suspension
GCD: global cessation of dreaming
GER: gastroesophageal reflux
GHB: gamma-hydroxybutyrate
GHRH: growth hormone-releasing hormone
GWA: genome wide association
5-HIAA: 5-hydroxyindole acetic acid
5-HT: hydroxytryptamine (serotonin)
HAPE: high-altitude pulmonary edema
Hcrt: hypocretin
HDI: hypnotic-dependent insomnia
HDL: high density lipoprotein
HIF: hypoxia inducible factor
HIV: human immunodeficiency virus
HLA: human leukocyte antigen
HOMA: homeostasis model assessment
HPA: hypothalamic-pituitary-adrenal axis
HRV: heart rate variability
HVA: homovanillic acid
HWHSGPS: Harvard Work Hours and Safety Group Police Study
IAPT: Improving Access to Psychological Therapies (program)
ICD: International Classification of Diseases
ICD-9-CM: *International Classification of Diseases*, ninth revision, Clinical Modification
ICD-10: *International Classification of Diseases*, tenth revision
ICSD3: *International Classification of Sleep Disorders*, third edition
ICV: intracerebroventricular

IEG: immediate early gene
IGL: intergeniculate leaflet
IL: interleukin
ILD: interstitial lung disease
IPSP: inhibitory postsynaptic potential
IRLS: International Restless Legs Scale
ISI: Insomnia Severity Index
iVAPS: intelligent volume assured pressure support
kd: kilodalton
KSS: Karolinska Sleepiness Scale
LAUP: laser-assisted uvulopalatoplasty
LD: light-dark
LDL: low density lipoprotein
l-dopa: l-dihydroxyphenylalanine, levodopa
LG: lateral geniculate
LL: constant light
LOC: left outer canthus
LPA (or LPOA): lateral preoptic area
LSAT: lowest oxyhemoglobin saturation
LTIH: long-term intermittent hypoxia
MAO: monoamine oxidase
MAOI: monamine oxidase inhibitor
MCTQ: Munich Chronotype Questionnaire
MDA: methylenedioxyamphetamine
MDD: major depressive disorder
MDMA: methylenedioxymethamphetamine ("ecstasy")
MDP-LD: muramyl dipeptide *N*-actyl-muramyl-l-alanyl-d-isoglutamine
MEG: magnetoencephalography
MEQ: Morningness-Eveningness Questionnaire
MI: myocardial infarction
MMC: migrating motor complex
MMO: maxillary and mandibular osteotomy
MMSE: Mini-Mental State Examination
MnPN: median preoptic nucleus
MNSA: muscle nerve sympathetic vasomotor activity
MPA (or MPOA): medial preoptic area
MPA: medroxyprogesterone acetate
mPFC: medial prefrontal cortex
MRA: mandibular repositioning appliance
MSA: multiple system atrophy
MSF: midpoint of sleep on free days
MSLT: Multiple Sleep Latency Test
MWT: Maintenance of Wakefulness Test
NAD: nicotinamide adenine nucleotide
NAMPT: nicotinamide phosphoribosyltransferase
NASH: nonalcoholic steatohepatitis
NCEP: National Cholesterol Education Program
NCSDR: National Center on Sleep Disorders Research
NE: norepinephrine
NET: norepinephrine transporter
NFLD: nonalcoholic fatty liver disease
NFLE: nocturnal frontal lobe epilepsy
NFκB: nuclear factor kappa B
NHANES: National Health and Nutrition Examination Survey
NIH: National Institutes of Health
NIPPV: nasal intermittent positive-pressure ventilation
NK: natural killer (cell)
NMDA: *N*-methyl-d-aspartate
NO: nitric oxide
NPPV: noninvasive positive-pressure ventilation
NPT: nocturnal penile tumescence
NREM: non–rapid eye movement, non-REM
OCD: obsessive-compulsive disorder
OFC: orbitofrontal cortex
6-OHDA: 6-hydroxydopamine
OHS: obesity-hypoventilation syndrome
OR: odds ratio
OSA: obstructive sleep apnea
OSAHS: obstructive sleep apnea–hypopnea syndrome
OSAS: obstructive sleep apnea syndrome
PACU: postanesthesia care unit
PCOS: polycystic ovary syndrome
PEEP: positive end-expiratory pressure
PER: period
PET: positron emission tomography
PGO: ponto-geniculo-occipital (spike)
PIA: pontine inhibitory area
PLMS (or PLM): periodic limb movements during sleep
PMDD: premenstrual dysphoric disorder
PNI: people not having insomnia
POA: preoptic area
POMS: Profile of Mood States
POSSR: Patrol Officers Shift Schedule Review
PR: prevalence ratio
PRC: phase–response curve
PSG: polysomnography, polysomnographic
PSQI: Pittsburgh Sleep Quality Index
PTSD: posttraumatic stress disorder
PVN: paraventricular nucleus
PVT: psychomotor vigilance test
PWI: people with insomnia
PWOP: people who did not report having the medical problem
PWP: people who reported have the medical problem
QTL: quantitative trait loci (or locus)
QW: quiet wakefulness
RBD: REM sleep behavior disorder
RDC: research diagnostic criteria
RDI: respiratory disturbance index
REM: rapid eye movement
RERA: respiratory effort related arousal
RFA: radiofrequency ablation
RHT: retinohypothalamic tract
RI: recombinant inbred
R_{in}: membrane input resistance
RIP: respiratory inductive plethysmography
RLS: restless legs syndrome
RMMA: rhythmic masticatory motor activity
ROC: right outer canthus
ROS: reactive oxygen species
RR: risk ratio
RSWA: REM sleep without atonia
RT: reaction time
RYGB: Roux-en-Y gastric bypass
SAFTE: Sleep, Activity, Fatigue, and Task Effectiveness (model)
SCD: stearoyl coenzyme A desaturase
SCID: Structured Clinical Interview for Diagnosis
SCN: suprachiasmatic nucleus
SCT: sleep compression therapy
SDB: sleep-disordered breathing
SE%: sleep efficiency percentage

SEMs: small eye movements
SIDS: sudden infant death syndrome
SIT: suggested immobilization test
SND: synucleinopathic disorders
SNP: single nucleotide polymorphism
SOL: sleep-onset latency
SOREM: sleep-onset REM
SOREMP: sleep-onset REM period
SP: sleep paralysis
SPM: statistical parametric mapping
SRE: sleep-related erection
SREBP: sterol regulatory element binding protein
SRED: sleep-related eating disorder
SRT: sleep restriction therapy
SSEP: somatosensory evoked potential
SSS: Stanford Sleepiness Scale
SSRI: selective serotonin reuptake inhibitor
STREAM: supra-threshold REM EMG activity metric
SWA: slow wave activity
SWAI: Sleep-Wake Activity Inventory
SWD: shift work disorder, shift work sleep disorder
SWS: slow wave sleep
T_a: ambient temperature
TAT: time above threshold
TCA: tricyclic antidepressant
tDCS: transcranial direct current stimulation
THH: terrifying hypnagogic hallucination
TIB: total time in bed
TLR: Toll-like receptor
TMJ: temporomandibular joint
TNF: tumor necrosis factor
TRD: tongue-retaining device
TST: total sleep time
UARS: upper airway resistance syndrome
UNS: Ullanlinna Narcolepsy Scale
UPF: uvulopalatal flap
UPPP: uvulopalatopharyngoplasty
V-EEG-PSG: video-electroencephalography—PSG
VIP: vasoactive intestinal peptide
VLDL: very low density lipoprotein
VLPO: ventrolateral POA (preoptic area)
VMAT2: vascular monoamine transporter-2
VTA: ventral tegmental area
WASO: wake after sleep onset
WSN: warm-sensitive neuron
ZCM: zero crossing mode

Continuing Medical Education (CME) and Maintenance of Certification (MOC) for PPSM, Sixth Edition

Atlanta Progressive CME has developed an online activity that is eligible for CME based on the *Principles and Practice of Sleep Medicine* (PPSM), sixth edition. In addition, this activity may be eligible for MOC toward recertification in sleep medicine through the American Board of Internal Medicine. Please visit http://www.sleepschool.com/ppsm6 for full details.

PART I

Principles of Sleep Medicine

7 Psychobiology and Dreaming

Section 1

Normal Sleep and Its Variants

Chapter 1

History of Sleep Physiology and Medicine

Rafael Pelayo; William C. Dement

Chapter Highlights

- Interest in sleep and dreams has probably existed since the dawn of humanity. Some of history's greatest figures have attempted to explain the physiologic and psychological bases of sleep and dreaming.
- The modern scientific study of sleep began with the discovery of the electrical activity in the brain. Further progress was marked by the discovery of and distinction between REM and NREM sleep. Identifying sleep pathology eventually led to the creation of sleep clinics.
- Sleep medicine as a medical specialty has existed for fewer than 50 years. The evolution of the field required clinical research, development of clinical services, training programs, and changes in the insurance industry and public policy that recognized the impact of sleep disorders on society.
- The field is still evolving as new disorders are being discovered, new treatments are being delivered, and basic science helps elucidate the complexity of sleep and its disorders. As sleep medicine faces new challenges, an understanding of its history can provide researchers with important insights for shaping the future of this discipline.

SLEEP AS A PASSIVE STATE

> Sleep is the intermediate state between wakefulness and death; wakefulness being regarded as the active state of all the animal and intellectual functions, and death as that of their total suspension.[1]

The foregoing is the first sentence of *The Philosophy of Sleep*, a book by Robert MacNish, a member of the faculty of physicians and surgeons of Glasgow; the first American edition was published in 1834 and the Scottish edition somewhat earlier. This sentence exemplifies the overarching historical conceptual dichotomy of sleep research and sleep medicine, which is sleep as a passive process versus sleep as an active process. Until the discovery of rapid eye movements and the duality of sleep, sleep was universally regarded as an inactive state of the brain. With one or two exceptions, most thinkers regarded sleep as the inevitable result of reduced sensory input, with the consequent diminishment of brain activity and the onset of sleep. Waking up and being awake were considered a reversal of this process, mainly as a result of bombardment of the brain by environmental stimuli. No real distinction was seen between sleep and other states of quiescence such as coma, stupor, intoxication, hypnosis, anesthesia, and hibernation.

The passive-versus-active historical dichotomy also is given great weight by the contempory investigator J. Allan Hobson.[2] As he noted in his book *Sleep*, published in 1989, "more has been learned about sleep in the past 60 years than

in the preceding 6,000." He went on, "In this short period of time, researchers have discovered that sleep is a dynamic behavior. Not simply the absence of waking, sleep is a special activity of the brain, controlled by elaborate and precise mechanisms."[2]

Dreams and dreaming were regarded as transient, fleeting interruptions of this quiescent sleep state. Because dreams seem to occur spontaneously and sometimes in response to environmental stimulation (e.g., the well-known alarm clock dreams), the notion of a stimulus that produces the dream was generalized by postulating internal stimulation from the digestive tract or some other internal source. Some anthropologists have suggested that notions of spirituality and the soul arose from primitive peoples' need to explain how their essence could leave the body temporarily at night in a dream and permanently at death.[3,4] How else to better explain seeing deceased loved ones in a dream than to imagine a spirit world and an afterlife? There should be no doubt that dreams influenced primitive cultures.

Sleep-promoting and sleep-inhibiting substances were part of ancient pharmacopeias. It had been observed in antiquity that alcohol would induce a sleeplike state. More than 5000 years ago the opium poppy was cultivated in Mesopotamia. Hippocrates in the 4th century BCE acknowledged its usefulness as a narcotic. Somewhat later, in Ethiopia, coffee consumption was thought to have begun when its power to prevent sleep was recognized. Coffee was historically associated with Sufism in Yemen, and it may have been used in religious activities. It was cultivated in the Arabian Peninsula in the 15th century, whence it spread to Europe and later the Americas.

In addition to the mere reduction of stimulation, a host of less popular theories were espoused to account for the onset of sleep. Vascular theories were proposed from the notion that the blood left the brain to accumulate in the digestive tract, and from the opposite idea that sleep was due to pressure on the brain by blood. Around the end of the 19th century, various versions of a "hypnotoxin" hypothesis were formulated in which fatigue products (toxins and the like) were accumulated during the day, finally causing sleep, during which they were gradually eliminated. This was an early mirror of current concepts on the role of adenosine accumulation leading to sleepiness.

The hypnotoxin theory reached its zenith in 1907, when the French physiologists Legendre and Pieron showed that blood serum from sleep-deprived dogs could induce sleep in other dogs that were not sleep-deprived.[5] The notion of a toxin causing the brain to sleep has gradually given way to the recognition that a number of endogenous "sleep factors" actively induce sleep by specific mechanisms.

In the 1920s, the University of Chicago physiologist Nathaniel Kleitman carried out a series of sleep deprivation studies and made the simple but brilliant observation that people who stayed up all night generally were less sleepy and impaired the next morning than in the middle of their sleepless night. Kleitman argued that this observation was incompatible with the notion of a continual buildup of a hypnotoxin in the brain or blood. In addition, he suggested that humans were about as impaired as they would get, that is, very impaired, after approximately 60 hours of wakefulness, and that longer periods of sleep deprivation would produce little additional change. In the 1939 (first) edition of his comprehensive landmark monograph *Sleep and Wakefulness*, Kleitman summarized his thinking as follows:

> It is perhaps not sleep that needs to be explained, but wakefulness, and indeed, there may be different kinds of wakefulness at different stages of phylogenetic and ontogenetic development. In spite of sleep being frequently designated as an instinct, or global reaction, an actively initiated process, by excitation or inhibition of cortical or subcortical structures, there is not a single fact about sleep that cannot be equally well interpreted as a let down of the waking activity.[6]

This statement succinctly provides insight into the historical adoption of the yin-yang symbol, ☯, as a symbol of sleep medicine.

THE ELECTRICAL ACTIVITY OF THE BRAIN

As the 20th century got under way, Camillo Golgi and Santiago Ramón y Cajal had demonstrated that the nervous system was not a mass of fused cells sharing a common cytoplasm but rather a highly intricate network of discrete cells that had the key property of signaling to one another. Luigi Galvani had discovered that the nerve cells of animals produce electricity, and Emil duBois-Reymond and Hermann von Helmholtz found that nerve cells use their electrical capabilities for signaling information to one another. In 1875, the Scottish physiologist Richard Caton demonstrated electrical rhythms in the brains of chickens. (In view of present-day concerns about the ethics of animal research, it bears mention that the key tool used today in neuroscience to monitor sleep both clinically and for research in humans was first demonstrated in such a model.) The centennial of his achievement was commemorated at the 15th annual meeting of the Association for the Psychophysiological Study of Sleep convening at the site of the discovery, Edinburgh.

It was not until 1928, however, when the German psychiatrist Hans Berger recorded electrical activity of the human brain and clearly demonstrated differences in these rhythms when subjects were awake versus asleep that a real scientific interest commenced.[7] Berger correctly inferred that the signals he recorded, which he called "electroencephalograms," were of brain origin. For the first time, the presence of sleep could be conclusively established without disturbing the sleeper, and more important, sleep could be continuously and quantitatively measured without disturbing the sleeper.

All of the classic major elements of sleep brain wave patterns were described by Loomis, Harvey, Hobart, Davis, and others at Harvard University in a series of influential papers published in 1937, 1938, and 1939.[8-10] Alfred Lee Loomis is a historically interesting figure who played a pivotal role in World War II. He developed amplifier systems to record sleep, and for reasons that are seemingly lost to history, he coined the term *K-complex*.[11] Blake, Gerard, and Kleitman added to this work from their studies at the University of Chicago. On the human electroencephalogram (EEG), sleep was characterized by high-amplitude slow waves and spindles, whereas wakefulness was characterized by low-amplitude waves and alpha rhythm.[12,13] The image of the sleeping brain completely "turned off" gave way to the image of the sleeping brain engaged in slow, synchronized, "idling" neuronal activity. Although their significance was not widely recognized at the time, these findings constituted some of the most critical developments in sleep research. Indeed, Hobson dated the

turning point of sleep research to 1928, when Berger began his work on the human EEG.[2] Used today in much the same way as they were in the 1930s, brain wave recordings with paper and ink, or more recently on computer screens, have been extraordinarily important to sleep research and sleep medicine.

Also in the 1930s, a series of investigations by Frederick Bremer seemed to establish conclusively both the passive theory of sleep and the notion that it occurred in response to reduction of stimulation and activity.[14,15] These studies were made possible by the aforementioned development of electroencephalography. Bremer studied brain wave patterns in two cat preparations. One, which Bremer called "encéphale isolé," was made by cutting a section through the lower part of the medulla. The other, "cerveau isolé," was made by cutting the midbrain just behind the origin of the oculomotor nerves. The first preparation permitted the study of cortical electrical rhythms under the influence of olfactory, visual, auditory, vestibular, and musculocutaneous impulses; in the second preparation, the field was narrowed almost entirely to the influence of olfactory and visual impulses.

In the first preparation, the brain continued to show manifestations of wakeful activity alternating with phases of sleep, as indicated by the EEG. In the second preparation, however, the EEG pattern assumed a definite deep sleep character and remained in this condition. In addition, the eyeballs immediately turned downward, with a progressive miosis. Bremer concluded that a functional (reversible, of course) deafferentation of the cerebral cortex occurs in sleep. The cerveau isolé preparation results in a suppression of the incessant influx of nerve impulses, particularly cutaneous and proprioceptive, which are essential for the maintenance of the waking state of the telencephalon. Apparently, olfactory and visual impulses are insufficient to keep the cortex awake. It probably is misleading to assert that physiologists assumed the brain was completely turned off, whatever this metaphor might have meant, because blood flow and, presumably, metabolism continued. However, Bremer and others certainly favored the concept of sleep as a reduction of activity—idling, slow, synchronized, "resting" neuronal activity.

THE RETICULAR ACTIVATING SYSTEM

After World War II, insulated, implantable electrodes were developed, and sleep research on animals began in earnest. In 1949, one of the most important and influential studies dealing with sleep and wakefulness was published: Moruzzi and Magoun's classic paper "Brain Stem Reticular Formation and Activation of the EEG."[16] These authors concluded that

> transitions from sleep to wakefulness or from the less extreme states of relaxation and drowsiness to alertness and attention are all characterized by an apparent breaking up of the synchronization of discharge of the elements of the cerebral cortex, an alteration marked in the EEG by the replacement of high voltage, slow waves with low-voltage fast activity.[16]

High-frequency electrical stimulation with electrodes implanted in the brainstem reticular formation produced EEG activation and behavioral arousal. These findings seemed to indicate that EEG activation, wakefulness, and consciousness were at one end of a continuum, and EEG synchronization, sleep, and lack of consciousness were at the other end. This view, as can be seen, is hardly different from that in MacNish's definition quoted at the beginning of this chapter.

The demonstration by Starzl and coworkers that sensory collaterals discharge into the reticular formation suggested that a mechanism was present by which sensory stimulation could be transduced into prolonged activation of the brain and sustained wakefulness.[17] By attributing an amplifying and maintaining role to the brainstem core and the conceptual ascending reticular activating system, it was possible to account for the fact that wakefulness outlasts, or occasionally is maintained in the absence of, sensory stimulation.

Chronic lesions in the brainstem reticular formation produced persisting slow waves in the EEG and immobility. The usual animal for this research was the cat, because excellent stereotaxic coordinates of brain structures had become available in this model.[18] These findings appeared to confirm and extend Bremer's observations. The theory of the reticular activating system was an anatomically based passive theory of sleep or an active theory of wakefulness. Figure 1-1 is from the proceedings of a symposium, *Brain Mechanisms and Consciousness,* which was published in 1954 and probably (other than arguably Freud's works) is the first genuine neuroscience bestseller.[19] Horace Magoun had extended his studies to the monkey, and this illustration represents the full flowering of the ascending reticular activating system theory.

EARLY OBSERVATIONS OF SLEEP PATHOLOGY

Insomnia has been described since the dawn of recorded history and attributed to many causes, including a recognition of the association between emotional disturbance and sleep disturbance. Scholars and historians have a duty to bestow credit accurately. Many discoveries, however, lie fallow for want of a contextual soil in which they may be properly understood and in which they may extend the understanding of more general phenomena. Important early observations were those of von Economo on "sleeping sickness" and of

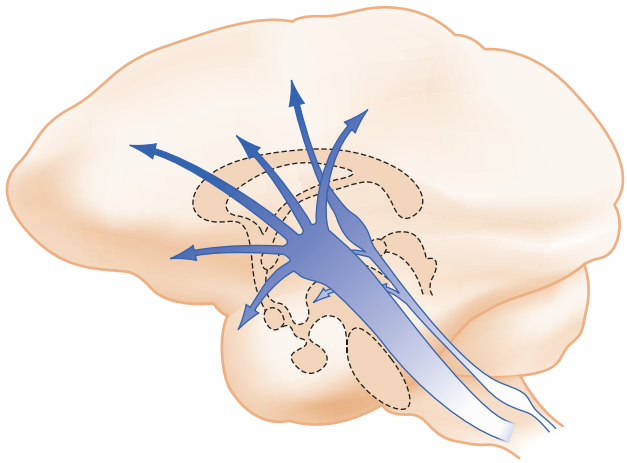


Figure 1-1 Lateral view of the monkey's brain, showing the ascending reticular activating system in the brainstem receiving collaterals from direct afferent paths and projecting primarily to the associational areas of the hemisphere. (Redrawn from Magoun HW: The ascending reticular system and wakefulness. In: Adrian ED, Bremer F, Jasper HH, editors. *Brain mechanisms and consciousness. A symposium organized by the Council for International Organizations of Medical Sciences,* 1954. Courtesy Charles C Thomas, Publisher, Springfield, Illinois.)

Pavlov, who observed dogs falling asleep during conditioned reflex experiments.[3]

Two early observations about sleep research and sleep medicine stand out. The first is the description in 1880 of narcolepsy by Jean Baptiste Edouard Gélineau, who derived the term from the Greek words *narkosis* ("a benumbing") and *lepsis* ("to overtake"). He was the first to clearly describe the collection of components that constitute the syndrome, although the term *cataplexy* for the emotionally induced muscle weakness was subsequently coined in 1916 by Richard Henneberg.

Obstructive sleep apnea syndrome (OSAS), which may be called the leading sleep disorder of the 20th century, was famously described, in 1836, not by a clinician but by the novelist Charles Dickens. In a series of papers entitled the "Posthumous Papers of the Pickwick Club," Dickens described Joe, a boy who was obese and always excessively sleepy. Joe, a loud snorer, was called "young dropsy," possibly as a result of having right-sided heart failure. Of note, Joe is praised for his ability to fall asleep instantaneously after drinking alcohol! Meir Kryger and Peretz Lavie published scholarly accounts of many early references to snoring and conditions that were most certainly manifestations of OSAS.[20-22] Professor Pierre Passouant provided an account of the life of Gélineau and his landmark description of the narcolepsy syndrome.[23]

SIGMUND FREUD AND THE INTERPRETATION OF DREAMS

By far the most widespread interest in sleep by health professionals and the general public was engendered by the theories of Sigmund Freud, specifically about dreams.[24] *The Interpretation of Dreams* was first published in German in 1895 and translated into English in 1913, with several subsequent revisions.[24] Of course, the real interest was in dreaming, with sleep as a necessary concomitant. Freud developed psychoanalysis, the technique of dream interpretation, as part of his therapeutic approach to emotional and mental problems. As the concept of the ascending reticular activating system dominated behavioral neurophysiology, so the psychoanalytic theories about dreams dominated the psychological side of the coin. Dreams were thought to be the guardians of sleep and to occur in response to a disturbance, to obviate waking up, as exemplified in the classic alarm clock dream. Freud's concept that dreaming discharged instinctual energy led directly to the notion of dreaming as a safety valve of the mind. At the time of the discovery of rapid eye movements during sleep (circa 1952), academic psychiatry was dominated by psychoanalysts, and medical students all over America were interpreting one another's dreams.

From today's vantage point, the dream deprivation studies of the early 1960s, engendered and reified by the belief in psychoanalysis, may be regarded by some as a digression from the mainstream of sleep medicine. On the other hand, because the medical-psychiatric establishment had begun to take dreams seriously, it also was ready to support sleep research fairly generously under the guise of dream research.

CHRONOBIOLOGY

Most, but not all, sleep specialists share the opinion that what has been called *chronobiology,* or the study of biologic rhythms, is a legitimate part of sleep research and sleep medicine. The 24-hour rhythms in the activities of plants and animals have been recognized for centuries. These biologic rhythms were quite reasonably assumed to be a direct consequence of the periodic environmental fluctuation of light and darkness. However, in 1729, Jean Jacques d'Ortous de Mairan described an experiment in which a heliotrope plant opened its leaves during the day even after it had been moved so that sunlight could not reach it. The plant opened its leaves during the day and folded them for the entire night even though the environment was constant. This was the first demonstration of the persistence of circadian rhythms in the absence of environmental time cues. Figure 1-2, which represents de Mairan's original experiment, is reproduced from *The Clocks That Time Us,* by Moore-Ede and colleagues.[25]

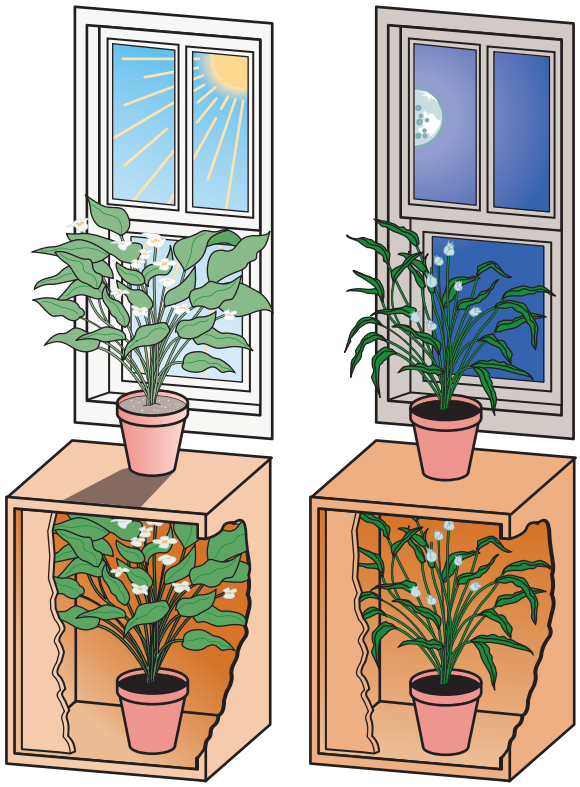


Figure 1-2 Representation of de Mairan's original experiment. When exposed to sunlight during the day (*upper left*), the leaves of the plant were open; during the night (*upper right*), the leaves were folded. De Mairan showed that sunlight was not necessary for these leaf movements by placing the plant in total darkness. Even under these constant conditions, the leaves opened during the day (*lower left*) and folded during the night (*lower right*). (Redrawn from Moore-Ede MC, Sulzman FM, Fuller CA. *The clocks that time us: physiology of the circadian timing system.* Cambridge [Mass.]: Harvard University Press; 1982. p. 7.)

Chronobiology and sleep research developed separately. Three factors appear to have contributed to this divergence:

1. The long-term studies commonly used in biologic rhythm research precluded continuous recording of brain wave activity. Certainly, in the early days, the latter was far too difficult and not really necessary. The measurement of wheel-running activity was a convenient and widely used method for demonstrating circadian rhythmicity.
2. The favorite animal of sleep research from the 1930s through the 1970s was the cat, and neither cats nor dogs demonstrate clearly defined circadian activity rhythms.

Figure 1-3 Nathaniel Kleitman (circa 1938), Professor of Physiology, University of Chicago School of Medicine.

3. The separation between chronobiology and sleep research was further maintained by the tendency for chronobiologists to know very little about sleep, and for sleep researchers to remain ignorant of such biologic clock mysteries as phase response curves, entrainment, and internal desynchronization.

THE DISCOVERY OF RAPID EYE MOVEMENT SLEEP

The characterization of rapid eye movement (REM) sleep as a discrete organismic state should be distinguished from the recognition that rapid eye movements occur during sleep. The historical threads of the discovery of rapid eye movements can be identified. Nathaniel Kleitman (Figure 1-3; Video 1-1), a professor of physiology at the University of Chicago, had long been interested in cycles of activity and inactivity in infants and in the possibility that this cycle ensured that the infant would have an opportunity to respond to hunger. He postulated that the times infants awakened to nurse on a self-demand schedule would be integral multiples of a basic rest-activity cycle. The second historical thread was Kleitman's interest in eye motility as a possible measure of "depth" of sleep. The reasoning behind this potential application was that eye movements had a much greater cortical representation than that of almost any other observable motor activity, and that slow, rolling, or pendular eye movements had been described at the onset of sleep, with a gradual slowing and disappearance as sleep deepened.[26]

In 1951, Kleitman assigned the task of observing eye movement to a graduate student in physiology named Eugene Aserinsky. Watching the closed eyes of sleeping infants was tedious, and Aserinsky soon found that it was easier to designate successive 5-minute epochs as "periods of motility" if he observed any movement at all, usually a writhing or twitching of the eyelids, versus "periods of no motility." Among the infants studied was his own child. In 1952, William C. Dement, at the time a second-year medical student at the University of Chicago, joined the research effort. The first task he was assigned was looking at the closed eyes of the research subjects, using a flashlight in the dark when electrical potentials were detected in the recording instruments in the adjacent room.

After describing an apparent rhythm in eye motility, Kleitman and Aserinsky decided to look for a similar phenomenon in adults. Again, watching the eyes during the day was tedious, and at night it was even worse. Casting about, they came upon the method of electrooculography and decided (correctly) that this would be a good way to measure eye motility continuously and would relieve the researcher of the tedium of direct observations. Sometimes in the course of recording electrooculograms (EOGs) during sleep, they saw bursts of electrical potential changes that were quite different from the slow movements at sleep onset.

When they were observing infants, Aserinsky and Kleitman had not differentiated between slow and rapid eye movements. On the EOG, however, the difference between the slow eye movements at sleep onset and the newly discovered rapid motility was obvious. Initially, there was a great deal of concern that these potentials were electrical artifacts. With their presence on the EOG as a signal, however, it was possible to watch the subject's eyes simultaneously, permitting easy detection of the distinct rapid movements of the eyes beneath closed lids.

At this point, Aserinsky and Kleitman made two assumptions:

1. These eye movements represented a "lightening" of sleep.
2. Because they were associated with irregular respiration and accelerated heart rate, they might represent dreaming.

The basic sleep cycle was not yet identified at this time, primarily because the EOG and other physiologic measures, notably the EEG, were not recorded continuously but rather were "sampled" during a few minutes of each hour or half-hour. The sampling strategy was designed to conserve paper (in the absence of research grants!); morever, no clear reason to record continuously had been identified. This schedule also made it possible for the researcher to nap between sampling episodes.

Aserinsky and Kleitman initiated a small series of awakenings, both when rapid eye movements were present and when they were not, for the purpose of eliciting dream recall. These workers did not apply sophisticated methods of dream content analysis, but the descriptions of dream content from the two conditions generally were quite different, with awakenings during periods of rapid eye movements often yielding vivid complex stories, in contrast with awakening periods, when rapid eye movements were not present, yielding nothing at all or very sparse accounts. This distinction made it possible to hypothesize that rapid eye movements were associated with dreaming. This was, indeed, a breakthrough in sleep research.[27,28] Although Dement participated in this research as a medical student, he was not credited in these early articles. His recollection is that he later coined the abbreviations REM and NREM to simplify the typing of subsequent manuscripts and publications (Dement, personal communication, 2014). These terms appear for the first time in the literature in a footnote by Dement and Kleitman in 1957.[29]

The occurrence of the eye movements was quite compatible with the contemporary dream theories that dreams occurred when sleep lightened, to prevent or delay awakening. In other words, dreaming could still be regarded as the "guardian" of sleep. It could no longer be assumed, however, that dreams were fleeting and evanescent. This recognition put an end to the concept that sleep was a passive state.

ALL-NIGHT SLEEP RECORDINGS AND THE BASIC SLEEP CYCLE

The seminal paper by Aserinsky and Kleitman, published in 1953,[27]attracted little attention, and no publications on the subject appeared from any other laboratory until 1959.[30] Staying up at night to study sleep remained an undesirable occupation by any standards. In the early 1950s, most previous research on the EEG patterns of sleep, like most approaches to sleep physiology generally, had either equated short periods of sleep with all sleep or relied on infrequent sampling during the night. Obtaining continuous records throughout typical nights of sleep seemed highly extravagant—owing in no small part to the cost of the blocks of paper required.

However, motivated by the desire to expand and quantify the description of rapid eye movements, Dement and Kleitman did just this: They recorded EEGs over a total of 126 nights with 33 subjects and, by means of a simplified categorization of EEG patterns, scored the paper recordings in their entirety.[31] On examining these 126 records, they found a predictable sequence of patterns, over the course of the night, that had been overlooked in all previous EEG studies of sleep. This sequence has now been observed throughout the world, and the original description remains essentially unchanged.

The usual sequence was that after the onset of sleep, the EEG pattern progressed fairly rapidly to slow wave sleep, which persisted for a variable period, generally approximately 30 minutes, and then a "lightening" took place. Whereas the progression from wakefulness to slow wave sleep at the beginning of the cycle almost invariably occurred through a continuum of change, the lightening usually was abrupt and coincident with a body movement or series of body movements. After the termination of stage 4, there generally was a short period of stage 2 or stage 3 sleep, which gave way to stage 1 and rapid eye movements. When the first eye movement period ended, the EEG again progressed through a continuum of change to slow wave sleep, which persisted for a time and then lightened, often abruptly, with body movement to stage 2, which again gave way to stage 1 and the second rapid eye movement period (as detailed in Dement and Kleitman's report[31]).

Dement and Kleitman found that this cyclic variation of EEG pattern occurred repeatedly throughout the night at intervals of 90 to 100 minutes from the end of one eye movement period to the end of the next. The regular occurrences of REM periods and dreaming strongly suggested that dreams did not occur in response to chance disturbances.

At the time of these observations, sleep was still considered to be a single state. Dement and Kleitman characterized the EEG pattern during REM periods as "emergent stage 1," as opposed to "descending stage 1" at the onset of sleep. The percentage of the total sleep time occupied by REM sleep was between 20% and 25%, and the periods of REM sleep tended to be shorter in the early cycles of the night. This pattern of all-night sleep has been seen over and over in normal humans of both sexes, in widely varying environments and cultures, and across the life span.

RAPID EYE MOVEMENT SLEEP IN ANIMALS

The developing knowledge of the nature of sleep with rapid eye movements was in direct opposition to the ascending reticular activating system theory and constituted a paradigmatic crisis. The following observations were crucial:

- Arousal thresholds in humans were much higher during periods of REM sleep associated with a low-amplitude, relatively fast (stage 1) EEG pattern than during similar "light sleep" periods at the onset of sleep.
- Rapid eye movements during sleep were discovered in cats; the concomitant brain wave patterns (low-amplitude, fast) were indistinguishable from those in active wakefulness.[32]
- By discarding the sampling approach and recording continuously, a basic 90-minute cycle of sleep without rapid eye movements, alternating with sleep with rapid eye movements, was discovered.[31] This basic sleep cycle characterized all episodes of nocturnal sleep. Continuous recording also revealed a consistent, low-amplitude EEG pattern during a precise interval of sleep always associated with bursts of REM, which were additionally established as periods of vivid dreaming.
- Observations of motor activity in both humans and animals revealed the unique occurrence of an active suppression of spinal motor activity and muscle reflexes.

Thus sleep consists not of one state but rather of two distinct organismic states, as different from one another as both are from wakefulness. It had to be conceded that sleep could no longer be thought of as a time of brain inactivity and EEG slowing. By 1960, this fundamental change in thinking about the nature of sleep was well established; it exists as fact that has not changed in any way since then.

The discovery of rapid eye movements during sleep in humans, plus the all-night sleep recordings that revealed the regular recurrence of lengthy periods during which rapid eye movements occurred and during which brain wave patterns resembled those of light sleep, prepared the way for the discovery of REM sleep in cats, despite the extremely powerful bias that an "activated" EEG pattern could not be associated with sleep. In the first study in cats, maintaining the insulation and hence the integrity of implanted electrodes had not yet been solved, so an alternative, placement of small pins in the scalp, was used. With this approach, the waking EEG was totally obscured by the electromyogram from the large temporal muscles of the cat. However, when the animal fell asleep, slow waves could be seen, and the transition to REM sleep was clearly observed because muscle potentials were completely suppressed. The cat's rapid eye movements and also the twitching of the whiskers and paws could be directly observed.

It is very difficult now, in the 21st century, to understand and appreciate the exceedingly controversial nature of these findings. The following personal account from Dement[33] illustrates both the power and the danger of scientific dogma:

> I wrote them [the findings] up, but the paper was nearly impossible to publish because it was completely contradictory to the totally dominant neurophysiological theory of the time. The assertion by me that an activated EEG could be associated with unambiguous sleep was considered to be

absurd. As it turned out, previous investigators had observed an activated EEG during sleep in cats[28,29] but simply could not believe it and ascribed it to arousing influences during sleep. A colleague who was assisting me was sufficiently skeptical that he preferred I publish the paper as sole author. After four or five rejections, to my everlasting gratitude, Editor-in-Chief Herbert Jasper accepted the paper without revision for publication in *Electroencephalography and Clinical Neurophysiology*.[33]

Of note, however, many early researchers (Dement included) did not recognize the significance of the absence of muscle potentials during the REM periods in cats. It remained for Michel Jouvet, working in Lyon, France, to insist on the importance of electromyographic suppression in his early papers, the first of which was published in 1959.[30,34] Hodes and Dement began to study the "H-reflex" in humans in 1960, finding complete suppression of reflexes during REM sleep, and Octavio Pompeiano and others in Pisa, Italy, worked out the basic mechanisms of REM atonia in the cat.[35,36]

DUALITY OF SLEEP

Even though the basic NREM sleep cycle was well established, the realization that REM sleep was qualitatively different from that in the remainder of the sleep cycle took years to evolve. Jouvet and colleagues performed an elegant series of investigations on the brainstem mechanisms of sleep that forced the inescapable conclusion that sleep consists of two fundamentally different organismic states.[37] Among their many early contributions were clarification of the role of pontine brainstem systems as the primary anatomic site for REM sleep mechanisms and the clear demonstration that electromyographic activity and muscle tonus are completely suppressed during REM periods and *only* during REM periods. These investigations began in 1958 and were carried out during 1959 and 1960.

It is now well established that atonia is a fundamental characteristic of REM sleep and is mediated by an active and highly specialized neuronal system. The pioneering microelectrode studies of Edward Evarts in cats and monkeys, and observations on cerebral blood flow in the cat by Reivich and Kety, provided convincing evidence that the brain during REM sleep is very active.[38,39] Certain areas of the brain appear to be more active in REM sleep than in wakefulness. By that time, the notion of sleep as a passive process was totally demolished, although a persistent attitude that NREM sleep was essentially inactive and quiet lingered for many years. By 1960, it was possible to define REM sleep as a completely separate organismic state characterized by cerebral activation, active motor inhibition, and, of course, an association with dreaming. The fundamental duality of REM versus NREM sleep was established fact. It is of historical interest that the fascination with dreaming influenced the naming of REM and everything else the rather dismissive term NREM, even though NREM took up a larger part of the sleep cycle. This rudimentary distinction may be historically analogous to early descriptions of portions of the genome as "junk DNA."

PRECURSORS OF SLEEP MEDICINE

Sleep research, which emphasized all-night sleep recordings, burgeoned in the 1960s and was the legitimate precursor of sleep medicine and particularly of its core clinical test, polysomnography. Much of the research at that time emphasized studies of dreaming and REM sleep and had its roots in a psychoanalytic approach to mental illness, which strongly implicated dreaming in the psychotic process. After sufficient numbers of all-night sleep recordings had been carried out in humans to demonstrate a highly characteristic "normal" sleep architecture, investigators noted a significantly shortened REM latency in association with endogenous depression.[40] This phenomenon has been intensively investigated ever since. Other important precursors of sleep medicine were the following:

1. Discovery of sleep-onset REM periods in patients with narcolepsy
2. Interest in sleep, epilepsy, and abnormal movement—primarily in France
3. Introduction of benzodiazepines and the use of sleep laboratory studies in defining hypnotic drug efficacy

Sleep-Onset REM Periods and Cataplexy

In 1959, a patient with narcolepsy came to the Mount Sinai Hospital in New York City to see Drs. Charles Fisher and Dement. At Fisher's suggestion, a nocturnal sleep recording was begun. Within seconds after he fell asleep, the patient showed the dramatic and characteristic rapid eye movements and sawtooth waves of REM sleep. The first paper documenting sleep-onset REM periods in a specific patient was published in 1960 by Gerald Vogel, at the time working in Chicago.[41] In a collaborative study between the University of Chicago and the Mount Sinai Hospital, data on nine narcoleptic patients with sleep-onset REM periods at night were reported in 1963.[42] Subsequent research showed that sleepy patients who did not have cataplexy did not have sleep-onset REM periods (SOREMPs), and those with cataplexy always had SOREMPs.[43] For the first time, a clinical role for the polysomnogram as a potential diagnostic tool was being identified! Sleep research was becoming sleep medicine.

The Narcolepsy Clinic: A False Start

In January 1963, after leaving Mount Sinai and moving to Stanford University, Dement was eager to test the hypothesis of an association between cataplexy and SOREMPs. However, not a single narcoleptic patient was located in the San Francisco Bay area. In desperation, the investigators placed a brief "want ad," requesting such subjects, in a daily newspaper, the *San Francisco Chronicle*. More than 100 people responded; approximately 50 of these patients were bona fide narcoleptics afflicted with both sleepiness and cataplexy.

The response to the ad was a noteworthy event in the development of sleep disorders medicine. With one or two exceptions, none of the narcoleptics had ever been correctly diagnosed. Responsibility for their clinical management had to be assumed in order to facilitate their participation in the research. The late Dr. Stephen Mitchell, who had completed his neurology training and was entering a psychiatry residency at Stanford University, joined Dement in creating a narcolepsy clinic in 1964, and soon they were managing well over 100 patients. This program involved seeing the patients at regular intervals and adjusting their medication. Nonetheless, it constituted a precursor to the typical sleep disorders clinic, because at least one daytime polygraphic sleep recording was performed in all patients to establish the presence of SOREMPs.

Patients were questioned comprehensively about their sleep. When possible, an all-night sleep recording also was carried out. Unfortunately, insurance companies declared that such recordings in narcoleptic patients were experimental. This ruling forced the closure of the clinic because of insufficient funds—foreshadowing how third party payers have influenced the practice of clinical sleep medicine in the United States.

European Interest

In Europe, a genuine research interest in sleep problems had arisen, and it achieved its clearest expression in a 1963 symposium held in Paris, organized by Professor H. Fischgold, with its proceedings published as *La Sommeil de Nuit Normal et Pathologique* in 1965.[44] The primary clinical emphasis in this symposium was the documentation of sleep-related epileptic seizures and analyses of a number of related studies on sleepwalking and night terrors. Investigators from France, Italy, Belgium, Germany, and the Netherlands took part. The important role of European sleep scientists in the establishment of clinical sleep medicine is discussed further on.

Benzodiazepines and Hypnotic Efficacy Studies

In parallel with the discoveries being made in narcolepsy, a renewed interest in the pharmacologic treatment of insomnia was emerging. Benzodiazepines were introduced in 1960 with the marketing of chlordiazepoxide (Librium). This compound offered a significant advance in terms of safety over barbiturates for the purpose of tranquilizing and sedating. It was quickly followed by diazepam (Valium) and the first benzodiazepine introduced specifically as a hypnotic, flurazepam (Dalmane). Although a number of studies had been done on the effects of drugs on sleep, the first use of the sleep laboratory to evaluate sleeping pills may have been the 1965 study by Oswald and Priest.[45] An important series of studies establishing the role of the sleep laboratory in the evaluation of hypnotic efficacy was carried out by Anthony Kales and colleagues at the University of California Los Angeles.[46] This group also carried out pioneering studies of patients with hypothyroidism, asthma, Parkinson disease, and somnambulism.[47-50]

THE DISCOVERY OF SLEEP APNEA

The original description of sleep apnea often is attributed to independent publications by Gastaut, Tassinari, and Duron in France and by Jung and Kuhlo in Germany.[51,52] Both of these groups reported their findings in 1965. Scholars have found references to this phenomenon in many places, but these publications allowed for a clear-cut recognition of the phenomenon, and they had a direct causal continuity to sleep disorders medicine as we know it today. Earlier work in this area deserves mention (Christian Guilleminault, personal communication, 2014). In a report published in German, a group from Heidelberg University Hospital in 1960 described a patient who had come to the hospital for investigation of recurring morning headaches and was observed to have respiratory pauses during sleep, with recovery breathing associated with a loud snore. A polygraphic recording during a nap was included in the publication.[53] From the National Institutes of Health (NIH), a publication by Drachman and Gumnit described evaluation of an obese woman using electroencephalography and blood gas analysis, which identified repetitive stoppage of air exchange despite persistence of thoracoabdominal movements. The patient was placed on a strict diet and after a significant weight loss saw her sleepiness disappear.[54] No further publications in the area of sleep from this group are available, so it appears their work was not appreciated at the time. Peretz Lavie has detailed the historical contributions made by scientists and clinicians around the world in helping to describe and elucidate this disorder.[21]

These important findings were widely ignored in America (Video 1-2). What should have been an almost inevitable discovery by either the otolaryngologic surgery community or the pulmonary medicine community did not occur because neither specialty included a tradition for carefully observing breathing during sleep. The well-known and frequently cited study by Burwell and colleagues—although impressive in a literary sense in its evoking of the somnolent boy Joe from *The Pickwick Papers*—erred badly in evaluating their somnolent obese patients only during waking, and in attributing the cause of the somnolence to hypercapnia.[55]

The popularity of this paper further reduced the likelihood of discovery of sleep apnea by the pulmonary community. The term *pickwickian* was an instant success as a neologism, and its colorful connotations may have played a role in stimulating interest in this syndrome by the European neurologists who also were interested in sleep.

A small group of French neurologists who specialized in clinical neurophysiology and electroencephalography were in the vanguard of sleep research. Among these was Christian Guilleminault, who was instrumental in later establishing the specialty of clinical sleep medicine at Stanford and throughout the world. Guilleminault also was the first to describe obstructive sleep apnea as a clinical syndrome.[56,57]

One of the collaborators in the French discovery of sleep apnea, C. Alberto Tassinari, joined the Italian neurologist Elio Lugaresi in Bologna in 1970. These clinical investigators, along with Giorgio Coccagna and a host of others, including Guilleminault, over the years performed a crucial series of clinical sleep investigations and, indeed, provided a complete description of the sleep apnea syndrome, including the first observations of the occurrence of sleep apnea in nonobese patients, an account of the cardiovascular correlates, and a clear identification of the importance of snoring and hypersomnolence as diagnostic indicators. These studies are recounted in Lugaresi's book, *Hypersomnia with Periodic Apneas,* published in 1978.[58]

ITALIAN SYMPOSIA

In 1967, Henri Gastaut and Elio Lugaresi (Figure 1-4) organized a symposium, the proceedings of which were published as *The Abnormalities of Sleep in Man,* that encompassed issues across a full range of pathologic sleep in humans.[59] This meeting took place in Bologna, Italy, and the papers presented covered many of what are now major topics in the sleep medicine field: insomnia, sleep apnea, narcolepsy, and periodic leg movements during sleep. It was an epic meeting from the standpoint of the clinical investigation of sleep; the only major issues not represented were clear concepts of clinical practice models and hard data on the high population prevalence of sleep disorders. However, the event that may have finally triggered a serious international interest in sleep apnea syndromes was a symposium organized by Lugaresi in 1972, which took place in Rimini, a small resort on the Adriatic coast.[54]

Figure 1-4 Elio Lugaresi, Professor of Neurology, University of Bologna, at the 1972 Rimini symposium.

BIRTH PANGS

Despite all the clinical research, the concept of all-night sleep recordings as a clinical diagnostic test did not emerge unambiguously. It is worth considering the reasons for this failure, partly because they continue to operate today as impediments to the expansion of the sleep medicine field, and partly to elucidate the field's long-overdue development.

The first important reason was the unprecedented burdensome nature of an all-night diagnostic test, particularly if it was conducted on an outpatient basis. The cost of all-night polygraphic recording, in terms of its basic expense, was high enough without adding the cost of hospitalization, although hospitalization would have legitimized the patient's spending the night in a testing facility. To sleep in an outpatient clinic for a diagnostic test was a totally unprecedented, time-intensive and labor-intensive enterprise, and completely in conflict with the brief time required for accepted test protocols such as reporting to the chemistry laboratory to give a blood sample, breathing into a pulmonary function testing apparatus, and undergoing a screening radiographic examination.

A second important barrier was the reluctance of nonhospital clinical professionals to work at night. Although medical house staff physicians are very familiar with night work, they do not generally enjoy it; furthermore, clinicians could not work 24-hour days, first seeing patients and ordering tests, and then conducting the tests themselves.

Finally, only a very small number of people in relevant fields understood that complaints of daytime sleepiness and nocturnal sleep disturbance represented something of clinical significance. Even narcolepsy, which was by the early 1970s fully characterized as an interesting and disabling clinical syndrome requiring sleep recordings for diagnosis, was not recognized in the larger medical community and had too low a prevalence to warrant creating a medical subspecialty. A study carried out in 1972 documented a mean of 15 years from onset of the characteristic symptoms of excessive daytime sleepiness and cataplexy to diagnosis and treatment by a clinician. The study also showed that a mean of 5.5 different physicians were consulted without benefit throughout that long interval.[60]

EARLY DEVELOPMENT OF STANFORD SLEEP MEDICINE CLINICAL PRACTICE

Creation of the sleep disorders clinic at Stanford University was in many ways a microcosm of how sleep medicine evolved throughout the world. Dement arrived at Stanford in 1963 to establish a sleep research program. A need for clinical application of the knowledge being acquired soon became obvious. By 1964 subjects in narcolepsy trials also were managed as patients. Patients complaining of insomnia were enrolled in hypnotic efficacy research studies. This arrangement brought the Stanford group into contact with many patients afflicted with insomnia and demolished the notion that a majority of such patients had psychiatric problems. An early concern was the reliability of the subjective descriptions of their sleep. The classic all-night sleep recording gave an answer and yielded a great deal of information. Throughout the second half of the 1960s, as a part of their research, the Stanford group continued to manage patients with narcolepsy and insomnia. As the group's reputation for expertise grew, it began to receive referrals for evaluation from physicians all over the United States. Vincent Zarcone, a psychiatrist, joined this effort to develop the field of clinical sleep medicine at Stanford. In 1970 a sleep clinic was formally established at Stanford. Not surprisingly, the fledging clinic immediately struggled with reimbursement issues.

When the clinic was opened in 1970, the central role of obstructive sleep apnea as a mechanism of sleep-related pathology was not yet appreciated. It took an international meeting in Bruges, Belgium, for the Stanford group to recognize the importance of this entity. At that meeting, Dr. Zarcone was particularly impressed with Christian Guilleminault, a neurologist with knowledge of sleep apnea who had previously performed sleep research at Stanford with Dr. Steve Hendrickson. At that Bruges meeting, Dr. Zarcone suggested to Dement that they try to recruit Guilleminault. Dement had already been considering recruiting a neurologist to Stanford. Guilleminault welcomed the opportunity to strengthen the clinical sleep medicine program at Stanford. The synergy of these three physicians set in motion the creation of the first successful sleep medicine clinic, which served as a model for the rest of the world.

In January 1972, Christian Guilleminault formally joined the Stanford group. He had extensive knowledge of the European studies of sleep apnea. Until his arrival, the Stanford group had not routinely used respiratory and cardiac sensors in their all-night sleep studies. Starting in 1972, these measurements became a routine part of the all-night diagnostic test. This test was given the permanent name of *polysomnography* in 1974 by Dr. Jerome Holland, a member of the Stanford group. Publicity about narcolepsy and excessive sleepiness resulted in a small flow of referrals to the Stanford sleep clinic, usually with the presumptive diagnosis of narcolepsy. During the first year or two, the goal for the Stanford practice was to see at least four new patients per week. To foster financial viability, the group did as much as possible (within ethical limits) to publicize its services. As a result, the clinic also acquired a small number patients, often self-referred, with chronic insomnia. The diagnosis of obstructive sleep apnea in patients with profound excessive daytime sleepiness was nearly always completely unambiguous.

Toward the end of 1972, the basic concepts and formats of sleep disorders medicine were sculpted to the extent that it was possible to offer a daylong course through Stanford University's Division of Postgraduate Medicine. In this course, titled "The Diagnosis and Treatment of Sleep Disorders," the topics covered were normal sleep architecture; the diagnosis and treatment of insomnia, with drug-dependent insomnia, pseudoinsomnia, central sleep apnea, and periodic leg movement as diagnostic entities; and the diagnosis and treatment of excessive daytime sleepiness or hypersomnia, with narcolepsy, NREM narcolepsy, and obstructive sleep apnea as diagnostic entities.

The cardiovascular complications of severe sleep apnea were alarming and often completely disabling. Unfortunately, the treatment options at this time were limited to often ineffective attempts to lose weight and chronic tracheostomy. The dramatic results of chronic tracheostomy in ameliorating the symptoms and complications of obstructive sleep apnea had been reported by Lugaresi and coworkers in 1970.[61] The notion of using such a treatment, however, was strongly resisted at the time by the medical community. One of the first patients referred to the Stanford sleep clinic for investigation of this severe somnolence and who eventually had a tracheostomy was a 10-year-old boy. From the very beginning of the development of clinical sleep medicine, children and adults were treated together.

In addition to medical skepticism, a major obstacle to the practice of sleep disorders medicine was the retroactive denial of payment by insurance companies, including the largest insurance company in the United States at the time. At a meeting with insurance company officials, Dr. Dement was even accused of being a "charlatan" when he tried to convey the importance of obstructive sleep apnea. A 3-year period of dogged educational efforts directed toward third party payers finally culminated in the recognition of polysomnography in California as a reimbursable diagnostic test in 1974. This landmark event opened the doors for the practice of sleep medicine throughout the nation. In retrospect, it seems clear that educational effort exerted and resulting policy decisions have undoubtedly saved countless lives and improved the health and well-being of perhaps millions of people worldwide.

CLINICAL SIGNIFICANCE OF EXCESSIVE DAYTIME SLEEPINESS

Christian Guilleminault, in a series of studies, had clearly shown that excessive daytime sleepiness was a major clinical complaint in several sleep disorders, as well as a pathologic phenomenon unto itself.[62] It was recognized, however, that methods to quantify this symptom and the underlying condition were not adequate to quantify the treatment outcome. The subjective Stanford Sleepiness Scale, developed by Hoddes and colleagues, did not give reliable results.[63] With the creation of sleep medicine clinics, a new problem emerged: how to objectively quantitate sleepiness.

The apparent lack of interest in daytime sleepiness by individuals who were devoting their careers to the investigation of sleep at that time has always been puzzling. Unquestionably, the current active investigation of this phenomenon is the result of the early interest of sleep disorders specialists. The neglect of sleepiness in previous research is all the more difficult to understand today, when sleepiness and the tendency to fall asleep during the performance of hazardous tasks are now widely recognized as important public health problems affecting our society. A number of reasons have been put forward. One is that sleepiness and drowsiness are negative qualities. A second is that the societal failure to confront the issue was fostered by language ambiguities in identifying sleepiness. A third is that the early sleep laboratory studies focused almost exclusively on REM sleep and other nighttime procedures, with little concern for the daytime except for psychopathology. Finally, the focus with regard to sleep deprivation was on performance from the perspective of human factors, rather than on sleepiness as representing a homeostatic response to sleep reduction.

An early attempt to develop an objective measure of sleepiness was that of Yoss and coworkers, who observed pupil diameter directly by video monitoring and described changes in sleep deprivation and narcolepsy.[64] Subsequently designated *pupillometry,* this technique has not been widely accepted. Dr. Mary Carskadon, while at Stanford, deserves most of the credit for the development of the latter-day standard approach to the measurement of sleepiness, called the Multiple Sleep Latency Test (MSLT).[65] She noted that subjective ratings of sleepiness made before a sleep recording frequently predicted the sleep latency. In the spring of 1976, she undertook to establish sleep latency as an objective measurement of the state of "sleepiness-alertness" by measuring sleep tendency before, during, and after 2 days of total sleep deprivation.[66] The protocol designed for this study has become the standard protocol for the MSLT. The choices of a 20-minute duration of a single test and a 2-hour interval between tests were essentially arbitrary and dictated by the practical demands of that study. This test was then formally applied to the clinical evaluation of sleepiness in patients with narcolepsy and, later, in patients with OSAS.[67,68] Of note, Dr. Gary Richardson as a student at Stanford coauthored these publications.

Carskadon and colleagues then undertook a monumental study of sleepiness in children by following them longitudinally across the second decade of life, which happens to also be the decade of highest risk for the development of narcolepsy. Using the new MSLT measure, these investigators found that 10-year-old children were completely alert in the daytime, but by the time the subjects reached sexual maturity, they were no longer fully alert even though they obtained almost the same amount of sleep at night as that in the period of childhood studied. Results of this remarkable decade of work and other studies are summarized in an important review.[69] In an effort spearheaded by Dr. Rafael Pelayo, Stanford University acknowledged the importance of this historic work by installing a permanent plaque in 2012 at the dormitory that housed this research.

Early MSLT research established the following important advances in thinking:

1. Daytime sleepiness and nighttime sleep are components of an interactive continuum, and the adequacy of nighttime sleep absolutely cannot be understood without a complementary measurement of the level of daytime sleepiness or its antonym, alertness.
2. Excessive sleepiness, also known as impaired alertness, was sleep medicine's most important symptom.

FURTHER DEVELOPMENT OF SLEEP MEDICINE

As the decade of the 1970s drew to a close, the consolidation and formalization of the practice of sleep disorders medicine were largely completed. What is now the American Academy of Sleep Medicine was formed and provided a home for professionals interested in sleep and, particularly, in the diagnosis and treatment of sleep disorders. This body, the Association of Sleep Disorders Centers (ASDC), began with five members in 1975. The organization then was responsible for the initiation of the scientific journal *Sleep.* It fostered the setting of standards through center accreditation and an examination for practitioners by which they were designated Accredited Clinical Polysomnographers.

The first international symposium on narcolepsy took place in the French Languedoc in the summer of 1975, immediately after the Second International Congress of the Association for the Physiological Study of Sleep (APSS) in Edinburgh. The former meeting, in addition to being scientifically productive, was of landmark significance because it produced the first consensus definition of a specific sleep disorder, drafted, revised, and unanimously endorsed by 65 narcoleptologists of international reputation.[70] The first sleep disorders patient volunteer organization, the American Narcolepsy Association, also was formed in 1975. The ASDC/APSS Diagnostic Classification of Sleep and Arousal Disorders was published in fall 1979 after 3 years of extraordinary effort by a small group of dedicated persons who made up the "nosology" committee chaired by Dr. Howard Roffwarg.[71] This early nosology was the precursor to the subsequent versions of the International Classification of Sleep Disorders.

Before the 1980s, the only effective treatment for severe OSAS was chronic tracheostomy. This highly effective but personally undesirable approach was replaced by two new procedures—one surgical, the other mechanical.[72,73] The first was uvulopalatopharyngoplasty(UPPP), which at the time was considered an advance, eventually fell into disfavor owing to its being both painful and often ineffective. UPPP did pave the way for more sophisticated and effective surgical options. The second was the widely used and highly effective continuous positive nasal airway pressure (CPAP) technique introduced by the Australian pulmonologist Colin Sullivan (Video 1-3). The first CPAP machines were very loud and uncomfortable. Fortunately, as the technology improved, CPAP devices entered the medical mainstream. The combination of the high prevalence of OSAS and, at the time, newly effective treatments fueled a strong expansion of sleep centers and clinicians. The ramifications of this growth are still being felt today.

The decade of the 1980s was capped by the publication of sleep medicine's first textbook, the first edition of *Principles and Practice of Sleep Medicine*.[74] For many years only one medical journal devoted to sleep existed; today, several are in publication, including *Sleep, Journal of Clinical Sleep Medicine, Journal of Sleep Research, Sleep and Biological Rhythms, Sleep & Breathing, Sleep Medicine, Sleep Medicine Reviews,* and *Sleep Research Online.* Articles about sleep are now routinely published in the major pulmonary, neurology, ear-nose-throat (ENT), pediatric, primary care, and psychiatric journals.

The 1990s saw an acceleration in the acceptance of sleep medicine throughout the world. Nonetheless, adequate sleep medicine services are still not readily available everywhere.[75,76]

In the United States, the National Center on Sleep Disorders Research (NCSDR) was established by statute as part of the National Heart, Lung, and Blood Institute of the National Institutes of Health.[77] The mandate of the NCSDR is to support research, promote educational activities, and coordinate sleep-related activities throughout various branches of the U.S. government. It is perhaps too easy to criticize any government body or to decry insufficient research funding, yet the mere recognition by the federal government of the importance of sleep by establishing the NCSDR is a huge achievement when taken in the perspective of how the sleep field began. This government initiative led to the development of large research projects dealing with various aspects of sleep disorders and the establishment of awards to develop educational materials at all levels of training.

The 1990s also saw the establishment of the National Sleep Foundation, as well as other organizations for patients. This foundation points out to the public the dangers of sleepiness and sponsors the annual National Sleep Awareness Week.

As the Internet increases exponentially in size, so does the availability of sleep knowledge for physicians, patients, and the general public. The average person today knows a great deal more about sleep and its disorders than the average person did at the end of the 1980s. It is perhaps unique to the sleep field that the Internet, on the one hand, has increased the availability of information on sleep. On the other hand, it seems self-evident that the Internet also has accelerated humanity's march toward a sleepless 24-hour society and has increased the pressure for sleep deprivation and poor hygiene, in particular among the young.

THE 21ST CENTURY AND BEYOND

The historical early development of clinical sleep medicine culminated with its acceptance in 2003 by the Accreditation Council on Graduate Medical Education (ACGME) as a formal training program. The field emerged from its embryonic origins to worldwide acceptance in a relatively short period of time, owing in no small part to the great public need for healthier sleep and alertness. The recognition of the importance of sleep as a health and wellness component was exemplified by the appointment of Dr. Mark Rosekind to the National Transportation Safety Board (NTSB) in 2010. For the first time in its history, the NTSB had a trained sleep scientist as a board member (Figure 1-5). The impact of this recognition is likely to be very far-reaching for public safety.

From today's vantage point, the greatest challenge for the future is the cost-effective expansion of sleep medicine to provide benefit to the increasing number of patients in society. The management of sleep deprivation and its serious consequences in the workplace, particularly in those industries that depend on sustained operations, continues to need increased attention. Healthy sleep needs to be a priority for all.

The education and training of all health professionals have far to go. This situation was highlighted by the report of the Institute of Medicine.[76] These problems also represent grand opportunities for research. Sleep medicine has come into its own. It has made concern for health a truly 24-hours-a-day

Figure 1-5 Dr. Mark Rosekind is sworn in by Dr. William Dement as the first sleep scientist at the National Transportation Safety Board (NTSB). Drs. Mary Carskadon and Deborah Babcock look on. In 2015 Dr. Rosekind was appointed administrator to the National Highway Traffic Safety Administration. (With permission from Dr. Rosekind and the NTSB.)

enterprise, and it has energized a new effort to reveal the secrets of the healthy and unhealthy sleeping brain.

Looking back at the history of sleep medicine forces the medical profession, and society as a whole, to look forward to the future. The future of sleep research indeed promises to be exciting. Finally answering the ancient questions about the basic functions of sleep and dreaming may be within the grasp of the current generation of young scientists. They would not be poised for these future discoveries if not for the early work described in this chapter.

Many times the young sleep medicine field seemed to be doomed to fail, yet the huge need to understand sleep and its disorders continued to push it forward. Currently, as the field faces new challenges with changes in health care and reimbursement policies, it is easy to be pessimistic about its future. Yet such challenges constitute part of a natural process of change. The forces that have driven the field forward are, if anything, expanding. The population is growing and getting older. Increasingly, people are expected to be alert and productive in a 24-hour society. Consequently, sleep medicine needs to continue to adapt to these societal changes. All practitioners in both sleep medicine and sleep research should keep in mind that millions of people have benefited from their work, and that billions more still need their help.

We remain realistically optimistic about the future of sleep medicine.

CLINICAL PEARL

Recent advances in sleep science, sleep medicine, public policy, and communications will foster an educated public that will know a great deal about sleep and its disorders. Clinicians should expect that their patients may have already learned about their sleep disorders from the information sources that are readily available. They also may have received considerable misinformation from these same sources. Sleep professionals need to know the history of sleep medicine for proper perspective and useful insights as the field evolves.

SUMMARY

Interest in sleep dates to antiquity and has influenced all cultures and religions. Ancient medical texts describe treatments for sleep problems such as insomnia. Just over a hundred years ago, sleep was thought of as a passive state. The discovery of electroencephalography led to concept of sleep as an active state. The discovery of REM sleep in the 1950s allowed the empirical challenge to the previously held beliefs. The formal study of sleep disorders using polysomnography progresses in the 1960s. Obstructive sleep apnea was described mostly by researchers based in Europe at that time. Despite a series of false steps, clinical sleep medicine was established at Stanford University in 1970 and shortly thereafter in other institutions. The organization of these groups led to the creation of professional sleep societies and further worldwide growth and recognition of sleep medicine. Sleep medicine was recognized in 2003 by the Accreditation Council on Graduate Medical Education (ACGME) as a formal training program. The field continues to evolve. As sleep medicine faces new challenges, an appreciation of its historical background can provide practitioners with insights for shaping the future of the discipline.

Selected Readings

Carskadon MA, Dement WC, Mitler MM, et al. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 1986;**9**:519–24.

Dement W, Kleitman N. The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *J Exp Psychol* 1957;**53**:339–46.

Dement WC, Vaughan CC. *The promise of sleep: a pioneer in sleep medicine explores the vital connection between health, happiness, and a good night's sleep*. New York: Delacorte Press; 1999.

Freud S. *The interpretation of dreams*. 3rd ed. New York: The Macmillan Company; 1913.

Gastaut H, Lugaresi E, Berti-Ceroni G, Coccagna G. *The abnormalities of sleep in man*. Bologna (Italy): Aulo Gaggi Editore; 1968.

Guilleminault C, Dement WC, Kroc Foundation. *Sleep apnea syndromes*. New York: Alan R. Liss; 1978.

Guilleminault C, Dement WC, Passouant P. *Narcolepsy: proceedings of the First International Symposium on Narcolepsy, July 1975, Montpellier, France*. New York: SP Books, division of Spectrum Publications (distributed by Halstead Press); 1976.

Hobson JA. *Sleep*. New York: Scientific American Library: (distributed by W.H. Freeman); 1989.

Kleitman N. *Sleep and wakefulness*. Rev. and enl. ed. Chicago: University of Chicago Press; 1963.

Mignot EJ. History of narcolepsy at Stanford University. *Immunol Res* 2014;**58**:315–39.

Rosekind M. Awakening a nation: a call to action. *Sleep Health* 2015;**1**:9–10.

A complete reference list can be found online at ExpertConsult.com.

Chapter 2

Normal Human Sleep: An Overview

Mary A. Carskadon; William C. Dement

Chapter Highlights

- Normal human sleep comprises two states—rapid eye movement (REM) and non–REM (NREM) sleep—that alternate cyclically across a sleep episode. State characteristics are well defined: NREM sleep includes a variably synchronous cortical electroencephalogram (EEG; including sleep spindles, K-complexes, and slow waves) associated with low muscle tonus and minimal psychological activity; the REM sleep EEG is desynchronized, muscles are atonic, and dreaming is typical.
- A nightly pattern of sleep in mature humans sleeping on a regular schedule includes several reliable characteristics: Sleep begins in NREM and progresses through deeper NREM stages (stages 2, 3, and 4 using the classic definitions, or stages N2 and N3 using the American Academy of Sleep Medicine Scoring Manual definitions) before the first episode of REM sleep occurs about 80 to 100 minutes later. Thereafter, NREM sleep and REM sleep cycle with a period of about 90 minutes. NREM stages 3 and 4 (or stage N3) concentrate in the early NREM cycles, and REM sleep episodes lengthen across the night.
- Age-related changes are also predictable: Newborn humans enter REM sleep (called active sleep) before NREM (called quiet sleep) and have a shorter sleep cycle (about 50 minutes); coherent sleep stages emerge as the brain matures during the first year. At birth, active sleep is about 50% of total sleep and declines over the first 2 years to about 20% to 25%. NREM sleep slow waves are not present at birth but emerge in the first 2 years. Slow wave sleep (stages 3 and 4) decreases across adolescence by about 40% from preteen years and continues a slower decline into old age, particularly in men and less so in women. REM sleep as a percentage of total sleep is about 20% to 25% across childhood, adolescence, adulthood, and into old age, except in dementia.
- Other factors predictably alter sleep, such as previous sleep-wake history (e.g., homeostatic load), phase of the circadian timing system, ambient temperature, medications and drugs, and sleep disorders.

A clear appreciation of the normal characteristics of sleep provides a strong background and template for understanding clinical conditions in which "normal" characteristics are altered as well as for interpreting certain consequences of sleep disorders. In this chapter, the normal young adult sleep pattern is described as a working baseline pattern. Normative changes associated with aging and other factors are summarized with that background in mind. Several major sleep disorders are highlighted by their differences from the normative pattern.

WHAT CHARACTERISTICS AND MEASURES ARE USED TO DEFINE SLEEP?

According to a simple behavioral definition, sleep is a reversible behavioral state of perceptual disengagement from and unresponsiveness to the environment. It is also true that sleep is a complex amalgam of physiologic and behavioral processes. Sleep is typically (but not necessarily) accompanied by postural recumbence, behavioral quiescence, closed eyes, and all the other indicators commonly associated with sleeping. In the unusual circumstance, other behaviors can occur during sleep. These behaviors can include sleepwalking, sleeptalking, teeth grinding, and other physical activities. Anomalies involving sleep processes also include intrusions of sleep—sleep itself, dream imagery, or muscle weakness—into wakefulness, for example.

Within sleep, two separate states have been defined on the basis of a constellation of physiologic parameters. These two states, rapid eye movement (REM) and non–rapid eye movement (NREM), exist in virtually all mammals and birds yet studied, and they are as distinct from one another as each is from wakefulness. (See Box 2-1 for a discussion of sleep stage nomenclature.)

NREM (pronounced *non-REM*) sleep is conventionally subdivided into four stages defined along one measurement axis, the electroencephalogram (EEG). The EEG pattern in NREM sleep is commonly described as synchronous, with such characteristic waveforms as sleep spindles, K-complexes, and high-voltage slow waves (Figure 2-1). The four NREM stages (stages 1, 2, 3, and 4) roughly parallel a depth-of-sleep continuum, with arousal thresholds generally lowest in stage 1 and highest in stage 4 sleep. NREM sleep is usually

Box 2-1 SLEEP MEDICINE METHODOLOGY AND NOMENCLATURE

In 2007, the American Academy of Sleep Medicine (AASM) published a new manual* for scoring sleep and associated events. This manual recommends alterations to recording methodology and terminology that the AASM will demand of clinical laboratories in the future. Although specifications of arousal, cardiac, movement, and respiratory rules appear to be value-added to the assessment of sleep-related events, the new rules, terminology, and technical specifications for recording and scoring sleep are not without controversy.

The current chapter uses the traditional terminology and definitions on which most descriptive and experimental research has been based since the 1960s.[1] Thus where the AASM uses the terms *N* for NREM sleep stages and *R* for REM sleep stages, *N1* and *N2* are used instead of stage 1 and stage 2; *N3* is used to indicate the sum of stage 3 and stage 4 (often called *slow wave sleep* in human literature); and *R* is used to name REM sleep. Another change is to the nomenclature for the recording placements. Therefore calling the auricular placements *M1* and *M2* (rather than *A1* and *A2*) is unnecessary and places the sleep EEG recording terminology outside the pale for EEG recording terminology in other disciplines. Although these are somewhat trivial changes, changes in nomenclature can result in confusion when attempting to compare with previous literature and established data sets and are of concern for clinicians and investigators who communicate with other fields.

Of greater concern are changes to the core recording and scoring recommendations that the AASM manual recommends. For example, the recommended scoring montage requires using a frontal (*F3* or *F4*) EEG placement with visual scoring of the recordings, rather than the central (*C3* or *C4*) EEG placements recommended in the standard manual. The rationale for the change is that the frontal placements pick up more slow wave activity during sleep. The consequences, however, are that sleep studies performed and scored with the frontal EEG cannot be compared with normative or clinical data and that the frontal placements also truncate the ability to visualize sleep spindles. Furthermore, developmental changes to the regional EEG preclude the universal assumption that sleep slow wave activity is a frontal event.

Other issues are present in this new AASM approach to human sleep; however, this is not the venue for a complete description of such concerns. In summary, the AASM scoring manual has not yet become the universal standard for assessing human sleep and might not achieve that status in its current form. Specifications for recording and scoring sleep are not without controversy.[2-7]

*See Iber C, Ancoli-Israel S, Quan SF. For the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications.* 1st ed. Westchester (IL): American Academy of Sleep Medicine, 2007. [Revised in 2013]; and Berry RB, Brooks R, Gamaldo CE, et al. For the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Version 2.1. www.aasmnet.org. Darien (IL): American Academy of Sleep Medicine; 2014.

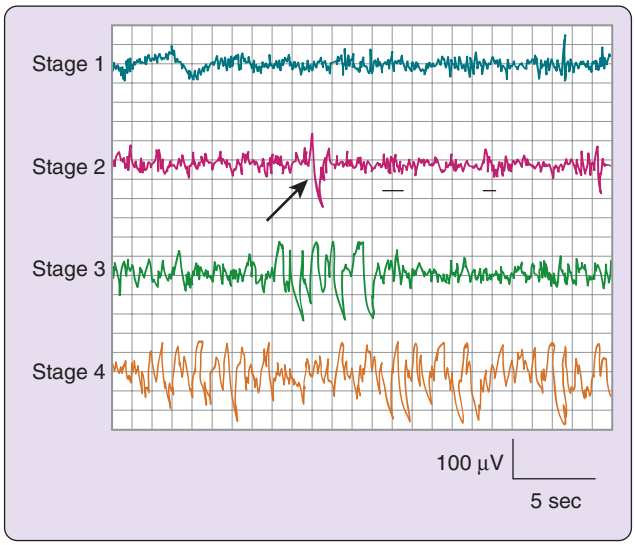


Figure 2-1 The Stages of NREM Sleep. The four electroencephalogram tracings depicted here are from a 19-year-old female volunteer. Each tracing was recorded from a referential lead (C3/A2) on a Grass Instruments (West Warwick, RI) Model 7D polygraph with a paper speed of 10 mm/sec, time constant of 0.3 sec, and ½-amplitude high-frequency setting of 30 Hz. On the second tracing, the *arrow* indicates a K-complex and the *underlining* shows two sleep spindles.

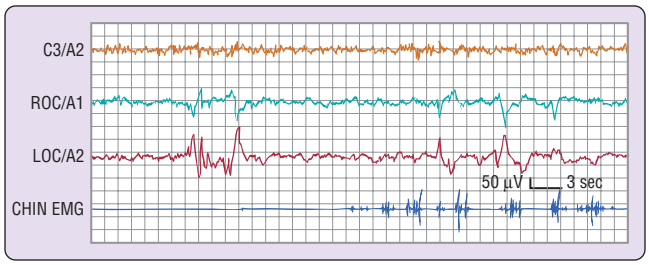


Figure 2-2 Phasic Events in Human REM Sleep. On the *left side* is a burst of several rapid eye movements (out-of-phase deflections in right outer canthus [ROC]/A1 and left outer canthus [LOC]/A2). On the *right side,* there are additional rapid eye movements as well as twitches on the electromyographic (EMG) lead. The interval between eye movement bursts and twitches illustrates tonic REM sleep.

associated with minimal or fragmentary mental activity. A shorthand definition of NREM sleep is a relatively inactive yet actively regulating brain in a movable body.

REM sleep, by contrast, is defined by EEG activation, muscle atonia, and episodic bursts of rapid eye movements. REM sleep usually is not divided into stages, although tonic and phasic types of REM sleep are occasionally distinguished for certain research purposes. The distinction of tonic versus phasic is based on short-lived events such as eye movements that tend to occur in clusters separated by episodes of relative quiescence. In cats, REM sleep phasic activity is epitomized by bursts of ponto-geniculo-occipital (PGO) waves, which are accompanied peripherally by rapid eye movements, twitching of distal muscles, middle ear muscle activity, and other phasic events that correspond to the phasic event markers easily measurable in humans. As described in Chapter 164, PGO waves are not usually detectable in humans. Thus the most commonly used marker of REM sleep phasic activity in humans is, of course, the occurrence of rapid eye movements (Figure 2-2); muscle twitches and cardiorespiratory irregularities often accompany the REM bursts. The mental activity of human REM sleep is associated with dreaming, based on vivid dream recall reported after about 80% of arousals from this state of sleep.[8] Inhibition of spinal motor neurons by

brainstem mechanisms mediates suppression of postural motor tonus in REM sleep. A shorthand definition of REM sleep, therefore, is an activated brain in a paralyzed body.

SLEEP ONSET

The onset of sleep under normal circumstances in normal adult humans is through NREM sleep. This fundamental principle of normal human sleep reflects a highly reliable finding and is important in considering normal versus pathologic sleep. For example, the abnormal entry into sleep through REM sleep can be a diagnostic sign in adult patients with narcolepsy.

Definition of Sleep Onset

The precise definition of the onset of sleep has been a topic of debate, primarily because there is no single measure that is 100% clear-cut 100% of the time. For example, a change in EEG pattern is not always associated with a person's perception of sleep, yet even when subjects report that they are still awake, clear behavioral changes can indicate the presence of sleep. To begin a consideration of this issue, let us examine the three basic polysomnographic measures of sleep and how they change with sleep onset. The electrode placements are described in Chapter 165.

Electromyogram

The electromyogram (EMG) may show a gradual diminution of muscle tonus as sleep approaches, but rarely does a discrete EMG change pinpoint sleep onset. Furthermore, the presleep level of the EMG, particularly if the person is relaxed, can be entirely indistinguishable from that of unequivocal sleep (Figure 2-3).

Electrooculogram

As sleep approaches, the electrooculogram (EOG) shows slow, possibly asynchronous eye movements (see Figure 2-3) that usually disappear within several minutes of the EEG changes described next. Occasionally, the onset of these slow eye movements coincides with a person's perceived sleep onset; more often, subjects report that they are still awake.

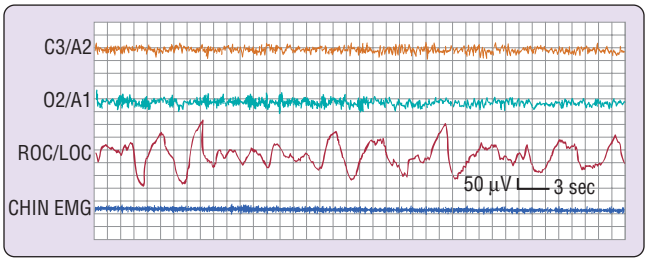


Figure 2-3 The Transition from Wakefulness to Stage 1 Sleep. The most marked change is visible on the two electroencephalographic (EEG) channels (C3/A2 and O2/A1), where a clear pattern of rhythmic alpha activity (8 cps) changes to a relatively low-voltage, mixed-frequency pattern at about the middle of the figure. The level of electromyographic (EMG) activity does not change markedly. Slow eye movements (right outer canthus [ROC]/left outer canthus [LOC]) are present throughout this episode, preceding the EEG change by at least 20 seconds. In general, the change in EEG patterns to stage 1 as illustrated here is accepted as the onset of sleep.

Electroencephalogram

In the simplest circumstance (see Figure 2-3), the EEG changes from a pattern of clear rhythmic alpha (8 to 13 cycles per second [cps]) activity, particularly in the occipital region, to a relatively low-voltage, mixed-frequency pattern (stage 1 sleep). This EEG change usually occurs seconds to minutes after the start of slow eye movements. With regard to introspection, the onset of a stage 1 EEG pattern may or may not coincide with perceived sleep onset. For this reason, a number of investigators require the presence of specific EEG patterns—the K-complex or sleep spindle (i.e., stage 2 sleep)—to acknowledge sleep onset. Even these stage 2 EEG patterns, however, are not unequivocally associated with perceived sleep.[9] A further complication is that sleep onset often does not occur all at once; instead, there may be a wavering of vigilance before "unequivocal" sleep ensues (Figure 2-4). Thus, it is difficult to accept a single variable as marking sleep onset. As Davis and colleagues[10] wrote many years ago (p. 35):

> Is "falling asleep" a unitary event? Our observations suggest that it is not. Different functions, such as sensory awareness, memory, self-consciousness, continuity of logical thought, latency of response to a stimulus, and alterations in the pattern of brain potentials all go in parallel in a general way, but there are exceptions to every rule. Nevertheless, a reasonable consensus exists that the EEG change to stage 1, usually heralded or accompanied by slow eye movements, identifies the transition to sleep, provided that another EEG sleep pattern does not intervene. One might not always be able to pinpoint this transition to the millisecond, but it is usually possible to determine the change reliably within several seconds.

Behavioral Concomitants of Sleep Onset

Given the changes in the EEG that accompany the onset of sleep, what are the behavioral correlates of the wake-to-sleep transition? The following material reviews a few common behavioral concomitants of sleep onset. Keep in mind that "different functions may be depressed in different sequence and to different degrees in different subjects and on different occasions" (p. 35).[10]

Simple Behavioral Task

In the first example, sleepy volunteers sitting at desks were asked to tap two switches alternately at a steady pace. As shown in Figure 2-5, this simple behavior continues after the onset of slow eye movements and may persist for several

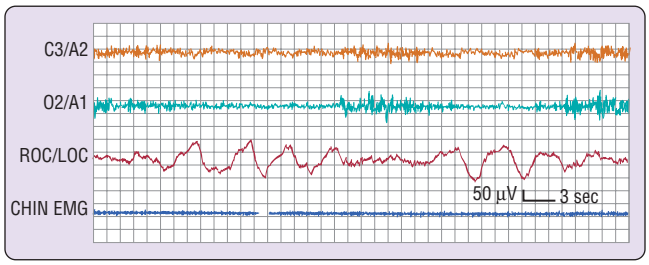


Figure 2-4 A Common Wake-to-Sleep Transition Pattern. Note that the electroencephalographic pattern changes from wake (rhythmic alpha) to stage 1 (relatively low-voltage, mixed-frequency) sleep twice during this attempt to fall asleep. EMG, Electromyogram; LOC, left outer canthus; ROC, right outer canthus.

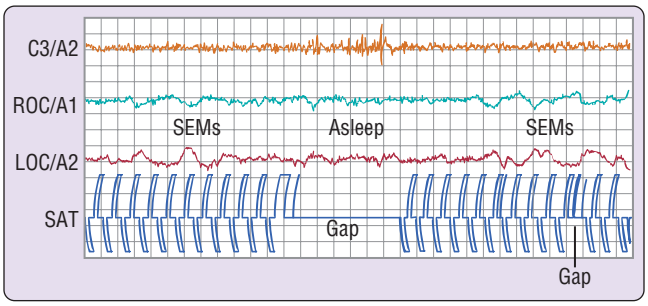


Figure 2-5 Failure to Perform a Simple Behavioral Task at the Onset of Sleep. The volunteer had been deprived of sleep overnight and was required to tap two switches alternately, shown as pen deflections of opposite polarity on the channel labeled SAT. When the electroencephalographic (EEG; C3/A2) pattern changes to stage 1 sleep, the behavior stops, returning when the EEG pattern reverts to wakefulness. LOC, left outer canthus; ROC, right outer canthus; SEMs, slow eye movements. (From Carskadon MA, Dement WC. Effects of total sleep loss on sleep tendency. *Percept Mot Skills* 1979;48:495–506.)

seconds after the EEG changes to a stage 1 sleep pattern.[11] The behavior then ceases, usually to recur only after the EEG reverts to a waking pattern. This is an example of what one may think of as the simplest kind of "automatic" behavior pattern. That such simple behavior can persist past sleep onset and as one passes in and out of sleep may explain how impaired, drowsy drivers are able to continue down the highway.

Visual Response

A second example of behavioral change at sleep onset derives from an experiment in which a bright light is placed in front of the subject's eyes and the subject is asked to respond when a light flash is seen by pressing a sensitive microswitch taped to the hand.[12] When the EEG pattern is stage 1 or stage 2 sleep, the response is absent more than 85% of the time. When volunteers are queried afterward, they report that they did not see the light flash, not that they saw the flash but the response was inhibited. This is one example of the perceptual disengagement from the environment that accompanies sleep onset.

Auditory Response

In another sensory domain, the response to sleep onset is examined with a series of tones played over earphones to a subject who is instructed to respond each time a tone is heard. One study of this phenomenon showed that reaction times became longer in proximity to the onset of stage 1 sleep, and responses were absent coincident with a change in EEG to unequivocal sleep.[13] For responses in both visual and auditory modalities, the return of the response after its sleep-related disappearance typically requires the resumption of a waking EEG pattern.

Olfactory Response

When sleeping humans are tasked to respond when they smell something, the response depends in part on sleep state and in part on the particular odorant. In contrast to visual responses, one study showed that responses to graded strengths of peppermint (strong trigeminal stimulant usually perceived as pleasant) and pyridine (strong trigeminal stimulant usually perceived as extremely unpleasant) were well maintained during initial stage 1 sleep.[14] As with other modalities, the response in other sleep stages was significantly poorer. Peppermint simply was not consciously smelled in stages 2 and 4 NREM sleep or in REM sleep; pyridine was never smelled in stage 4 sleep, and only occasionally in stage 2 NREM and in REM sleep.[14] On the other hand, a tone successfully aroused the young adult participants in every stage. One conclusion of this report was that the olfactory system of humans is not a good sentinel system during sleep.

Response to Meaningful Stimuli

One should not infer from the preceding studies that the mind becomes an impenetrable barrier to sensory input at the onset of sleep. Indeed, one of the earliest modern studies of arousability during sleep showed that sleeping humans were differentially responsive to auditory stimuli of graded intensity.[15] Another way of illustrating sensory sensitivity is shown in experiments that have assessed discriminant responses during sleep to meaningful versus nonmeaningful stimuli, with meaning supplied in a number of ways and response usually measured as evoked K-complexes or arousal. The following are examples.

- A person tends to have a lower arousal threshold for his or her own name versus someone else's name.[16] In light sleep, for example, one's own name spoken softly will produce an arousal; a similarly applied nonmeaningful stimulus will not. Similarly, a sleeping mother is more likely to hear her own baby's cry than the cry of an unrelated infant.
- Williams and colleagues[17] showed that the likelihood of an appropriate response during sleep was improved when an otherwise nonmeaningful stimulus was made meaningful by linking the absence of response to punishment (a loud siren, flashing light, and the threat of an electric shock).

From these examples and others, it seems clear that sensory processing at some level does continue after the onset of sleep. Indeed, one study has shown with functional magnetic resonance imaging that regional brain activation occurs in response to stimuli during sleep and that different brain regions (middle temporal gyrus and bilateral orbitofrontal cortex) are activated in response to meaningful (person's own name) versus nonmeaningful (beep) stimuli.[18]

Hypnic Myoclonia

What other behaviors accompany the onset of sleep? If you awaken and query someone shortly after the stage 1 sleep EEG pattern appears, the person usually reports the mental experience as one of losing a direct train of thought and of experiencing vague and fragmentary imagery, usually visual.[19] Another fairly common sleep-onset experience is hypnic myoclonia, which is experienced as a general or localized muscle contraction very often associated with rather vivid visual imagery. Hypnic myoclonias are not pathologic events, although they tend to occur more commonly in association with stress or with unusual or irregular sleep schedules.

The precise nature of hypnic myoclonias is not clearly understood. According to one hypothesis, the onset of sleep in these instances is marked by a dissociation of REM sleep components, wherein a breakthrough of the imagery component of REM sleep (hypnagogic hallucination) occurs in the absence of the REM motor inhibitory component. A response by the individual to the image, therefore, results in a movement or jerk. The increased frequency of these events in association with irregular sleep schedules is consistent with the

increased probability of REM sleep occurring at the wake-to-sleep transition under such conditions (see later). Although the usual transition in adult humans is to NREM sleep, the REM portal into sleep, which is the norm in infancy, may become partially opened under unusual circumstances or in certain sleep disorders.

Memory Near Sleep Onset

What happens to memory at the onset of sleep? The transition from wake to sleep tends to produce a memory impairment. One view is that it is as if sleep closes the gate between short-term and long-term memory stores. This phenomenon is best described by the following experiment.[20] During a presleep testing session, word pairs were presented to volunteers over a loudspeaker at 1-minute intervals. The subjects were then awakened either 30 seconds or 10 minutes after the onset of sleep (defined as EEG stage 1) and asked to recall the words presented before sleep onset. As illustrated in Figure 2-6, the 30-second condition was associated with a consistent level of recall from the entire 10 minutes before sleep onset. (Primacy and recency effects are apparent, although not large.) In the 10-minute condition, however, recall paralleled that in the 30-second group for only the 10 to 4 minutes before sleep onset and then fell abruptly from that point until sleep onset.

In the 30-second condition, therefore, both long-term (4 to 10 minutes) and short-term (0 to 3 minutes) memory stores remained accessible. In the 10-minute condition, by contrast, words that were in long-term stores (4 to 10 minutes) before sleep onset were accessible, whereas words that were still in short-term stores (0 to 3 minutes) at sleep onset were no longer accessible; that is, they had not been consolidated into long-term memory stores. One conclusion of this experiment is that sleep inactivates the transfer of storage from short- to long-term memory. Another interpretation is that encoding of the material before sleep onset is of insufficient strength to allow recall. The precise moment at which this deficit occurs is not known and may be a continuing process, perhaps reflecting anterograde amnesia. Nevertheless, one may infer that if sleep persists for about 10 minutes, memory is lost for the few minutes before sleep. The following experiences represent a few familiar examples of this phenomenon:

- Inability to grasp the instant of sleep onset in your memory
- Forgetting a telephone call that had come in the middle of the night
- Forgetting the news you were told when awakened in the night
- Not remembering the ringing of your alarm clock
- Experiencing morning amnesia for coherent sleeptalking
- Having fleeting dream recall

Patients with syndromes of excessive sleepiness can experience similar memory problems in the daytime if sleep becomes intrusive.

Learning and Sleep

In contrast to this immediate sleep-related "forgetting," the relevance for sleep to human learning—particularly for consolidation of perceptual and motor learning—is of growing interest.[21,22] The importance of this association has also generated some debate and skepticism.[23] Nevertheless, a spate of recent research is awakening renewed interest in the topic, and mechanistic studies explaining the roles of REM and NREM sleep and particular components of the sleep EEG pattern (e.g., sleep spindles) more precisely have shown compelling evidence that sleep plays an important role in learning and memory (see Chapter 22).

PROGRESSION OF SLEEP ACROSS THE NIGHT

Pattern of Sleep in a Healthy Young Adult

The simplest description of sleep begins with the ideal case, the healthy young adult who is sleeping well and on a fixed schedule of about 8 hours per night (Figure 2-7). In general, no consistent male versus female distinctions have been found in the normal pattern of sleep in young adults. In briefest summary, the normal human adult enters sleep through NREM sleep, REM sleep does not occur until 80 minutes or longer thereafter, and NREM sleep and REM sleep alternate through the night, with about a 90-minute cycle (see Chapter 165 for a full description of sleep stages).

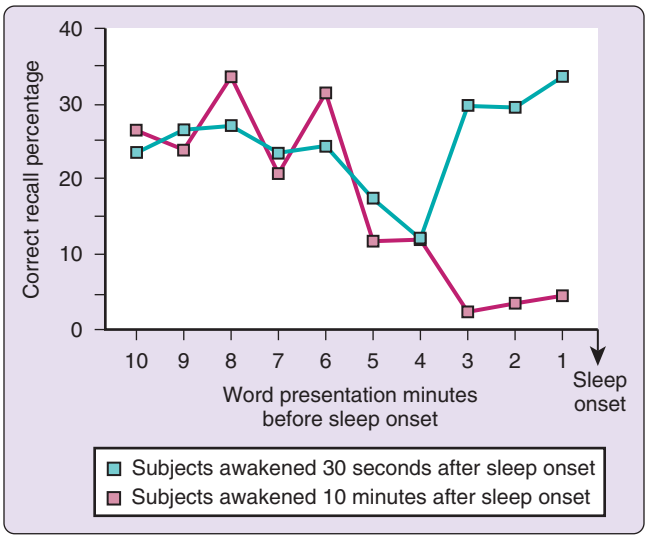


Figure 2-6 Memory is impaired by sleep, as shown by the study results illustrated in this graph. Refer to text for explanation.

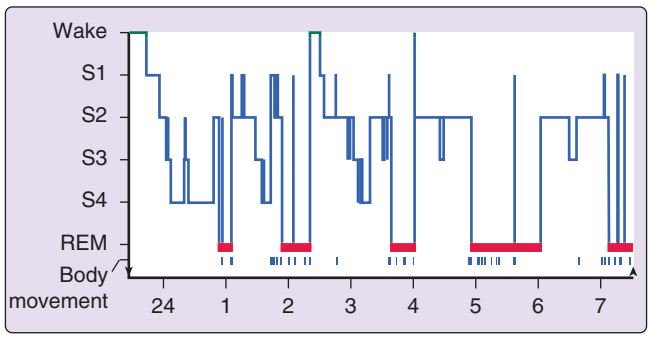


Figure 2-7 The progression of sleep stages across a single night in a normal young adult volunteer is illustrated in this sleep histogram. The text describes the ideal or average pattern. This histogram was drawn on the basis of a continuous overnight recording of electroencephalogram, electrooculogram, and electromyogram in a normal 19-year-old man. The record was assessed in 30-second epochs for the various sleep stages. REM, rapid eye movement.

First Sleep Cycle

The first cycle of sleep in the normal young adult begins with stage 1 sleep, which usually persists for only a few (1 to 7) minutes at the onset of sleep. Sleep is easily discontinued during stage 1 by, for example, softly calling a person's name, touching the person lightly, quietly closing a door, and so forth. Thus, stage 1 sleep is associated with a low arousal threshold. In addition to its role in the initial wake-to-sleep transition, stage 1 sleep occurs as a transitional stage throughout the night. A common sign of severely disrupted sleep is an increase in the occurrences and percentage of stage 1 sleep.

Stage 2 NREM sleep, signaled by sleep spindles or K-complexes in the EEG (see Figure 2-1), follows this brief episode of stage 1 sleep and continues for about 10 to 25 minutes in the first sleep cycle. In stage 2 sleep, a more intense stimulus is required to produce arousal. The same stimulus that produced arousal from stage 1 sleep often results in an evoked K-complex but no awakening in stage 2 sleep.

As stage 2 sleep progresses, high-voltage slow wave activity gradually begins to appear in the EEG. Eventually, this activity meets the criteria[1] for stage 3 NREM sleep, that is, high-voltage (at least 75 μV) slow wave (2 cps) activity accounting for more than 20% but less than 50% of the EEG activity. Stage 3 sleep usually lasts only a few minutes in the first cycle and is transitional to stage 4 as more and more high-voltage slow wave activity occurs. Stage 4 NREM sleep—identified when the high-voltage slow wave activity comprises more than 50% of the record—usually lasts about 20 to 40 minutes in the first cycle of a healthy young adult. An incrementally larger stimulus is usually required to produce an arousal from stage 3 or 4 sleep than from stage 1 or 2 sleep. (Investigators often refer to the combined stages 3 and 4 sleep as slow wave sleep [SWS], delta sleep, or deep sleep, or N3 in the newer nomenclature.)

A series of body movements usually signals an "ascent" to lighter NREM sleep stages. A brief (1- or 2-minute) episode of stage 3 sleep might occur, followed by perhaps 5 to 10 minutes of stage 2 sleep interrupted by body movements preceding the initial REM episode. REM sleep in the first cycle of the night is usually short-lived (under 10 minutes). The arousal threshold in this REM episode is variable, as is true for REM sleep throughout the night. Theories to explain the variable arousal threshold of REM sleep have suggested that at times, the person's selective attention to internal stimuli (i.e., dreaming) precludes a response, or that the arousal stimulus is incorporated into the ongoing dream story rather than producing an awakening. Certain early experiments examining arousal thresholds in cats found highest thresholds in REM sleep, which was then termed *deep sleep* in this species. Although this terminology is still often used in publications about sleep in animals, it should not be confused with human NREM stages 3 and 4 sleep, which is also often called *deep sleep*. In addition, the term *SWS* is sometimes used (as is *synchronized sleep*) as a synonym for all of NREM sleep in other species and is thus distinct from SWS (stages 3 and 4 NREM) in humans.

NREM-REM Cycle

NREM sleep and REM sleep continue to alternate through the night in cyclic fashion. REM sleep episodes usually become longer across the night. Stages 3 and 4 sleep occupy less time in the second cycle and might disappear altogether from later cycles as stage 2 sleep expands to occupy the NREM portion of the cycle. The average length of the first NREM-REM sleep cycle is about 70 to 100 minutes; the average length of the second and later cycles is about 90 to 120 minutes. Across the night, the average period of the NREM-REM cycle is about 90 to 110 minutes. Across the night, stage 1 sleep will account for about 2% to 5%, stage 2 about 45% to 55%, SWS about 10% to 20%, and REM sleep about 20% to 25% of sleep in a healthy young adult.

Distribution of Sleep Stages Across the Night

In the young adult, SWS dominates the NREM portion of the sleep cycle toward the beginning of the night (about the first one third); REM sleep episodes are longest in the last one third of the night. Brief episodes of wakefulness tend to intrude later in the night, usually near REM sleep transitions, and they usually do not last long enough to be remembered in the morning. The preferential distribution of REM sleep toward the latter portion of the night in normal human adults is linked to a circadian oscillator, which can be gauged by the oscillation of body temperature.[24,25] The preferential distribution of SWS toward the beginning of a sleep episode is not thought to be mediated by circadian processes but shows a marked response to the length of prior wakefulness,[26] thus reflecting the homeostatic sleep system, highest at sleep onset and diminishing across the night as sleep pressure wanes or as "recovery" takes place. Thus these aspects of the normal sleep pattern highlight features of the two-process model of sleep as elaborated on in Chapter 36.

Length of Sleep

The length of nocturnal sleep depends on a great number of factors—of which volitional control is among the most significant in humans—and it is thus difficult to characterize a "normal" pattern. Most young adults report sleeping about 7.5 hours a night on weekday nights and slightly longer, 8.5 hours, on weekend nights. The variability of these figures from person to person and from night to night, however, is quite high. Sleep length also depends on genetic determinants,[27] and one may think of the volitional determinants (staying up late, waking by alarm, and so on) superimposed on the background of a genetic sleep need. Length of prior waking also affects how much one sleeps, although not in a one-for-one manner. Indeed, the length of sleep is also determined by processes associated with circadian rhythms. Thus *when* one sleeps helps to determine *how long* one sleeps. In addition, as sleep is extended, the amount of REM sleep increases because the occurrence of REM sleep depends on the persistence of sleep into the peak circadian time.

Generalizations About Sleep in the Healthy Young Adult

A number of general statements can be made regarding sleep in the healthy young adult who is living on a conventional sleep-wake schedule and who is without sleep complaints:

- Sleep is entered through NREM sleep.
- NREM sleep and REM sleep alternate with a period near 90 minutes.
- SWS predominates in the first third of the night and is linked to the initiation of sleep and the length of time awake (i.e., sleep homeostasis).

- REM sleep predominates in the last third of the night and is linked to the circadian rhythm of body temperature.
- Wakefulness in sleep usually accounts for less than 5% of the night.
- Stage 1 sleep generally constitutes about 2% to 5% of sleep.
- Stage 2 sleep generally constitutes about 45% to 55% of sleep.
- Stage 3 sleep generally constitutes about 3% to 8% of sleep.
- Stage 4 sleep generally constitutes about 10% to 15% of sleep.
- NREM sleep, therefore, is usually 75% to 80% of sleep.
- REM sleep is usually 20% to 25% of sleep, occurring in four to six discrete episodes.

Factors Modifying Sleep Stage Distribution

Age

The strongest and most consistent factor affecting the pattern of sleep stages across the night is age (Figure 2-8). The most marked age-related differences in sleep from the patterns described earlier are found in newborn infants. For the first year of life, the transition from wake to sleep is often accomplished through REM sleep (called *active sleep* in newborns). The cyclic alternation of NREM-REM sleep is present from birth but has a period of about 50 to 60 minutes in the newborn compared with about 90 minutes in the adult. Infants also only gradually acquire a consolidated nocturnal sleep cycle, and the fully developed EEG patterns of the NREM sleep stages are not present at birth but emerge over the first 2 to 6 months of life. When brain structure and function achieve a level that can support high-voltage slow wave EEG activity, NREM stages 3 and 4 sleep become prominent.

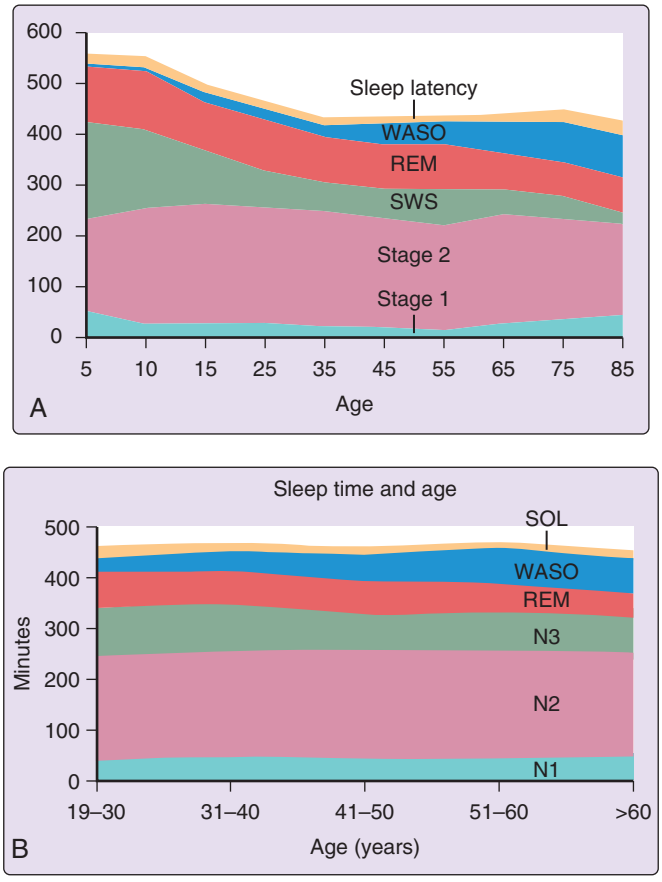


Figure 2-8 Changes in Sleep with Age. **A,** Time (in minutes) for sleep latency and wake time after sleep onset (WASO) and for REM sleep and NREM sleep stages 1, 2, and slow wave sleep (SWS). Summary values are given for ages 5 to 85 years. **B,** Changes in sleep in adults using the current AASM scoring standards. Time (in minutes) for sleep latency and WASO and for REM sleep and NREM sleep stages N1, N2, and N3. Values are medians. (**A,** From Ohayon M, Carskadon MA, Guilleminault C, et al. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep* 2004;27:1255–73; **B,** Data from Mitterling T, Högl B, Schönwald SV, et al. Sleep and respiration in 100 healthy Caucasian sleepers—a polysomnographic study according to American Academy of Sleep Medicine standards. *Sleep* 2015;38:867–75.)

SWS is maximal in young children and decreases markedly with age. The SWS of young children is both qualitatively and quantitatively different from that of older adults. For example, it is nearly impossible to wake youngsters in the SWS of the night's first sleep cycle. In one study,[28] a 123-dB tone failed to produce any sign of arousal in a group of children whose mean age was 10 years. In addition, children up to midadolescence often "skip" their first REM episode, perhaps because of the quantity and intensity of slow wave activity early in the night. A similar, although less profound qualitative difference distinguishes SWS occurring in the first and later cycles of the night in older humans. A marked quantitative change in SWS occurs across adolescence, when SWS decreases by about 40% during the second decade, even when length of nocturnal sleep remains constant.[29] Feinberg[30] hypothesized that the age-related decline in nocturnal SWS, which parallels loss of cortical synaptic density, is causally related to this cortical resculpting. More recent findings by de Vivo and colleagues in an animal model question that hypothesis.[30a] By midadolescence, youngsters no longer typically skip their first REM, and their sleep resembles that described earlier for young adults. By age 60 years, SWS is quite diminished, particularly in men; women maintain SWS later into life than men.

REM sleep as a percentage of total sleep is maintained well into healthy old age; the absolute amount of REM sleep at night has been correlated with intellectual functioning[31] and declines markedly in the case of organic brain dysfunctions in elderly people.[32]

Arousals during sleep increase markedly with age. Extended wake episodes of which the individual is aware and can report, as well as brief and probably unremembered arousals, increase with aging.[33] The latter type of transient arousals may occur with no known correlate but are often associated with occult sleep disturbances, such as periodic limb movements during sleep (PLMS) and sleep-related respiratory irregularities, which also become more prevalent in later life.[34,35]

Perhaps the most notable finding regarding sleep in elderly people is the profound increase in interindividual variability,[36] which thus precludes generalizations such as those made for young adults.

Prior Sleep History

A person who has experienced sleep loss on one or more nights shows a sleep pattern that favors SWS during recovery (Figure 2-9). Recovery sleep is also usually prolonged and deeper—that is, having a higher arousal threshold throughout—than basal sleep. REM sleep tends to show a rebound on the second or subsequent recovery nights after an episode of sleep loss. Therefore, with total sleep loss, SWS tends to be

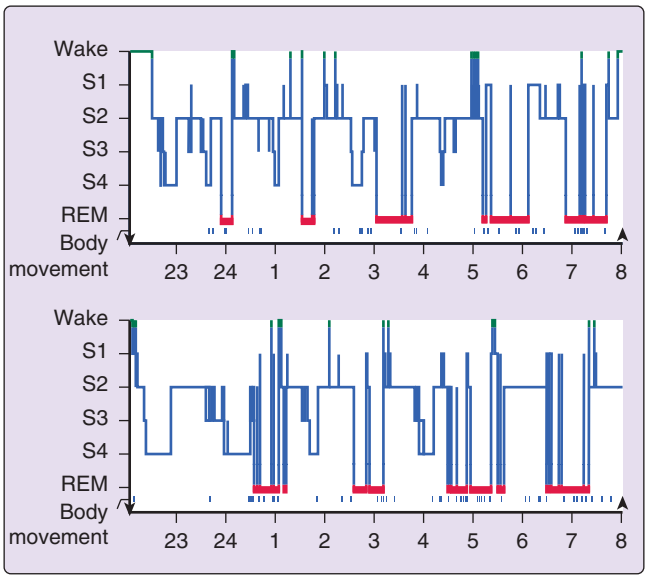


Figure 2-9 The *upper histogram* shows the baseline sleep pattern of a normal 14-year-old female volunteer. The *lower histogram* illustrates the sleep pattern in this volunteer for the first recovery night after 38 hours without sleep. Note that the amount of stage 4 sleep on the lower graph is greater than on baseline and that the first REM sleep episode is markedly delayed.

preferentially recovered compared with REM sleep, which tends to recover only after the recuperation of SWS. Thus both states of sleep show evidence of homeostatic regulation.

Cases in which a person is differentially deprived of REM or SWS—either operationally, by being awakened each time the sleep pattern occurs, or pharmacologically (see later)—show a preferential rebound of that stage of sleep when natural sleep is resumed. This phenomenon has particular relevance in a clinical setting, in which abrupt withdrawal from a therapeutic regimen may result in misleading diagnostic findings (e.g., sleep-onset REM periods [SOREMPs] as a result of a REM sleep rebound when REM suppressant medication is withdrawn) or could conceivably exacerbate a sleep disorder (e.g., if sleep apneas tend to occur preferentially or with greater intensity in the rebounding type of sleep).

Chronic restriction of nocturnal sleep, an irregular sleep schedule, or frequent disturbance of nocturnal sleep can result in a peculiar distribution of sleep states, most commonly characterized by premature REM sleep, that is, SOREMPs. Such episodes can be associated with hypnagogic hallucinations, sleep paralysis, or an increased incidence of hypnic myoclonia in persons with no organic sleep disorder.

Although not strictly related to prior sleep history, the first night of a laboratory sleep evaluation is commonly associated with more frequent arousals and a disruption of the normal distribution of sleep states, characterized chiefly by a delayed onset of REM sleep.[37] Often this delay takes the form of skipping the first REM episode of the night. In other words, the NREM sleep stages progress in a normal fashion, but the first cycle ends with an episode of stage 1 or a brief arousal instead of the expected brief REM sleep episode. In addition, REM sleep episodes are often disrupted, and the total amount of REM sleep on the first night in the sleep laboratory is also usually reduced from the normal value.

Circadian Rhythms

The circadian phase at which sleep occurs affects the distribution of sleep stages. REM sleep, in particular, occurs with a circadian distribution that peaks in the morning hours coincident with the trough of the core body temperature rhythm.[24,25] Thus, if sleep onset is delayed until the peak REM phase of the circadian rhythm—that is, the early morning—REM sleep tends to predominate and can even occur at the onset of sleep. This reversal of the normal sleep onset pattern may be seen in a healthy person who acutely undergoes a phase shift, either as a result of a work shift change or as a change resulting from jet travel across a number of time zones. Studies of persons sleeping in environments free of all cues to time show that the timing of sleep onset and the length of sleep occur in association with circadian phase.[38,39] Under these conditions, sleep distribution with reference to the circadian body temperature phase position shows that sleep onset is likeliest to occur on the falling limb of the temperature cycle. A secondary peak of sleep onsets, corresponding to afternoon napping, also occurs; the offset of sleep occurs most often on the rising limb of the circadian body temperature curve.[40]

Temperature

Extremes of temperature in the sleeping environment tend to disrupt sleep. REM sleep is commonly more sensitive to temperature-related disruption than is NREM sleep. Accumulated evidence from humans and other species suggests that mammals have only minimal ability to thermoregulate during REM sleep; in other words, the control of body temperature is virtually poikilothermic in REM sleep.[41] This inability to thermoregulate in REM sleep probably affects the response to temperature extremes and suggests that such conditions are less of a problem early during a night than late, when REM sleep tends to predominate. It should be clear, as well, that such responses as sweating or shivering during sleep under ambient temperature extremes occur in NREM sleep and are limited in REM sleep.

Drug Ingestion

The distribution of sleep states and stages is affected by many common drugs, including those typically prescribed in the treatment of sleep disorders as well as those not specifically related to the pharmacotherapy of sleep disorders and those used socially or recreationally. Whether changes in sleep stage distribution have any relevance to health, illness, or psychological well-being is unknown; however, particularly in the context of specific sleep disorders that differentially affect one sleep stage or another, such distinctions may be relevant to diagnosis or treatment. A number of generalizations regarding the effects of certain of the more commonly used compounds on sleep stage distribution can be made.

- Benzodiazepines tend to suppress SWS and have no consistent effect on REM sleep.
- Tricyclic antidepressants, monoamine oxidase inhibitors, and certain selective serotonin reuptake inhibitors tend to suppress REM sleep. An increased level of motor activity during sleep occurs with certain of these compounds, leading to a pattern of REM sleep without motor inhibition or an increased incidence of PLMS. Fluoxetine is also associated with rapid eye movements across all sleep stages ("Prozac eyes").

- Withdrawal from drugs that selectively suppress a stage of sleep tends to be associated with a rebound of that sleep stage. Thus, acute withdrawal from a benzodiazepine compound is likely to produce an increase of SWS; acute withdrawal from a tricyclic antidepressant or monoamine oxidase inhibitor is likely to produce an increase of REM sleep. In the latter case, this REM rebound could result in abnormal SOREMPs in the absence of an organic sleep disorder, perhaps leading to an incorrect diagnosis of narcolepsy.
- Acute presleep alcohol intake can produce an increase in SWS and suppression of REM sleep early in the night, which can be followed by REM sleep rebound in the latter portion of the night as the alcohol is metabolized. Low doses of alcohol have minimal effects on sleep stages, but they can increase sleepiness in the late evening.[42,43]
- Acute effects of marijuana (tetrahydrocannabinol [THC]) include minimal sleep disruption, characterized by a slight reduction of REM sleep. Chronic ingestion of THC produces a long-term suppression of SWS.[44]

Pathology

Sleep disorders, as well as other nonsleep problems, have an impact on the structure and distribution of sleep. As suggested before, these distinctions appear to be more important in diagnosis and in the consideration of treatments than for any implications about general health or illness resulting from specific sleep stage alterations. A number of common sleep-stage anomalies are associated with sleep disorders.

Narcolepsy. Narcolepsy is characterized by an abnormally short delay to REM sleep, marked by SOREMPs. This abnormal sleep-onset pattern occurs with some consistency, but not exclusively; that is, NREM sleep onset can also occur. Thus one diagnostic test consists of several opportunities to fall asleep across a day (see Chapter 173). If REM sleep occurs abnormally on two or more such opportunities, narcolepsy is extremely probable. The occurrence of this abnormal sleep pattern in narcolepsy is thought to be responsible for a number of the characteristic symptoms of this disorder. In other words, dissociation of components of REM sleep into the waking state results in hypnagogic hallucinations, sleep paralysis, and, most dramatically, cataplexy.

Other conditions in which a short REM sleep latency can occur include infancy, in which sleep-onset REM sleep is normal; sleep reversal or jet lag; acute withdrawal from REM-suppressant compounds; chronic restriction or disruption of sleep; and endogenous depression.[45] Reports have indicated a relatively high prevalence of REM sleep onsets in young adults[46] and in adolescents with early rise times.[47] In the latter, the REM sleep onsets on morning (8:30 AM and 10:30 AM) naps were related to a delayed circadian phase as indicated by later onset of melatonin secretion.

Sleep Apnea Syndromes. Sleep apnea syndromes may be associated with suppression of SWS or REM sleep secondary to the sleep-related breathing problem. Successful treatment of this sleep disorder, as with nocturnal continuous positive airway pressure, can produce large rebounds of SWS or REM sleep when first implemented (Figure 2-10).

Sleep Fragmentation. Fragmentation of sleep and increased frequency of arousals occur in association with a number of sleep disorders as well as with medical disorders involving physical pain or discomfort. PLMS, sleep apnea syndromes, chronic fibrositis, and so forth may be associated with tens to hundreds of arousals each night. Brief arousals are prominent in such conditions as allergic rhinitis,[48,49] juvenile rheumatoid arthritis,[50] and Parkinson disease.[51] In upper airway resistance syndrome,[52] EEG arousals are important markers because the respiratory signs of this syndrome are less obvious than in frank obstructive sleep apnea syndrome, and only subtle indicators may be available.[53] In specific situations, autonomic changes, such as transient changes of blood pressure,[54] can signify arousals; Lofaso and colleagues[55] indicated that autonomic changes are highly correlated with the extent of EEG arousals. Less well studied is the possibility that sleep fragmentation may be associated with subcortical events not

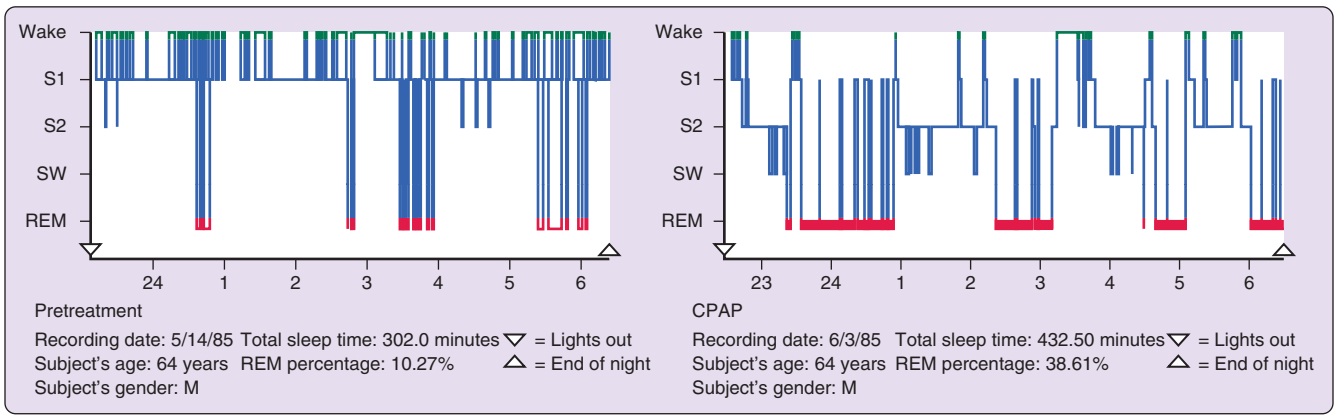


Figure 2-10 These sleep histograms depict the sleep of a 64-year-old male patient with obstructive sleep apnea syndrome. The *left graph* shows the sleep pattern before treatment. Note the absence of slow wave (SW) sleep, the preponderance of stage 1 (S1), and the very frequent disruptions. The *right graph* shows the sleep pattern in this patient during the second night of treatment with continuous positive airway pressure (CPAP). Note that sleep is much deeper (more SW sleep) and more consolidated and that REM sleep in particular is abnormally increased. The pretreatment REM percentage of sleep was only 10%, compared with nearly 40% with treatment. (Data supplied by G. Nino-Murcia, Stanford University Sleep Disorders Center, Stanford, CA.)

visible in the cortical EEG signal. These disorders also often involve an increase in the absolute amount of and the proportion of stage 1 sleep.

CLINICAL PEARLS

- The clinician should expect to see less slow wave sleep (stages 3 and 4; N3) in older persons, particularly men.
- Clinicians or colleagues might find themselves denying middle of the night communications (nighttime calls) because of memory deficits that occur for events proximal to sleep onset. This phenomenon might also account for memory deficits in excessively sleepy patients.
- Many medications (even if not prescribed for sleep) can affect sleep stages, and their use or discontinuation alters sleep. For example, REM-suppressant medications can result in a rebound of REM sleep when they are discontinued.
- Certain patients have sleep complaints (insomnia, hypersomnia) that result from attempts to sleep or be awake at times not in synchrony with their circadian phase.
- Patients who wake with events early in the night might have a disorder affecting NREM sleep; patients who wake with events late in the night may have a disorder affecting REM sleep.
- When using sleep restriction to build sleep pressure, treatment will be more effective if sleep is scheduled at the correct circadian phase. The problem of napping in patients with insomnia is that naps diminish the homeostatic drive to sleep.

SUMMARY

This chapter provides an overview of human sleep, with a focus on the healthy young adult as a template against which to evaluate and understand the expected changes that can occur as well as unusual circumstances and clinical conditions. Thus we find that maturational changes from infancy through old age carry different associations with the sleep of a healthy young adult and frame the first questions we should ask when confronted with an unknown case: what is the age? We also learn that sleep and the stages of sleep have important concomitants for cognitive function, perception, and the internal milieu. Later chapters catalog many specific properties of sleep physiology, neurochemistry, and sleep disorders; this chapter provides a foundation to support integration of that detailed information.

Selected Readings

Abel T, Havekes R, Saletin JM, Walker MP. Sleep plasticity and memory from molecules to whole-brain networks. *Curr Biol* 2013;**23**:R774–88.

Berry RB, Brooks R, Gamaldo CE, et al. For the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Version 2.1. Darien (IL): American Academy of Sleep Medicine; 2014 <www.aasmnet.org>.

Carskadon M. Sleep in adolescents: the perfect storm. *Pediatr Clin North Am* 2011;**58**:637–47.

Eisermann M, Kaminska A, Moutard M-L, et al. Normal EEG in childhood: from neonates to adolescents. *Neurophysiol Clin* 2013;**43**:35–65.

Foley D, Ancoli-Israel S, Britz P, Walsh J. Sleep disturbances and chronic disease in older adults: results of the 2003 National Sleep Foundation Sleep in America Survey. *J Psychosom Res* 2004;**56**(5):497–502.

Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health* 2015;**1**:40–3.

Jenni O, LeBourgeois M. Understanding sleep-wake behavior and sleep disorders in children: the value of a model. *Curr Opin Psychiatry* 2005;**19**:282–7.

Lesku JA, Roth TC, Rattenborg NC, et al. History and future of comparative analyses in sleep research. *Neurosci Biobehav Rev* 2009;**33**:1024–36.

Mitterling T, Högl B, Schönwald SV, et al. Sleep and respiration in 100 healthy Caucasian sleepers—a polysomnographic study according to American Academy of Sleep Medicine standards. *Sleep* 2015;**38**:867–75.

Ohayon M, Carskadon MA, Guilleminault C, et al. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep* 2004;**27**:1255–73.

Roenneberg T, Kuehnle T, Pramstaller PP, et al. A marker for the end of adolescence. *Curr Biol* 2004;**14**:R1038–9.

A complete reference list can be found online at ExpertConsult.com.

Chapter 3

Normal Aging

Donald L. Bliwise; Michael K. Scullin

Chapter Highlights

- The integrity of sleep with advancing years is challenged not only by changes in circadian and homeostatic processes but also by medical, cognitive, and psychiatric morbidities.
- Sleep-disordered breathing and restless legs syndrome show age dependence and may contribute to poor sleep quality in old age.
- Both descriptive and interventional data suggest that sleep disturbances of all types in aging may contribute to a wide array of morbidities and, possibly, mortality and should not be dismissed by the sleep medicine specialist.
- Basic science implies that the breakdown of sleep processes in the aged organism may reflect physiologic aging at the system, cellular, and molecular levels.

As the populations of industrialized societies age, knowledge of defining how sleep is affected by age will assume greater importance. Within the United States, where the average current life expectancy is 78.7 years, the fastest growing segment of the population is those who are 85 years and older. These huge numbers force the sleep medicine specialist to confront the definition of what is "normal." Researchers often use the term *normal* to connote a variety of meanings. In sleep medicine, confusion often occurs because the term is used descriptively, to indicate representativeness, as well as clinically, to indicate absence of disease.

Aging is also subject to semantic confusion. Chronologic age has been shown repeatedly only to approximate physiologic (biologic) age. The decline in slow wave sleep (SWS), for example, can occur at a chronologic age (at least in men) far earlier than most age-related declines in other biologic functions. Some researchers in gerontology have noted that distance from death may be a far better approximation of the aging process, but too few longitudinal sleep studies in humans exist to yield these types of findings. However, studies of invertebrates have shed new light on relationships between physiologic age and sleep that can affect the functional significance of age-dependent changes (see Basic Science Considerations, later).

In addition to the issue of physiologic age, subjective age must be considered. Because the practice of sleep disorders medicine in geriatrics relies heavily on the increased self-reports of sleep disturbance seen in aging, subjective appraisal of the older person's symptoms must be considered. Whether an aged person views 75% sleep efficiency as insomnia or merely accepts this as a normal part of aging may depend largely on that person's perspective on growing old and what that means to him or her. It has been reported that older people are more likely to perceive themselves as having sleep problems if they have difficulty falling asleep rather than staying asleep, even though the latter continues to be a generally more commonly endorsed symptom (see Bliwise[1] for review). In addition, some have suggested that self-reports of sleep (relative to sleep measured by polysomnography [PSG]) are inherently less accurate and valid in older relative to younger subjects, although evidence for such age differences in other studies is decidedly mixed and varies according to the variables under consideration or the subject's sex.

Finally, normal aging must be viewed in counterpoint to pathologic aging (see Chapter 96). Although the prevalence of dementing illnesses is high in late life, determination of the number of normal elderly persons who may be in incipient stages of dementia has seldom been addressed. Additionally, recognition of mental impairments in the more limited domains of memory, executive function, language, attention, and visuospatial ability characterized as of lesser severity has led to the use of an intermediate diagnostic category termed *mild cognitive impairment* (MCI).[2] Few sleep studies of normal aging rely on extensive diagnostic work to eliminate persons in the earliest stages of mental impairment, even though PSG studies in well-defined MCI patients are now appearing.[3]

The point here is not to dismiss all that is known about sleep patterns in normal aging as inadequate but rather to point out the complexities of defining normal aging. Normal aging can never be defined without some arbitrary criteria. Throughout this chapter, we will refer to aging across several species, encompassing both what in humans may be considered "middle-aged" (approximately 40 to 65 years) and "elderly" (older than 65 years). We recognize fully the otherwise arbitrary nature of these verbal and numeric descriptors of processes that are most assuredly gradual and continuous and vary widely across individuals. It is also important to recognize that the age-dependent alterations in sleep may simply be secondary manifestations of senescence.

As in all areas of medicine, genetics are becoming increasingly recognized as affecting physiology, and this seems

particularly true for age-dependent changes in sleep patterns. In mice, age-dependent changes were strain dependent, and rebound effects (particularly for slow wave activity) from sleep deprivation were moderated by genotype.[4] In large populations of elderly persons, various actigraphic measures of sleep continuity were associated with several novel single nucleotide polymorphisms.[5]

SLEEP ARCHITECTURE

Although age-dependent alterations in sleep architecture have been described for many years,[6] only recently have attempts been made to summarize this large body of cross-sectional data using meta-analytic techniques.[7,8] Results from the first of these analyses[7] indicated that although sleep efficiency showed clear age-dependent declines up to and beyond age 90 years, most age-dependent changes in sleep architecture occurred before the age of 60 years, with few changes in SWS (now referred to as *N3 sleep* in the revised American Academy of Sleep Medicine [AASM] nomenclature[9]; see later), rapid eye movement (REM) sleep, and stage 1 percentage (N1) occurring after that.[7] Some variables (total sleep time, REM) appeared best characterized as linear decline, whereas others (SWS, wake after sleep onset) followed a more exponential course. Sleep latency showed no clear age effect after age 60 years, although it increased up to that point. A second meta-analysis focused only on REM percentage and noted a cubic trend, with REM apparently increasing after age 75 years and then demonstrating an even steeper drop after age 90 years.[8] The meaning of the latter data is unclear and raises many questions as to the extent of the precision of chronologic age to capture biologic processes in these upper age ranges. Published population-based longitudinal data on sleep architecture would assist in addressing many of these uncertainties.

Although meta-analyses can provide cumulative information on age-dependent values across many laboratories, enormous variability in parameter values exists across studies,[7] and much of the sleep architecture was not scored blindly to the patient's chronologic age or sex. This might limit the value of meta-analytic approaches for extrapolation of readily usable, age-dependent laboratory norms. By contrast, the systematically collected, rigorously acquired data derived from the centralized scoring center for the Sleep Heart Health Study (SHHS), although subject to survivor effects and based on single-night data derived from composite cohorts, offer detailed appreciation of how comorbidities, demographics, and sleep-disordered breathing (SDB) can affect observed sleep architecture values employing traditional Rechtschaffen and Kales rules.[10] Some have viewed the SHHS sleep architecture data as broadly representative of the elderly population generally because persons with a wide variety of medical conditions were not excluded.[11]

Percentage of Time Spent in Each Sleep Stage

Table 3-1 provides sleep architecture values for 2685 SHHS participants aged 37 to 92 years, excluding persons who use psychotropic medications and who have high alcohol intake, restless legs syndrome symptoms, and systemic pain conditions. About one third of these participants were hypertensive, and about 10% had a history of cardiovascular disease or chronic pulmonary disease. Results clearly show that although age effects were apparent in some measures, gender occupied a far more dramatic role in sleep architecture, in some cases showing considerable divergence when comparing women and men. Most notable in this regard was percentage of time spent in N3 (sleep stages 3 plus 4), which showed enormous gender differences at every age and, in fact, showed no appreciable decline with aging in women, relative to men.

Men demonstrate a marked cross-sectional decline with aging, as well as huge individual differences in every age group. In fact, the extent of these individual differences is emphasized by the fact that even within men as a group, coefficients of variation (ratio of variance to mean) in percentage of time spent in N3 far exceeded those for all other sleep variables in both men and women. Although gender differences in SWS have been noted previously (see Bliwise[6] for review), the fact that the age-dependent decline may be confined to men suggests a more limited utility of this often-characterized aging biomarker for women. Confirmatory results from another normative database including exceptionally healthy older adults and including only second-night data also showed stronger age-associated decline in SWS in men, though at higher absolute levels of N3.[12] The higher values may have

Table 3-1 Sleep Architecture as a Function of Age

	Percentage of Time Spent in Stage—Mean (95% CI)							
	Stage 1		Stage 2		Stages 3 + 4		REM Sleep	
Age (yr)	Men	Women	Men	Women	Men	Women	Men	Women
37-54	5.8 (5.2–6.5)	4.6 (4.1–5.3)	61.4 (60.0–62.8)	58.5 (57.1–60.0)	11.2 (9.9–12.6)	14.2 (12.7–15.9)	19.5 (18.8–20.2)	20.9 (20.0–21.8)
55-60	6.3 (5.6–7.0)	5.0 (4.4–5.7)	64.5 (63.2–65.9)	56.2 (54.5–57.8)	8.2 (7.1–9.5)	17.0 (15.2–18.9)	19.1 (18.4–19.8)	20.2 (19.3–21.1)
61-70	7.1 (6.4–7.9)	5.0 (4.4–5.7)	65.2 (63.9–66.5)	57.3 (55.7–58.9)	6.7 (5.7–7.7)	16.7 (14.8–18.6)	18.4 (17.8–19.1)	19.3 (18.4–20.2)
>70	7.6 (6.8–8.5)	4.9 (4.3–5.6)	66.5 (65.1–67.8)	57.1 (55.6–58.7)	5.5 (4.5–6.5)	17.2 (15.5–19.1)	17.8 (17.1–18.5)	18.8 (18.0–19.6)

CI, Confidence interval; REM, rapid eye movement.
From Redline S, Kirchner HL, Quan SF, et al. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch Intern Med* 2004;164:406–18.

reflected second-night adaptation effects, not available in SHHS, which relied only on single night data.

In contrast to the results of the SHHS, these gender differences in SWS were not confirmed meta-analytically.[7] At least one study has proposed that gender differences in delta activity are more likely to be a function of overall lower electroencephalogram (EEG) amplitude in men relative to women.[13] When corrected for overall amplitude, the decreased growth hormone secretion seen in postmenopausal women was accompanied by lower delta amplitude than in comparably aged men.[14] A decline in the amplitude and the incidence of the evoked K-complex over the age range of 19 to 78 years has been reported in both women and men, suggesting that similar deficits in delta synchronization processes operate equally in both sexes.[15] Gender did not play a significant role when assessed as the homeostatic response of nighttime delta power to daytime napping in either young or elderly subjects.[16] Furthermore, in a group of 20- to 60-year-olds, increased age was associated with lower slow wave density in both men and women; however, the specific characteristics of slow waves differed across genders such that women tended to show higher amplitude, faster frequency, steeper slope, and shorter positive phases for slow waves than men.[17]

Percentage of time spent in sleep stage 1 also showed similar gender-related effects in SHHS, and age-dependent increases in this sleep stage, usually considered to represent a feature of fragmented, transitional sleep, were confined to men. By contrast, percentage of time spent in REM sleep showed a modest decline with age, but the effect was detected in both men and women. REM percentages of 18% to 20% in 75- to 85-year-olds were derived from curve smoothing in a meta-analysis focused on only REM sleep measures in normal aging,[8] which were slightly lower than, but essentially similar to, SHHS data (see Table 3-1). In SHHS, sleep efficiency also declined with age, with mean values of 85.7 (standard deviation [SD] = 8.3) in the 37- to 54-year-old group, 83.3 (SD = 8.9) in the 55- to 60-year-old group, 80.6 (SD = 11.7) in the 61- to 70-year-old group, and 79.2 (SD = 10.1) in the older-than-70-years group, but without differential effects of gender, findings corroborated meta-analytically in persons older than 60 years.[7] However, the declines in percentage of time spent in REM sleep and the (male-specific) increases in percentage of time spent in sleep stage 1 seen in SHHS were not confirmed meta-analytically in persons older than 60 years.[7] The density of eye movements in REM is reduced with aging,[18] but lack of standardization across laboratories precludes examination of this aspect of REM using meta-analytic techniques.

Arousals during Sleep

Brief arousals during sleep, representing one component of the microarchitecture of sleep, continue to attract considerable interest as a metric, with particular relevance for the aged population. When examined in the laboratory, healthy older persons wake up from sleep more frequently than younger persons do, regardless of circadian phase, but they have no greater difficulty falling back to sleep.[19] Failure to maintain continuous sleep has, as its basic science counterpart, short bout lengths, a feature highly characteristic of sleep in many aged lower mammalian species (see Bliwise[6] for review) as well as nonhuman primates.[20] In elderly persons without SDB, arousal indexes from 18 to 27 events per hour have been reported.[21] Among the predominantly elderly subjects (mean age, 61 years) in SHHS, the mean (SD) arousal index showed significant but relatively small increases with age: 16.0 (8.2) for 37- to 54-year-olds, 18.4 (10.0) for 55- to 61-year-olds, 20.3 (10.5) for 62- to 70-year-olds, and 21.0 (11.6) for subjects older than 70 years.[10] Values approximating these have been reported[22] in another group of subjects without sleep apnea or periodic leg movements, thus further corroborating these SHHS values. Greater arousal index during N3 discriminated healthy older adults from patients with mild cognitive impairment.[23]

Other phasic events of non–rapid eye movement (NREM) sleep, such as K-complex and spindle density, also decrease with age.[24] Spindle density is thought to reflect, at least partially, the corticothalamic functional integrity of gamma-aminobutyric acid–ergic (GABAergic) systems. Using an automated spindle detector, one study that included adults aged 20 to 73 years found that middle-aged and older adults had reduced spindle density, amplitude, and duration, particularly in anterior derivations (Fp1 and F3 channels) that were independent of gender.[25]

Although, like other metrics of impaired sleep quality, brief arousals show a male predominance (also seen meta-analytically using wake after sleep onset[7]), the influences of age and gender are not as pronounced as the effects of breathing events (Table 3-2). In fact, when accounting for the presence of brief arousals in elderly persons, the respiratory disturbance index (RDI) predicts 10-fold more variance than age and 5-fold more variance than gender. Higher levels of RDI were also associated with slightly lower percentage of time spent in REM sleep in both men and women and with lower percentages of time spent in N3 in men.

Murine models have suggested that age differences in ability to maintain consistency of sleep state (defined with 4-second epochs) is more likely to reflect transitions involving NREM, rather than REM, sleep[26]—an effect also noted in some studies of older humans, using 2-minute bout durations.[27] Novel correlates of sleep fragmentation in elderly persons have been noted. For example, beta activity (but not delta activity) in the sleep EEG correlates strongly with sleep fragmentation regardless of circadian phase.[28] Visually

Table 3-2 Brief Arousal Index in Elderly Subjects as a Function of Sleep-Disordered Breathing

	Arousal Index: Brief Arousals per Hour of Sleep (±SD)	
RDI	Men	Women
≤5	16.7 (7.7)	14.7 (7.1)
>5 to 15	20.5 (8.7)	17.9 (7.8)
>15 to 30	25.2 (10.3)	23.2 (10.4)
>30*	39.4 (14.7)	29.7 (13.6)

*Estimated weighted values.
RDI, Respiratory disturbance index (apneas plus hypopneas per hour of sleep), a measure of sleep-disordered breathing; SD, standard deviation.
From Redline S, Kirchner HL, Quan SF, et al. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch Intern Med* 2004;164:406–18.

scored arousals in older persons have been shown to be preceded by relatively lower and more temporally limited increments in delta band power relative to similarly scored arousals in middle-aged subjects.[29] Within a population of women and men aged 55 to 100 years who wore wrist actigraphy for an average of nine 24-hour periods, chronologic age was strongly correlated with fragmentation of the rest-activity rhythms,[30] the effect being more pronounced in men and for the transition from rest to activity than from activity to rest.

Comorbidities

Insofar as comorbidities are concerned, SHHS sleep architecture data showed substantial convergence with meta-analytically derived data. In SHHS, selected medical comorbidities (e.g., a positive history of cardiovascular disease, hypertension, and stroke) were associated with disturbed sleep architecture. Consistent with results suggesting that reduced sleep amounts or quality might predispose one to the metabolic syndrome in old age, diabetic patients had smaller percentages of time in stages 3 plus 4 sleep, lower sleep efficiencies, and higher numbers of brief arousals and percentage of time spent in sleep stage 1. In most cases, however, these effects appeared to be less salient (i.e., predicted less variance) for sleep architecture than demographic variables such as gender, age (to a lesser extent), and, in some cases, ethnicity,[10] except for the arousal index, for which RDI was by far the single most powerful predictor. Less disease-specific moderator effects from meta-analytic approaches also suggested that across the entire life span, age effects were reduced substantially when persons with medical and psychiatric conditions were included.[7] The inclusion of persons with sleep apnea showed some evidence of reducing the effects of age in sleep efficiency, wake after sleep onset, and SWS when considered across the entire adult life span,[7] data that are compatible with SHHS.

Slow Wave Sleep

The gender differences in SWS reported by SHHS notwithstanding, several aspects of these data must be viewed in the context of prior literature on age-dependent changes in architecture. When analyzed with period-amplitude analyses, the major change in SWS ascribed to aging has been a decline in delta wave amplitude rather than wavelength (Figure 3-1) (see Bliwise[6] for review). The decrease in delta amplitude simply may be a more readily identifiable visual change of the sleep EEG, which is present at frequencies up to about 10 Hz, though it is difficult to see above this.[31] When scored visually using central derivations and employing a 75-μV threshold, typical figures for the amount of stages 3 plus 4 sleep in elderly persons have often been considered to fall in the 5% to 10% range. Thus the figures reported by SHHS, particularly for women, are somewhat higher than these conventionally accepted figures. Whether these values represent a more precise rendering of delta activity within sleep, perhaps engendered by the visual analyses of EEG waveforms on digital display or the simultaneous availability of precise calibration of the 75-μV criterion for delta waves stipulated by the Rechtschaffen and Kales guidelines, is unclear. Nonetheless, the controlled visual analyses conducted by SHHS are likely to represent a standard of PSG technology aspired to by the field of sleep medicine, thus arguing that these metrics may well represent how sleep architecture measures should be benchmarked.

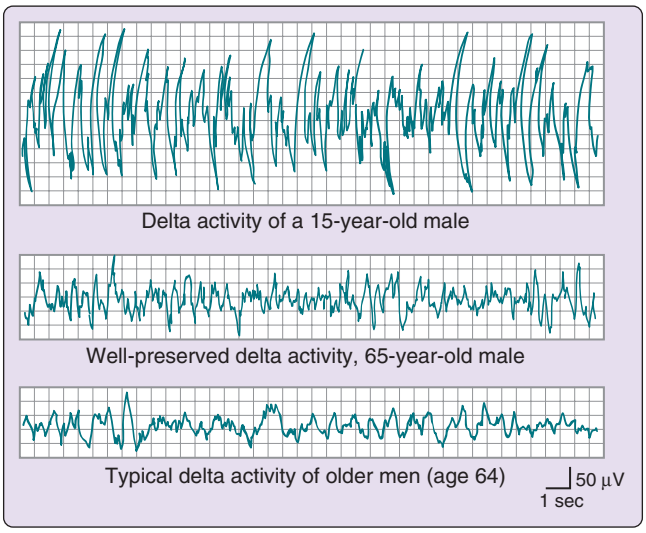


Figure 3-1 Age differences in delta activity. The *top tracing* shows typically abundant high-amplitude delta in an adolescent. The *middle tracing* shows particularly well-preserved delta in an older man. Note the marked decrease in amplitude relative to the adolescent. The *bottom tracing* is a more typical example of delta activity in an older man. Note the number of waves failing to meet the 75-μV amplitude criterion. (From Zepelin H. Normal age related change in sleep. In: Chase MH, Weitzman ED, editors. *Sleep disorders: basic and clinical research*. New York: Spectrum; 1983. p. 431–45.)

Given the current AASM guidelines for sleep stage scoring,[9] much of the foregoing normative data on sleep architecture may have limited relevance for laboratories that elect to adopt such changes. For example, slow wave activity has higher amplitude when recorded from frontal derivations relative to central derivations. This would be expected to result in increased levels of visually scored slow wave (i.e., N3) sleep. One study comparing recordings scored with both revised AASM and traditional Rechtschaffen and Kales criteria have shown a number of significant differences in resulting measures.[32] Predictably, particularly in older persons, the revised scoring system resulted in higher percentages in N3 sleep. Given that middle-aged subjects also show decreases in delta activity, most pronounced frontally but also in central and parietal and occipital derivations,[17] such effects are probably not limited to elderly persons. Beyond creating the need to establish new normative data, the mechanistic and functional significance or the diagnostic and therapeutic importance of such a revisionary approach remain obscure. Much the same effect could be obtained by adopting alternative scoring thresholds of less than 75 μV for defining delta wave activity. Such proposals were put forth in the 1990s (see Bliwise[6] for review), but have not led to enhanced understanding of the age-dependent changes in SWS. Eventually, digitized indexes of delta activity (e.g., fast Fourier transform, zero-crossing, or hybrid techniques) might come to replace such conventional measures; however, considerable controversy regarding filtering, sampling rates, and data storage formatting leaves formal adoption of such approaches dubious for routine clinical purposes at this time,[33] though such efforts at signal processing are yielding important new clues regarding the significance of sleep-related delta activity for aging. Appreciation of individual differences in slow wave activity must also take into

account that the amount of delta demonstrated may reflect, at least in part, the variable number of tandem repeats in the PER3 polymorphism, an association first demonstrated in younger, and now in elderly, persons.[34]

Slow wave activity during sleep may represent synaptic downscaling and memory consolidation processes, which are viewed as critical for neural efficiency and memory retention[35,36] (see Chapter 22). Given at least some data suggesting decreased SWS with age, such findings might fit with mild impairments in cognition that characterize normal aging. Extremely low-frequency (<1.0 Hz) slow wave activity in NREM sleep has been thought to hold particular significance as a more immediate reflection of cellular processes than conventionally defined delta activity.[37] In dementia, the integrity of the normal auditory evoked response (K-complex) may be impaired,[38] whereas in normal aging, there is some suggestion that spectral power at frequencies below 0.7 Hz might demonstrate fewer age differences than at between 0.7 and 3.0 Hz.[39] Other attempts to examine NREM sleep in old age using nonlinear, dynamic EEG approaches yield compatible findings. The sleep EEG of healthy older adults in the first NREM period may resemble patterns of compensatory activation similar to those seen in young adults during sleep subsequent to sleep deprivation.[40] Whether such phenomenologic parallels of altered functional connectivity in sleep deprivation and aging hold prognostic or practical significance at the individual case level is unknown but certainly plausible. Nonlinearity of the NREM sleep EEG increases with normal aging[40] and sample entropy, a novel measure reflecting dynamic probabilities of state,[41] also changes with age. Less SWS in older humans was related to greater cortical atrophy as determined neuopathologically.[42]

CIRCADIAN RHYTHMS IN AGING

In humans, most descriptive data collected under entrained conditions have suggested that the amplitude of the sleep-wake rhythm, body temperature (Figure 3-2),[43] and some hormones decrease with aging. Sex differences in such phenomena have also been reported (see Bliwise[44] for review). However, in exceptionally healthy older adults such differences might not always appear,[45] and a study of centenarians indicated relatively robust neuroendocrine profiles.[46] A phase advance of aging has often been ascribed to the timing of sleep patterns in older adults. Many sleep medicine specialists use this designation as an abbreviation for indicating that older persons go to bed earlier in the evening and wake up earlier in the morning than younger persons, findings corroborated in age differences in the timing of bedtimes and wake-up times in dozens of cross-sectional surveys (see Bliwise[6] for review) and even longitudinally.[47]

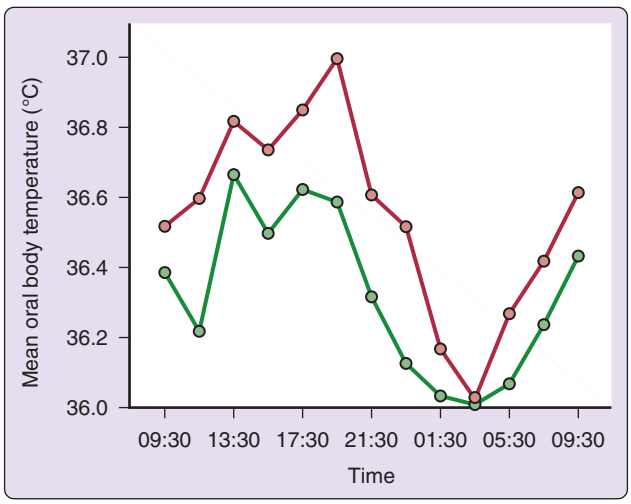


Figure 3-2 Oral temperatures in young (*red circles*) and old (*green circles*) subjects, showing apparent decreased amplitude and earlier phase in body temperature cycle as a function of aging. Data were obtained under entrained conditions. (From Richardson GS, Carskadon MA, Orav EJ. Circadian variation of sleep tendency in elderly and young adult subjects. *Sleep* 1982;5[Suppl 2]:S82–94.)

However, the reader should be aware that fundamental changes in phase relationships in human circadian rhythms might not always substantiate the notion of such an advance. For example, in the constant routine protocol, older subjects' typical earlier bedtimes and wake-up times relative to younger subjects actually were more phase delayed, rather than phase advanced, relative to their peaks in melatonin.[48] The implication of this finding is that the earlier bedtimes and wake-up times may be due to homeostatic factors. Studies using 28-hour forced desynchrony, an experimental protocol that requires subjects to sleep on a 2:1 wake-to-sleep ratio outside the limits of entrainment of the circadian system, have shown absence of age differences in estimates of tau, the endogenous period length of the human core temperature rhythm.[49] Again, these results implicate factors other than circadian ones that may be responsible for the earlier bedtimes of older humans. Further evidence of interaction between circadian and homeostatic factors comes from a study of age differences in the melatonin rhythm under the constant routine protocol, during which younger subjects showed elevated melatonin levels during the higher homeostatic pressure induced by this procedure, whereas older subjects did not.[50]

The interaction between homeostatic and circadian influences in older humans has been described elegantly in the forced desynchrony protocol, which has shown that throughout the assigned (9.33-hour) sleep period, elderly subjects awakened more frequently than younger subjects, regardless of circadian phase.[51] The duration of awakenings was virtually identical in young and old subjects, but the largest differences in the frequency of awakenings between young and older subjects were detected early, rather than late, in the sleep period. These results appear incompatible with the broadly defined phase-advanced hypothesis of sleep in elderly persons, in which differences in sleep consolidation would be predicted to be most pronounced late in the sleep period (corresponding to early morning awakening) and least pronounced just after sleep onset (corresponding to early evening sleepiness). This study also analyzed sleep structure immediately before awakenings from consolidated periods of sleep. With circadian phase controlled, older subjects were far less likely to awaken from stage 1 and far more likely to awaken from stage 2 than were young subjects, suggesting that awakenings in the older subjects probably represented abrupt transitions from NREM sleep rather than gradual lightening of sleep.[51] These findings can be interpreted as an indication of a reduced homeostatic pressure for continuous bouts of sleep independent of circadian phase in the sleep of older persons.

A well-described feature of the circadian system in the aged organism is the relative impairment in the ability to

phase-shift. This may in part reflect loss of rhythmic function within the suprachiasmatic nucleus.[52] In humans, the sleep-wake system appears particularly vulnerable to such changes in phase shifting[53] and might account for some of the apparent self-selection out of shift work typically seen in older people (see Bliwise[54] for review). In rodents, such impairments in phase shifting have also been described.[55] Although both photic and nonphotic influences on impaired phase-shifting ability have been described in aging (see Bliwise[54] for review), the ability to phase-shift and entrain to light in old age might be expected to be particularly impaired because of challenges to the visual system that occur as a part of human aging (e.g., cataracts, macular degeneration). Perhaps as a consequence, in epidemiologic studies, elderly persons with visual impairments were 30% to 60% more likely to have impaired nighttime sleep relative to visually unimpaired elderly subjects.[56] In the laboratory setting, however, disagreement exists as to whether, among persons without visual impairment, responsiveness of the circadian system, as indexed by melatonin rhythms, is decreased[57] or unchanged[58] by light exposure. Finally, another study examined the ability of a nonphotic stimulus (evening exercise) to phase-shift melatonin onset in young and elderly persons and found comparable phase delays subsequent to exercise in both groups, arguing for the importance of entraining factors other than bright light to affect rhythms in elderly persons.[59] Despite this, neuroimaging indicated that alternate exposure to darkness and blue light indicated that aging was associated with impaired reactivity in an array of visual, alertness, and executive function–related brain regions in older relative to younger adults.[60]

Invertebrate models in studies of sleep and aging have been broadly confirmatory for many of the changes described in mammals, such as the decline in nocturnal sleep durations.[61] Several studies in the fruit fly (*Drosophila melanogaster*)[62] and in the honeybee (*Apis mellifera*)[63] have indicated a dispersal of sleep around the 24-hour day occurring with aging. Perhaps more noteworthy is that these studies have also implied that total amount of sleep per 24 hours does not merely redistribute (by shorter bout length and increased number of bouts) but also might increase in quantity with advanced physiologic age. These changes appeared particularly pronounced in male fruit flies[62] and female honeybees[63] near the end of life. Because an avowed function of the mammalian master clock, located in the suprachiasmatic nucleus, is to provide clock-dependent alerting, these data are consistent with early studies that demonstrated such a role for the suprachiasmatic nucleus in primates using lesion models.[64] Although studies involving neurodegenerative disease in aged humans have shown increased sleep per 24 hours consistent with level of dementia (see Chapter 96), the invertebrate data imply that increased rather than decreased sleep durations would be likely to occur in aged humans normatively on a population-wide basis. In this regard, it is noteworthy that the largest single compilation of human sleep durations ever published (more than 1 million subjects) shows that reported sleep durations increase, rather than decrease, with age.[65] Taken together, these data suggest that increased sleep durations in aged humans might fundamentally represent biologic rather than sociocultural factors.

Among lower vertebrates, aged zebrafish (*Danio rerio*) have been shown to demonstrate reduced rest-activity rhythm and shorter nocturnal sleep durations across the life span (from 1 to 4 years), and these were associated with decreased brain melatonin during the dark period.[66] Melatonin administration was shown to partially restore sleep durations and simple measures of learning in this species. Zebrafish aging was also shown to be associated with gene expression in several clock genes (*Bmal1, PER1*),[66] although the translation of this finding back to older humans might not be clear.[67] Additionally, in rodents, age effects in the expression of several clock genes may occur equally in both wild-type and mutant lines.[68]

CAUSES AND CONSEQUENCES OF POOR SLEEP IN OLD AGE

Causes

The prevalence of insomnia in older adults varies across studies, but figures between approximately 20% and 40% are typically reported. In one of the largest surveys of an American population (the Established Populations for Epidemiologic Studies of the Elderly, with more than 9000 participants), 29% of the older-than-65-years population reported difficulty maintaining sleep.[69] Relative to sleep maintenance, sleep latency is less likely to be problematic for the elderly population, with prevalence figures on the order of 10% to 19% typically seen, although one study[70] reported a relatively high prevalence of sleep latency problems (36.7%) in a largely rural aged population. In nearly all studies, elderly women have a greater probability of sleep complaints and sedative-hypnotic use than elderly men (see Bliwise[1] for review). Conflicting data exist on racial differences in sleep complaints,[71] though health disparities in sleep durations did not appear to depend on age.[72] Some large-scale survey data have questioned whether reports of sleep as a "problem"[73] or having "insufficient" sleep[74] show any age dependence whatsoever,[75] but such approaches ultimately confound elderly persons' reports of their health behaviors with their judgments of well-being, which generally improve with advancing chronologic age. Social science research has shown that as people age, perceived life satisfaction increases, reflecting what has been termed a "positivity effect"[76] or a "paradox of aging"[77] in many domains of life.[78] These types of biases are likely operating when populations are asked about their sleep using questions such as these.

Figures for regular use of sedative-hypnotics in elderly populations range from as low as 5% per year to as high as 16% over 4.5 years, 29% over 8 years, 34% over 1 year, and 62% over 3 years (see Bliwise[1] for review). Although subjective reports of poor sleep invariably increased with age, and numerous changes in sleep architecture have been frequently noted (see earlier), the relationship between these two domains is significant, though relatively weak, with women showing slightly higher levels of association relative to men.[79]

Disturbed sleep is thought to both contribute to and reflect allostatic load of illness in old age and might reflect physiologic age of the organism. When other causes of poor sleep are taken into account, chronologic age might explain little of the observed higher prevalence in elderly persons. Psychiatric conditions have always been considered to play a major role in the insomnia of old age. A study of more than 3000 older men reported that depressive symptoms were associated with difficulties in falling asleep, but not lower sleep efficiency or total sleep time.[80] Among women, although the caregiving

role per se was not associated with poor sleep, for women who were depressed, caregiving was associated with reported sleep problems.[81] Regardless of caregiving status, anxiety appeared to play a larger role in predicting poor sleep quality in these older women, even relative to depressed mood per se.[82] However, poor sleep, left untreated, appears to be a risk factor for incident depression in elderly persons.[83]

Limitations in mobility, visual impairment, lack of regular exercise, alcohol use, and smoking all contribute to declining sleep quality in older persons, as do chronic pain conditions, such as arthritis, hip fracture, fibromyalgia, headache, and back pain; cardiovascular diseases such as hypertension, myocardial infarction, stroke, congestive heart failure, and angina; respiratory conditions such as asthma and bronchitis; and other systemic diseases such as diabetes, gastroesophageal reflux, and duodenal ulcer (see Carrier and Bliwise[84] for review). When persons with such comorbidities are eliminated from consideration, the resulting insomnia prevalence in elderly populations may be only 1% to 3%,[85] reiterating the decreased significance of chronologic age per se in predicting poor sleep. In women, some evidence suggests that menopause may also be associated with declining sleep quality (see Chapter 159). Retirement from work has been shown to be related to fewer sleep problems.[86]

Additionally, in both women and men, the otherwise mundane occurrence of nocturia (nightly awakenings to void) appears to be associated with poor sleep,[87] even when other factors such as pain and medical comorbidity are taken into account. In fact, nocturia may be the single most common factor associated with poor sleep in elderly persons.[87] The duration of the first uninterrupted sleep period, otherwise referred to as the time to first void, has been shown to correlate with nearly all measures of whole-night sleep quality,[88] and lengthening of the first uninterrupted sleep period in a population containing a substantial portion of elderly persons was associated with improvements in many such measures.[89] Effects were independent from age. Nocturia is an important consideration in the lives of older people because it has been related to lower quality of life,[90] depression,[91] and mortality, even with some control over comorbidities.[92] Some have questioned the basis of the causal association between nocturia and poor sleep, arguing that the effects on disturbed sleep appear of small magnitude[93] when measured actigraphically[94] or that awakenings precede the voiding episodes. The latter explanation appears incompatible with evidence that elevated detrusor activity during sleep temporally precedes such episodes,[95] as well as improvements in sleep when nocturnal diuresis is treated.[96]

SDB was associated with the presence of nocturia in the Sleep Heart Health Study,[97] perhaps through increased levels of brain naturietic peptide or decreased levels of arginine vasopressin (see Bliwise[98] for discussion), though the former appeared more likely in one case series of men older than 60 years.[99] A Danish case-control study found an association between nocturia and SDB only when nocturnal polyuria (i.e., >33% of urine volume produced at night) was present.[100] Treatment with continuous positive airway pressure (CPAP) has been reported to reduce the number of nocturnal voids in some[101] but not all[102] studies, though interestingly, behavioral treatments for poor sleep in the absence of SDB can also reduce nocturia.[103] Although the relation between nocturia and lower sleep quality has been noted to be independent of age, and individuals of all ages report being "bothered" by the phenomenon,[90] an association between nocturia and SDB was reported to weaken with age in one study,[104] which might account for the negative findings in a recent major CPAP trial in elderly persons.[102]

Several longitudinal studies examining the incidence (development of new cases) of insomnia over periods of up to 10 years have been reported. The single best predictor of insomnia continuing longer than 10 years was insomnia at a previous time, although cardiovascular and pulmonary comorbidities conferred risk as well in the older-than-65-years population.[105] Reported remissions were less likely in older subjects than in younger ones.[106] The Established Populations for Epidemiologic Studies of the Elderly data indicated a yearly incidence of insomnia complaints in the aged population of about 5%, with a spontaneous remission rate of about 50% over 3 years.[107] In these data, incident insomnia was related to heart disease, stroke, hip fracture, and new-onset depression. Spontaneous remission of insomnia was related to the resolution of depression, physical illness, and physical disability affecting activities of daily living,[107] whereas in the Cardiovascular Health Study, persistence of insomnia was associated with unresolved depression.[108] Important from the standpoint of prevention, another study reported that higher levels of physical activity were protective for incident insomnia over an 8-year period.[109] Two Scottish cohorts of different ages at entry (36 to 57 years and 56 to 76 years), including both men and women followed over 20 years for sleep complaints, found strong evidence that manual labor was associated with chronic (unremitting) and incident insomnia.[110]

Potential Consequences

A major question regarding the frequent complaints of poor sleep among elderly persons involves whether these have an impact on their health. If the poor sleep of old age, although annoying and distressing for many, represents primarily a quality-of-life issue, albeit one modifiable by medical or behavioral interventions, it might cast a different perspective on this problem, than would a medical disorder such as SDB, for which negative outcomes may be better defined and quantified. There is no question that almost universally, poor nocturnal sleep is distressing and related to lower quality of life of many older persons. This has been demonstrated in elderly populations in the United States,[111] Canada,[112] Europe,[113] Asia,[114] and Africa.[115]

Most work-relating poor sleep quality or duration to putative adverse outcomes has been observational, which, although often provocative, lacks the definitive element of proof of causation that is afforded by randomized clinical trials. Unfortunately, with few exceptions, most pharmacologic and nonpharmacologic randomized clinical trials attempting to treat poor sleep in elderly persons seldom rely on outcomes other than conventional subjective and PSG measures of nocturnal sleep per se. Rare exceptions to the latter have been several insomnia treatment studies among older adults that have demonstrated increases in selected quality-of-life measures such as the SF-36[116] or decreased daytime napping.[117] Data relevant to interventions for poor sleep and their effects on other medical outcomes in old age (e.g., hypertension, insulin resistance) have yet to be published.

Among observational studies, the association between nocturia and insomnia has led to speculation that the more likely

a person is to rise from bed during the night to use the bathroom, the more likely the person is to fall.[87] Considerable evidence for this association exists at the population level, where studies have shown associations between insomnia and falls.[118] Sleep durations of less than 5 hours were associated with an increased risk for falls of more than 50%.[119] Risk for increased falls with short duration of sleep or poor quality of sleep (or both) is also consistent with data suggesting that insomnia is associated with impaired physical function. For example, lower sleep efficiencies were associated with lower grip strength and slower walking speed in a population of elderly men.[120] Short sleep durations were associated with a slightly different set of markers of physical impairment in elderly women, primarily consisting of chair-to-stand speed.[121] Although the increased risk for falls in older populations has been typically ascribed to psychotropic and sedative-hypnotic medications, reanalyses of some of these databases have suggested that poor sleep per se may be a more relevant predictive factor.[122]

Relative to poor-quality sleep, at least some data suggest adverse outcomes associated with short sleep durations in older populations. Gangwisch and colleagues[123] reported that sleep durations of less than 5 hours were associated with higher rates of all-cause mortality in subjects 60 years and older, a finding that was not present in persons ages 32 to 59 years. Sleep durations of less than 5 hours per night in even older populations (67 to 99 years old) were associated with obesity as well,[124] a finding otherwise well acknowledged in populations younger than 65 years, in women in one study[125] and in men and women in another.[126] Hypertension has also been associated with short sleep durations in elderly persons,[127] and with decreased N3,[128] but other studies of older populations indicated that neither short sleep durations[129] nor complaints of poor sleep[130] were associated with this morbidity. Diabetes and impaired glycemic control were associated with sleep durations of less than 6 hours (but not insomnia complaints) across the age range of 53 to 93 years[131] and were independent of age across an even broader age range.[132] As mentioned earlier, these are all observational studies, which, although impressive by size of the samples studied and control over confounding variables, did not manipulate sleep quality or sleep duration to demonstrate improvement in any of these putative adverse outcomes in older persons.

An area of emerging interest is whether changes in sleep with aging have ramifications for memory and cognitive changes.[133] In young adults, sleep is theorized to benefit learning, memory, and cognition through synaptic downscaling,[134] memory reactivation,[135] and increased alertness (see Chapter 22). With increasing age and disease, there are not only changes in sleep quality (as described previously) but also declines in memory, attention, executive function, and processing speed cognitive domains.[136] Thus it is possible that poor sleep might contribute to declining cognition in aging adults and possibly to the development of cognitive disorders such as MCI and Alzheimer disease.

A substantial literature has now reported significant correlations between sleep measures (self-report, actigraphy, and PSG) and performance on cognitive tests in middle-aged and older adults, after correcting for demographic and health-related confounding variables. In cross-sectional studies, middle-aged adults consistently demonstrate detrimental relationships between cognitive performance and self-reported short sleep, long sleep, difficulty falling asleep, and nighttime awakenings; by contrast, older adults only consistently show cross-sectional associations between cognitive performance and self-reported long sleep duration or delayed sleep-onset latency.[133] Cause and effect are difficult to distinguish in such correlational studies, but several longitudinal studies have now connected short and long (self-reported) sleep at baseline in middle-aged adults to subsequent cognitive decline up to 28 years later.[137] For example, in the Whitehall II middle-aged cohort that included more than 5000 participants, reported short and long sleep at baseline, as well as longitudinal changes in sleep duration, was predictive of accelerated cognitive decline on a battery of tests 5.4 years later.[138] Some large cross-sectional studies have suggested that different aspects of reported lower sleep quality, rather than measured SDB, were most strongly associated with both amnestic and nonamnestic types of MCI.[139] Epidemiology studies—including Study of Osteoporotic Fractures (SOF) and Osteoporotic Fractures in Men (MrOS)—that have use actigraphy to define sleep-wake state, have tended to provide converging results using more abbreviated cognitive test batteries and typically implicate wake after sleep onset or sleep efficiency rather than total sleep duration[140] as relevant correlates of impaired cognition. Few studies have evaluated age-dependent cognitive changes in relation to PSG variables since Irwin Feinberg's seminal work,[141] but lower REM sleep quantity and density have been related to cognitive decline in both 3-year[142] and 14-year[143] longitudinal studies. It remains unclear whether declining REM sleep causes cognitive functioning to worsen or if age- and disease-related declines in cholinergic neurotransmission drive both REM sleep declines and cognitive declines.

Experimental studies that have manipulated sleep duration or specific features of sleep have also produced noteworthy findings. Almost every sleep deprivation study that examined cognitive outcomes in relation to aging (see Scullin and Bliwise[133] for review) has shown that sleep deprivation has less of an impact on cognitive performance in older adults than in young adults (e.g., see Duffy and colleagues[144]). This pattern might indicate that older adults need sleep less than young adults, but defining age-dependent change in sleep need and the overall meaning of sleep deprivation effects in older adults is still vigorously debated. Another experimental approach is to attempt to increase total sleep duration (e.g., by having adults take afternoon naps). When measuring cognitive performance before and after a single nap (relative to a wake condition), nap-related benefits to cognitive performance are consistently observed in middle-aged adults.[145] Similar nap-related benefits to cognition also seem to emerge in interventional studies in which participants are asked to attempt to take an afternoon nap every day for a month.[146] However, perhaps corresponding to the sleep deprivation literature, napping studies that have used older age groups have often failed to observe any cognitive benefits of napping.[147] Pharmacologic enhancement of sleep, specifically defined as increases in spindles[148] and slow wave activity, seems to benefit cognition in young and middle-aged adults[149] but does not greatly benefit older adults.[150] A rodent study suggested that sleep deprivation resulted in increased dendritic branching in prefrontal cortex in older animals but decreased density in the hippocampal CA1 layer,[151] which is consistent with the divergence of results shown in the human behavioral studies cited previously. The mechanisms involved in these differential

results by age are uncertain, though some speculation has focused on age differences in the adenosine A(2A) receptor gene in response to sleep loss.[152]

One mechanism by which poor sleep could affect cognitive functioning with advancing age is by reducing the efficiency of sleep-dependent memory consolidation (i.e., the stabilization and integration of learned experiences during the day). Animal studies have found reduced hippocampal reactivation of place cells during sleep (i.e., memory "replay").[153] In humans, sleep-dependent memory consolidation also appears to decline steadily across the life span. Memory consolidation effects are larger in children than in young adults,[154] larger in young adults than in middle-aged adults,[155] and larger in middle-aged adults than in older adults.[156] Aging appears to detrimentally affect both procedural (motor) memory and declarative (episodic or explicit) memory consolidation, and adults in the eighth and ninth decades of life may show no memory consolidation effects at all.[157]

Another, not mutually exclusive mechanistic influence whereby poor sleep could accelerate cognitive decline involves amyloid deposition. Basic science studies have found that sleep facilitates the clearance of brain metabolites including ß-amyloid[158] and that sleep deprivation in rodents is linked to increased amyloid plaque burden.[159] In humans, self-reported short sleep duration and actigraphy-measured sleep efficiency have been associated with higher ß-amyloid burden, as measured with positron emission tomography[160] or cerebrospinal fluid (CSF).[161] It is therefore perhaps not surprising that epidemiology studies have found that poor sleep quality can predict development of MCI[162] and Alzheimer disease.[163]

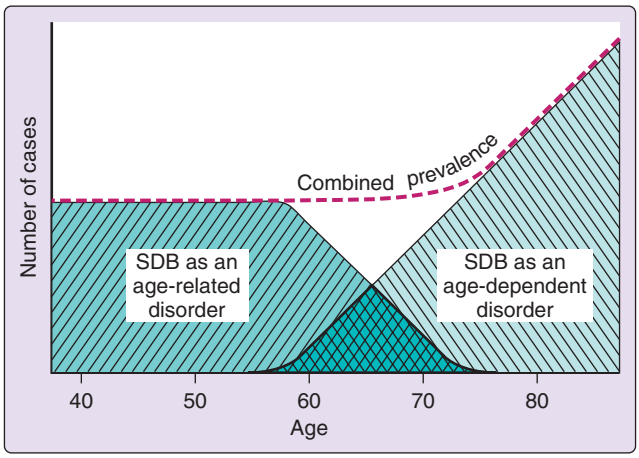


Figure 3-3 Heuristic model suggesting that sleep-disordered breathing (SDB) is both an age-related and an age-dependent condition with potential overlap of distributions in the 60- to 70-year-old age range. Cross-sectionally, note that the number of cases observed can remain high and increase with age, despite a presumed decrease in age-related SDB.

RESTLESS LEGS SYNDROME AND PERIODIC LIMB MOVEMENTS DURING SLEEP

One specific cause of insomnia in elderly persons is restless legs syndrome (RLS) (see Chapter 95). This condition, characterized by an urge to move the legs, which is usually accompanied by sensations of discomfort, aggravation of symptoms by rest and temporary relief of symptoms by movement, and worsening during the evening or nocturnal hours, is exceedingly common in elderly populations. Estimates vary, but the condition appears to be more prevalent in northern European[164] relative to Asian populations,[165] and several genotypes have been identified (see Chapter 95). Peak prevalence was noted in the group aged 60 to 69 years for women (16.3%) and 50 to 59 years for men (7.8%),[164] though the population sampled included persons up to age 90 years. Another European study including subjects up to their 80s also showed similar gender differences (14.7% in women; 6.8% in men), with peak prevalence for both genders in the 50- to 59-year-old range.[166] Thus, in some respects, RLS prevalence appears to be more age related than age dependent (see later and Figure 3-3), although at least one study of a Dutch population noted that prevalence was highest in the oldest subjects (80 to 100 years of age).[167]

A Finnish longitudinal study of individuals in their 60s at baseline reported decreased prevalence over 10 years, but a survivorship phenomenon could not be ruled out.[168]

Periodic limb movements during sleep (PLMS) are stereotypic, repetitive, nonepileptiform movements of the legs usually consisting of dorsiflexion of the ankle but occasionally limited to flexion of the great toe or incorporating flexion at the level of knee or hip. They often, but not invariably, occur in conjunction with RLS. The intermovement interval has been reported to decrease with age from about 24 to 28 seconds before the age of 55 years to about 14 to 16 seconds after the age of 65 years.[169] Age-dependent increases in the occurrence of PLMS have been noted cross-sectionally in series without a drop in the oldest (e.g., >80 years) groups.[170] Curiously, longitudinal follow-up of elderly subjects did not show increases consistently,[171] perhaps owing to inherent variability in PLMS. Prevalence, defined as a periodic leg movements index of 15 movements or more per hour, has been estimated as high as 52% in a population of older women.[172] When measured with wrist actigraphy in that population, the presence of PLMS was associated with sleep durations of less than 5 hours of sleep,[173] though earlier population-based studies have presented conflicting data as to whether PLMS, in the absence of frank RLS symptoms, were associated with poor sleep.[170] One study of men aged 40 to 60 years noted poorer sleep quality when the periodic leg movements index exceeded 10.[174] Other studies demonstrating the mixed pattern of results correlating PLMS with sleep complaints have been reviewed elsewhere.[175] Noteworthy is a large study reporting on simultaneous PSG measurements of sleep architecture and piezoelectric recordings of leg movement activity in 2872 men that reported strong associations between higher levels of PLMS or PLMS with arousals and multiple measures of poor sleep quality (lower sleep efficiency, higher N1 percentage, lower N3 percentage).[176] In a normative study of more than 1000 individuals across a broad age range, PLMS increased with aging in both men and women, but the age effect was three times as strong in men.[177]

One possible explanation for the variability of results across studies is that PLMS may vary considerably from night to night. A 15-night study suggested that estimated prevalence for PLMS might stabilize only after multiple nights of measurement.[178] The discrepancy between the higher prevalence of PLMS relative to RLS and the failure of a number of studies to show associations between their presence and

specific symptoms suggests that in many elderly persons, PLMS may be an incidental finding.[175]

The worsening of RLS and PLMS with aging suggests that this syndrome may be associated with other conditions known to be common in older populations. Given the likelihood of anemia among elderly persons, current attention has focused largely on iron transport and storage deficiencies.[179] Elderly RLS patients with serum ferritin levels of less than 45 mg/mL showed subjective improvement following use of ferrous sulfate, although their total iron levels were no different.[179] These findings were later replicated by the same research group.[180] Because iron represents a key component of production of dopamine, it could play a role in presence of RLS in some elderly subjects. One population-based study could not confirm the ferritin finding,[181] although another report indicated that higher serum-soluble transferrin receptor levels (often characteristic of early-stage anemia) and lower serum iron levels were associated with RLS.[164] An interesting perspective on iron metabolism and aging was based on examination of ferritin levels in the CSF of elderly RLS patients. Older patients had higher CSF ferritin levels than did younger patients; however, for elderly patients whose RLS had been long-standing, lower levels were associated with a more severe condition.[182]

Heightened awareness of both PLMS and RLS as phenomena associated with cardiovascular comorbidities (see Chapter 95) provides new impetus that these conditions are taken seriously. In a cross-sectional analysis of more than 500 patients in their mid-60s, the presence of a periodic limb movements index higher than 35/hour was associated with multiple measures of left ventricular dysfunction[183] and, in a subset of these patients, was also predictive of progression of atrial fibrillation over a median interval of nearly 3 years.[184]

SLEEP-DISORDERED BREATHING

Specific considerations related to diagnosis and treatments of SDB in elderly persons are covered in Chapter 152. This section deals with more general issues involving age dependence.

The previously proposed heuristic model for SDB (see Figure 3-3) posits that SDB represents both an age-related phenomenon (with a specific vulnerability confined to middle age) and an age-dependent phenomenon (with a prevalence that steadily increases throughout the human life course).[1] The articulation and differentiation of these two presumably separate but chronologically overlapping distributions represent a major challenge to clinicians. Practically, if the health consequences of SDB in elderly populations are diminished, the necessity to treat the enormous numbers of elderly persons who have the condition is reduced. Age dependence implies that SDB risk factors might be best considered markers of physiologic or biologic age.[185] Chronologic age may thus serve only as a proxy for other risk factors that are themselves age dependent.

Risk Factors

Risk factors for SDB in the older population may differ to some extent from those in middle-aged populations. In SHHS, several markers of obesity that were significant cross-sectional predictors of SDB in middle-aged populations (neck circumference and waist-to-hip ratio) were no longer significant predictors by age 70 years and 80 years, respectively,[186] although body mass index continued to be correlated with SDB, even past age 80 years, albeit with a somewhat diminished effect. Although the male predominance in SDB is thought to equalize in old age, this was not the case within SHHS.[186] Other cohort studies including older subjects suggested roughly equal prevalence in elderly men and women.[187,188]

The prevailing view for many decades was that most SDB in elderly persons consisted of central (i.e., diaphragmatic) events, whereas in the middle-aged population, obstructive events predominated; however, this is unsubstantiated by both descriptive studies, showing the predominance of obstructive apneas, and by pathophysiologic studies, which show increased tendency for upper airway collapse with aging.[189] Upper airway resistance has been reported to be higher in both REM and NREM sleep in older men relative to younger men,[190] and closing pressures during sleep were higher in older subjects in N2 sleep relative to younger subjects.[190a] Acute reduction of CPAP pressure during NREM sleep results in enhanced collapsibility of the upper airway in older, relative to younger, persons.[189] Aging has been associated with lengthening of the soft palate and with upper airway fat pad deposition, both of which may contribute to oropharyngeal collapse during sleep.[191] Lower lung volumes have been shown to predict incident SDB in elderly persons over time,[192] perhaps by providing less caudal traction on the trachea and hastening upper airway collapse during sleep. In older animals, the pharyngeal muscles appear to have a worse profile for endurance relative to the diaphragm, which may enhance susceptibility to collapse,[193] and a shift from type IIa to IIb fibers occurred in the genioglossus in 24-month-old rats, a finding interpreted as conferring susceptibility to fatigue.[194] Remodeling of the motor unit firing pattern of the genioglossus in older human subjects may be a counterpart to these histologic changes.[195] Associations between weight loss and SDB development over intervals of up to 30 years in aged humans represents proof of concept that generalized muscle weakness (sarcopenia) could underlie incident SDB in older adults.[196]

Predisposing influences on SDB in elderly human populations are not limited to neuromuscular factors. Ventilatory control instability,[197] which may be accentuated by the decrease of N3 sleep with age, might also predispose to SDB in elderly persons, though not all studies report high loop gain in older subjects.[198] Recent modeling studies using abrupt decreases of CPAP pressures have suggested diminished ventilatory feedback in the sleep of older, relative to younger, subjects, a finding shown to reflect decreased controller gain.[189] These and other potential age-dependent risk factors for SDB are shown in Figure 3-4.

Outcomes

Potential outcomes relevant to SDB in old age include mortality, cardiovascular and neurobehavioral morbidities, and morbidities related to other potential end-organ damage (see Figure 3-4). The prevailing viewpoint in sleep medicine has been that SDB demonstrates weakened associations with morbidities in elderly, relative to middle-aged, persons. Offered here is a brief description of studies in older populations suggesting otherwise.

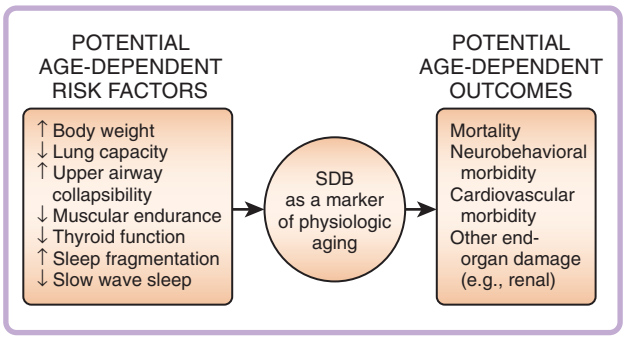


Figure 3-4 Sleep-disordered breathing (SDB) in older adults as an age-dependent condition. Other potentially associated age-dependent risk factors and outcomes are shown.

In elderly persons, SDB has been associated cross-sectionally with clinically defined hypertension,[199] a nondipping blood pressure pattern,[200] composite cardiovascular disease history (in men),[201] stroke,[202] reduced kidney function (in men),[203] poorer physical function (in men),[120] nocturia,[204] overactive bladder,[205] and impaired cognition (in women).[206] Longitudinal data have shown relationships between snoring and daytime sleepiness and incident cardiovascular disease[207] and between declining mental status test scores and the development of SDB,[208] though the effect sizes for such associations may be small for cognition.[209] Additionally, higher health care costs were associated with sleep apnea in both middle-aged and elderly persons.[210] An important association between SDB and frailty has also been noted in older women,[211] which is particularly important given the fact that, as a well-acknowledged geriatric syndrome, frailty is highly predictive of other morbidities and of mortality.[212] Moderate to severe SDB was related to all-cause, cancer, and stroke mortality in a middle-aged population studied over 20 years,[213] and, at least for all-cause mortality, the presence of daytime sleepiness conferred additive risk.[214] As outlined elsewhere,[1] whether age moderates the relationship between SDB and mortality is controversial. Several natural history studies of SDB in old age continue to report absence of associations between SDB and mortality[215] or present data interpreted as suggesting that SDB does not progress over a relatively short interval of 3 years.[216] Longitudinal data collected over 20 to 30 years suggest otherwise.[196]

In SHHS, when subjects with prevalent cardiovascular disease were excluded, relationships between SDB (as measured by quartiles of the apnea-hypopnea index) and various morbidities (including diabetes and hyperlipidemia) were clearly lower in the older-than-65-years population than in the younger-than-65-years population, but only in men, not women, where the associations were similar.[217] By contrast, Haas and colleagues[218] reported that isolated systolic hypertension was unrelated to SDB in any age range but that systolic and diastolic hypertension were related to SDB in only those younger than 60 years. In another report examining associations between multiple measures of SDB and more broadly defined cardiovascular disease (including coronary heart disease, congestive heart failure, and stroke), relationships with SDB, although reduced to some extent by age, were still age independent.[219] Echocardiography suggests that left ventricular diastolic and systolic dysfunction occurs in elderly patients with SDB.[220]

Despite this suggestive evidence, other studies continue to minimize the significance of SDB for elderly populations. For example, it has been contended that sleep apnea has little effect on quality of life in elderly persons,[221] and others have argued that ischemic preconditioning essentially renders the SDB of old age innocuous because some component of protective adaptation is likely to have occurred.[222] Associations with both cognition[223] and daytime sleepiness[224] have been questioned. Pulse transit time, often used as a proxy for SDB screening in middle-aged patients, was shown to be less valid as a marker for SDB in an aged population,[225] and electrophysiologic changes intrinsic to cardiac control (less variability and decreased entropy) during REM sleep in old age in the absence of SDB have also been described.[226] Autonomic changes with aging, including an age-dependent reduction of parasympathetic modulation (as measured from R-R intervals in REM) in patients with SDB were noted by others,[227] but such changes in cardiorespiratory coupling are conceptualized as indicative of more disease and as more permissive for sympathetic influence, rather than less.[228] Goff and colleagues[229] have shown that cardiovascular responses (elevations in heart rate and blood pressure) to auditory stimulation are reduced during normal sleep in older relative to younger persons, and they also noted a similar blunting of response during flow-limited breathing.[230] These findings were interpreted as consistent with reduced associations between SDB and systemic hypertension that have been reported in some studies of older persons. Other evidence indicates that aortic pulse wave velocities in relation to SDB were exaggerated rather than dampened in elderly subjects,[231] and elderly women with SDB were shown to be at higher risk for both clinic-measured and 24-hour elevations in blood pressure.[232] The consequences of SDB in older populations thus remain an area rife with controversy, and the sleep medicine specialist should be cognizant of these issues. For further discussion, the reader is directed elsewhere (see Bliwise[1] for review) and to other chapters in this volume (Chapter 152).

The emergence of several relatively large-scale recent intervention studies specifically administrating CPAP to geriatric populations with sleep apnea have suggested at least modest benefit of treatment for elderly patients. The United Kingdom–based PREDICT trial noted improvements in self-reported daytime sleepiness, as well as some health care cost savings in a population older than 65 years studied for 12 months.[102] Blood pressure reductions with CPAP were noted to be as marked in subjects older than 60 years as those seen in individuals younger than 60 years,[233] and a large clinical trial of nearly 1000 elderly patients from Spain noted a doubling of risk for cardiovascular mortality when comparing individuals with untreated severe SDB and those using CPAP.[234]

WHY DO OLDER PEOPLE NAP?

A time-honored question, asked by both professionals and the lay public, involves the significance of napping in old age. From the layperson's perspective, the question is most typically: "Is it normal to nap?" or "Are naps good or bad for my health?" The sleep medicine specialist may ask fundamentally similar, though more diagnostically inclined, questions, such as "What is the probability that daytime naps in a 75-year-old indicate SDB?" "Does excessive sleepiness during the day in

an older person portend dementia?" or "To what extent does daytime napping adversely affect sleep at night?" These are highly relevant questions that are made even more difficult to answer by cultural issues related to napping, the complexities in relying on self-reports to derive estimates of the physiologic tendency of sleep during the daytime hours, and the fact that, overarching all other issues, sleeping during the daytime hours in old age is most assuredly a multidetermined phenomenon.

Elsewhere we have reviewed the complex matrix of results that suggest that napping is both a beneficial and potentially protective event in the life of an older person as well as an identifiable risk factor for numerous morbidities and even mortality.[1] A case-control PSG study[235] comparing sleepy and nonsleepy older adults and measuring a wide array of variables including comorbid medical disease, psychopathology, medications, alcohol, smoking, measurements of SDB and PLMS, and physical pain showed that male sex, poor sleep quality, sleep interruptions because of nocturnal pain or bathroom trips, and medications known to induce sleepiness differentiated cases and controls. Only severe SDB (>30 events per hour) predicted sleepiness, but PLMS did not.[235] In contrast, at least one major cohort study failed to find any association between SDB and multiple sleep latency test–confirmed sleepiness in adults older than 60 years.[224] Taken together, these findings suggest there are many factors that predict why an older person may be sleepy during the day.

Napping has also been associated with falls,[236] CSF amyloid,[161] incipient cognitive decline,[237] and depression (in women)[238] and with nocturia,[239] diabetes,[240] and lower quality of life[241] in both men and women. On the other hand, evidence continues to accrue that naps may be protective for cardiovascular events,[242] might improve daytime function,[243] do not adversely affect nocturnal sleep,[244] and might even be associated with longer sleep duration the previous night.[245] Other studies suggest that naps and hypersomnolence portend mortality[246] or ischemic heart disease[247] and that daytime fatigue or anergy predict diverse morbidities[248] or all-cause mortality.[249] Again, however, not all population-based studies concur, and some suggest absence of excess mortality risk associated with napping,[250] particularly in elderly persons.[242] A study of a British population showed that napping conferred all-cause mortality risk only in individuals younger than 65 years.[251] Several observational, cross-sectional studies of longevous populations from the Mediterranean[252] and China[253] (the latter including more than 2700 individuals older than 100 years) imply that napping may indeed have survivorship benefits, though longitudinal data from a subset of the Chinese cohort also implied mortality risk in men.[254] Finally, one study of older individuals aged 75 to 94 years reported that, if short nighttime sleep durations were taken into account, daytime naps were indeed protective for mortality, but in the presence of nocturnal sleep longer than 9 hours, naps were associated with increased mortality risk.[255] The latter study is consistent with suggestions that data on sleep durations be balanced with the related but distinct notion of "insufficient" sleep as an independent risk.[256] Clearly, the many reasons for daytime napping and sleepiness in older populations continue to be elusive and outcomes associated with the phenomenon are disparate (see Bliwise[1] for review). Additionally, the methodologic issues involved in defining napping (e.g., some studies specify durations whereas others do not) are substantial and undoubtedly affect the lack of comparability across studies.[257]

BASIC SCIENCE CONSIDERATIONS

The mechanistic basis for the age-dependent decline in sleep includes likely interactions between hypocretin and degradation of proteins associated with neurodegeneration, such as β-amyloid (Aβ42) and tau. Hypocretin neurons are lost with senescence,[258] their expression is reduced in older animals,[259] and hypocretin administration, particularly of hypocretin-1, shows diminished effects on wake promotion with aging.[260] Widespread age-dependent loss of this neuropeptide-activating system in lower animals and humans[261] invites comparison with studies suggesting positive associations between impaired cognition and daytime sleepiness noted previously. Nevertheless, a feed-forward interaction between increased hypocretin and β-amyloid in mice also has been implied.[159] In Alzheimer disease, higher CSF hypocretin-1 levels were related to both higher Aβ42 and tau in CSF,[262] though neuropathologic studies in such patients have suggested lower counts of hypocretin-1 immunoreactive neurons in the hypothalamus.[263] Associations between Aβ42 burden and poor sleep quality were noted earlier in this chapter using both neuroimaging[159] and CSF markers.[160] Higher tau density (defined postmortem by relative presence of neurofibrillary tangles) was associated with worse antemortem sleep recorded actigraphically, with some evidence of effect modification by genotype in a community population,[264] and other data have confirmed a relationship between disturbed sleep and elevated CSF tau, but not Aβ42.[265] Taken together, these data suggest some disagreement regarding the role of hypocretin in the sleep disturbance accompanying the neurodegenerative diseases of old age. Perhaps the most fundamental question raised by these studies is whether the associations between disrupted wake-sleep function and markers of Alzheimer disease pathology are mediated by hypocretin-1 expression or whether loss of sleep per se (regardless of how this occurs) plays the more crucial role. At least for β-amyloid, basic science studies generally favor the latter. For example, genetic manipulation of the hypocretin system in mice that did not express this peptide suggested that Aβ deposition results from sleep loss.[266] Curiously, those same animals were noted to sleep longer and have lower Aβ pathology than transfected controls.

The ultimate significance of alterations in sleep with advancing years remains enigmatic. If one views such changes in sleep merely as epiphenomenal to components of the aging process, they may be merely a consequence of more fundamental changes in the biology of the organism operating at the systemic, cellular, or molecular levels. On the other hand, might age-dependent alterations in sleep and rhythms themselves be potential influences on physiologic aging? If that is indeed the case, then manipulations or interventions that alter sleep might modify disease course, change fundamental processes of aging, or perhaps even contribute to the longevity of the organism. Research in this exciting area is only just beginning, but certain provocative clues are emerging, particularly as we learn more about sleep's functions.

A particularly intriguing area involves what is now acknowledged to be a fundamental biomarker of cellular aging, telomere shortening, and how that may relate to sleep.

Telomeres are noncoding, repetitive segments of DNA that function to seal and protect the ends of chromosomes during mitosis. With successive cell divisions, telomere length decreases; hence the marker has seen widespread use in studies of aging. Among middle-aged and older men and women, telomere shortening was associated with poorer quality sleep,[267] though data from the (female) Nurses Health Study showed associations in younger women (<50 years).[268] Another study in the Whitehall II cohort demonstrated an association only in older men.[269] Given that SDB may represent a good marker for physiologic age (see earlier), it is not surprising that an association between short telomere length and a history of SDB has also been reported,[270] a finding equally strong in both men and women. Inferences of causality in such descriptive human studies must always be made circumspectly; however, associations between disturbed sleep, short sleep, or SDB and inflammatory markers, such as interleukin-6, tumor necrosis factor-α, or C-reactive protein, in many of these[271] and similar elderly population studies involving either only men[272] or men and women combined[273] provide a plausible biologic substrate whereby sleep could affect cellular aging. Associations between telomere length, aging, and inflammation have been well established.[274]

Invertebrate models also have proved invaluable in understanding more about how sleep is related to the aging process. The relative strength of the sleep-wake cycle and the length of sleep bouts have been shown to break down with age in *Drosophila* species, and the extent of the disruption was magnified under higher temperatures (29°C) relative to more moderate (25°C) or cooler (18°C) temperatures,[62] an important observation because *Drosophila* are known to have a longer life span in cooler temperatures. Additionally, incorporation of paraquat, an herbicide known to induce oxidative stress, into the flies' food supply produced similar results in shortening life span. The possibility of "rescue" of functional senescence in older *Drosophila* has recently been demonstrated using social enrichment manipulations that improve sleep, which appear to operate through dopamine and alerting signaling pathways.[275] Caloric restriction is known to extend life span across many species, and there is now evidence that excessive caloric intake increases sleep fragmentation, at least in one mutant line lacking a functional dopamine transporter gene.[276] The sirtuins, which are a class of proteins involved in energy transfer and linked to such nutrient restriction, have been shown to be associated with extended life span not only in *Drosophila* but also in mammals.[277] In mice, a wide array of hypocretinergic, wake-promoting neurons throughout brainstem regions were shown to express the sirtuin-1 protein, but older animals incurred a deficit in its expression, which was manifested as lipocuscin aggregation throughout the cytoplasm in neurons of these areas.[278] These data are important because they link a protein known to be involved in the longevity-promoting functions of caloric restriction to a neuropeptide-regulating wakefulness.

Examination of sleep in *Drosophila* mutants with exceedingly short sleep durations and rapid physiologic aging (i.e., shortened life span) also suggested that sleep may indeed play a vital role in survival.[279] In the first of these, the *minisleep (MNS)* line derived from an exhaustive search of 9000 mutant lines, homozygous flies showed reductions in sleep durations of 37% of female flies and 32% of male flies relative to wild-type flies.[280] Genotyping suggested that the *MNS* line was characterized by a point mutation in the *Shaker* gene, thought to regulate potassium channel repolarization and broadly defined neuronal excitability. Perhaps most important from the standpoint of aging, survivorship to very old age in these flies (defined in this study as 56 days) was substantially lower than in flies with other *Shaker* locus mutations or in wild-type controls; in some cases the effects approached a fivefold reduction in life span.[280] A different fly line, called *sleepless (sss),* which overexpresses another *Shaker*-related protein involved in the downregulation of potassium influx, has been shown to be associated with even more dramatic reductions in sleep amount (85% for males, 80% for females).[281] In this mutant, reductions in survivorship were even more profound, with virtually none of the *sss* flies surviving beyond 50 days (median, about 30 days) relative to the median survival of control flies (about 70 days).[281] These data certainly imply that absence of sleep is associated with more rapid aging, though they do not suggest what function ascribed to sleep may be related to acceleration of such aging processes. By contrast, studies of sleep deprivation in mammals might afford a broader perspective on such issues.

Sleep deprivation studies in humans during the past 30 years have shown repeatedly that sleep loss interacts with aging in several key ways. Homeostatic pressure for sleep may be somewhat diminished in humans and in some rodent species, particularly when recovery sleep is characterized by changes in both sleep duration and delta activity pressure. In humans, the behavioral consequences of sleep loss, framed in terms of both greater daytime sleepiness and more impaired waking performances, appear diminished, rather than enhanced, with aging (see Scullin and Bliwise[133] for a review of this area). These facts imply, but cannot prove, that sleep might have less significance and, as a corollary, loss of sleep may be of less consequence—or at the very least, no more consequence—as the organism ages. The data presented throughout this chapter notwithstanding, such results might be interpreted hastily as suggesting *less* of a need to intervene in the sleep of older persons.

By contrast, important new molecular evidence suggests that at the subcellular level, sleep deprivation may be *more* detrimental for function in aging animals than in young animals. More specifically, the unfolded protein response within the endoplasmic reticulum occurring after sleep deprivation[282] may be modified substantially by aging. Protein aggregation is a well-acknowledged feature of many neurodegenerative conditions in late life and appears to occur as one of the earliest signs of impending neuronal death. The unfolded protein response, which consists of a host of molecular changes involving attenuation of translation of adverse proteins, degradation of misfolded proteins, and promotion of factors (chaperones) protective for normal function of the endoplasmic reticulum, shows synergistic age and sleep-deprivation effects. For example, expression of a major chaperone, BiP/GRP78, is known to be downregulated in both the brain and liver of aged animals, but it showed major upregulation following sleep deprivation in young, but not old, mice.[282,283] Perhaps even more relevant were changes in proapoptotic proteins (e.g., caspase-12). Aging was associated with greater expression of these markers of apoptosis, and sleep deprivation in the mouse activated this pathway as well. Strikingly, sleep deprivation in the older animal further accentuated such markers of preprogrammed cell death beyond those seen in

normal aging and beyond those in sleep deprivation,[283] and there is evidence of greater unfolded protein response to sleep loss in pancreatic beta cells of older mice relative to younger mice.[284] If these findings were in any way translatable to humans, they would suggest a *greater*, rather than *lesser*, necessity to attend to the many sleep problems in old age.

CLINICAL PEARLS

- In old age, in particular, the sleep medicine specialist should always remember that there are numerous overlapping and contributing reasons that the elderly patient has disrupted sleep at night or may be sleepy during the day.
- Napping should never be assumed to be innocuous, nor should excessive sleepiness during the day be attributed unequivocally to active or subclinical disease.
- SDB may be associated with adverse outcomes in elderly populations as well as middle-aged populations and might warrant treatment.
- Poor sleep may not only reduce the quality of life of the older person but also portend adverse health consequences.

SUMMARY

Defining normality in elderly populations remains a challenging task. However, an ever-increasing database informs the sleep medicine specialist about potential morbidities that may be associated with poor sleep and various sleep disorders in old age. Multiple factors contribute to poor sleep at night and excessive sleepiness during the day in the older adult. Although most studies have been observational and descriptive, the examination of sleep in lower animals may allow greater understanding of age-dependent changes in sleep as an indicator of biologic or physiologic age and shed new light on the importance of treating sleep problems in human aging.

ACKNOWLEDGMENTS

This work is supported by National Institute on Aging grants AG-020269, AG-025688, and F32-AG041543 and by National Institute of Neurological Disorders and Stroke grant NS-050595.

Selected Readings

Bliwise DL. Cataloging nocturia (circa 2014). *Sleep* 2014;**37**:631–3.
Bliwise DL, Colrain IM, Swan GE, et al. Incident sleep disordered breathing in old age. *J Gerontol A Biol Sci Med Sci* 2010;**65**:997–1003.
Cohen-Mansfield J, Perach R. Sleep duration, nap habits and mortality in older persons. *Sleep* 2012;**35**:1003–9.
Edwards BA, Wellman A, Sands SA, et al. Obstructive sleep apnea in older adults is a distinctly different phenotype. *Sleep* 2014;**37**:1227–36.
Ferrie JE, Shipley MJ, Akbaraly TN, et al. Change in sleep duration and cognitive function: findings from the Whitehall II Study. *Sleep* 2011;**34**:565–73.
Fronczek R, van Geest S, Frolich M, et al. Hypocretin (orexin) loss in Alzheimer's disease. *Neurobiol Aging* 2012;**33**:1642–50.
Fung MM, Peters K, Redline S, et al. Decreased slow wave sleep increases risk of developing hypertension in elderly men. *Hypertension* 2011;**58**:596–603.
Hasan S, Dauvilliers Y, Mongrain V, et al. Age-related changes in sleep in inbred mice are genotype dependent. *Neurobiol Aging* 2012;**33**:e13–26.
Martinez-Garcia MA, Campos-Rodriguez F, Catalan-Serra P, et al. Cardiovascular mortality in obstructive sleep apnea in the elderly: role of long-term continuous positive airway pressure treatment: a prospective observational study. *Am J Respir Crit Care Med* 2012;**186**:909–16.
McMillan A, Bratton DJ, Faria R, et al. Continuous positive airway pressure in older people with obstructive sleep apnoea syndrome (PREDICT): a 12-month, multicentre, randomized trial. *Lancet Respir Med* 2014;**2**(10):804–12.
Naidoo N, Ferber M, Master M, et al. Aging impairs the unfolded protein response to sleep deprivation and leads to proapoptotic signaling. *J Neurosci* 2008;**28**:6539–48.
Nixon JP, Mavanji V, Butterick TA, et al. Sleep disorders, obesity, and aging: the role of orexin. *Ageing Res Rev* 2015;**20**:63–73.
Scullin MK, Bliwise DL. Sleep, cognition, and aging: integrating a half-century of multidisciplinary research. *Perspect Psychol Sci* 2015;**10**(1):97–137.
Spira AP, Gamaldo AA, An Y, et al. Self-reported sleep and β-amyloid deposition in community-dwelling older adults. *JAMA Neurol* 2013;**70**:1537–43.
Terpening Z, Lewis SJ, Yee BJ, et al. Association between sleep-disordered breathing and neuropsychological performance in older adults with mild cognitive impairment. *J Alzheimers Dis* 2015;**46**(1):157–65.
Yin D. The essential mechanisms of aging: what have we learnt in ten years? *Curr Top Med Chem* 2016;**16**(5):503–10.

A complete reference list can be found online at ExpertConsult.com.

Chapter 4

Daytime Sleepiness and Alertness

Timothy Roehrs; Mary A. Carskadon; William C. Dement; Thomas Roth

Chapter Highlights

- Sleepiness is a common problem in the general population that can be caused by reduced sleep time in otherwise healthy adults, by fragmented and disrupted sleep in primary sleep disorders, by sedating drugs, and by various neurologic and medical disorders.
- Sleepiness is a physiologic need state that is evident by rapid and unintended sleep onsets throughout the day, a relative irreversibility, and increased duration.
- Excessive and persistent sleepiness is life-threatening, but when its presence is recognized and its etiology identified, it can be successfully treated or at least minimized.

Sleepiness is a physiologic drive state basic to the survival of an organism. Although as yet no single or combination of biologic markers of sleepiness are known, its presence and intensity can be inferred by how readily sleep onset occurs, how easily sleep is disrupted, and how long sleep endures with an enforced bedtime and no sleep disorder or environmental interruption. Sleepiness is normally expressed in a 24-hour rhythm linked to the environmental light-dark cycle. Excessive daytime sleepiness (EDS) can be seen in otherwise healthy individuals who do not have an adequate opportunity or circumstances for sleep. Most important, EDS is a symptom associated with serious life-threatening medical conditions. Since the development of clinical sleep medicine, a growing scientific literature on the nature of sleepiness and its determinants in clinical populations, selected populations of healthy volunteers, and the general population has emerged.

This chapter reviews information regarding the prevalence of sleepiness in the population. The various methods used to measure sleepiness in the population and laboratory are described, and guidelines regarding clinical assessment of sleepiness are offered. The nature and neurobiologic substrates of sleepiness are discussed, and the known determinants of sleepiness are described. Finally, the clinical and public health significance of persistent complaints of sleepiness is discussed.

EPIDEMIOLOGY OF SLEEPINESS

Prevalence estimates of sleepiness in the population vary widely depending on the definition of sleepiness used and the type of population sampled. Surveys and questionnaires have queried the experience of a mood or feeling of a state of sleepiness, fatigue, or tiredness; about falling asleep unintentionally; or about struggling to stay awake and fighting sleep onset. Developments in the epidemiology of sleepiness have included using standardized sleepiness scales and physiologic assessments of sleepiness that measure the behavior of falling asleep, its estimated likely occurrence, or the speed of its actual occurrence.

In a study representative of the Finnish population, 11% of women and 7% of men reported daytime sleepiness almost every day.[1] In another survey, representative of a large geographic area in Sweden, 12% of respondents thought their sleep was insufficient.[2] In that survey, insufficient sleep, and not its consequent daytime sleepiness, was the focus of the questions. The U.S. Centers for Disease Control and Prevention analyzed results from the Behavioral Risk Factor Surveillance System in four states and found that 10.1% of U.S. adults reported receiving insufficient rest or sleep on all of the preceding 30 days.[3]

Two studies representative of the U.S. population assessed sleepiness with a physiologic measure, the Multiple Sleep Latency Test (MSLT). Given the necessary time commitment required of participants for MSLT studies, the representative integrity of study results is critically dependent on the recruitment response rate. From a large Southeastern Michigan random sample ($n = 1648$) representative of the U.S. population, a subsample ($n = 259$) with a 68% response rate was recruited to undergo a nocturnal polysomnogram and MSLT the following day. The prevalence of excessive sleepiness, defined as a MSLT average sleep latency of less than 6 minutes, was 13%.[4] In another probability sample of 6947 Wisconsin state employees, a subsample ($n = 632$), collected with a 52% response rate, slept at home and then completed an MSLT in the laboratory the next day. Twenty-five percent had an average sleep latency of less than 5 minutes.[5] These two studies also used the Epworth Sleepiness Scale (ESS), a validated self-report scale about the likelihood of falling asleep in various situations, to assess perceived sleepiness; in the Michigan study 20% had ESS scores higher than 10 and in the Wisconsin study 25% had scores higher than 11. The higher prevalence in the Wisconsin study, despite the more stringent definition of sleepiness (MSLT of 5 vs. 6 minutes and ESS of 11 vs. 10), could be attributed to an age difference

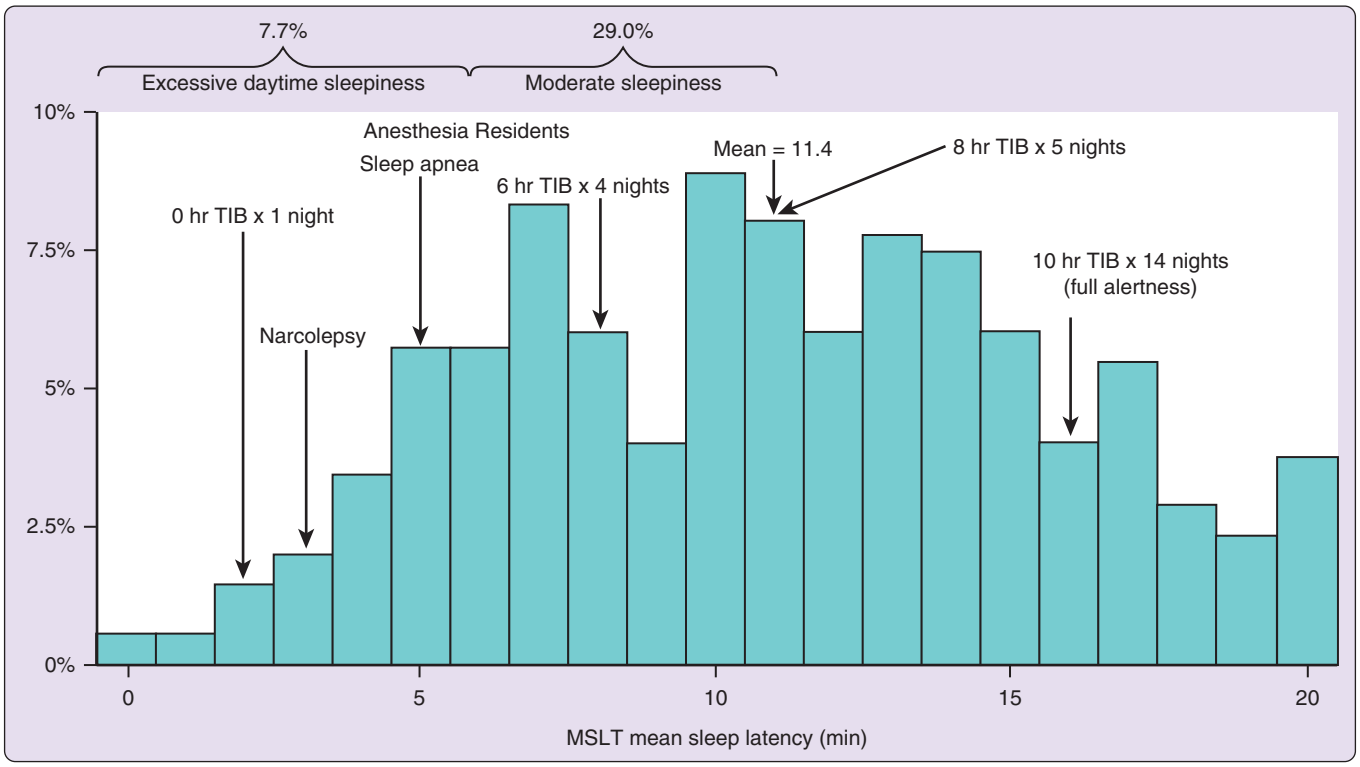


Figure 4-1 The distribution of mean daily sleep latency (minutes) on the Multiple Sleep Latency Test (MLST) in a subsample (n = 259) recruited (68% response rate) from a large southeastern Michigan random sample (n = 1648) representative of the U.S. population. The population mean is 11.4 minutes, and this is compared with means reported for various patient groups[59,69,70] and the means found in healthy normal individuals after various bedtime manipulations.[22,61] TIB, Time in bed.

in the samples (51 vs. 42 years on average) or the previous night's sleep time and circumstances (habitual at home, on average 7.1 hours vs. standard laboratory 8.5 hours). In Figure 4-1 the distribution of sleepiness, defined as average sleep latency on the MSLT, is illustrated for the Michigan population representative sample. The average sleep latencies (MSLT) of various clinical samples and experimental sleep time manipulations are provided for comparisons.

Risk Factors for Sleepiness

The risk factors for sleepiness identified in the various surveys include hours of daily sleep, employment status, marital status, snoring, and depression. Among 26- to 35-year-old members of a large health maintenance organization in Michigan, respondents reported 6.7 hours of sleep on weekdays and 7.4 hours on weekend days, on average.[6] The hours of sleep were inversely related to daytime sleepiness scores on the Sleep-Wake Activity Inventory. Both these variables were related to employment and marital status, with full employment and being single predicting less sleep time and greater sleepiness. Self-reported snoring and depression, as measured by a structured diagnostic interview, were also associated with increased sleepiness. In the Finnish study cited earlier, sleepiness was associated with moderate to severe depression and with snoring more than three times per week.[1]

Prevalence rates for sleepiness of 15% and greater also have been found for specific age groups, which are consistent with smaller laboratory studies using the physiologic measure of sleepiness, the MSLT, which is described later. Young adults were sleepier, on average, than a comparison group of middle-aged adults, and about 20% of the young adults had mean daily sleep latencies of less than 5 minutes, a level of sleepiness considered pathologic.[7] Healthy elderly persons also were found to be physiologically sleepier than middle-aged adults.[8] In surveys of the workforce engaged in shift or night work, complaints of excessive sleepiness during waking hours are more frequent than among day workers, and continuous ambulatory electroencephalographic (EEG) field monitoring has confirmed the sleepiness.[9]

NATURE OF SLEEPINESS

Physiologic Need State

Sleepiness, according to a consensus among sleep researchers and clinicians, is a basic physiologic need state.[10] Sleepiness has been likened to hunger or thirst, which are physiologic need states basic to the survival of the individual organism. The presence and intensity of this state can be inferred by how readily sleep onset occurs, how easily sleep is disrupted, and how long uninterrupted sleep endures. Deprivation or restriction of sleep increases sleepiness, and as hunger or thirst is reversible by eating or drinking, sleep reverses sleepiness in healthy individuals. In the organism's daily homeostatic economy, severe deprivation states do not normally occur and hence are not routinely responsible for regulating eating or drinking; other factors (i.e., taste, smell, time-of-day, social factors, biologic variables) modulate these behaviors before

severe deprivation states develop. Similarly, routine consumption of sleep is not purely homeostatic but is greatly influenced by the circadian light-dark cycle and by social (i.e., job, family, and friends) and environmental (i.e., noise, light, and bed) factors.

The subjective experience of sleepiness and its behavioral indicators (yawning, eye rubbing, nodding) can be reduced under conditions of high motivation, excitement, exercise, and competing needs (e.g., hunger, thirst); that is, physiologic sleepiness may not necessarily be manifest. The expression of mild to moderate sleepiness can be masked by any number of factors that are alerting, including motivation, environment, posture, activity, light, or food intake. Studies have shown that average sleep latency on the MSLT is increased by 6 minutes when sitting compared with lying in bed and also by 6 minutes when immediately preceded by a 5-minute walk.[11] When physiologic sleepiness is most severe and persistent, the ability to reduce its impact on overt behavior wanes. The likelihood of sleep onset increases, and the intrusion of microsleeps into ongoing behavior occurs. On the other hand, a physiologically alert (*sleepiness* and *alertness* are used here as antonyms) person does not experience sleepiness or appear sleepy even in the most soporific situations. Heavy meals, warm rooms, boring lectures, and the monotony of long-distance automobile driving unmask physiologic sleepiness when it is present, but they *do not* cause it.

Within a conventional 24-hour sleep and wake schedule, maximum sleepiness ordinarily occurs in the middle of the night when the individual is sleeping, and consequently this sleepiness typically is not experienced or remembered. When forced to be awake in the middle of the night, one experiences loss of energy, fatigue, weariness, difficulty concentrating, and memory lapses. When significant physiologic sleepiness (as a result of reduced sleep quantity or quality) intrudes on one's usual waking activities during the day, similar symptoms are experienced.

Adaptation to the *experience* of chronic sleepiness most probably occurs. Clinicians have reported anecdotally that successfully treated patients will frequently comment that they had forgotten the experience of complete alertness. Reduced sensitivity to chronic sleepiness is a likely explanation for the disparities between subjective assessments, even when done with validated scales, and the MSLT[8,11] Typically, it is the most sleepy individuals that show the greatest disparity in subjective versus objective assessments.[8,11] Such individuals deny sleepiness despite significant objective indicators of sleepiness. On the other hand, basally alert individuals (ESS mean = 5.6, standard error of the mean = 0.3) after a one-night acute sleep restriction were quite accurate in estimating their sleepiness as gauged by increases in EEG theta activity shown during a simulated driving task.[12] Studies have also shown that compensation occurs to the cognitive and behavioral effects of experimental sleep restriction and increased sleepiness, particularly when the sleep loss is mild and accumulates at a slow rate.[13] The absence of a readily apparent behavioral deficiency to the sleepy individual probably also contributes to the subjective-objective disparity seen in chronically sleepy individuals. Finally, findings from a general population study indicate that subjective sleepiness has multiple dimensions, beyond an increased tendency to fall asleep.[14] Consequently, patients often mistake chronic debilitating fatigue for sleepiness.[15]

The specific nature of this physiologic need state is unclear. Whether sleepiness is unidimensional, varying only in severity, or multidimensional, varying as to etiology or chronicity, has been discussed.[16] If it is unidimensional, whether or not sleepiness and alertness are at opposite poles of the dimension is also an issue. Earlier, it was noted that sleepiness and alertness are being used as antonyms, which suggests a unipolar state. However, it is possible that sleepiness varies from presence to absence and is distinct from alertness. It was noted that sleepiness may be multidimensional, and among the different types of sleepiness cited are rapid eye movement (REM) versus non–rapid eye movement (NREM) and core versus optional sleepiness.[16] A complete discussion of the heuristic value and evidence to support these distinctions is beyond the scope of this chapter. Nonetheless, the point must be made that these theoretical perspectives may be colored by different measures, experimental demands, populations studied, and subject or patient motivations (i.e., sensitivity to and capacity to counteract sleepiness).

Neural Substrates of Sleepiness

The substrates of sleepiness have yet to be determined. It is assumed that sleepiness is a central nervous system (CNS) phenomenon with identifiable neural mechanisms and neurochemical correlates. Various electrophysiologic events suggestive of incipient sleep processes appear in behaviorally awake organisms undergoing sleep deprivation. In sleep-deprived animals, ventral hippocampal spike activity, which normally is a characteristic of NREM sleep, increases during behavioral wakefulness and in the absence of the usual changes in cortical EEG indicative of sleep.[17] Humans deprived, or restricted, of sleep show identifiable microsleep episodes (brief intrusions of EEG indications of sleep) and increased amounts of alpha and theta activity while behaviorally awake.[18] The evidence suggests that these electrophysiologic events are indicants of sleepiness.

An emerging literature of neuroimaging studies, both structural and functional, has suggested specific brain systems that may be involved in sleepiness. Sleep deprivation in young healthy volunteers reduced regional cerebral glucose metabolism, as assessed by positron emission tomography, in thalamic, basal ganglia, and limbic regions of the brain.[19] Functional magnetic resonance imaging (fMRI) after chlorpheniramine (a sedating antihistamine) compared with placebo showed increased frontal and temporal activation.[20] Because fMRI was conducted while the subject was performing cognitive tasks, the authors interpreted the observed brain activation to have resulted from the increased mental effort necessary to counteract sleepiness required to perform the task. Two groups of patients with severe or slight hypersomnia associated with paramedian thalamic stroke on an MRI showed lesions involving dorso- and centromedial thalamic nuclei, bilateral lesions in the severe group and unilateral in the slight group.[21] As yet, these imaging data are not conclusive. They do suggest it may be possible to identify brain regions and functions that vary with sleepiness. But the nature of the alteration may depend on the behavioral load imposed on the sleepy subject as well as the cause of the sleepiness.

The neurochemistry of sleepiness-alertness involves critical and complex issues that have not yet been fully untangled (see Chapters 7, 8, and 9 for a complete discussion).

First, a basic issue concerns whether sleepiness-alertness has a neurochemistry specific and unique from that associated with the sleep process. Second, it is not clear whether sleepiness and alertness are controlled by separate neurochemicals or by a single substance or system. Third, the relation of the neurochemistry of sleepiness-alertness to circadian mechanisms has not yet been determined. Given the number of questions, it should be of no surprise that these are areas of active research.

Neurophysiologic studies of sleep and wake mechanisms have implicated adenosine, gamma-aminobutyric acid (GABA), histamine, serotonin, the catecholamines, and acetylcholine in control of wake and sleep.[22] Evidence from animal studies suggests extracellular adenosine is the homeostatic sleep factor reflecting sleep drive, with brain levels accumulating during prolonged wakefulness and declining during sleep.[23] Adenosine may be a biomarker of sleepiness.[23] It stimulates both $A_1$and A_2 receptors, and a positron emission tomography study in humans has shown upregulation of A_1 receptors in cortical and subcortical regions of the brain during 24 hours of sleep deprivation.[24] On the other hand, a microdialysis study in patients undergoing depth electrode implantation for control of epileptic seizures failed to find an accumulation of adenosine in the hippocampus, amygdala, and cortex during sleep deprivation.[25] These data raise fundamental questions: Does adenosine act globally or locally? Does it accumulate or are receptors sensitized? The basal forebrain, the ventral lateral preoptic area, and the lateral hypothalamus are all potential sites, owing to their known involvement in the control of sleep and wake, at which to assess levels and receptor upregulation.

The peptide hypocretin, also called orexin, has received much attention for its role in the pathophysiology of narcolepsy.[26] It is considered to be a major wake-promoting hypothalamic neuropeptide, and a hypocretin/orexin deficiency has been found in human narcolepsy; however, its interactive role in the homeostatic control of sleep and sleepiness has yet to be determined. Hypocretin/orexin is discussed in greater detail in Chapters 8 and 89.

Pharmacologic studies provide other interesting hypotheses regarding the neurochemistry of sleepiness-alertness. For example, the benzodiazepines induce sleepiness and facilitate GABA function at the $GABA_A$ receptor complex, thus implicating this important and diffuse inhibitory neurotransmitter.[24] Another example involves histamine, which is now considered a CNS neurotransmitter and is thought to have CNS-arousing activity.[24] Antihistamines that penetrate the CNS produce sleepiness.[27] A recent functional neuroimaging study of histamine H_1 receptors in human brain found that the degree of sleepiness associated with cetirizine (20 mg) was correlated to the degree of H_1 receptor occupancy.[28]

Stimulant drugs indicate the involvement of several other transmitters and neuromodulators. The mechanism of action of one class of drugs producing psychomotor stimulation and arousal, the amphetamines, is blockade of catecholamine uptake.[29] Another class of stimulants, the methylxanthines, which include caffeine and theophylline, are adenosine receptor antagonists. Adenosine has inhibitory activity on the two major excitatory neurotransmitters acetylcholine and glutamate. In conclusion, although it is widely held that sleepiness is a physiologic state, its physiologic substrates are as yet not fully defined.

ASSESSMENT OF SLEEPINESS

Quantifying Sleepiness

The behavioral signs of sleepiness include yawning, ptosis, reduced behavioral activity, lapses in attention, and head nodding. An individual's subjective report of his or her level of sleepiness also can be elicited. As noted earlier, a number of factors such as motivation, stimulation, and competing needs can reduce the behavioral manifestation of sleepiness. Thus behavioral and subjective indicators often underestimate physiologic sleepiness.

Assessment problems were evident early in research on the daytime consequences of sleep loss. Sleep loss compromises daytime functions; virtually everyone experiences dysphoria and reduced performance efficiency when not sleeping adequately. Nevertheless, many of the tasks used to assess the effects of sleep loss are insensitive.[30] In general, only long and monotonous tasks are reliably sensitive to changes in the quantity and quality of nocturnal sleep. An exception is a 10-minute visual vigilance task, completed repeatedly across the day, during which lapses (response times ≥500 msec) and declines in the best response times are increasingly observed as sleep is lost, either during total deprivation or cumulatively over nights of restricted bedtimes.[31]

In various measures of mood, including factor analytic scales, visual analogue scales, and scales for specific aspects of mood, subjects have shown increased fatigue or sleepiness with sleep loss. Among the various subjective measures of sleepiness, the Stanford Sleepiness Scale is the best validated.[32] Yet clinicians have found that chronically sleepy patients may rate themselves alert on the Stanford Sleepiness Scale even while they are falling asleep behaviorally.[33] Such scales are state measures that query individuals about how they feel at the present moment. Another perspective is to view sleepiness behaviorally, as in the likelihood of falling asleep, and thus ask individuals to rate that likelihood in different social circumstances and over longer periods. The ESS has been validated in clinical populations, showing a 74% sensitivity and 50% specificity relative to the MSLT in a study of sleep disorders patients.[34] It asks about falling asleep in settings in which affected patients typically report falling asleep (e.g., while driving, at church, in social conversation). The time frame over which ratings are made is typically 2 to 4 weeks.

The standard physiologic measure of sleepiness is the MSLT, which has been studied extensively, with more than 7500 citations to "MSLT and sleep" in a Google Scholar August 2014 search. The MSLT similarly conceptualizes sleepiness as the tendency to fall asleep by measuring the speed of falling asleep. The MSLT has gained wide acceptance within the field of sleep and sleep disorders as the standard method of quantifying sleepiness.[35] Using standard polysomnographic techniques, this test measures, on repeated opportunities at 2-hour intervals throughout the day, the latency to fall asleep while lying in a quiet, dark bedroom. The MSLT is based on the assumption, as outlined earlier, that sleepiness is a physiologic need state that leads to an increased tendency to fall asleep. The metric typically used to express sleepiness has been average daily sleep latency (i.e., mean of the four or five tests conducted), but survival analyses have also been successfully used.[36] The reliability and validity of this measure have been documented in a variety of experimental and clinical situations.[37] In contrast to tests of performance, motivation

does not seem to reduce the effect of sleep loss as measured by the MSLT. After total sleep deprivation, subjects can compensate for impaired performance, but they cannot stay awake long while in bed in a darkened room, even if they are instructed to do so.[38]

An alternative to the MSLT, suggested by some clinical investigators, is the Maintenance of Wakefulness Test (MWT). This test requires that subjects lie in bed or sit in a chair in a darkened room and try to remain awake.[39] Like the MSLT, the measure of ability to remain awake is the latency to sleep onset. The test has not been standardized: there are 20-minute and 40-minute versions, and the subject is variously sitting upright in a chair, lying in bed, or semirecumbent in bed. The reliability of the MWT has not been established either. One study reported sensitivity to the therapeutic effects of continuous positive airway pressure (CPAP) in patients with sleep apnea,[40] and several studies reported sensitivity to the therapeutic effects of stimulants in narcolepsy.[41] A study attempted to tease apart the critical factors being measured by the MWT and concluded that, unlike the MSLT, which measures level of sleepiness, the MWT measures the combined effects of level of sleepiness and the degree of arousal as defined by heart rate.[42]

The rationale for the MWT is that clinically the critical issue for patients is how long wakefulness can be maintained. A basic assumption underlying this rationale, however, may not be valid: it assumes that a set of circumstances can be evaluated in the laboratory that will reflect an individual's probability of staying awake during daily activities. Such a circumstance is not likely because environment, motivation, circadian phase, and any competing drive states all affect an individual's tendency to remain awake. Stated simply, an individual crossing a congested intersection at midday is more likely to stay awake than an individual driving on an isolated highway in the middle of the night. The MSLT, on the other hand, addresses the question of the individual's risk for falling asleep by establishing a setting to maximize the likelihood of sleep onset: all factors competing with falling asleep are removed from the test situation. Thus the MSLT identifies sleep tendency or clinically identifies maximum risk for the patient. Obviously, the actual risk will vary from individual to individual, from hour to hour, and from environment to environment.

Relation of Sleepiness to Behavioral Functioning

Given that the MSLT is a valid and reliable measure of sleepiness, the question arises as to how this measure relates to an individual's capacity to function. Direct correlations of the MSLT with other measures of performance under normal conditions have not been very robust. Several studies have found, however, that when sleepiness is at maximal levels, correlations with performance are high. For example, MSLT scores after sleep deprivation,[43] after administration of sedating antihistamines,[44] and after benzodiazepine administration[45] correlate with measures of performance and even prove to be the most sensitive measure. Studies also have compared levels of sleepiness to the known performance-impairing effects of alcohol.[46] A study relating performance lapses on a vigilance task to the cumulative effects of sleep restriction found a function comparable to that of the MSLT under a similar cumulative sleep restriction[47] (Figure 4-2). The reason many studies have found weak correlations between performance and MSLT at normal or moderate levels of sleepiness is that laboratory performance and MSLT are differentially affected by variables such as age, education, and motivation.

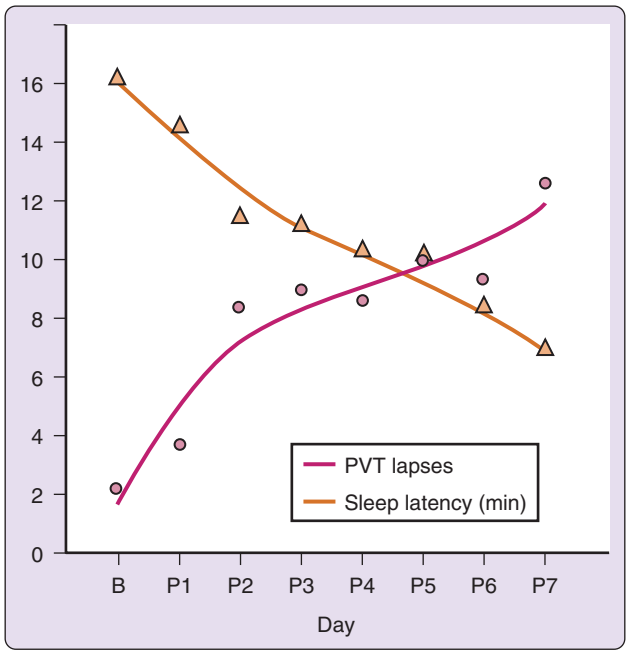


Figure 4-2 Similar functions relating mean daily sleep latency on the Multiple Sleep Latency Test (MSLT) and mean daily lapses on the visual psychomotor vigilance test (PVT) to the cumulative effects of sleep restriction (about 5 hours of bedtime nightly) across 7 consecutive nights (P1 to P7). (Modified from Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep* 1997;20:275.)

For the most part, the literature relating sleepiness and behavioral functioning has focused on psychomotor and attention behaviors, with the major outcomes being response slowing and attentional lapses. These impairments can be attributed to slowed processing of information and microsleeps, that is, intrusion of sleep-preparatory and sleep-onset behaviors. Research has focused on other behavioral domains not as clearly associated with sleep-mediated behaviors, including decision making and pain sensation. Several studies have shown that increased sleepiness is associated with poor risk-taking decisions.[48] Sleep loss and its associated sleepiness have also been shown to increase pain sensitivity.[49]

Clinical Assessment of Sleepiness

Assessing the clinical significance of a patient's complaint of excessive sleepiness can be complex for an inexperienced clinician. The assessment depends on two important factors: chronicity and reversibility. Chronicity can be explained simply. Although a healthy normal individual may be acutely sleepy, the patient's sleepiness is persistent and unremitting. As to reversibility, unlike the healthy normal person, increased sleep time may not completely or consistently ameliorate a patient's sleepiness. Patients with excessive sleepiness may not complain of sleepiness per se but rather its consequences: loss of energy, fatigue, lethargy, weariness, lack of initiative, memory lapses, or difficulty concentrating.

Table 4-1 Sleep-Inducing Situations for Patients with Apnea*

Situation	Percentage of Patients
Watching television	91
Reading	85
Riding in a car	71
Attending church	57
Visiting friends and relatives	54
Driving	50
Working	43
Waiting for a red light	32

*n = 384 patients.

To clarify the patient complaint, it is important to focus on soporific situations in which physiologic sleepiness is more likely to be manifest, as was discussed earlier. Such situations might include watching TV, reading, riding in a car, listening to a lecture, or sitting in a warm room. Table 4-1 presents the commonly reported "sleep-inducing" situations for a large sample of patients with sleep apnea syndrome. After clarifying the complaint, one should ask the patient about the entire day: morning, midday, and evening. In the next section, it will become clear that most adults experience sleepiness over the midday. However, patients experience sleepiness at other times of the day as well, and often throughout the day. When necessary (i.e., patients nonresponsive to treatment or in jobs carrying personal or public risk) objective documentation of sleepiness and its severity should be sought. As indicated earlier the standard and accepted method to document sleepiness objectively is the MSLT.

Guidelines for interpreting the results of the MSLT are available[35] (see Figure 4-1). A number of case series of patients with disorders of excessive sleepiness have been published, with accompanying MSLT data for each diagnostic classification.[50] These data provide the clinician with guidelines for evaluating the clinical significance of a given patient's MSLT results. Although these data cannot be considered norms, a scheme for ranking MSLT scores to indicate degree of pathology has been suggested.[51] An average daily MSLT score of 5 minutes or less suggests pathologic sleepiness, a score of more than 5 minute but less than 10 minutes is considered a diagnostic gray area, and a score of more than 10 minutes is considered to be in the normal range (see Figure 4-1 for MSLT results in the general population). The MSLT is also useful in identifying sleep-onset REM periods (SOREMPs), which are common in patients with narcolepsy.[51] The American Academy of Sleep Medicine Standards of Practice Committee has concluded that the MSLT is indicated in the evaluation of patients with suspected narcolepsy.[51] However, the prevalence of multiple SOREMPs in the general population is 4% to 9%, and excessive sleepiness as found in shift workers, young adults, and patients with apnea is predictive of multiple SOREMPs.[52,53] MSLT results also must be evaluated with respect to the conditions under which the testing was conducted. Standards have been published for administering the MSLT, which must be followed to obtain a valid, interpretable result.[35]

It should be emphasized that neither the MSLT nor the MWT is diagnostic of a sleep disorder. These tests merely confirm the presence and intensity of sleepiness. In fact, studies have failed to find differences between samples of patients with sleep disorders and age-matched controls.[54] As illustrated in Figure 4-1 the distribution of MSLT latencies in the general population includes individuals who would be considered excessively sleepy. As noted previously they differ from patients in that increased nightly bedtime and a consequent increase in sleep time reverses the sleepiness, which does not occur in patients until treated.

DETERMINANTS OF SLEEPINESS

Quantity of Sleep

The degree of daytime sleepiness is directly related to the amount of nocturnal sleep. The performance effects of acute and chronic sleep deprivation are discussed in Chapters 5. As to sleepiness, partial or total sleep deprivation in healthy normal subjects is followed by increased daytime sleepiness the following day.[37] Therefore modest nightly sleep restriction accumulates over nights to progressively increase daytime sleepiness and performance lapses (see Figure 4-1).[55] However, the speed at which sleep loss is accumulated is critical, and studies have shown adaptation to a slow accumulation of 1 to 2 hours nightly sleep loss occurs, which then increases the duration of the subsequent recovery process.[13] Increased sleep time in healthy, but sleepy, young adults by extending bedtime beyond the usual 7 to 8 hours per night produces an increase in alertness (i.e., reduction in sleepiness).[56] Further, the pharmacologic extension of sleep time by an average of 1 hour in elderly persons produces an increase in mean sleep latency on the MSLT (i.e., increased alertness).[57]

Reduced sleep time explains the excessive sleepiness of several patient and nonpatient groups. For example, a subgroup of sleep clinic patients has been identified whose excessive daytime sleepiness can be attributed to chronic insufficient sleep.[58] These patients show objectively documented excessive sleepiness and "normal" nocturnal sleep with unusually high sleep efficiency (time asleep/time in bed), and they report about 2 hours more sleep on each weekend night than each weekday night. Regularizing bedtime and increasing time in bed produces a resolution of their symptoms and normalized MSLT results.[59] The increased sleepiness of healthy young adults also can be attributed to insufficient nocturnal sleep. When the sleepiest 25% of a sample of young adults is given extended time in bed (10 hours) for as long as 5 to 14 consecutive nights, their sleepiness is reduced to a level resembling the general population.[56]

Individual differences in tolerability to sleep loss have been reported.[60] These differences can be attributed to a number of possible factors. A difference in the basal level of sleepiness at the start of a sleep time manipulation is quite possible given the range of sleepiness in the general population (see Figure 4-1). The basal differences may reflect insufficient nightly sleep relative to a person's sleep need.[56] Also there may be differences in the sensitivity and responsivity of the sleep homeostat to sleep loss (i.e., how large a sleep deficit the system can tolerate and how robustly the sleep homeostat produces sleep when detecting deficiency). Finally, genetic differences in sleep need, the set point around which the sleep homeostat regulates daily sleep time, have long been

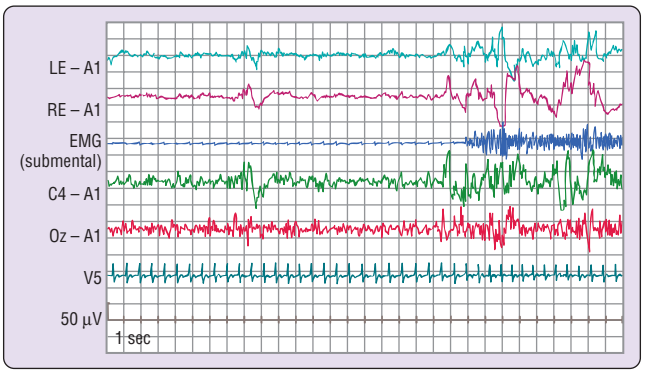


Figure 4-3 A transient arousal (*right side* of figure) fragmenting sleep. The preexistence of sleep is evident by the K-complex at second 9 of the epoch preceding the arousal. C4-A1, Electroencephalogram referenced to A1 from C4 placement; EMG, electromyogram from submental muscle; LE-A1, left electrooculogram referenced to A1; Oz-A1, electroencephalogram referenced to A1 from Oz placement; RE-A1, right electrooculogram referenced to A1; V5, electrocardiogram from V5 placement. (Modified from American Sleep Disorders Association. EEG arousals: scoring rules and examples. *Sleep* 1992;15:173–84.)

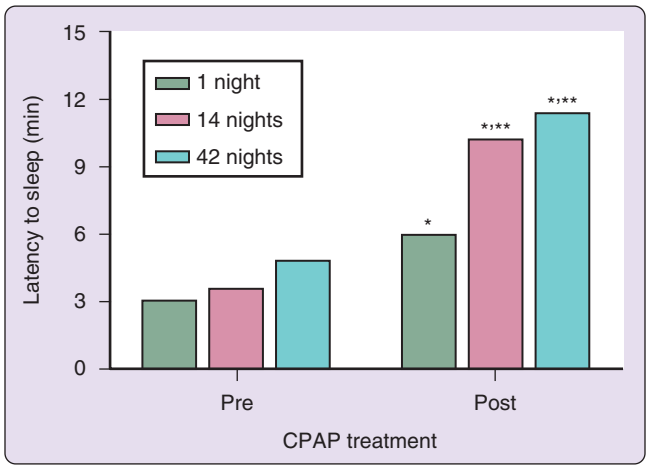


Figure 4-4 Mean daily sleep latency on the Multiple Sleep Latency Test in patients with obstructive sleep apnea syndrome before (pre) and after (post) 1, 14, and 42 nights of continuous positive airway pressure (CPAP) treatment. *$P < .05$; **$P < .01$. (Modified from Lamphere J, Roehrs T, Wittig R, et al. Recovery of alertness after CPAP in apnea. *Chest* 1989;96:1364–7.)

hypothesized, and one study has suggested that a gene polymorphism may mediate vulnerability to sleep loss.[61] These all are fertile areas for research.

Quality of Sleep

Daytime sleepiness also relates to the quality and the continuity of a previous night's sleep. Sleep in patients with a number of sleep disorders is punctuated by frequent, brief arousals of 3 to 15 seconds' duration. These arousals are characterized by bursts of EEG speeding or alpha activity and, occasionally, transient increases in skeletal muscle tone. Standard scoring rules for transient EEG arousals have been developed.[62] A transient arousal is illustrated in Figure 4-3. These arousals typically do not result in awakening by either Rechtschaffen and Kales sleep staging criteria or behavioral indicators, and the arousals recur in some conditions as often as 1 to 3 times per minute. The arousing stimulus differs in the various disorders and can be identified in some cases (apneas, leg movements, pain) but not in others. Regardless of etiology, the arousals generally do not result in shortened sleep but rather in fragmented or discontinuous sleep, and this fragmentation produces daytime sleepiness.[63]

Correlation evidence suggests a relation between sleep fragmentation and daytime sleepiness. Fragmentation, as indexed by number of brief EEG arousals, number of shifts from other sleep stages to stage 1 sleep or wake, and the percentage of stage 1 sleep, correlates with EDS in various patient groups.[63] Treatment studies also link sleep fragmentation and excessive sleepiness. Patients with sleep apnea syndrome who are successfully treated by surgery (i.e., number of apneas are reduced) show a reduced frequency of arousals from sleep as well as a reduced level of sleepiness, whereas those who do not benefit from the surgery (i.e., apneas remain) show no decrease in arousals or sleepiness, despite improved sleeping oxygenation.[64] Similarly, CPAP, by providing a pneumatic airway splint, reduces breathing disturbances and consequent arousals from sleep and reverses EDS.[65] The reversal of daytime sleepiness following CPAP treatment of sleep apnea syndrome is presented in Figure 4-4. The hours of nightly CPAP use predicts both subjective and objective measures of sleepiness.[66]

Experimental fragmentation of the sleep of healthy normal subjects has been produced by inducing arousals with an auditory stimulus. Several studies have shown that subjects awakened at various intervals during the night demonstrate performance decrements and increased sleepiness on the following day.[67] Studies have also fragmented sleep without awakening subjects by terminating the stimulus on EEG signs of arousal rather than on behavioral response. Increased daytime sleepiness (shortened latencies on the MSLT) resulted from nocturnal sleep fragmentation in one study,[68] and in a second study, the recuperative effects (measured as increased latencies on the MSLT) of a nap following sleep deprivation were compromised by fragmenting the sleep on the nap.[69]

One nonclinical population in which sleep fragmentation is an important determinant of excessive sleepiness is the elderly population. Many studies have shown that even elderly individuals without sleep complaints show an increased number of apneas and periodic leg movements during sleep.[70] As noted earlier, elderly people as a group are sleepier than other groups.[8] Furthermore, it has been demonstrated that elderly people with the highest frequency of arousal during sleep have the greatest daytime sleepiness.[71]

Circadian Rhythms

A biphasic pattern of objective sleep tendency was observed when healthy normal young adult and elderly subjects were tested every 2 hours over a complete 24-hour day.[72] During the sleep period (11:30 PM to 8 AM) the latency testing was accomplished by awakening subjects for 15 minutes and then allowing them to return to sleep. Two troughs of alertness—one during the nocturnal hours (about 2 to 6 AM) and another during the daytime hours (about 2 to 6 PM)—were observed.

Other research protocols have yielded similar results. In constant routine studies, in which external environmental stimulation is minimized and subjects remain awake,

superimposed on the expected increase in self-rated fatigue resulting from the deprivation of sleep is a biphasic circadian rhythmicity of self-rated fatigue similar to that seen for sleep latency.[73] In another constant routine study in which EEG was continuously monitored, a biphasic pattern of "unintentional sleep" was observed.[74] In studies with sleep scheduled at unusual times, the duration of sleep periods has been used as an index of the level of sleepiness. A pronounced circadian variation in sleep duration is found, with the termination of sleep periods closely related to the biphasic sleep latency function in the studies cited earlier.[75] If individuals are permitted to nap when they are placed in time-free environments, this biphasic pattern becomes quite apparent in the form of a midcycle nap.[76]

This circadian rhythm in sleepiness is part of a circadian system in which many biologic processes vary rhythmically over 24 hours. The sleepiness rhythm parallels the circadian variation in body temperature, with shortened latencies occurring in conjunction with temperature troughs.[72] Nevertheless, these two functions, sleep latency and body temperature, are not mirror images of each other; the midday body temperature decline is relatively small compared with that of sleep latency. Further, under free-running conditions, the two functions become dissociated.[77] However, no other biologic rhythm is as closely associated with the circadian rhythm of sleepiness as is body temperature.

Earlier, it was noted that shift workers are unusually sleepy, and jet travelers experience sleepiness acutely in a new time zone. The sleepiness in these two conditions results from the placement of sleep and wakefulness at times that are out of phase with the existing circadian rhythms. Thus not only is daytime sleep shortened and fragmented but also wakefulness is required at the peak of sleepiness or trough of alertness. Several studies have shown that pharmacologic extension and consolidation of out-of-phase sleep can improve daytime sleepiness[78] (see Chapter 40 for more detail). Yet the basal circadian rhythm of sleepiness remains, although the overall level of sleepiness has been reduced. In other words, the synchronization of circadian rhythms to the new sleep-wake schedule is not hastened.

Central Nervous System Drugs

Sedating Drug Effects

CNS depressant drugs, as expected, increase sleepiness. Many of these drugs act as agonists at the $GABA_A$ receptor complex. The benzodiazepine hypnotics hasten sleep onset at bedtime and shorten the latency to return to sleep after an awakening during the night (which is their therapeutic purpose), as demonstrated by a number of objective studies.[79] Long-acting benzodiazepines continue to shorten sleep latency on the MSLT the day following bedtime administration.[79] Finally, ethanol administered during the daytime (9 AM) reduces sleep latency in a dose-related manner as measured by the MSLT.[80]

Second-generation antiepileptic drugs, including gabapentin, gabatrol, vigabatrin, pregabalin, and others, enhance GABA activity through various mechanisms that directly or indirectly involve the $GABA_A$ receptor.[81] The sedating effects of these various drugs have not been thoroughly documented, but some evidence indicates they do have sedative activity. $GABA_B$ receptor agonists are being investigated as treatments for drug addictions, and the preclinical animal research suggests these drugs may have sedative activity as well.[82]

Antagonists acting at the histamine-1 (H_1) receptor also have sedating effects. One of the most commonly reported side effects associated with the use of H_1 antihistamines is daytime sleepiness. Several double-blind, placebo-controlled studies have shown that certain H_1 antihistamines, such as diphenhydramine, increase sleepiness using sleep latency as the objective measure of sleepiness, whereas others, such as terfenadine or loratadine, do not.[83] The difference among these compounds relates to their differential CNS penetration and binding. Others of the H_1 antihistamines (e.g., tazifylline) are thought to have a greater peripheral compared with central H_1 affinity, and, consequently, effects on daytime sleep latency are found only at relatively high doses.[83]

Antihypertensives, particularly β-adrenergic receptor blockers, are also reported to produce sedation during the daytime.[84] These CNS effects are thought to be related to the differential liposolubility of the various compounds. However, we are unaware of any studies that directly measure the daytime sleepiness produced by beta blockers; the information is derived from reports of side effects. As noted earlier, it is important to differentiate sleepiness from tiredness or fatigue. Patients may be describing tiredness or fatigue resulting from the drugs' peripheral effects (i.e., lowered cardiac output and blood pressure), not sleepiness, a presumed central effect.

Sedative effects of dopaminergic agonists used in treating Parkinson disease have been reported as adverse events in clinical trials and in case reports as "sleep attacks" while driving.[85] It is now clear these "sleep attacks" are not attacks per se but are the expression of excessive sleepiness. Although the dose-related sedative effect of these drugs has been established, the mechanism by which the sedative effects occur is unknown. The dopaminergic agonists are also known to disrupt and fragment sleep.[86] Thus the excessive sleepiness may be secondary to disturbed sleep or to a combination of disturbed sleep and direct sedative effects.

Alerting Drug Effects

Stimulant drugs reduce sleepiness and increase alertness. The drugs in this group differ in their mechanisms of action. Amphetamine, methylphenidate, and pemoline block dopamine reuptake and to a lessor extent enhance the release of norepinephrine, dopamine, and serotonin. The mechanism of modafinil is not established; some evidence suggests that modafinil has a mechanism distinct from the classic stimulants. Amphetamine, methylphenidate, pemoline, and modafinil are used to treat the EDS associated with narcolepsy, and some have been studied as medications to maintain alertness and vigilance in normal subjects under conditions of sustained sleep loss (e.g., military operations). Studies in patients with narcolepsy using MSLT or MWT have shown improved alertness with amphetamine, methylphenidate, modafinil, and pemoline.[87] There is dispute as to the extent to which the excessive sleepiness of narcoleptics is reversed and the comparative efficacy of the various drugs. In healthy normal persons restricted or deprived of sleep, both amphetamine and methylphenidate increase alertness on the MSLT and improve psychomotor performance.[88,89] Caffeine is an adenosine receptor antagonist. In doses equivalent to 1 to 3 cups of coffee, caffeine reduced daytime sleepiness on the MSLT in normal subjects after 5 hours of sleep the previous night.[90]

Influence of Basal Sleepiness

The preexisting level of sleepiness-alertness interacts with a drug to influence the drug's behavioral effect. In other words, a drug's effect differs when sleepiness is at its maximum compared with its minimum. As noted previously, the basal level of daytime sleepiness can be altered by restricting or extending time in bed[60]; this in turn alters the usual effects of a stimulating versus a sedating drug. A study showed comparable levels of sleepiness-alertness during the day following 5 hours in bed and morning (9 AM) caffeine consumption compared with 11 hours in bed and morning (9 AM) ethanol ingestion.[91] Follow-up studies explored the dose relations of ethanol's interaction with basal sleepiness.[92] Dose-related differences in daytime sleepiness following ethanol and 8 hours of sleep were diminished after even 1 night of 5-hour sleep, although the measured levels of ethanol in breath were consistent day to day. In other words, sleepiness enhanced the sedative effects of ethanol. In contrast, caffeine and methylphenidate produced a similar increase in alertness, regardless of the basal level of sleepiness. Clinically, these findings imply, for example, that a sleepy driver with minimal blood ethanol levels may be as dangerous as an alert driver who is legally intoxicated.[92]

The basal state of sleepiness also influences drug-seeking behavior. The likelihood that a healthy normal person without a drug abuse history will self-administer methylphenidate is greatly enhanced after 4 hours of sleep the previous night compared with 8 hours of sleep (see Figure 4-4). Although not experimentally demonstrated as yet, self-administration of caffeine also is probably influenced by basal state of sleepiness. The high volume of caffeine use in the population probably relates to the high rate of self-medication for sleepiness due to chronic insufficient sleep in the population.

Central Nervous System Pathologies

Pathology of the CNS is another determinant of daytime sleepiness. The previously noted hypocretin/orexin deficiency is thought to cause excessive sleepiness in patients with narcolepsy.[93] Another sleep disorder associated with excessive sleepiness due to an unknown pathology of the CNS is idiopathic CNS hypersomnolence. A report of a series of rigorously diagnosed cases ($n = 77$) found moderate MSLT scores (8.5-minute mean latency) relative to narcoleptics (4.1-minute mean latency).[94] As yet hypocretin/orexin deficiency has not been shown in this disorder. These two conditions are described in detail in Chapters 89 and 91.

Excessive sleepiness is reported in other neurologic diseases. A study in patients with myotonic dystrophy type 1 reported excessive sleepiness on the MSLT and reduced cerebrospinal levels of hypocretin/orexin.[95] "Sleep attacks" have been reported in Parkinson disease, and assessment with the MSLT suggests these "attacks" are the expression of excessive daytime sleepiness.[96] What remains unresolved in the excessive sleepiness of Parkinson disease is the relative contribution of the disease itself, the fragmentation of sleep due to periodic leg movements or apnea, and the dopaminergic drugs used in treating Parkinson disease.[97] The previously cited study found no differences in sleepiness as a function of prescribed drug or sleep fragmentation, although further assessment in larger unselected samples is necessary to confirm this finding.

CLINICAL AND PUBLIC HEALTH SIGNIFICANCE OF SLEEPINESS

Although the patients at sleep disorders centers are not representative of the general population, they do provide some indications regarding the clinical significance of sleepiness. Their sleep-wake histories directly indicate the serious impact excessive sleepiness has on their lives.[91] Nearly half the patients with excessive sleepiness report automobile accidents; half report occupational accidents, some life threatening; and many have lost jobs because of their sleepiness. In addition, sleepiness is considerably disruptive of family life.[98] An elevated automobile accident rate (i.e., sevenfold) among patients with excessive sleepiness has been verified through driving records obtained from motor vehicle agencies.[99]

Population-based information regarding traffic and industrial accidents also suggests a link between sleepiness and life-threatening events. Verified automobile crashes occurred more frequently in a representative sample of people with MSLT scores of 5 minutes or less.[100] The highest rate of automobile crashes occurs in the early morning hours, which is notable because the fewest automobiles are on the road during these hours. Also during these early morning hours, the greatest degree of sleepiness is experienced.[101] Long-haul truck drivers have crashes most frequently (even corrected for hours driving before the accident) during the early morning hours, again when sleepiness reaches its zenith.[102]

Workers on the graveyard shift were identified as a particularly sleepy subpopulation. In 24-hour ambulatory EEG recordings of sleep and wakefulness, workers (20% in one study) were found to actually fall asleep during the night shift.[2] Not surprisingly, the poorest job performance consistently occurs on the night shift, and the highest rate of industrial accidents is usually found among workers on this shift.[103] Medical residents are another particularly sleepy subpopulation. In surveys, subjects reporting 5 or fewer hours of sleep per night were more likely to make medical errors and report serious accidents, and they were two times more likely to be named in medical malpractice suits.[104,105] In a survey of medical house staff, 49% reported falling asleep while driving and 90% of the episodes occurred postcall compared with 13% of fall-asleep episodes reported by the medical faculty, and 20 of the 70 house staff were involved in automobile crashes compared with 11 of the 85 faculty.[105]

Cognitive function is also impaired by sleepiness. Adults with various disorders of excessive sleepiness have cognitive and memory problems.[106] The memory deficiencies are not specific to a certain sleep disorder but rather specific to the sleepiness associated with the disorder.[107] When treated adequately, sleepiness is rectified and the memory and cognitive deficits similarly improve.[108] Results of sleep deprivation studies in healthy normal patients support the relation between sleepiness and memory deficiency. Even modest reductions of sleep time are associated with cognitive deficiencies.[109]

Sleepiness also depresses arousability to physiologic challenges: 24-hour sleep deprivation decreases upper airway dilator muscle activity[110] and decreases ventilatory responses to hypercapnia and hypoxia.[111] In a canine model of sleep apnea, periodic disruption of sleep with acoustic stimuli (i.e., sleep fragmentation, in contrast to sleep deprivation) resulted in lengthened response times to airway occlusion, greater oxygen desaturation, increases in inspiratory pressures, and

surges in blood pressure.[112] Depressed physiologic responsivity due to sleepiness is clinically significant for patients with sleep apnea and other breathing disorders because they are all exacerbated by sleepiness. The emerging data on sleepiness and pain threshold, cited earlier, are also clinically significant in the management of both acute and chronic pain conditions.

Finally, life expectancy data directly link excessive sleep (not specifically sleepiness) and mortality. A 1976 study found that men and women who reported sleeping more than 10 hours a day were about 1.8 times more likely to die prematurely than those sleeping between 7 and 8 hours daily.[113] This survey, however, associated hypersomnia and increased mortality and not necessarily EDS, for which the relation is currently unknown.

CLINICAL PEARL

Sleepiness, when most excessive and persistent, is a signal to the individual to stop operating: It is dangerous and life-threatening to continue without sleep. To the clinician, that signal warns that there may be some underlying pathology that can be successfully treated or in the very least minimized as to its vital, life-threatening impact.

SUMMARY

Sleepiness is a problem reported by 10% to 25% of the population, depending on the definition of sleepiness used and the population sampled. It is most common in young adults and elderly persons. Sleepiness is a physiologic need state, with its intensity evident by how rapidly sleep onset occurs, how easily sleep is disrupted, and how long sleep endures. Validated self-rated scales and physiologic measures are available to assess the presence and degree of sleepiness. Relative to sleepy, healthy adults, the chronicity and irreversibility of sleepiness indicate its clinical and pathologic significance. Sleepiness is caused by reduced sleep time as often seen in otherwise healthy adults, by fragmented and disrupted sleep as found in patients with primary sleep disorders, by administration of sedating drugs and discontinuation of alerting drugs, and by various neurologic disorders. Sleepiness has a normal circadian rhythm that is increased in circadian rhythm misalignments such as those occurring in shift work or jet lag. Excessive and persistent sleepiness is life threatening, but when its presence is recognized and its etiology identified, it can be successfully treated or at least minimized.

Selected Readings

Carskadon MA, Dement WC, Mitler MM, et al. Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. *Sleep* 1986;**9**:519–24.

Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep* 1997;**20**:267–77.

Elmenhorst D, Meyer PT, Winz OH, et al. Sleep deprivation increases A1 adenosine receptor binding in the human brain: a positron emission tomography study. *J Neurosci* 2007;**27**:2410–15.

Fernandez-Mendoza J, Vgontzas AN, Kritikou I, et al. Natural history of excessive daytime sleepiness: role of obesity, weight loss, depression, and sleep propensity. *Sleep* 2015;**38**:351–60.

Gonçalves M, Amici R, Lucas R, et al. Sleepiness at the wheel across Europe: a survey of 19 countries. *J Sleep Res* 2015;**24**(3):242–53.

Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;**17**:703–10.

Kim H, Young T. Subjective sleepiness dimensions and correlates in the general population. *Sleep* 2005;**28**:625–34.

Mignot E. A commentary on the neurobiology of the hypocretin/orexin system. *Neuropsychopharmacology* 2001;**25**:S5–13.

Punjabi NM, Bandeen-Roche K, Young T. Predictors of objective sleep tendency in the general population. *Sleep* 2003;**26**:678–83.

Weaver TE, Maislin G, Dinges Bloxham T, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. *Sleep* 2007;**30**:711–19.

A complete reference list can be found online at ExpertConsult.com.

Chapter 5

Sleep Deprivation

Siobhan Banks; Jill Dorrian; Mathias Basner; David F. Dinges

Chapter Highlights

- Surveys indicate that 35% to 40% of the adult U.S. population report sleeping less than the usually recommended 7 to 8 hours on weekday nights, and about 15% report sleeping less than 6 hours.
- Neurobehavioral deficits accumulate across days of sleep restriction to levels equivalent to those found after 1 to 3 nights of total sleep deprivation.
- Neurobehavioral responses to sleep deprivation have been found to be stable and consistent, suggesting they are trait-like and possibly genetic.
- Recovery appears to be affected by the type of sleep loss (acute versus chronic), the recovery sleep duration, and the number of days allowed for recovery. In addition, although an individual may report feeling recovered, performance may remain impaired, increasing risk for accidents and injury.

Both sufficient sleep continuity and sleep duration are prerequisites for the recuperative effects of sleep.[1] Acute total sleep deprivation refers to wake periods that extend beyond the typical 16 to 18 hours, whereas sleep restriction refers to not getting enough sleep per 24 hours for one or multiple nights. Sleep restriction (i.e., too little sleep for prolonged periods) occurs frequently and results from a number of factors, including medical conditions (e.g., pain), sleep disorders, work demands (including extended work hours and shift work), and social and domestic responsibilities.[2] Shift work affects one of five working Americans and is projected to increase in the future. Shift work includes working evenings and nights or rotating shifts, and because sleep is often scheduled at an adverse circadian phase, it is shorter and more disrupted.[3]

Both types of sleep restriction (acute and chronic) affect sleep recuperation and neurobehavioral performance. Neurobehavioral deficits accumulate as sleep loss increases. This can then result in an increased risk for workplace errors and injuries, traffic accidents, personal conflicts, health complaints, and drug and alcohol use. Well-controlled sleep studies found that restriction of sleep to between 3 and 7 hours of time in bed per 24 hours for a period of 1 to 2 weeks resulted in declining behavioral alertness and cognitive performance across days in a near-linear and dose-response fashion.[4,5] The cumulative deficits in performance on the Psychomotor Vigilance Test (PVT) that developed during 2 weeks of sleep restricted to 4 hours per night were comparable to those after 3 nights of total sleep deprivation,[5] showing that sleep restriction can induce waking brain deficits equivalent to the most severe total sleep deprivation.

Short-term sleep restriction occurs when one fails to obtain a usual amount of sleep.[6] A half-century ago, Kleitman[7] first used the phrase "sleep debt" to describe the circumstances of delaying sleep onset time while holding sleep termination time constant. He described the increased sleepiness and decreased alertness in individuals on such a sleep-wake pattern and proposed that those subjects who were able to reverse these effects by extending their sleep on weekends were able to "liquidate the debt."

Based on the manner in which the concept of sleep debt is most often used, it refers to increased pressure for sleep that results from an inadequate amount of physiologically normal sleep. To determine the effects of sleep loss on a range of neurobehavioral and physiologic variables, a variety of paradigms have been used, including controlled, restricted time in bed for sleep opportunities in both continuous and distributed schedules (e.g., nocturnal anchor sleep with a daytime nap),[8] gradual reductions in sleep duration over time,[9] selective deprivation of specific sleep stages,[10] and situations in which the time in bed is individualized, such that it is reduced to a percentage of the individual's habitual time in bed.[11] These studies have ranged from 24 hours[12] to 8 months[9] in length, but they typically involved subjects outside the laboratory in uncontrolled conditions.

Many of the early experimental reports and reviews published before 1997 concluded that sleep restriction in the range commonly experienced by the general population (i.e., sleep durations of 4 to 7 hours) resulted in some increased subjective sleepiness but had little or no effect on cognitive performance capabilities. Consequently, there was a widely held belief that individuals could "adapt" to reductions in sleep duration, to 4 to 5 hours per day. However, nearly all of these early reports of adaptation to sleep loss were limited either by lack of experimental control over a considerable number of

critical variables, including sleep actually obtained at night and during the day (i.e., naps) and the use of stimulants (caffeine, nicotine), or by a reliance on subjective reports of sleepiness, which have since been shown to be a poor reflection of actual neurobehavioral impairment due to sleep restriction. Most of these early studies also used small sample sizes, failed to include a control group, lacked physiologic measures of sleep and waking, focused on only a few assays of performance or used performance assessments contaminated by practice effects, had infrequent assessment times within and between days, and failed to test quantitatively for cumulative performance changes relative to the cumulative sleep loss experienced by the subjects.[11] They are a reminder that inadequate scientific methods result in erroneous results. Since 1997, experiments that have corrected for these methodologic weaknesses have found markedly different results from those earlier studies and have documented cumulative objective changes in outcomes as sleep restriction progressed.[6]

At the same time, epidemiologic studies have found that habitual short sleep duration is associated with negative health consequences, including obesity,[13] diabetes,[14] hypertension,[15] cardiometabolic risk factors[16] and cardiovascular disease,[17] declines in cognitive function,[18] and all-cause mortality.[19] The finding is particularly important when it is observed in prospective population studies. Evidence from experimental studies shows that both acute total sleep deprivation and sleep restriction cause inadequate pancreatic insulin secretion,[20] decreased insulin sensitivity,[21] changes in the appetite-regulating hormones leptin and ghrelin,[22] attenuated immune response to vaccination,[23] and increased sympathetic activity and venous endothelial dysfunction in healthy adults.[24] These effects provide biologic plausibility for a causal relationship between short sleep and negative health outcomes as well as all-cause mortality when considered with the well-documented safety risks of reduced sleep (e.g., the increased risk for motor vehicle crashes).

This chapter reviews the cognitive and neurobehavioral consequences of acute total sleep deprivation and sleep restriction in healthy individuals and the theoretical explanations for these effects.

INCIDENCE OF SLEEP DEPRIVATION

Human sleep need, or more precisely, the duration of sleep needed to prevent feeling sleepy during daytime, elevated sleep propensity, and cognitive deficits, has been a long-standing controversy central to whether sleep restriction may compromise health and behavioral functions. Despite the scientifically documented benefits of sufficient sleep for cognitive performance, safety, and health, current representative surveys indicate that 35% to 40% of the adult U.S. population report sleeping less than the usually recommended 7 to 8 hours on weekday nights, and about 15% report sleeping less than 6 hours.[25] Importantly, these are self-reported sleep time estimates that have been shown to overestimate actigraphically or polysomnographically measured sleep by up to 1 hour.[26,27] Less is known about the prevalence of acute total sleep deprivation in the population, but given the ever-increasing pervasiveness of shift work in our 24/7 society, it is likely that the prevalence of acute total sleep deprivation experienced while transitioning into and out of night shift work is increasing as well.

EFFECTS OF SLEEP DEPRIVATION

Acute Total Sleep Deprivation

The first total sleep deprivation and cognition study was conducted in the late nineteenth century. Studies at this time involved substantial sleep deprivation periods (36 to 90 hours) and demonstrated that memory and response time were significantly impaired.[28] Over time, hundreds of studies of total sleep loss have found significant cognitive deficits (for review, see[29]). Many facets of waking function are affected by sleep deprivation. These include reduced cognitive processing speed,[30] constructive thinking,[31] verbal memory,[32] and spatial working memory[33] as well as increased tendency for false memories.[34] Further, there are changes in subjective sleepiness and in mood and emotional processing (including reading positive emotional expressions), reduced threshold for stress, and elevated stress reactions.[35,36] Interestingly, reasoning and complex cognitive tasks appear to remain substantively unaffected by acute total sleep deprivation.[37]

One of the most sensitive and widely used cognitive tasks in studies of sleep loss is the PVT,[38] which is a measure of vigilant attention.[39] Studies have consistently shown that sleep deprivation increases PVT response slowing and lapses,[40] which are thought to reflect microsleeps.[7,41] As sleep deprivation accumulates, microsleeps or brief lapses of a half-second can increase to 10 seconds and longer.[7,41,42] It has been suggested that these lapses involve shifts in neuronal activity in frontal, thalamic, and secondary sensory processing areas of the brain.[43] In sleep-deprived subjects, cognitive performance contains lapses of attention that occur unpredictably, and they increase in frequency and duration as a function of the length of the sleep deprivation. This has led to the "wake state instability" hypothesis.[39,40,42,43] This instability appears to involve moment-to-moment fluctuations in the relationship between neurobiologic systems mediating wake maintenance and sleep initiation.[43] State instability is manifested in increased failures to respond to stimuli (errors of omission) as well as increased responses in the absence of stimuli (errors of commission).[29,42]

It has been demonstrated that the negative impact of total sleep deprivation captured in these laboratory studies is not only measurable but also meaningful. The performance effects of sleep deprivation and alcohol intoxication have been demonstrated to be qualitatively and quantitatively similar.[44] Dawson and Reid[44] found that performance impairment after 17 hours awake was equivalent to that produced by a blood alcohol concentration of 0.05%.

An overarching finding from the studies of total sleep deprivation is that waking function is influenced by a sleep homeostatic process that builds up during wakefulness and declines during sleep in a nonlinear fashion (as measured by slow wave energy or delta power in the non–rapid eye movement sleep electroencephalogram) and a circadian process, with nearly 24-hour periodicity.[45] The combined influences of these factors are described as the two-process model.[46] Since its inception, the two-process model has gained widespread acceptance for its explanation of the timing and structure of sleep. Its use has extended to predictions of waking alertness and neurobehavioral functions in response to different sleep-wake scenarios.[46] This extension of the two-process model was based on observations that as sleep pressure accumulated with increasing time awake, so

did waking neurobehavioral or neurocognitive impairment; as sleep pressure dissipated with time asleep, performance capability improved during the following period of wakefulness. In addition, forced-desynchrony experiments revealed that the sleep homeostatic and circadian processes interacted to create periods of stable wakefulness and consolidated sleep during normal 24-hour days.[47]

The two-process model has been used in the development of tools to model changes in sleep, alertness, and performance on the basis of recent sleep-wake (or work) history. These are discussed in detail in Chapter 74. In general, these models accurately predict waking performance and self-reported responses to total sleep deprivation. However, they often fail to adequately predict sleepiness and cognitive performance responses during sleep restriction.[48] Recent model extensions have started to deal with these shortcomings.[49] It is clear that the two-process model has had a profound theoretical influence on predictions of sleepiness based on total sleep deprivation data, but its inability to capture all the dynamic changes in PVT performance or cognitive tasks during sleep restriction suggests there are additional biologic factors relevant to the brain's response to sleep restriction.

Sleep Restriction

The effects of sleep loss may be quantified by physiologic, behavioral, cognitive, and subjective tools. Two large-scale, controlled laboratory studies identified dose-related effects of short-term sleep restriction on neurobehavioral performance measures.[4,5] In one study, truck drivers were randomized to 7 nights of 3, 5, 7, or 9 hours of time in bed for sleep per night and performance was assessed with the PVT.[4] Subjects in the 3- and 5-hour time-in-bed groups experienced a decrease in performance across days of the sleep restriction protocol, with increases in the mean reaction time, in the number of lapses, and in the speed of the fastest reaction on the PVT.[4] A significant decrease in response speed was seen in subjects who had 7 hours of time in bed per night. Performance in the group allowed 9 hours of time in bed was stable across the 7 days.

In another major sleep restriction experiment,[5] young adults had their sleep duration confined to 4, 6, or 8 hours of time in bed per night for 14 nights. Deficits in cognitive functions were observed, including reduced vigilant attention (Figure 5-1), impaired memory task, and slowed mental processing speed. These performance deficits accumulated across the experimental protocol in those subjects allowed less than 8 hours of time in bed for sleep per night.[5] Data from this study demonstrate that sleep restriction–induced deficits continued to accumulate beyond the 7 nights of restriction used in other experiments.[4,11] On day 14 of the sleep restriction schedule, subjects in the 4-hour time-in-bed condition had performance similar to that after 1 or 2 nights without any sleep[5] (Figure 5-1).

It also appears that the neurocognitive effects of restricting nocturnal sleep to 6 or 4 hours per night for multiple nights are fundamentally the same as when sleep is split each day into two sleep opportunities.[50] Cognitive performance deficits also accumulate across consecutive days in which restricted sleep occurs during the daytime and wakefulness occurs at night.[51] The primary difference between the nocturnally[5] and diurnally[51] placed restricted sleep periods is that the magnitude of neurobehavioral impairment is significantly greater with daytime sleep.

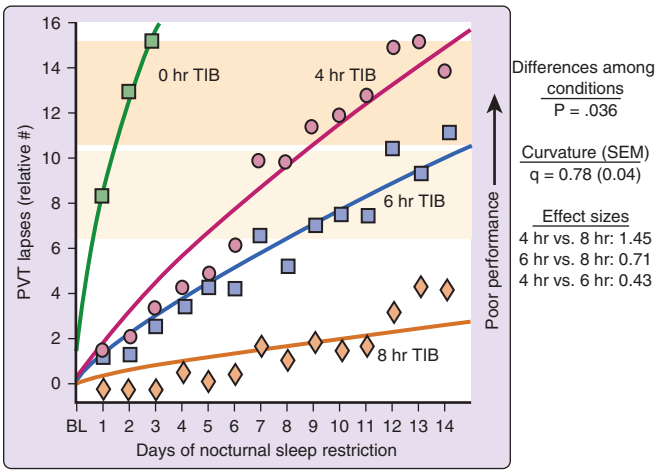


Figure 5-1 Psychomotor Vigilance Test (PVT) performance lapses under varying doses of daily sleep. Displayed are group averages for subjects in the 8-hour (*diamond*), 6-hour (*blue square*), and 4-hour (*circle*) sleep period time in bed (TIB) across 14 days and in the 0-hour (*green square*) sleep condition across 3 days. Subjects were tested every 2 hours each day; data points represent the daily average (0730 hours to 2330 hours) expressed relative to baseline (BL). The curves through the data points represent statistical nonlinear model-based best-fitting profiles of the response to sleep deprivation for subjects in each of the four experimental conditions. The mean ± standard error of the mean (SEM) ranges of neurobehavioral functions for 1 and 2 days of 0 hours of sleep (total sleep deprivation) are shown as *light and dark bands*, respectively, allowing comparison of the 3-day total sleep deprivation condition and the 14-day sleep restriction conditions. (Modified from Van Dongen HP, Maislin G, Mullington JM, et al. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from sleep restriction and total sleep deprivation. *Sleep* 2003;26:117–26.)

All these studies suggest that when time in bed for sleep is restricted to less than 7 hours per night in healthy adults (aged 21 to 64 years) for multiple nights, cumulative deficits in a variety of cognitive performance functions can accumulate. In addition, these deficits can be similar to those seen after 1 night or even 2 nights of total sleep deprivation. Collectively, they suggest that there is a neurobiologic integrator that accumulates either homeostatic sleep drive or the neurobiologic consequences of excess wakefulness.[4,5] No definitive evidence exists yet as to what this neurobiologic integrator might be, but one hypothesis suggests that it may involve extracellular adenosine in the basal forebrain.[52]

In contrast to reports of continuing accumulation of cognitive deficits associated with sleep restriction, subjective assessments of sleepiness and alertness demonstrate near-saturating functions across periods of sleep restriction.[5] Self-reported mood, sleepiness, and fatigue do not parallel the continuing decline in cognitive performance associated with nightly restriction of sleep to 7 hours or less.[5] This suggests that individuals frequently underestimate the impact of sleep deprivation on their performance. Experiments using driving simulators have found similar results, with drivers unable to accurately recognize crash risk.[53]

There is increasing evidence of physiologic and health-related consequences of sleep loss. Alterations in physiologic parameters, such as endocrine and immune function, have

been recognized and have implications for health status and risk. An elevated mortality risk in those individuals who reported sleeping less than 6.5 hours per night has been found.[54] As outlined before, the literature points to a strong relationship between insufficient sleep and poor metabolic health characterized by conditions such as cardiovascular disease,[55] type 2 diabetes,[56] and obesity.[57] A meta-analysis study found a consistent increased risk of obesity among short sleepers in both children (sleeping <10 hours per night) and adults (sleeping <5 hours per night), but as the authors pointed out, causal inference was difficult because of the lack of control of confounders and inconsistency in the methodologies used.[58] On the other hand, a recent laboratory-based experimental study of the effects of sleep restriction to 4 hours per night for 5 nights in 198 healthy adults did find that relative to control subjects allowed up to 10 hours of sleep per night, sleep-restricted adults gained an average of nearly 1 kg of weight during the sleep restriction period.[59]

Extended wakefulness appears to have a metabolic cost that triggers neuroendocrine and immune function changes.[60] In a seminal experimental study that compared sleep restriction (4 hours per night for 6 nights) with sleep extension (12 hours per night for 6 nights), evening cortisol was elevated, sympathetic activation was increased, and thyrotropin activity and glucose tolerance were decreased.[61] It has been long recognized that a relationship exists between sleep and immune function. Changes in natural killer cell activity,[62,63] lymphokine-activated killer cell activity,[62] interleukin-6,[64] and soluble tumor necrosis factor-alpha receptor 1[65] have been reported with total sleep deprivation and sleep restriction. C-reactive protein is a predictive inflammatory marker of increased risk for cardiovascular disease, and increased C-reactive protein has been reported after both total sleep deprivation[66] and sleep restriction (4 hours of time in bed for sleep per night) in healthy individuals.[67]

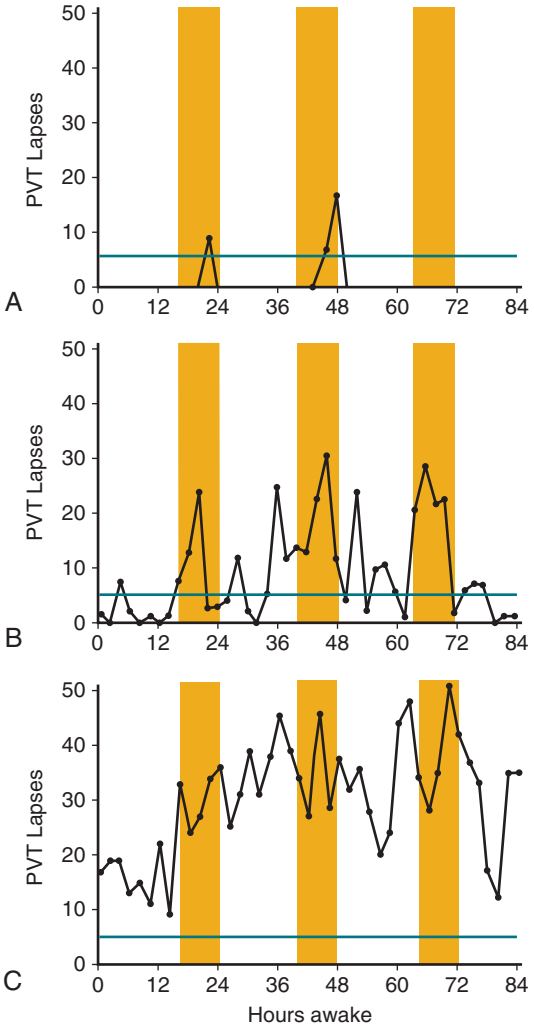


Figure 5-2 Individual differences in the response to sleep deprivation (D.F. Dinges, unpublished data). The three subjects in **A** to **C** performed a 10-minute Psychomotor Vigilance Test (PVT) every 2 hours during 88 hours of acute total sleep deprivation. The *horizontal line* reflects 5 PVT lapses (number of reaction times >500 ms), and *bars* indicate the period from 0000 hours to 0800 hours. **A,** A resilient response. **B,** Affected by sleep deprivation but with daytime improvement. **C,** A vulnerable response to sleep deprivation. These individual differences in the response to sleep deprivation on the PVT were not accounted for by demographic factors, IQ, or sleep need. (From Basner M, Rao H, Goel N, Dinges DF. Sleep deprivation and neurobehavioral dynamics. *Curr Opin Neurobiol* 2013;23:854–63, with permission.)

INDIVIDUAL DIFFERENCES IN RESPONSES TO SLEEP LOSS

Whereas the majority of healthy adults experience sleepiness and related cognitive deficits under conditions of total sleep deprivation or sleep restriction, interindividual variability in the neurobehavioral and physiologic responses is substantial.[7,29,42,68] There are certain individuals who display minimal impairment during sleep loss, some who are moderately affected, and others who are particularly vulnerable (Figure 5-2).[69] To date, research has not supported the suggestion that differences in vulnerability reflect baseline scores (following adequate rest), nor have studies found that demographic factors, such as IQ, habitual sleep amount, or personality, account for these differences.[70] Studies investigating whether the same individuals are vulnerable to total sleep deprivation and sleep restriction have yielded mixed results, likely due to small samples and varied methods.[71,72]

Importantly, however, within subjects repeatedly exposed to the same sleep loss intervention, neurobehavioral responses to sleep deprivation have been found to be stable and consistent, suggesting they are trait-like.[73] This also seems to be true for total sleep deprivation and sleep restriction.[68] It has therefore been suggested that these trait-like differences in vulnerability may reflect underlying genetic differences.[70] A number of possible genetic contributors have been considered in the literature. These include *PERIOD3* variable number tandem repeat (*PER3*)[74,75] and *ADORA2A* polymorphisms.[75] Whereas some work has suggested a connection between *PER3* and performance vulnerability to total sleep deprivation[76] and sleep restriction,[75] other researchers have found that sleep homeostatic responses, rather than performance responses during sleep restriction, were related to *PER3*.[74] It has also been suggested that the importance of *PER3* and *ADORA2A* polymorphisms as they relate to performance variability during sleep restriction may be influenced by the length and severity of the sleep loss.[75] Currently, this work requires larger replicate samples to untangle some of these relationships.

DETECTION OF SLEEPINESS

Sleepiness is often defined as the "propensity to fall asleep" and can be measured both objectively and subjectively[77]; however, as mentioned before, self-reported sleepiness has been shown to be unreliable (i.e., not reflecting performance capability) during sleep loss.[78] This stresses the need for brief, objective, and unobtrusive measures of sleepiness that can be readily applied under operational conditions.

The measures that best track the subtle changes in performance caused by sleep deprivation are neural state indicators (such as electroencephalography, electrooculography, and functional magnetic resonance imaging) or behavioral indicators of attention stability, like the PVT.[79] The PVT has been shown to be among the most sensitive measures of sleep loss, partly as it prevents compensatory stimulation and is not confounded by aptitude and learning effects like other performance measures.[39,79] It also reflects performance that has ecologic validity (i.e., vigilant attention is needed for learning, safe driving). The PVT is thus often used as a "gold standard" measure for the neurobehavioral consequences of sleep loss against which other fatigue detection technologies are measured.

Chua et al[80] investigated how well several objective physiologic variables as well as subjective measures of sleepiness correlated with the PVT during 1 night of acute total sleep deprivation. The highest correlation was found for the percentage of slow eye closures, in agreement with several authors who found oculomotor responses to be sensitive to sleep restriction.[81] Eyelid closure and slow rolling eye movements are part of the initial transition from wake to drowsiness; they are associated with vigilance lapses and have been found to be a sign of drowsiness while driving.[82] Increased slow eye movements attributed to attentional failures have been reported to be increased by reduced sleep time in medical residents.[83] Sleep restriction has also been found to decrease saccadic velocity and to increase the latency to pupil constriction in subjects allowed only 3 hours or 5 hours of time in bed for sleep over 7 nights.[84] Other variables that correlated highly with PVT performance included the power spectral density of the electrocardiogram RR interval in the range of 0.02 to 0.08 Hz, electroencephalogram delta power in the range of 1 to 4.5 Hz, followed by self-reported measures of sleepiness.[80] The high correlation between heart rate variability and the PVT has recently been replicated for sleep restriction as well.[85]

SLEEP DEPRIVATION AND BRAIN METABOLISM

Sleep deprivation induces changes in brain metabolism and neural activation that involve distributed networks and connectivity.[1] Early positron emission tomography sleep deprivation studies found metabolic rate reductions in thalamic, parietal, and prefrontal regions associated with prolonged sleep loss.[4,86] More recent blood oxygenation level–dependent functional magnetic resonance imaging (fMRI) studies showed significant declines in regional brain activation during cognitive task performance after 1 night of total sleep deprivation. These changes included reduced frontoparietal activation during lapses on a visual selective attention task[87] and were mainly observed in vulnerable subjects with larger performance deficits, whereas resilient subjects demonstrated a trend toward increased parietal activation during performance lapses,[87] suggesting a potential compensatory mechanism after sleep loss. Recent positron emission tomography studies have observed downregulation of striatal dopamine receptors[88] and increased cerebral serotonin receptor binding with sleep deprivation.[89] This may reflect a complex adaptive response of the brain to sleep deprivation.

Poudel et al[90] used arterial spin-labeled perfusion fMRI to measure resting cerebral blood flow (CBF) changes after 1 night with 4-hour sleep opportunity. Only drowsy participants (established with video of the eyes during the scan) exhibited a significantly reduced frontoparietal CBF after sleep deprivation, whereas nondrowsy subjects maintained frontoparietal CBF and increased CBF in basal forebrain and cingulate regions. These findings support a compensatory mechanism after sleep loss,[90] which may explain some of the variance between those resilient and those vulnerable to sleep loss. Recent resting-state functional connectivity fMRI (FC-fMRI) studies have consistently shown an organized mode of resting brain function.[91] Two recent FC-fMRI studies reported reduced functional connectivity within the default mode network and between the default mode network and its anti-correlated network associated with sleep deprivation,[92] suggesting that functional connectivity of the brain changes as a result of sleep loss. A recent meta-analysis on the effects of acute total sleep deprivation on the attending brain found decreases in brain activation in the frontoparietal attention network (prefrontal cortex and intraparietal sulcus) and in the salience network (insula and medial frontal cortex) but increases in thalamic activation (the salience network segregates the most relevant internal and external stimuli to guide behavior). The authors speculated that the latter may reflect a complex interaction between the dearousing effects of sleep loss and the arousing effects of task performance on thalamic activity.[93]

RECOVERY FROM SLEEP LOSS

An emerging area of research has begun to focus on recovery processes from acute total sleep deprivation and sleep restriction. Results from these studies suggest that the recovery process may be slower and more complex than originally thought. Recovery appears to be affected by the type of sleep loss (acute versus chronic), the recovery sleep duration, and the number of days allowed for recovery. In addition, aspects of neurobehavioral functioning appear to recover at different rates, and although an individual may report feeling recovered, performance may remain impaired, increasing risk for accidents and injury. It is therefore critically important to understand how much sleep people need to recover from periods of sleep loss.

Recovery following Sleep Restriction

Few studies have examined recovery sleep after periods of sleep restriction. Two important sleep restriction studies that also included a short in-laboratory recovery phase were those by Dinges et al[11] and Belenky et al.[4] In the Dinges study, sleep was restricted below habitual sleep duration by 33% (average, 4.98 hours per night; standard deviation, 0.57) for a consecutive 7 nights, after which participants were allowed one or two

10-hour recovery sleeps.[11] In the Belenky study, participants were permitted 3, 5, 7, or 9 hours in bed each night, for 7 nights, followed by three 8-hour recovery opportunities.[4] These studies showed that either two 10-hour or three 8-hour sleep opportunities are sufficient to recover performance to baseline levels. Although participants felt that their functioning was restored, with subjective reports of sleepiness and performance recovering to baseline, subjective measures do not appear to accurately parallel objective measures of neurobehavioral recovery. These findings suggest that more than 2 or 3 nights of extended sleep may be needed to return neurobehavioral functions to baseline levels. This may be especially important in situations in which individuals are not able to choose or to extend the length of their recovery sleep period.

Recent studies have extended this work by manipulating the amount of sleep obtained before a period of sleep restriction in a prophylactic manner, the duration of the recovery period (longer than 3 days), and the length of the recovery sleep opportunity. Rupp et al[94] extended the recovery period to five 8-hour sleep opportunities, following a week of sleep restricted to 3 hours of time in bed each night.[94] Despite the extra recovery nights, performance still failed to return to baseline levels. In a second group that extended their sleep prophylactically to 10 hours of time in bed before the period of sleep restriction, performance declined at a slower rate and recovered to baseline more quickly, suggesting that sleep restriction and recovery vary as a function of prior sleep.[94] Therefore, sleep extension or supplemental nap sleeps can be used as a prophylactic measure or countermeasure to lessen the performance effects of sleep deprivation during periods of extended wakefulness or sleep deprivation.

In the first study to systematically investigate different doses of recovery sleep (0, 2, 4, 6, 8, and 10 hours of time in bed) after sleep restriction of 4 hours per night for 5 nights, Banks et al[30] found that the extended 10-hour recovery sleep opportunity was not enough for full recovery of sustained attention, subjective sleepiness, or fatigue. Importantly, it was observed that each recovery sleep dose had equivalent restorative value (i.e., linear) for sleep and objective sleepiness measures. However, in contrast, recovery of performance outcomes, sustained attention, and subjective sleepiness was exponential, indicating that the restorative value of the recovery sleep doses decreased with increasing sleep time.

Overall, work to date suggests that complete recovery from a period of sleep restriction may necessitate a sleep opportunity of more than 10 hours or more than 3 days if sleep is restricted to 8 hours a night. In addition, different aspects of performance and neurobehavioral function appear to recover at different rates, with different trajectories.

Recovery following Acute Total Sleep Deprivation

Both the number of nights needed for recovery and extended recovery sleep opportunities following total sleep deprivation have been investigated.[95,96] In one particular study, participants were deprived of sleep for 1 or 2 nights, followed by a consecutive 5 nights for recovery that were either restricted to 6 hours or extended to 9 hours of time in bed. Independent of the severity of the prior sleep loss or the length of recovery opportunity, slow wave sleep returned to baseline levels after the first recovery sleep, and participants in each of the groups obtained comparable amounts of slow wave sleep. Performance was also complete after 1 night of extended recovery sleep following 1 night of sleep deprivation. However, when the period of sleep deprivation was more severe (i.e., 2 nights), cognitive performance remained significantly below baseline even after 5 nights of extended recovery sleep. Neither performance nor subjective sleepiness recovered to baseline when recovery sleep time was 6 hours of time in bed per night.

These studies underscore the importance of not restricting recovery sleep opportunities to less than 6 hours. Indeed, it is apparent that adding just 1 hour extra can significantly increase the rate of recovery from sleep deprivation. Moreover, it appears that the more severe the sleep loss, the longer the recovery sleep opportunity required.

CLINICAL PEARL

Physicians should be aware that sleep deprivation can be caused by sleep disorders, work schedules, and modern lifestyles. The consequences of sleep deprivation include excessive daytime sleepiness and poor mood, impaired cognitive performance, and driving safety and appear to play a role in chronic health diseases, such as cardiovascular disease, obesity, and type 2 diabetes.

SUMMARY

Sleep is important for healthy functioning, but sleep deprivation is common in many sectors of the community. Sleep deprivation causes a multitude of changes to cognitive and behavioral functioning and also increases the risk for chronic diseases, such as type 2 diabetes, obesity, and cardiovascular disease. There are large individual differences in the cognitive responses to sleep deprivation. These individual differences are stable across multiple exposures and types of sleep loss, suggesting a trait-like (possibly genetic) basis. The negative effects of sleep deprivation can be recovered by extending sleep, but recovery appears to be dependent on the type of sleep loss (acute versus chronic), the recovery sleep duration, and the number of days allowed for recovery. Biomathematical models and the development of unobtrusive technologies to detect fatigue while working are ways in which the deleterious effects of sleep deprivation can be managed in our 24/7 society.

ACKNOWLEDGMENT

This chapter was written with support from the National Institutes of Health grant R01 NR004281 (D.F.D.) and the National Space Biomedical Research Institute through NASA cooperative agreement NCC 9-58 (D.F.D., M.B.).

Selected Readings

Banks S, Dinges DF. Behavioral and physiological consequences of sleep restriction. *J Clin Sleep Med* 2007;**3**:519–28.

Banks S, Van Dongen HP, Maislin G, Dinges DF. Neurobehavioral dynamics following chronic sleep restriction: dose-response effects of one night for recovery. *Sleep* 2010;**33**:1013–26.

Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res* 2003;**12**:1–12.

Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep* 2010;**33**:585–92.

Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida C, editor. *Sleep deprivation*. New York: Marcel Dekker; 2005.

Goel N, Basner M, Dinges DF. Phenotyping of neurobehavioral vulnerability to circadian phase during sleep loss. *Methods Enzymol* 2015;**552**: 285–308.

Goel N, Dinges DF. Behavioral and genetic markers of sleepiness. *J Clin Sleep Med* 2011;**7**(Suppl.):S19–21.

Lamond N, Jay SM, Dorrian J, et al. The dynamics of neurobehavioural recovery following sleep loss. *J Sleep Res* 2007;**16**:33–41.

Ma N, Dinges DF, Basner M, Rao H. How acute total sleep loss affects the attending brain: a meta-analysis of neuroimaging studies. *Sleep* 2015; **38**(2):233–40.

Roach GD, Lamond N, Dawson D. Feedback has a positive effect on cognitive function during total sleep deprivation if there is sufficient time for it to be effectively processed. *Appl Ergon* 2016;**52**:285–90.

Van Dongen HP, Baynard MD, Maislin G, Dinges DF. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep* 2004;**27**:423–33.

Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;**26**:117–26.

A complete reference list can be found online at ExpertConsult.com.

Chapter
6

Genetics of Normal Human Sleep

Hyun Hor; Mehdi Tafti

Chapter Highlights

- Twin studies indicate that 20% to 50% of variance in sleep habits, duration, and organization can be explained by genetic factors, whereas the heritability of the electroencephalogram (EEG), both waking and sleep, is among the highest in humans.
- Single genes with large effects on sleep are unknown. Such genes affecting the sleep EEG are found in animal models but not yet in humans.
- Recent studies concentrated on the genetics of subjective (self-reported) sleep duration, but so far genetic variants with small effect sizes have been identified. These findings have not been replicated to date and must therefore be considered preliminary.
- Large-scale objectively (EEG-based) measured sleep studies are needed to discover the molecular bases of sleep and the sleep EEG.

The first edition of *Principles and Practice of Sleep Medicine* contained no information about how genes contribute to sleep and circadian biology. Sleep is such a universal component of life that it is not surprising that recent research has shown that genetics plays a pivotal role in its regulation, that humans share many elements of genetic control of sleep and circadian rhythms with other life forms (see Chapters 16 to 30), and that genetics plays an important role in some sleep disorders (see Chapter 31).

Genetics of sleep in humans can be divided in two main categories: the genetics of normal sleep and the genetics of abnormal sleep. Whereas the study of sleep disorders from the genetic point of view follows approaches similar to those in the study of many other rare and common diseases (see Chapter 31), the study of normal human sleep is a challenging task for different reasons. First of all, human sleep can be considered a complex trait with a wide phenotypic variability. Genetic factors have been suggested to play an important role as certain sleep characteristics can be repeatedly found within families, but numerous environmental factors, especially those that arise from today's lifestyle, are major confounders in clearly determining a specific sleep pattern or phenotype, which is an important prerequisite of identifying genetic factors of human sleep. Furthermore, the identification of genetic factors contributing to sleep phenotypes in humans is even more difficult because the target tissue, namely, the brain, is inaccessible for in-depth genetic and molecular studies, in contrast with animal model studies of sleep (rats, mice, and fruit flies), in which, for example, transcriptome analyses of the brain have identified major genetic contributors to the regulation of sleep.[1-3]

Taking into account these limitations, we summarize the recent advances in this field from both the epidemiologic and the genetic points of view.

GENETIC EPIDEMIOLOGY OF SLEEP WITH FOCUS ON TWIN STUDIES

Substantial evidence suggesting that sleep is a complex trait with a strong genetic component has arisen from twin studies. In the field of human genetics, twin studies are a major tool to characterize the genetic and environmental contribution to a certain trait or phenotype because monozygotic twins share the same genetic background, whereas dizygotic twins share only half of the segregating alleles. The characterization of specific phenotypes, taking into account the underlying genetic background in twin studies, thus allows us to quantify the additive genetic effects, the contribution of environmental factors, or the interaction of both genetic and environmental factors.

Sleep studies in twins cover both subjective reports on sleep habits and patterns, including sleep quality and quantity, and objective measurements of sleep phenotypes based on polysomnography analyses and electroencephalographic patterns of sleep.

Numerous twin studies have been published to date focusing on subjective sleep characteristics. Sleep duration constitutes the largest amount of data, and studies yield partially conflicting results. Some studies showed low or modest genetic contribution in children and school-aged twins[4,5] as well as overall heritability estimates of 0.44[6] and 0.3[7] in large-scale studies in adults, indicating a stronger genetic contribution to sleep duration with increasing age; other studies in monozygotic and dizygotic young adult twins living either together or apart showed low or no evidence of genetic influence but a strong environmental contribution to sleep duration.[6,8] These conflicting results have been explained by the potentially important role of voluntary control influenced by lifestyle factors, such as staying out late and sleeping in late.[9] In

contrast, the genetic influence on subjective sleep quality appeared stable across different age groups in twins, in whom heritability estimates ranged from 33% to 46%.[6,8,10]

Twin studies using subjective methods have also been useful to assess the heritability of the human circadian rhythm or, more specifically, the diurnal preference in humans (better known as morningness and eveningness types). These studies also indicate a consistent strong genetic effect on this trait with heritability estimates ranging from 44% to 50%, a finding that was also observed in monozygotic and dizygotic twins who had been raised apart.[8,11-13] Also, this effect seemed stable across all age groups, although the diurnal preference was shown to shift from evening-type in adolescence to morning-type at an advanced age. In this context, we should point out the identification of *PER3* as a gene associated with diurnal preference in an age-dependent manner (see later section on candidate gene analysis).

Objective measurements take advantage of electroencephalographic sleep recordings used in polysomnographic investigations of human sleep. The sleep electroencephalogram (EEG) offers the possibility of gaining detailed information about sleep patterns, including sleep duration, latency, onset, and efficiency, as well as more global measures, such as the sleep architecture reflected by the EEG power spectra and the sleep stage organization. A number of such twin studies revealed a substantial genetic influence on various sleep patterns, such as different sleep stages (sleep stage 2, slow wave sleep) and spindle density, as well as the time of rapid eye movement (REM) sleep and wake after sleep onset.[9,14-20] A strong genetic effect on non-REM (NREM) sleep EEG has been consistently described in a recent study,[21] and other studies even demonstrate that NREM sleep EEG can be reliably distinguished between subjects, irrespective of sleep pressure.[22-25] These "individual fingerprints" of NREM sleep EEG were therefore claimed as one of the most heritable traits in humans. In contrast, whether human REM sleep is under genetic control remains controversial. Whereas some studies did not find clear differences in REM sleep patterns between monozygotic and dizygotic twins,[16,17,19] others estimated a strong heritability of REM sleep amount and especially REM density.[14,16-19,26] Overall, sleep duration and structure have moderate to high heritability (up to 60%), whereas the sleep EEG is strongly heritable (up to 98%). The latter is supported by mouse studies[27-29] and was used to map and to identify major genes regulating the theta oscillations in REM and delta oscillations in NREM sleep.[30,31] These studies strongly suggest that electroencephalographic features of sleep, more than its amount or distribution, are the best candidates for genetic studies because these phenotypes are the most heritable sleep traits.

CANDIDATE GENE ANALYSES FOR NORMAL SLEEP AND CIRCADIAN PHENOTYPES

Twin studies have successfully demonstrated that many human sleep traits are under genetic control. However, what are the genes regulating these traits, and more important, what methods are currently available to address this question and to identify the underlying genetic factors? In contrast to animal studies, in which genes can be identified by sleep recordings and transcriptome analysis of brain tissue samples (e.g., by expression quantitative trait locus analysis of different mouse inbred strains after sleep deprivation), the identification of genes regulating sleep or circadian rhythms in humans was to date mainly restricted to candidate gene analysis. However, the choice of candidate genes was in turn restricted by previous findings that were mainly based on the knowledge of the physiology and genetics of sleep and circadian rhythm regulation in other species.

Candidate Gene Analyses of Circadian Clock Genes in Humans

Genes regulating circadian rhythms have been subject to genetic analysis to identify regulators of human sleep and diurnal preference. As discussed in detail in Chapters 26 to 30, the circadian clock consists of a group of genes that undergo an autoregulatory negative feedback loop. Here we discuss the genes of this loop, which plays an important role in the regulation of human sleep with distinct consequences on the timing and duration of sleep. Whereas we aim to focus on genes regulating normal human sleep, we need to stress that some genes have been identified in traits that show marked deviations from the average sleep patterns and therefore have been in some cases defined as "disorders." Nevertheless, we consider them as natural variations of normal human sleep presenting as an extreme phenotype of a specific trait.

CLOCK

The *CLOCK* gene was one of the first genes shown to have an effect on diurnal preference. Katzenberg et al[32] identified a single nucleotide polymorphism (SNP) in the 3′ untranslated region associated with a delayed sleep phase. However, this initial finding was only partially replicated in follow-up studies.[32-36] *CLOCK* was also associated with sleep duration in an extended candidate gene association study based on linkage disequilibrium.[37] For this purpose, 194 SNP markers in 19 "clock" genes were studied in a discovery and subsequent replication study comprising a total number of 1294 subjects from two different populations (South Tyrol and Estonia) with either short (<7 hours) or long (>8.5 hours) habitual sleep habits.

Period2 (PER2)

A syndrome characterized by an advanced sleep phase was first reported in 1999 by Jones et al,[38] who described a large family from Utah presenting with a syndrome with an approximately 4- to 6-hour advanced sleep phase without major impact on the quality or quantity of sleep. Subsequent linkage and positional cloning analyses by Toh et al[39] finally identified a mutation in the casein kinase I–binding domain of *PER2* as the cause of this syndrome.

Casein Kinase I (CSNK1D)

Since the initial identification of a mutation in *PER2*, a candidate gene association study identified novel mutations in the *CSNK1D* gene as a cause of familial advanced sleep phase syndrome.[40] This finding was further supported by functional analyses in transgenic mice and fruit flies that displayed an altered diurnal preference indicating the identification of a central regulator of the circadian clock.

DEC2

In a small family with two affected members, He et al[41] identified a mutation in the *DEC2* gene as a cause of the short

sleeper phenotype. This gene is a transcriptional repressor within the mammalian circadian clock and was therefore one of the subjects of investigation in a candidate gene approach, followed by the engineering of transgenic mice carrying the identified mutation, which displayed shorter sleep, corroborating the association in humans. Sequencing of the *DEC2* gene identified an additional heterozygous variant in one twin of a dizygotic pair among a cohort of sleep-deprived individuals (either acute or partially chronic).[42] This finding was confirmed by polysomnography recordings revealing reduced sleep duration and less recovery sleep following sleep deprivation compared with the other twin carrying the homozygous major allele. However, the amount of NREM sleep remained the same between twins.

Period3 (PER3)

Various studies have shown that SNPs and a variable number tandem repeat (VNTR) polymorphism in the promoter and coding region of *PER3* are associated with diurnal preference or delayed sleep phase syndrome (DSPS).[43-47] Ebisawa et al[44] were the first to report on a possible association of structural polymorphisms of *PER3* with DSPS, and particularly the VNTR polymorphism, represented by either a short (4 repeats) or a long (5 repeats) allele, was the subject of further investigation. Whereas homozygosity for the longer variant (5/5) was associated with the morningness type, the shorter one was associated with eveningness type. In addition, individuals homozygous for the long allele (5/5) presented a higher amount of slow wave sleep in NREM and an increased theta and alpha activity in wakefulness and REM sleep.[43] Interestingly, 75% of individuals diagnosed with DSPS were carriers of the homozygous short variant (4/4).[45] However, one independent study so far did not successfully replicate the VNTR association with diurnal preference,[48] suggesting that further investigations with larger sample sizes are needed to clarify this putative association.

Candidate Gene Analyses Related to Adenosinergic Neurotransmission in Humans

Adenosine plays a major role in the regulation of sleep and wakefulness.[49] Early evidence in mice and subsequent genetic and pharmacogenetic studies in humans confirmed its important role in sleep regulation. Adenosine deaminase is an enzyme that degrades adenosine to inosine. A known SNP (c.22G>A) in the gene encoding adenosine deaminase has a major functional effect; carriers of c.22G>A display a lower enzymatic activity than homozygous carriers of the major allele, which thus leads to a lower adenosine degradation rate. As a consequence, this SNP has been further studied for its role in sleep regulation. This study revealed an increased amount of slow wave sleep as well as higher delta power in NREM sleep, a marker of sleep need, in individuals carrying c.22G>A.[50] This finding was further supported by a sleep deprivation study (40 hours of prolonged wakefulness) that showed increased slow wave sleep and delta activity in NREM sleep during the recovery night in carriers of c.22G>A, suggesting that this genetic variation leads to an enhanced sleep pressure through reduced degradation of adenosine.[51]

Additional evidence for the importance of the adenosinergic neurotransmission in sleep regulation came from the genetic study of adenosine receptors. Four receptor subtypes mediate the cellular effects of adenosine, with the two subtypes A_1 and A_{2A} mainly involved in mediating the effect of adenosine in sleep.[49] The A_{2A} receptor was principally studied from the genetic point of view because of its involvement in the response to caffeine, which mainly acts as a competitive antagonist at adenosine receptors.[52] Regardless of the effect of caffeine, an SNP (c.1083T>C) in *ADORA2A*, the gene that encodes the A_{2A} receptor, was shown to affect the duration of slow wave sleep as well as sleep intensity.[53] In fact, homozygous carriers of the G/G genotype showed fewer awakenings, longer time spent in slow wave sleep, and higher delta power during sleep. More important, in a follow-up study on the effect of and sensitivity to caffeine in different genotype and haplotype carriers of *ADORA2A*,[54] the minor C allele of the c.1083T>C SNP conferred higher sensitivity to caffeine-induced sleep disturbances. Interestingly, a specific haplotype comprising eight SNPs in this gene was associated with a missing effect of caffeine on NREM sleep in recovery sleep, whereas carriers of other haplotypes presented reduced rebound in slow wave sleep on caffeine intake. The missing effect correlated with a failure of caffeine to rescue the vigilance decline after sleep loss; thus, reduced slow wave sleep reflected the successful counteraction to vigilance decline by caffeine. A recent genome-wide association study confirmed the association of *ADORA2A* with caffeine-related sleep disturbances.[55]

The results of these studies strengthened the hypothesis that adenosine and its receptors play an important role in sleep regulation, particularly in NREM sleep homeostasis.

Candidate Gene Analyses Related to Glutaminergic and Dopaminergic Neurotransmission

GRIA3

In a population-based study in humans, an SNP in the inotropic glutamate receptor gene (*GRIA3*), located on the X chromosome, was previously identified to be associated with depressive disorder. As sleep disturbances are potential precipitating factors for the initiation of depressive disorders,[56] this gene was tested, among others, as a candidate gene for sleep duration.[57] In this study, a significant association was found between rs687577 and sleep duration in healthy women, a finding that still needs to be independently replicated. *GRIA3* encodes the glutamate receptor 3 (GluR3) subunit, which is one of the four AMPA receptor subunits and is expressed, among others, in the thalamus.[58] Electroencephalographic recordings in *GluR3* knockout mice have revealed marked changes in the EEG, particularly during NREM sleep, suggesting an important role of the GluR3 subunit in the generation of cortical slow oscillations,[59] as well as consistent changes of the expression level of GluR3, which increases in the cortex under sleep deprivation and decreases during recovery sleep.[60]

COMT

The dopaminergic system is considered to play an important role in sleep regulation as many stimulants and wake-promoting drugs are known to act through dopaminergic neurotransmission.[61] Therefore, a functional SNP leading to Val158Met in catechol-*O*-methyltransferase (COMT), an enzyme metabolizing cerebral dopamine, has been studied in connection with subjective and objective measures of sleep homeostasis. Two studies[62,63] revealed no difference in subjective sleepiness after sleep deprivation between Val and Met

homozygous genotypes. However, the objective measures collected by Goel et al showed larger declines of slow wave sleep after sleep deprivation in homozygous Met carriers compared with homozygous Val carriers, whereas the study of Bodenmann et al revealed no major change in slow wave sleep activity between both genotypes but an increase in power of certain NREM frequency bands during recovery sleep after modafinil administration in Val/Val genotype carriers.[62,64] Moreover, both studies showed no difference between cognitive and executive functioning at baseline condition between genotypes, but Bodenmann et al found that modafinil is able to maintain these baseline performances after sleep deprivation in Val/Val but not in Met/Met carriers. These results highlight the important role of COMT genetic variation, enzyme activity, and dopamine levels, particularly in the prefrontal cortex, in the regulation of sleep and wakefulness in normal subjects.

GENOME-WIDE ASSOCIATION STUDIES OF NORMAL SLEEP PHENOTYPES

Candidate gene analyses led to the discovery of genes associated with certain sleep traits and phenotypes. However, these approaches did not yield replicable results in a considerable number of cases. During the past years, genome-wide association studies (GWAS) have become a well-established tool to identify genetic variants and genes associated with different disorders and traits, and they allow us to study the genetics of a given trait in a hypothesis-free manner. This methodology has proved successful in identifying true genetic variants that have mostly been successfully replicated in follow-up studies.

The first large-scale GWAS on sleep phenotypes was published in 2007 by Gottlieb et al, who took advantage of the Framingham Heart Study, which was initially founded to investigate the epidemiology of cardiovascular diseases,[65] but numerous other phenotypes were collected by questionnaires.[66] In a subset of 749 subjects, both phenotype data on sleepiness, usual bedtime, and usual sleep duration and genotype data for 100,000 SNPs were available. Family-based association tests revealed a linkage peak on chromosome 16 including the *CSNK2A2* gene to usual bedtime, whereas sleep duration was linked to the region encompassing the *PROK2* gene on chromosome 3. *CSNK2A2* and *PROK2* reached LOD scores above 2 and are known components of the circadian clock. Other SNPs in various genes with lower LOD scores (<2) were identified, among which one should be noted that is linked to usual bedtime, the *CLOCK* gene on chromosome 4. Population-based association tests, on the other hand, identified an intronic SNP in the phosphodiesterase 4D gene (*PDE4D*) to be associated with sleepiness (as measured by the Epworth Sleepiness Scale score), reaching genome-wide significance. Of note, the results of the population- and family-based association tests were poorly correlated. The only SNP in a coding region with the highest significance in both association tests was located in *NPSR1*, a component of the neuropeptide S receptor, and associated with usual bedtime. One of the major limitations of this study is the fact that many genes previously linked to various sleep phenotypes (see earlier), especially those forming part of the circadian clock, such as *PER2*, *CSNK1D*, and *PER3*, were either not or only poorly represented on the genotyping platform used. Although promising, the results presented in this study have not been replicated in an independent study. In this context, an attempt to replicate the *NPSR1* association by objective measures using actigraphy failed to replicate. However, this study including 393 subjects found an association of the studied variant with duration of sleep and rest, indicating that objective measurements such as actigraphy could most probably lead to more reliable and precise association results.[67]

Another GWAS with focus on sleep duration collected data from seven different European populations[68] and comprised a discovery phase with GWAS data using up to 2.5 Mio SNPs in 4251 subjects of European ancestry and a replication phase involving 5949 subjects. The study revealed a genome-wide significant association of sleep duration with an intronic variant in the *ABCC9* gene (ATP-binding cassette, subfamily C, member 9). However, this association became significant only after the analysis in a subgroup of the replication sample that was chosen for the season of entry and chronotype, as the season is known to influence the circadian entrainment and sleep behavior. The associated variant explained 5% (10 minutes) of the variation in sleep duration. Nevertheless, functional analysis in *Drosophila* of the ATP-sensitive potassium channel subunit (SUR2), which is the ortholog of *ABCC9* in *Drosophila*, confirmed its important influence on sleep duration. In fact, knock-down experiments of *dSUR* in neurons resulted in a marked reduction of nighttime sleep with a delay of sleep onset of 3 hours, whereas there was little effect on daytime sleep. Given that earlier studies in *Drosophila* indicated an involvement of potassium channel activity in the regulation of sleep duration,[69,70] this finding strengthens the hypothesis that potassium channel regulatory proteins in the brain might play an important role in the modulation of sleep across species. The *ABCC9* variant was also associated with sleep duration in the British G1912 cohort, although only in a recessive model given a rare homozygous genotype of the minor allele.[71]

A GWAS using samples and data from the Australian Twin Registry collected six sleep measures phenotyped by questionnaires, including sleep onset time, latency, quality, depth, and duration, in a total of 2323 subjects from whom both genotype and phenotype data were available.[72] The discovery phase using 2.3 Mio imputed SNPs did not reveal any genome-wide significant associations, but the authors pointed out SNPs in the *CACNA1C* gene on chromosome 12, a gene that had been previously linked to bipolar disorder, schizophrenia,[73,74] and narcolepsy,[75] that showed some evidence of association with sleep latency and quality. However, the replication study did not confirm this association. More importantly, none of the variants identified by Gottlieb et al[66] replicated in this study. Nonetheless, a replication of variants associated with sleepiness was not possible because no data on sleepiness (as measured, for example, by the Epworth Sleepiness Scale) were included in this study. However, Parsons et al[71] finally reported on a successful replication of a variant in the *CACNA1C* gene in the British G1219 cohort with sleep quality, using an additive model of inheritance.

A GWAS published by Ollila et al[76] used a Finnish discovery sample of 1941 subjects and replication sample of 6834 searching for genes influencing sleep duration. Similar to previous GWAS, no genome-wide significant association was identified and therefore 31 variants with suggestive association were included in the replication study, which were derived

from various study settings and designs. Nevertheless, variants in three genes nominally reached significance at $P < 5 \times 10^{-5}$ and were further analyzed by an expression quantitative trait locus analysis of sleep duration in 207 samples. *KLF6* was the only gene with an expression level correlating with sleep duration, a finding that was then subjected to replication in a small sleep restriction study including nine cases and four controls. In cases carrying the *KLF6* allele associated with shorter sleep duration, the gene expression after sleep restriction (4 hours during 5 nights) as well as after 2 nights of recovery sleep was significantly higher than in controls. Of note, previously identified associations by GWAS, notably *ABCC9* with sleep duration[68] and *PDE4D* with sleepiness,[66] were not replicated in this study.

The latest GWAS published by Gottlieb et al[77] included a total number of 47,180 individuals of European ancestry from 18 different population-based cohorts representing the largest GWAS studying subjective sleep duration to date. Of note, some of these cohorts have already been used in the aforementioned studies. Variants in two genes reached genome-wide significance; however, only one of them replicated in an independent cohort of African American origin comprising a sample of 4747 individuals. The associated variants in this gene were located in an intergenic region upstream of *PAX8*, which encodes a transcription factor regulating thyroid development and function. Although replicated, the strongest associated SNP contributed to a marginal effect characterized by a prolongation of 3.1 minutes per copy of its minor allele.

CONCLUSIONS AND FUTURE OUTLOOK

Epidemiologic twin studies clearly demonstrated that human sleep, including various sleep patterns and measures, is genetically determined. Candidate gene analyses and, more recently, GWAS have been the major tools to identify underlying genes and genetic factors controlling human sleep. Our summary of the recent findings revealed obvious weaknesses in both applied methods that substantially lower the chances of identifying true genetic causes and associations with sleep phenotypes. One of the main problems in genetic studies of normal sleep is the accuracy and common consensus of the studied sleep phenotypes. Many sleep patterns, such as sleep duration, latency, and quality, are quantitative traits and have been to date mainly studied by subjective measurements using questionnaires. Although many association studies have used standardized questionnaires, such as the Epworth Sleepiness Scale or the Horne and Östberg questionnaire, the value of these measures can often be contested as they do not take into account many confounding and environmental factors, such as shift work, voluntary sleep restriction, underlying medical conditions, stress, diet, and other lifestyle factors. Furthermore, several studies, or even discovery and replication phases within the same association study, are hardly comparable as different questionnaires or different scales were used to measure the same phenotype. Hence, association studies are often underpowered, although up to a few thousand subjects have been included. As a consequence, even much larger sample sizes would be needed to overcome these inconsistencies and to increase the power to detect truly associated genetic variants. In this context, the more precise a phenotype is, the lower the population size needed in association studies to identify responsible genes for a given phenotype. So far, the phenotypes used, such as self-reported sleep duration or timing, are extremely noisy, leading to inconsistent associations. Therefore, standardized and validated questionnaires are needed to define specific sleep phenotypes, similar to diagnostic criteria for medical diseases.

If questionnaires, or, in other words, subjective assessments of sleep phenotypes, are not able to correct for confounding factors, the solution is to focus on objective measures, such as actigraphy or, even better, polysomnographic phenotypes with strict inclusion criteria of study subjects. This is all the more an important issue considering the known and frequent discrepancy between self-assessed and objective sleep characteristics.[78] Efforts are ongoing to study large cohorts including genetics and the characterization of sleep phenotypes by polysomnographic studies.

Replication is crucial in genetic studies, but many of the identified associations have not been independently replicated. Thus, especially associations of sleep patterns to genes that have been identified on the basis of questionnaires should be considered preliminary.

In analogy to studies of rare and common diseases, the genetics of normal human sleep will most probably also move toward large-scale sequencing efforts by exome or whole genome sequencing. These technologies should allow us to identify rare, common, and structural genetic variants implicated in extreme forms of specific phenotypes, such as short and long sleepers with regard to sleep duration. Nevertheless, apart from a direct search for genetic factors regulating normal human sleep, this research will remain strongly dependent on basic research on sleep in other species as well as on the study of human sleep disorders that allows us to link phenotypes of sleep disorders to mechanisms regulating physiologic sleep and its biologic substrates and underlying genes.

CLINICAL PEARL

Various sleep phenotypes have a strong genetic component, but the underlying genetic contributions to these phenotypes still need to be fully discovered. Therefore, in clinical terms, genetic factors have a major impact on determining and predicting a wide range of sleep patterns in humans. This includes pharmacogenetic interactions, as illustrated by the involvement of genetic variants in the adenosinergic and dopaminergic systems in sleep and wakefulness responses to caffeine or modafinil, respectively.

SUMMARY

Sleep and the EEG are complex phenotypes, poorly understood at the molecular level. Nevertheless, evidence indicates that several sleep phenotypes have heritability estimates in the range of 30% to 50% and that the EEG (both waking and sleep EEG) is highly heritable (up to 98%). The heritability of sleep duration and sleep timing has been documented for decades, but genes controlling these phenotypes remain largely unknown. As opposed to animal models such as mice and *Drosophila*, in which molecular genetics has proved extremely successful in discovering genes involved in sleep and the sleep EEG, the genetics of human sleep has made little progress during the past 20 to 30 years. Based on sleep physiology and pharmacology, the candidate gene approach found substantial

effect of genetic variations in several neurotransmitter pathways (e.g., adenosinergic and dopaminergic). Although a single gene variant (*DEC2*) was found to be causally involved in sleep duration (short sleep), recent genome-wide association studies found genetic variants with small effects (explaining just a small percentage of variance in sleep length). Because the sleep EEG is highly heritable and mouse genetics has identified several genes with major effects, it is expected that large-scale electroencephalography-based sleep studies in humans might also reveal major genes regulating electroencephalographic features characteristic of sleep (delta and theta oscillations).

Selected Readings

Andreani TS, Itoh TQ, Yildirim E, et al. Genetics of circadian rhythms. *Sleep Med Clin* 2015;**10**(4):413–21.

Andretic R, Franken P, Tafti M. Genetics of sleep. *Annu Rev Genet* 2008;**42**:361–88.

Barclay NL, Gregory AM. Quantitative genetic research on sleep: a review of normal sleep, sleep disturbances and associated emotional, behavioural, and health-related difficulties. *Sleep Med Rev* 2013;**17**:29–40.

Bedrosian TA, Fonken LK, Nelson RJ. Endocrine effects of circadian disruption. *Annu Rev Physiol* 2016;**78**:March 17.

Dauvilliers Y, Tafti M, Landolt HP. Catechol-O-methyltransferase, dopamine, and sleep-wake regulation. *Sleep Med Rev* 2015;**2**:47–53.

Sehgal A, Mignot E. Genetics of sleep and sleep disorders. *Cell* 2011;**146**:194–207.

A complete reference list can be found online at ExpertConsult.com.

Section 2

Sleep Mechanisms and Phylogeny

Chapter 7

Neural Control of Sleep in Mammals

Dennis McGinty; Ronald Szymusiak

Chapter Highlights

- Mammalian sleep and wake states are regulated by multiple neuronal systems located in the brainstem, diencephalon, and telencephalon. Although the isolated brainstem is capable of generating non–rapid eye movement (NREM)-like and REM-like states, lesions of several brainstem and forebrain regions in the otherwise intact brain can be associated with altered amounts of NREM and/or REM sleep.
- Arousal and waking are facilitated by chemically distinct neuronal groups localized in the midbrain, the posterior and lateral hypothalamus, and the basal forebrain. Functionally important arousal systems include histaminergic, orexinergic, serotoninergic, cholinergic, dopaminergic, and noradrenergic neurons.
- Arousal systems have widespread projections that regulate global aspects of arousal including changes in electroencephalographic, motor, sensory, autonomic, and integrative functions. Arousal systems control the excitability of thalamic and cortical neurons. Reduced activity in arousal systems promotes synchronous discharge of intracortical and thalamocortical circuits that underlie NREM sleep patterns on the electroencephalogram.
- The preoptic anterior hypothalamus—the preoptic area (POA)—contains gamma-aminobutyric acid (GABA)-ergic/galaninergic neurons that exhibit increased activity during NREM and REM sleep and respond to physiologic signals that increase sleep, such as warming, sustained wakefulness, and endogenous sleep factors. POA sleep-active GABAergic neurons project to histaminergic, orexinergic, serotoninergic, and noradrenergic arousal systems and, through coordinated inhibition of these arousal systems, promote transitions from waking to sleep.
- Sleep-promoting molecules are related to the multiple functions of sleep: adenosine to resupply brain energy reserves, cytokines to facilitate immune functions, growth hormone–releasing hormone to promote anabolic processes, unfolding protein response signals to prevent protein misfolding, and oxidative stress signaling molecules to prevent oxidative stress–induced cell damage.

DIVERSE BRAIN REGIONS MODULATE WAKING AND NON–RAPID EYE MOVEMENT SLEEP

Isolated Forebrain

The capacity of various brain regions to generate sleep and awake states was first studied by isolating or removing major regions. The physiology of the chronically maintained isolated forebrain, or chronic *cerveau isolé,* preparation has been examined in dogs and cats.[1] Immediately after complete midbrain transections, the isolated forebrain exhibits continuous slow waves and spindles on the electroencephalogram (EEG). Thus structures below the midbrain normally must facilitate awake-like EEG states. By contrast, if a brainstem transection was made at the midpontine level, an activated or wake-like forebrain EEG state predominated immediately after the transection, but with some residual episodes of EEG slow wave activity. In this preparation, the forebrain exhibited evidence of classical conditioning and other signs of an integrated waking state. These studies argue that neuronal groups localized between the midpons and upper midbrain are important for generating a waking (wakefulness)-like state. After 5 to 9 days of recovery, the chronic cerveau isolé rat preparation exhibits a circadian pattern of EEG activation and synchronization.[2] In this preparation, the creation of preoptic area (POA) lesions is followed by a continuously activated EEG pattern. Thus the isolated forebrain can generate a sustained wake-like state, and the POA must play a critical role in initiating the sleep-like EEG state of the isolated forebrain (see later). Wake-like and sleep-like EEG states appear to depend on a balance between wake-promoting and sleep-promoting systems. The terms *wake-like* and *sleep-like* are used here because these preparations cannot exhibit the full behavioral spectrum of sleep and wake states.

Diencephalon

Chronic diencephalic cats, in which the neocortex and striatum have been removed, exhibit behavioral waking with persistent locomotion and orientation to auditory stimuli, a quiet sleep-like or non–rapid eye movement (NREM)-like state with typical cat sleeping postures, and a REM-like state including muscle atonia, rapid eye movements, muscle twitches, and pontine EEG spikes.[3] EEG patterns recorded in the thalamus showed increased amplitude in conjunction with the NREM sleep–like state, although true spindles and slow waves were absent. The thalamic EEG exhibits desynchronization during the REM-like state.

In summary, the neocortex and striatum are not required for any behaviorally defined sleep-wake states, and an NREM-like state occurs in the absence of sleep spindles and slow waves.

Thalamus

Cats subjected to complete thalamectomy continue to exhibit episodes of EEG and behavioral sleep and waking, although spindles are absent on the EEG,[4] and the animals exhibit reductions in both NREM and REM sleep. Fatal familial insomnia[5] is a rare neurodegenerative disease in humans characterized by progressive autonomic hyperactivation, motor disturbances, loss of sleep spindles, and severe NREM sleep insomnia. Neuropathologic examination findings include severe cell loss and gliosis in the anterior medial thalamus, including the dorsomedial nucleus. However, patients with paramedian thalamic stroke, with magnetic resonance imaging-verified damage to the dorsomedial and centromedial nuclei, present with either severe hypersomnolence or increased daytime sleepiness, not insomnia.[6] In summary, the thalamus plays a critical role in regulating cortical EEG patterns during waking and sleep, and specific regions within this structure appear to have either wake-promoting or hypnogenic functions.

Lower Brainstem

After recovery from the acute effects of the complete midbrain transections in cats (as just described), the lower brainstem can generate rudimentary behavioral waking, an NREM-like state, and a REM-like state.[3] Waking is characterized by crouching, sitting, attempts to walk, dilated pupils, and head orientation to noises. In the first sleep-like state, these cats lie in a random position, pupils exhibit reduced but variable miosis, and eyes exhibit slow and nonconjugate movements, and the animals can be aroused by auditory or other stimuli. If this stage is not disturbed, these cats enter another stage, characterized by complete pupillary miosis, loss of neck muscle tone, and rapid eye movements, identifying a REM-like state. Additional studies support the hypothesis that the lower brainstem contains sleep-facilitating processes. Low-frequency electrical stimulation of the dorsal medullary reticular formation in the nucleus of the solitary tract produced neocortical EEG synchronization.[7] Lesioning or cooling of this site was followed by EEG activation.[8] Recently, a population of sleep-active neurons has been identified in the rostral medulla of rats and mice, located lateral and dorsal to the facial nerve.[9] Many of these sleep-active medullary neurons express inhibitory neurotransmitters such as gamma-aminobutyric acid (GABA) or glycine (see later). Excitotoxic lesions of the parafacial zone are associated with increased waking and decreased NREM sleep.[9]

In summary, widespread structures in the mammalian nervous system, from the neocortex to the lower brainstem, have the capacity to facilitate both sleep-like and waking-like states and to modulate the amounts of sleep.

RETICULAR ACTIVATING SYSTEM AND DELINEATION OF AROUSAL SYSTEMS

The transection studies just described support the concept of a pontomesencephalic wake-promoting or arousal system. No discovery was historically more significant than the description of the reticular activating system (RAS) by Moruzzi and Magoun.[10] Introduction of large lesions of the core of the rostral pontine and mesencephalic tegmentum was followed by persistent somnolence and EEG synchronization, and electrical stimulation of this region induced arousal from sleep. Interruption of sensory pathways did not affect EEG activation. It was hypothesized that cells in the RAS generated forebrain activation and wakefulness.

The concept of the RAS has been superseded by the finding that arousal is facilitated not by a single system but instead by several discrete neuronal groups localized within and adjacent to the pontine and midbrain reticular formation and its extension into the hypothalamus (Figure 7-1). These neuronal systems are differentiated by the expression of enzymes that synthesize specific neurotransmitters and neuromodulators. These include neurons that synthesize serotonin,

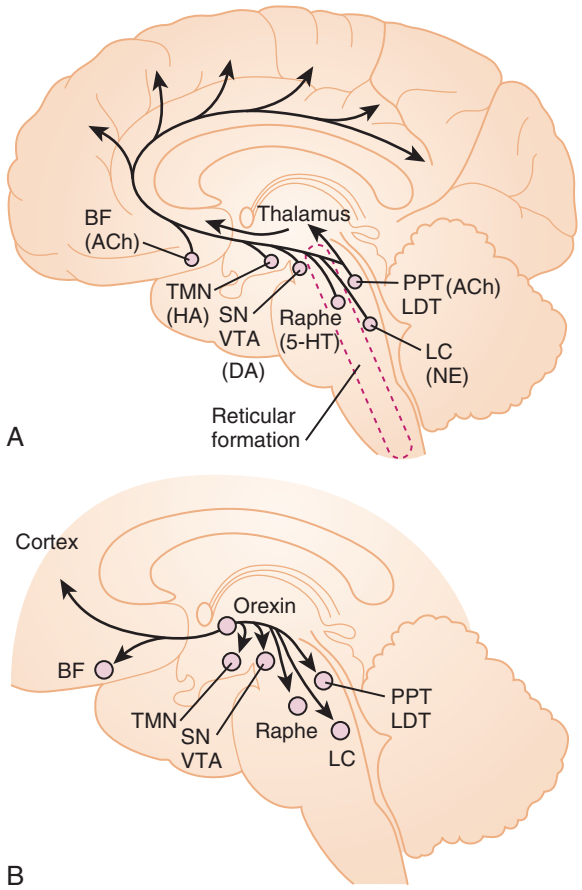


Figure 7-1 A, Sagittal view of a brain providing an overview of the wake-control networks described in the text. The upper brainstem, posterior and lateral hypothalamus, and basal forebrain (BF) contain groups of neurons with identified phenotypes with arousal-inducing properties. These clusters include neurons expressing serotonin (5-HT), norepinephrine (NE), acetylcholine (ACh) in both pontomesencephalic and basal forebrain clusters, dopamine (DA), and histamine (HA). **B,** Sagittal view of brainstem and diencephalon showing localization of orexin-containing neurons and their projections to both forebrain and brainstem. All of these groups facilitate EEG arousal (waking and REM) and/or motor-behavioral arousal (waking). The arousal systems facilitate forebrain EEG activation both through the thalamus and the basal forebrain and through direct projections to neocortex. Arousal systems also facilitate motor-behavioral arousal through descending pathways. EEG, Electroencephalogram; LC, locus coeruleus; LDT, lateral dorsal tegmental; PPT, pedunculopontine; SN, substantia nigra; TMN, tuberomammallary nucleus; VTA, ventral tegmental area.

norepinephrine (noradrenaline), histamine, acetylcholine, and orexin/hypocretin (herein called orexin). Each of these systems has been studied extensively in the context of the control of specific aspects of waking behaviors. Presented next is a brief overview of each of these *arousal systems,* focusing on their contribution to generalized brain arousal or activation. As background, some general properties of these neuronal systems are summarized in the following list:

1. Arousal is a global process, characterized by concurrent changes in several physiologic systems, including autonomic, motor, endocrine, and sensory systems, as well as in EEG tracings. All of the arousal systems share one critical property: Their neurons give rise to long, projecting axons with extensive terminal fields that impinge on multiple regions of the brainstem and forebrain. In this chapter, the emphasis is on the ascending projections from the brainstem and hypothalamus to the diencephalon, limbic system, and neocortex, because these are particularly germane to the generation of cortical arousal. Some arousal systems give rise to descending projections as well, which also are likely to play a role in regulating certain properties of sleep–wake states, such as changes in muscle tone and autonomic function.
2. Most arousal systems have been studied by recording the discharge patterns of neurons in "freely moving" animals, relative to those in spontaneously occurring wake and sleep states. Increased discharge during arousal or wake states compared with sleep constitutes part of the evidence for an arousal system.
3. The actions of a neurotransmitter on a target neuronal system are determined primarily by the properties of the receptors in the target. The neurotransmitters and neuromodulators underlying arousal systems each act on several distinct receptor types, with diverse actions. In addition, postsynaptic effects are regulated by transmitter-specific "reuptake" molecules, which transport the neurotransmitter out of the synaptic space, terminating its action. Pharmacologic actions are usually mediated by actions on specific receptor types or transporters (see examples further on).
4. Chronic lesions of individual arousal systems or genetic knockout of critical molecules have only small or sometimes no effect on sleep-wake patterns (with the exception of serotonin and orexin knockouts; see later), even though acute manipulations of these same systems have strong effects on sleep-wake. The absence of chronic lesion or knockout effects is likely to be due to the redundancy of the arousal systems, such that, over time, deficiency in one system is compensated for by other systems or by changes in receptor sensitivity.
5. Electrophysiologic studies show that the arousal systems normally are activated and deactivated within seconds of a change in behavioral state. Thus effects of acute experimental manipulations of neurotransmitters that mediate arousal, as generated by acute administration of drugs, may best mimic the normal physiologic pattern and be more informative than chronic lesions regarding their function.
6. REM sleep is a state associated with behavioral and muscle quiescence and intense cortical EEG arousal. In parallel with these two aspects of REM sleep, arousal systems can be classified into two types: ones that are "off" in the REM state, befitting its sleep-like property, and others that are "on" in the REM state, befitting its wake-like properties. Some arousal-promoting systems (summarized later) also play a role in REM sleep control. Detailed analyses of the control of the role of these systems in REM sleep are presented in Chapter 8.

WAKE-ON, RAPID-EYE-MOVEMENT–OFF AROUSAL SYSTEMS

Serotonin

Neurons containing serotonin, or 5-hydroxytryptamine (5-HT), are found in the dorsal raphe and median raphe nuclei of the midbrain. These neurons project to virtually all regions of the diencephalon, limbic system, and neocortex. It initially was hypothesized that serotonin might be a sleep-promoting substance,[11] but evidence shows that the immediate effect of serotonin release is arousal (as reviewed by Ursin[12]). Although some heterogeneity is typical, the

discharge rates of most dorsal and median raphe neurons are highest during waking and lower during NREM, with minimal discharge in REM. Release of serotonin in the forebrain is highest in waking. Because of the diversity of serotonin receptors (there are at least 14 types), the effects of serotonin on target neurons are complex. Some receptor types are inhibitory, some are excitatory. At least one class of receptors, 5-HT_{2A}, appears to facilitate NREM sleep, because 5-HT_{2A}–knockout mice have reduced NREM.[13] Another type, 5-HT_{1A}, is inhibitory to REM sleep: 5-HT_{1A}–knockout mice have increased REM.[14] Selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors, which augment the actions of serotonin, are used to treat a variety of medical and psychiatric problems, and some drugs in this class have arousing or alerting properties.

Norepinephrine

Norepinephrine-containing neuronal groups in mammals are found throughout the brainstem, but the primary nucleus giving rise to ascending projections is the locus coeruleus. Norepinephrine neurons in this nucleus project throughout the diencephalon, forebrain, and cerebellum. Most locus coeruleus neurons exhibit regular discharge during waking, reduced discharge during NREM sleep, and near-complete cessation of discharge in REM sleep, a pattern congruent with a role in behavioral arousal.[15] Acute inactivation of the locus coeruleus or introduction of a lesion in the ascending pathway from this nucleus increases slow wave EEG activity during sleep.[16] Distinct roles for alpha-1, alpha-2, and beta norepinephrine receptor types are established. Direct application of alpha-1 and beta agonists in POA and adjacent basal forebrain sites induces increased wakefulness (reviewed by Berridge[17]). The arousal-producing effects of psychostimulant drugs such as amphetamines depend partly on induction of increased norepinephrine release and inhibition of norepinephrine reuptake, as well as on enhanced dopamine action (see later).

Histamine

Histamine-containing neurons in mammals are discretely localized within the tuberomamillary nucleus (TMN) and adjacent posterior hypothalamus (PH). Histamine neurons project throughout the hypothalamus and forebrain, including to the neocortex, as well as to the brainstem and spinal cord (reviewed by Haas and colleagues[18]). Administration of histamine type 1 (H_1) receptor antagonists (antihistamines) that penetrate the blood-brain barrier can result in sedative effects.[18] Transient inactivation of the TMN region results in increased NREM sleep.[19] Histamine neurons exhibit regular discharge during waking, greatly reduced discharge during NREM sleep, and cessation of discharge in REM sleep.[20] These neurons express H_3-type histamine receptors that are inhibitory, so-called autoreceptors. Thus the neurons are inhibited by histamine. Administration of an antagonist of this receptor causes disinhibition of histamine neurons and increased waking.[21]

Orexin

The loss of orexin neurons is known to underlie the human disease narcolepsy, the major symptoms of which are cataplexy and excessive sleepiness.[22,23] Orexin-containing neurons are localized within the midlateral hypothalamus, and like other arousal systems, they give rise to projections to all brain regions including the brainstem.[24] Among the targets of orexin terminals are other arousal-promoting neurons including histamine, 5-HT, and norepinephrine neurons. Orexin-containing neurons are active in waking, and they are "off" in both NREM and REM sleep.[25,26] Local administration of orexin in several brain sites induces arousal.[27] (See also Chapter 8.)

WAKE-ON, RAPID-EYE-MOVEMENT–ON AROUSAL SYSTEMS

Acetylcholine

Acetylcholine (ACh)-containing neurons are localized within two regions: the dorsolateral pontomesencephalic reticular formation, including the pedunculopontine tegmental (PPT) and laterodorsal tegmental (LDT) nuclei, and the basal forebrain.[28] The pontomesencephalic ACh neuronal group projects to the thalamus, hypothalamus, and basal forebrain; the basal forebrain group projects to the limbic system and neocortex. Neurons in both groups exhibit higher rates of discharge in both waking and REM sleep than in NREM sleep,[29] and release of ACh also is increased in these states.[30] Pharmacologic blockade of ACh receptors induces EEG synchrony and reduces vigilance, and inhibition of the ACh-degrading enzyme cholinesterase enhances arousal.[31]

Dopamine

Dopamine-containing neurons are localized primarily within the substantia nigra and the adjacent ventral tegmental area of the midbrain and the basal and medial hypothalamus.[32] Putative dopaminergic neurons in the ventral tegmental area exhibit highest activity in waking and REM sleep. An additional population of dopamine-containing wake-active neurons is localized in the rat ventral periaqueductal gray matter, and destruction of ventral periaqueductal gray matter dopaminergic neurons with 6-hydroxy-dopamine increases daily sleep time by 20%.[33] Release of dopamine in the frontal cortex is higher during wakefulness than during sleep.[34] Dopamine is inactivated primarily through reuptake by the dopamine transporter. Stimulant drugs such as amphetamines and modafinil act primarily through dopamine receptors, particularly by binding to and suppressing the dopamine transporter, reducing reuptake.[35] The degeneration of the nigrostriatal dopamine system is a neuropathologic basis of Parkinson disease, which can be accompanied by excessive daytime sleepiness.[36] Patients with this disease, however, exhibit widespread neuronal loss, including of orexinergic neurons,[37] which also may contribute to their sleepiness. Dopamine agonists used to treat periodic limb movement in sleep and restless legs syndrome do not usually induce arousal, probably because they have effects on both presynaptic and postsynaptic dopamine receptors; presynaptic receptors inhibit transmitter release, counteracting postsynaptic stimulation.

Glutamate

Glutamate is the most widespread excitatory neurotransmitter of the brain. Glutamate-containing neurons are found throughout the brain, including in the core of the pontine and midbrain reticular formation.[38] Extracellular levels of glutamate in the cortex and in the hypothalamus are elevated during waking and REM sleep, compared with NREM sleep.[39,40] Arousal is increased by application of glutamate to

many sites.[41] Its actions are mediated by receptors controlling membrane ion flux, including the *N*-methyl-D-aspartate (NMDA) receptor, and "metabotrophic" receptors controlling intracellular processes. Humans may be exposed to systemic NMDA receptor antagonists in the form of anesthetics (e.g., ketamine) or recreational drugs (e.g., phencyclidine). The effects are dosage-dependent: Low dosages produce arousal, and high dosages produce sedation. In rats, exposure to NMDA antagonists induces a potent long-lasting enhancement of NREM slow wave activity.[42]

SLEEP-PROMOTING MECHANISMS

As noted in the foregoing review of evidence for multiple neurochemically specific arousal systems, the activity of each of these neuronal groups is reduced during NREM sleep. In most groups, the reduction in neuronal discharge precedes EEG changes that herald sleep onset. How is the process of sleep onset orchestrated?

Rostral Hypothalamic Sleep-Promoting System

Sleep-Active Neurons in the Preoptic Area

More than 70 years ago, Von Economo postulated a POA sleep-promoting area on the basis of his observation that postmortem examinations in patients with encephalitis who had severe insomnia showed inflammatory lesions in this area of the brain.[43] Patients with hypersomnia had lesions in the vicinity of the posterior hypothalamus (PH). To von Economo, these observations suggested the concept of opposing hypothalamic sleep-promoting and wake-promoting systems. In rats, symmetric bilateral transections of the POA resulted in complete sleeplessness, and symmetric bilateral transections of the PH caused continuous sleep.[44] Rats that underwent both POA and PH transections exhibited continuous sleep, as with PH transections alone. This finding was interpreted as showing that the POA normally inhibits the PH wake-promoting region. The PH wake-promoting system can now be understood as the rostral extension of arousal-promoting systems and pathways summarized earlier.

The existence of a sleep-promoting mechanism in the POA has been confirmed by a variety of methods. Experimental bilateral lesions of the POA with diameters of 1 to 2 mm in rats and cats induce partial sleep loss. Larger bilateral lesions (3 to 5 mm in diameter) that extend into the adjacent basal forebrain are associated with more severe insomnia (as described in our review of this work[45]). After introduction of POA lesions that result in partial sleep loss, residual sleep is characterized by reduced slow wave (delta) EEG activity.[46] Delta EEG activity in NREM sleep is augmented by POA warming (see later). Because delta activity is recognized as a marker of enhanced sleep drive, this finding suggests that POA output contributes to the regulation of sleep drive. Electrical or chemical stimulation of the POA evokes EEG synchronization and behavioral sleep.

The identification of sleep-active POA neurons has been advanced by the application of the c-Fos immunostaining method.[47] Rapid expression of the proto-oncogene *c-fos* has been identified as a marker of neuronal activation in many brain sites and in multiple cell types.[48] Thus c-Fos immunostaining permits functional mapping of neurons, identifying neurons that were activated in the preceding 30 to 60 minutes. After sustained sleep, but not waking, a discrete cluster of neurons exhibiting c-Fos is found in the ventrolateral preoptic area (VLPO).[47] The VLPO is located at the base of the brain, lateral to the optic chiasm. Sleep-related Fos immunoreactive neurons also are localized within the rostral and caudal median preoptic nucleus (MnPN).[49] The MnPN is a midline cell group that widens to form a "cap" around the rostral end of the third ventricle. Examples of c-Fos immunostaining and the correlations between c-Fos counts and sleep amounts are shown in Figure 7-2. The number of sleep-related Fos immunoreactive neurons in the MnPN is increased in rats exposed to a warm ambient temperature, in association with increased NREM sleep.[49]

The VLPO and the MnPN contain a high density of neurons with sleep-related discharge.[50,51] Most sleep-active neurons in these nuclei are more activated during both NREM and REM sleep than in awake states (Figure 7-3). VLPO neurons typically exhibit increases in activity during spontaneous transitions between waking and sleep and display a progressive increase in discharge rate from light to deep NREM sleep. MnPN neurons often show a gradual increase in firing rates *before* sleep onset, elevated discharge rates during NREM sleep, and a small but significant additional increase in discharge rate during REM sleep.

VLPO and MnPN sleep regulatory neurons are dynamically responsive to changes in homeostatic sleep pressure induced by sleep deprivation. In the MnPN, c-Fos expression in GABAergic neurons is increased after a brief (2- to 3-hour) period of sleep deprivation, even if no opportunity for recovery sleep is permitted.[52] Fos expression in GABAergic VLPO neurons is increased during recovery sleep subsequent to sleep deprivation.[52] When the neuronal discharge of individual sleep-active neurons in the MnPN and VLPO is continuously recorded across baseline sleep-wake, sleep deprivation, and recovery sleep, dynamic responses to changing homeostatic sleep pressure are evident[53] (Figure 7-4). MnPN and VLPO neurons identified as sleep-active during spontaneous baseline sleep exhibit progressive increases in waking discharge rate during sleep deprivation that are correlated with behavioral indices of sleep pressure (Figure 7-4). NREM sleep-related discharge is elevated in comparison with baseline NREM sleep early in the recovery period and then declines in association with the reduction in EEG delta power.[53] Thus sleep-active neurons in the MnPN/VLPO are involved in orchestrating spontaneous wake-sleep transitions (see further on) and function as components of the neuronal circuits that mediate homeostatic responses to sustained wakefulness.

Orchestration of Sleep by Sleep-Promoting Circuits of the Preoptic Area

How do POA sleep-active neurons initiate and sustain sleep? VLPO neurons that exhibit c-Fos immunoreactivity during sleep express glutamic acid decarboxylase, the synthetic enzyme that produces the *inhibitory* neurotransmitter GABA. A majority of MnPN neurons that exhibit sleep-related Fos-immunoreactivity also express glutamic acid decarboxylase.[54] VLPO neurons send anatomic projections to histamine neurons in the TMN.[55] Additional projections of the VLPO include the midbrain dorsal raphe and the locus coeruleus. The MnPN also projects to both the dorsal raphe and locus coeruleus.[56] Both MnPN and VLPO also project to the perifornical lateral hypothalamic area, the location of cell bodies of the orexin arousal system.[57] Thus discharge of VLPO and MnPN

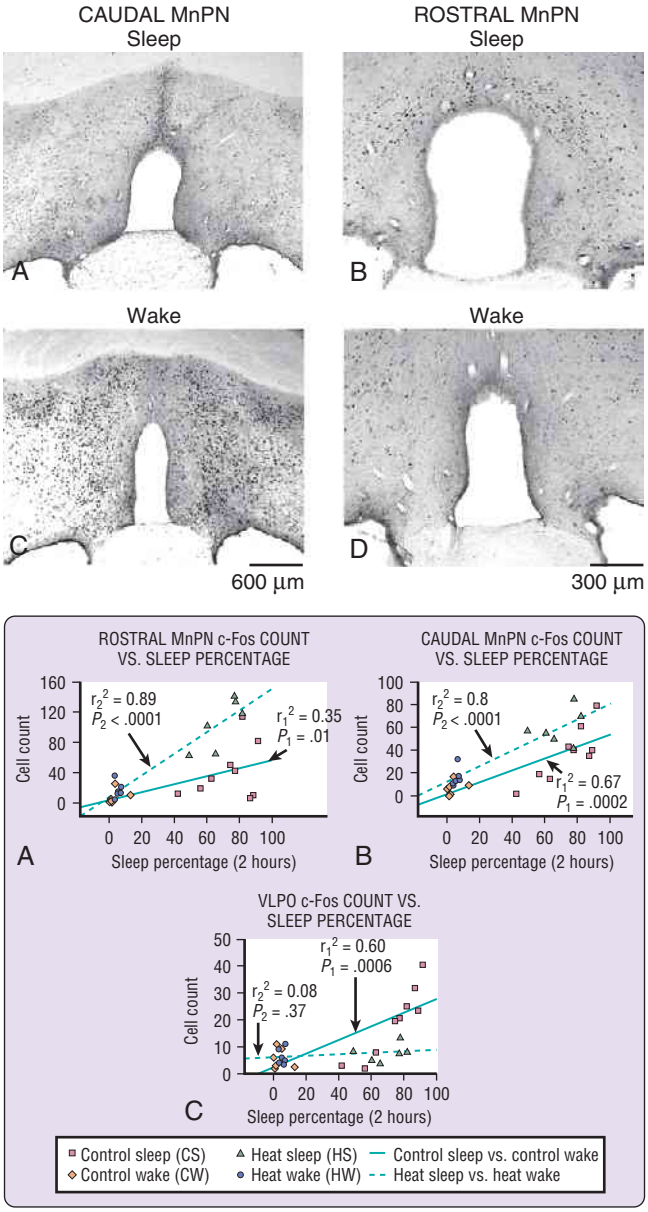


Figure 7-2 *Upper panel:* Examples of c-Fos immunostaining of preoptic area (POA) neuronal nuclei, identified by *dark spots* after either sustained spontaneous sleep or wakefulness. c-Fos immunostaining is a marker of neuronal activation and is a method for mapping the localization of sleep-active neurons in the brain. After sleep, increased staining was seen in the midline **(A)** or around the top of the third ventricle **(B)** relative to the wake samples (**C** and **D**). These sites correspond to the caudal and rostral median preoptic nucleus (MnPN). Similar results were seen in the ventrolateral preoptic area (VLPO). In other POA sites, c-Fos immunostaining was seen after both waking and sleep. *Lower panel:* Regression functions and correlations relating c-Fos counts and sleep amounts before sacrifice among individual animals. In all sites, high correlations were found between sleep amounts and c-Fos counts at a normal ambient temperature. Groups of animals were studied in both normal and warm ambient temperatures. In a warm ambient temperature, c-Fos counts after sleep and correlations between counts and sleep amounts were increased in MnPN sites (**A** and **B**), but they were suppressed in the VLPO **(C)**. (From Gong H, Szymusiak R, King J, et al. Sleep-related c-Fos expression in the preoptic hypothalamus: effects of ambient warming. *Am J Physiol Regul Integr Comp Physiol* 2000;279:R2079-88.)

GABAergic neurons during sleep is expected to release GABA at these sites. Indeed, GABA release is increased during NREM sleep and further increased in REM sleep in the PH, doral raphe, and locus coeruleus.[58-60] Sleep-active neurons in VLPO and MnPN exhibit discharge-rate-change profiles across the wake-NREM-REM cycle that are reciprocal to those of wake-promoting histamine, 5-HT, norepinephrine, and orexin neurons (Figure 7-5). These findings support a hypothesis that POA sleep-active neurons, through release of GABA, inhibit multiple arousal systems.

The PH and LH areas also facilitate both motor and autonomic activation, and these processes are under GABAergic control.[61] Sleep-related GABAergic inhibition of PH and LH provides a basis for sleep-related deactivation of autonomic and motor functions. POA lesions suppress REM as well as NREM sleep. Because sustained NREM sleep normally precedes REM sleep, REM sleep loss after POA damage may be a secondary consequence of NREM sleep disturbance. Alternatively, POA mechanisms may directly facilitate REM sleep as well as NREM sleep.

GABAergic neurons in the VLPO region are inhibited by ACH, 5-HT, and norepinephrine.[62] MnPN neurons are inhibited by norepinephrine.[63] Thus POA sleep-active neurons inhibit arousal systems, and arousal systems inhibit POA sleep-active neurons. These mutually inhibitory processes are hypothesized to underlie a bi-stable sleep-wake switch (see Figure 7-5).[64] Activation of arousal systems inhibits sleep-active neurons, thereby removing inhibition of arousal systems and facilitating stable episodes of waking. On the other hand, activation of sleep-promoting neurons would inhibit arousal-related neurons, thereby disinhibiting sleep-promoting neurons and promoting consolidated sleep episodes. This model provides a mechanism for the stabilization of both sleep and waking states.

Abnormalities in one or more of the components of this "flip-flop" system could result in less stable sleep and wake states, a possible explanation for the fragmentation of sleep in sleep disorders in which insomnia is an element, and for the fragmentation of waking in narcolepsy.

Sleep Regulatory Functions of Melanin-Concentrating Hormone Neurons

Neurons expressing the inhibitory peptide melanin-concentrating hormone (MCH) are located in the PH, where they are intermingled with orexin neurons (reviewed by Bittencourt[65]). A majority of MCH neurons also express the inhibitory neurotransmitter GABA. Rats possess approximately twice as many MCH neurons as orexin neurons, with a more widespread distribution throughout the hypothalamus. MCH groups include a medial cluster found adjacent to the third ventricle, a perifornical cluster, a far lateral cluster located medial to the internal capsule and extending to the zona incerta, and a more posterior-medial group localized in the supramammalary region.[65] MCH neurons have anatomic interconnectivity with hypothalamic and brainstem nuclei implicated in sleep-wake regulation.[65,66]

Intracerebroventricular injection of MCH was found to increase NREM and REM sleep, identifying a potential sleep regulatory role for the peptide.[67] Subsequent work pointed to a role for MCH neurons in REM sleep control. Increased c-Fos immunoreactivity was observed in MCH neurons during REM-enriched sleep after 72 hours of REM sleep

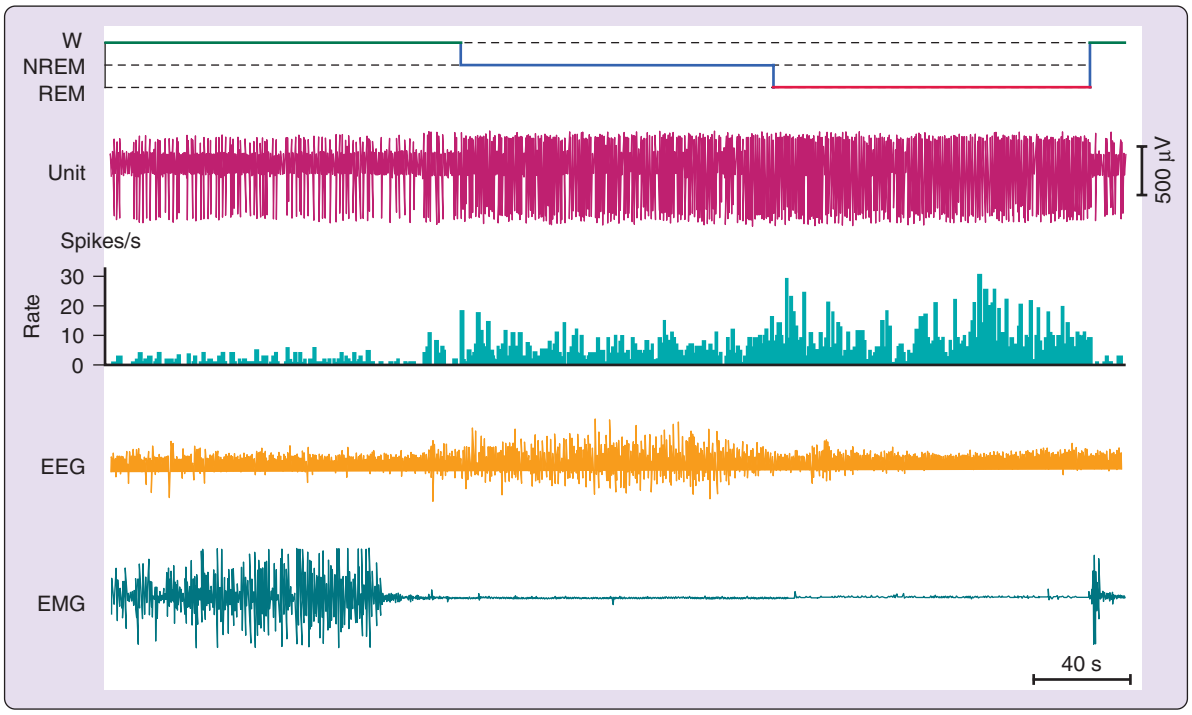


Figure 7-3 Example of sleep-active neurons in the median preoptic nucleus (MnPN). Shown is a continuous recording of discharge of an MnPN neuron during a wake (W)-NREM-REM cycle (*top*). Discharge rate increased at the onset of sleep, as indicated by the increased amplitude of the electroencephalogram (EEG). Discharge rate increased further in association with REM sleep (*right*). Such sleep-active neurons constituted a majority of those encountered in the median preoptic nucleus (MnPN) and the ventrolateral preoptic area (VLPO). The presence of sleep-active neurons is one critical piece of evidence for the importance of a brain region in the facilitation of sleep. EMG, Electromyogram. (From Suntsova N, Szymusiak R, Alam MN, et al. Sleep-waking discharge patterns of median preoptic nucleus neurons in rats. *J Physiol* 2002;543:665-77.)

deprivation.[68] REM sleep–related c-Fos expression was found in MCH neurons with descending projections to critical brainstem nuclei implicated in REM sleep control.[69] In the only published report of sleep-wake discharge pattern in identified MCH neurons, a small sample of neurons located in the perifornical area were activated during REM sleep relative to the status of such neurons during waking and NREM sleep,[70] consistent with a hypothesized REM sleep regulatory role. Targeted optogentic excitation of MCH neurons during NREM sleep was found to promote transitions to REM sleep,[71,72] whereas stimulation at REM sleep onset prolonged REM bout durations.[71] Release of GABA by MCH neurons in response to optogenetic stimulation may be responsible for REM sleep effects.[71] Optogenetic silencing of MCH neurons had no effect on REM sleep measures,[72] indicating that activation of MCH neurons is sufficient but not necessary to promote REM sleep from a background of NREM sleep.

Other selective genetic manipulations of MCH neuronal function point to a key role for these neurons in the control of sleep onset and NREM sleep. Continuous optogenetic excitation of MCH neurons (by light pulses delivered for 1 minute every 5 minutes at 10 Hz) at the start of the dark (active) phase in mice reduced sleep latency, decreased the duration of waking episodes by 50%, and increased NREM and REM sleep time.[73] In the human brain, MCH release, as measured in the amygdala, is maximal at sleep onset.[74] Widespread, conditional ablation of the MCH neuronal population by cell-specific expression of diphtheria toxin A increased wakefulness and decreased NREM sleep without affecting REM sleep.[72] Collectively, the findings suggest a heterogeneity of MCH neuronal function, with some neurons interacting with circuits that control REM sleep and another population of MCH neurons functioning to promote NREM sleep onset and maintenance.

Cortical Sleep-Active Neurons

A population of cortical GABAergic interneurons specifically activated during sleep has been described.[75] In addition to GABA, these neurons colocalize immunoreactivity for neuronal nitric oxide synthase (nNOS) and the substance P receptor, NK1.[75,76] Expression of c-Fos immunoreactivity in these neurons is correlated with time spent in NREM sleep and with NREM EEG delta power. Fos expression during recovery sleep is proportional to the duration of previous time awake, suggesting that sleep-active cortical nNOS neurons are inhibited during waking and activated in response to homeostatic sleep pressure.[76,77] Waking inhibition may be mediated by synaptic input from basal forebrain cholinergic neurons and from monoaminergic inputs arising from the TMN, dorsal raphe nucleus, and locus coeruleus.[76] It is unclear if activation during sleep is due to disinhibition caused by sleep-related silencing of subcortical input, or if sleep-related excitation from neuronal and/or neuromodulatory sources (e.g., adenosine, cytokines) also is involved.

Although sleep-active cortical nNOS neurons are found in multiple cortical areas in rats and mice, the density of these neurons is low. These cortical neurons appear to be a subset of type II nNOS interneurons—that is, those with the largest

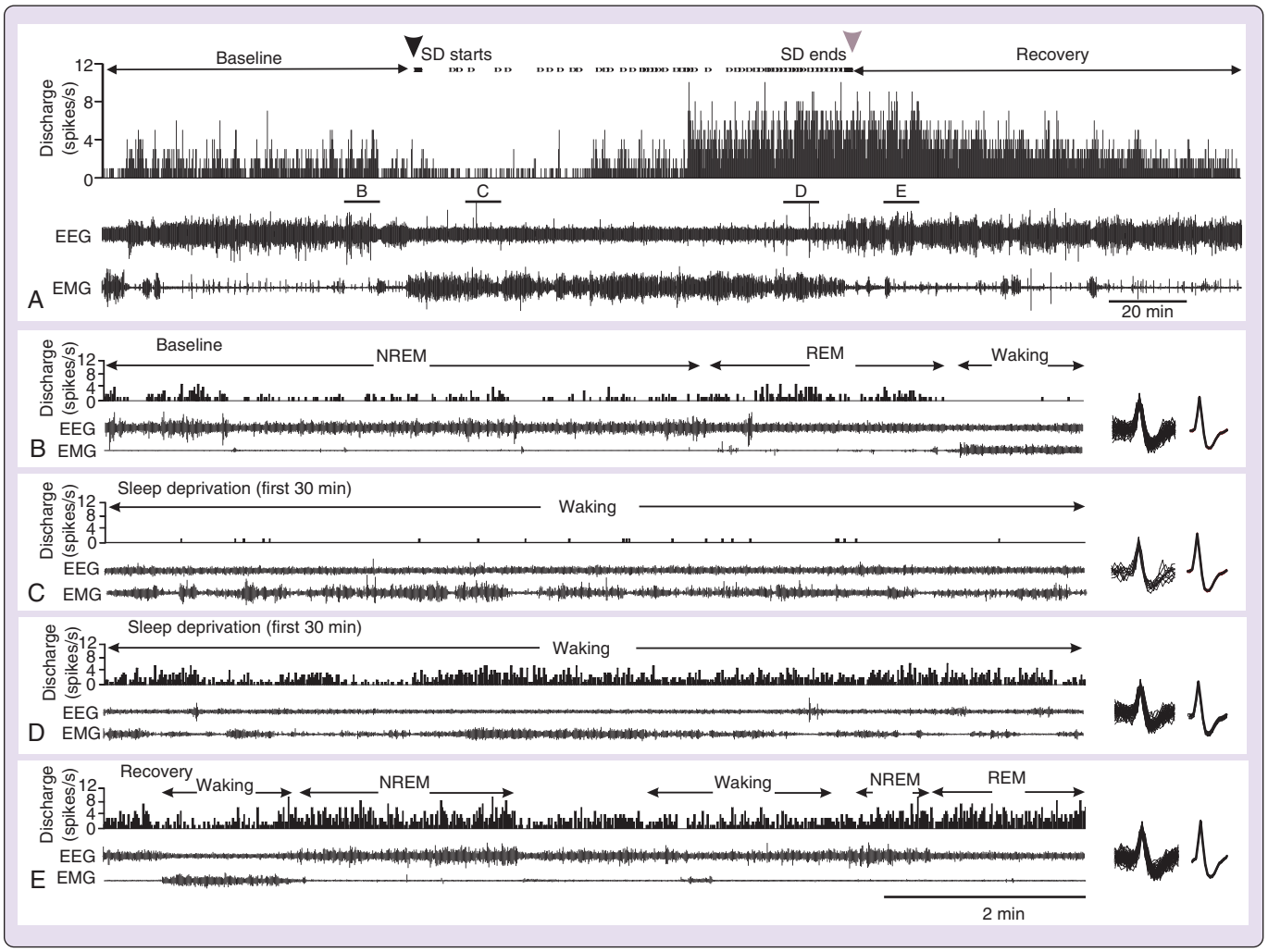


Figure 7-4 Responses of MnPN sleep-active neuron to sleep deprivation (SD) and during recovery sleep (RS). **A,** *From top to bottom:* discharge rate histogram (spikes/sec), cortical EEG, and neck muscle EMG recordings during baseline, SD, and RS. *Black* and *gray arrowheads* at the top indicate start and end of SD, respectively. The *dots* between the *arrows* indicate times when the animal began to fall sleep and the experimenter intervened to maintain wakefulness. **B-E,** Expansion of areas labeled in **A,** showing 10 minutes of recording during baseline **(B),** the first 30 minutes of SD **(C),** the last 30 minutes of SD **(D),** and early RS **(E).** The waveforms at the right of **B** to **E** are superimposed and averaged action potentials recorded the 10 minutes shown in each figure, demonstrating stability of unit recording across all three experimental conditions. During the baseline period, the cell exhibited elevated discharge rates during NREM and REM sleep compared with waking **(B).** At the start of SD, only infrequent interventions were required to keep the animal awake (**A,** *top trace*), and discharge rate of the cell was uniformly low at less than 1 spikes/second **(C).** As homeostatic sleep pressure increased with continuing SD, as evidenced by the increasing number of interventions required to maintain wakefulness **(A),** the discharge rate of the cell increased. By the end of 2 hours of SD, the waking discharge rates were approximately double that during baseline sleep (**B** and **D**). Discharge of the cell remained elevated in comparison with baseline during early RS (**B** and **E**) but returned to baseline levels after 2 hours of unrestricted sleep **(A).** EEG, Electroencephalogram; EMG, electromyogram. (From Alam MA, Kumar S, McGinty D, et al. Neuronal activity in the preoptic hypothalamus during sleep deprivation and recovery sleep. *J Neurophysiol* 2014;111:287-99.)

cell bodies, which are a source of cortico-cortical and other long-range projections (as reviewed by Wisor and colleagues[77]). In keeping with this anatomic feature, sleep-active nNOS neurons are well suited to organize synchronous discharge across large ensembles of cortical neurons that are characteristic of deep NREM sleep.[76,77] These neurons may be a component of local cortical circuits responsible for use-dependent augmentation of EEG slow wave activity.[77-79] The differential functional roles of GABA versus NOS signaling by cortical sleep active neurons are unknown.

THALAMIC-CORTICAL INTERACTIONS AND GENERATION OF THE SLEEP ELECTROENCEPHALOGRAM

Changes in cortical EEG patterns usually are considered to be the defining feature of NREM sleep in mammals. Briefly reviewed here is the current understanding of thalamocortical mechanisms underlying NREM sleep EEG patterns, and of how modulation of thalamocortical circuits by arousal and sleep regulatory neuronal systems has an impact on key features of the sleep EEG.

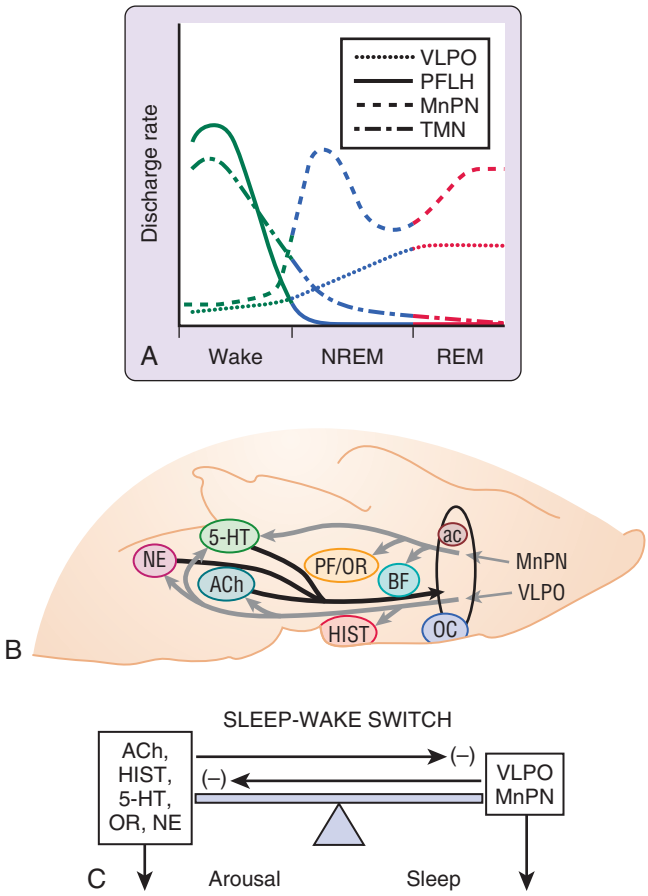


Figure 7-5 Interactions of the preoptic area (POA) sleep-promoting neuronal system with arousal systems that can account for the orchestration of the sleep process. **A,** The neuronal discharge rates across the wake-NREM-REM cycle of sleep-active neurons from the ventrolateral preoptic area (VLPO) and the median preoptic nucleus (MnPN), and from arousal-related (wake-active) neurons in the perifornical lateral hypothalamus (PFLH) and tuberomammillary nucleus (TMN). These neuronal groups generally have reciprocal discharge patterns, although MnPN and VLPO neurons exhibit peak activity at different times during NREM episodes. The wake-active, NREM-diminished, REM-off discharge pattern shown for TMN and a subgroup of PFLH neurons also is characteristic of putative serotoninergic neurons of the dorsal raphe nucleus (DR) and putative noradrenergic neurons of the locus coeruleus (see also **C**). **B,** Sagittal section of the diencephalon and upper brainstem of the rat showing anatomic interconnections of MnPN and VLPO neurons with arousal-related neuronal groups. The MnPN and VLPO distribute projections to sites of arousal-related activity including the (1) basal forebrain, (2) PFLH, which includes orexin-containing neurons, (3) histamine (HIST)-containing neurons of the tuberomammillary nucleus, (4) pontomesencephalic acetylcholine (ACh)-containing neurons, (5) pontomesencephalic serotonin (5-HT)-containing neurons, and (6) noradrenergic (norepinephrine [NE]-containing) neurons of the pons, particularly the locus coeruleus (LC). 5-Hydroxytryptamine (5-HT), NE, and ACh arousal-related neurons provide inhibitory feedback to sleep-active neurons. The arousal-related neuronal groups also have widespread additional ascending and descending projections that control state-related functions throughout the brain. **C,** Wakefulness and sleep are each facilitated by several neurotransmitters and neuromodulators. ACh-, 5-HT-, NE-, histamine (HA)-, and glutamate (Glu)-expressing neurons directly activate thalamocortical and cortical neurons as well as hypothalamic and basal forebrain nuclei to promote waking. Orexin (OR) neurons facilitate arousal through direct effects and through excitation of ACh, 5-HT, NE, and HA neurons. Sleep-promoting neurons expressed in the preoptic area are gamma-aminobutyric acid (GABA)-ergic and act by inhibiting wake-promoting neurons. Other sleep-promoting molecules, adenosine (AD), endoplasmic reticulum (ER) stress signaling molecules, growth hormone–releasing hormone (GHRH), certain cytokines, and signals of oxidative stress such as oxidized glutathione (GSSG) (see text), are thought to act indirectly, by either facilitating sleep-active neurons or inhibiting wake-promoting neurons, or both in the case of AD. Subsets of wake-promoting neurons and sleep-promoting neurons are mutually inhibitory. Accordingly, if one system is more strongly activated, the opposing system is inhibited, with consequent reduced feedback inhibition of the initiating neurons, thereby facilitating sleep-wake state stability, as in an electrical "flip-flop" switch. The circadian clock in the SCN regulates the timing of sleep by exciting arousal systems during the active phase of the day and by exciting or disinhibiting sleep-promoting neurons during the rest phase. Because of the redundancy of both wake-promoting and sleep-promoting systems, deficiency of any one system may have only small effects on the amount of sleep but may reduce the stability of sleep. ac, Anterior commissure. (Modified from McGinty D, Szymusiak R. Hypothalamic regulation of sleep and arousal. *Front Biosci* 2003;8: 1074-83.)

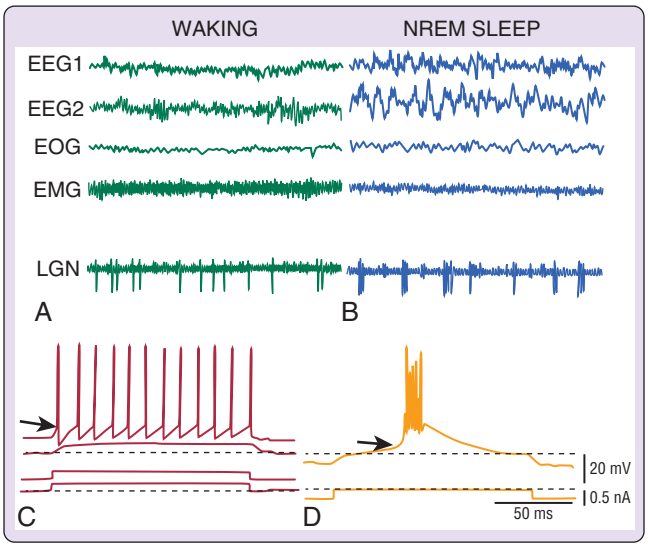


Figure 7-6 Thalamic neurons exhibit distinct patterns of action potential generation during waking/REM sleep and during NREM sleep. **A** and **B,** Typical extracellularly recorded discharge patterns of a neuron in the cat lateral geniculate nucleus during waking and NREM sleep. Note the change from tonic, single-spike firing during the awake state **(A)** to high-frequency-burst firing during NREM sleep **(B).** Tonic versus burst firing reflects intrinsic, voltage-dependent properties of thalamic neurons. These discharge patterns can be recorded in neurons from isolated slices of thalamus. **C** and **D,** In vitro intracellular recordings of a relay neuron from guinea pig thalamus. In **C,** direct current (DC) injections are labeled 1 and 2, and recordings of intracellular voltage, 3 and 4. Spontaneous resting potential for the cell is indicated by the *dashed line*. A depolarizing current step (1) delivered at resting potential (>–65 mV) evokes a tonic depolarizing response in the neuron (3) that is subthreshold for action potential generation. When membrane potential is rendered more positive by DC injection, the same depolarizing step (2) evokes tonic generation of fast action potentials (*arrow*) that persist for the duration of the depolarizing pulse. In **D,** the neuron has been hyperpolarized below resting potential (<–65 mV) by negative DC injection, and the low threshold Ca^{2+} current (I_t) is activated. When a depolarizing pulse is applied on the background of hyperpolarization, a slow Ca^{2+} spike is evoked (*arrow*), and it is crowned by a high-frequency burst of fast Na^+ action potentials. I_t is inactivated in response to the Ca^{2+}-mediated depolarization, and membrane potential sags toward the hyperpolarized level despite the continuance of the depolarizing current step. EOG, Electrooculogram; LGN, lateral geniculate nucleus. (**A** and **B** modified from McCarley RW, Benoit O, Barrionuevo G. Lateral geniculate nucleus unitary discharge during sleep and waking: state- and rate-specific effects. *J Neurophysiol* 1983;50:798-818. **C** and **D** modified from Jahnsen H, Llinas R. Electrophysiological properties of guinea-pig thalamic neurons: an in vitro study. *J Physiol* 1984;349:205-26.)

Thalamocortical circuits exhibit two fundamentally different modes of operation across the sleep-wake cycle: a state of tonic activation, or desynchrony, during waking and REM sleep and a state of rhythmic, synchronized activity that is characteristic of NREM sleep.[80] The two functional modes of thalamocortical activity are evident at the level of single neurons. During waking and REM sleep, thalamocortical neurons exhibit tonic firing of single action potentials (Figure 7-6, *A*) that are modulated by the levels of excitatory input from thalamic afferents, including specific sensory afferents.[80,81] During NREM sleep, relay neurons discharge in high-frequency bursts of action potentials, followed by long pauses (see Figure 7-6, *B*).

These two modes of action potential generation reflect the expression of intrinsic properties of thalamocortical neurons, and a specialized, voltage-sensitive Ca^{2+} current plays a critical role.[82] This Ca^{2+} current, known as the low-threshold or transient Ca^{2+} current (I_t), is inactivated (nonfunctional) when the membrane potential of thalamic relay neurons is relatively depolarized (less negative than –65 mV). Thus, when depolarizing input is delivered to a relay neuron that is resting at this level of membrane polarization, the cell responds with tonic single-spike firing (see Figure 7-6, *C*). When relay neurons are hyperpolarized (membrane potential more negative than –65 mV), I_t becomes activated, and depolarizing input evokes a slow Ca^{2+}-mediated depolarization (100 to 200 milliseconds in duration) that is crowned by a burst of three to eight fast Na^+-mediated action potentials (see Figure 7-6, *D*). There is a pause in the generation of fast action potentials after the burst, because I_t is self-inactivated by the Ca^{2+}-mediated depolarization, and membrane potential falls below the threshold for action potential generation and is restored back to the resting, hyperpolarized state (see Figure 7-6, *D*). Thus the properties of I_t equip thalamocortical neurons with the ability to generate action potentials in two different modes: (1) tonic firing when stimulated from a relatively depolarized resting state and (2) burst-pause firing from a hyperpolarized resting state.[80,81]

The major NREM sleep EEG rhythms—spindles, delta waves, and slow oscillations—all arise through a combination of intrinsic neuronal properties (e.g., I_t) and the synaptic organization of cortical and thalamic circuits.[83] A schematic of the core circuitry responsible for the generation of sleep rhythms in the EEG is shown in Figure 7-7. Thalamocortical relay neurons receive excitatory input from sensory neurons and from several of the brainstem arousal systems that function to promote depolarization and tonic firing in relay cells during waking. During waking, this excitation is faithfully conveyed by ascending thalamocortical axons to the functionally relevant area of cortex for processing and integration. Thalamic relay neurons also send a collateral projection to the adjacent portion of the thalamic reticular nucleus (RE), which is a thin band of neurons that surrounds most thalamic relay nuclei. RE neurons are GABAergic, and they send an inhibitory projection back to relay neurons. These reciprocal connections between relay and RE neurons are thought to be important for aspects of waking thalamic function.

The final critical piece of the basic circuitry for consideration is the feedback projection from layer VI pyramidal cells in cortex to both thalamic relay neurons and RE neurons. Corticothalamic projections are topographically organized so that each cortical column has connectivity with the same relay neurons from which they derive thalamic inputs, and with the corresponding sector of the RE. In this anatomic/functional scheme, a key feature is the central location of RE neurons, which receive copies of thalamocortical and corticothalamic activity and send an inhibitory projection back to the relay neurons. Although corticothalamic projections have excitatory effects on their postsynaptic targets in the thalamus, corticothalamic inputs to the RE are so powerful that the net response evoked in relay neurons by cortical stimulation often is inhibition.[83]

Sleep Spindles

In humans, EEG spindles are waxing and waning, nearly sinusoidal waves with a frequency profile of 10 to 15 Hz. Spindles are generated in the thalamus, as evidenced by the fact that thalamectomy eliminates spindles in the sleep EEG.[4] At the level of the thalamus, spindles are generated by an interplay between neurons in the RE and the relay nuclei.[80,81,83] RE neurons also possess I_t calcium channels and exhibit high-frequency-burst firing from a hyperpolarized background. A high-frequency burst in RE neurons will

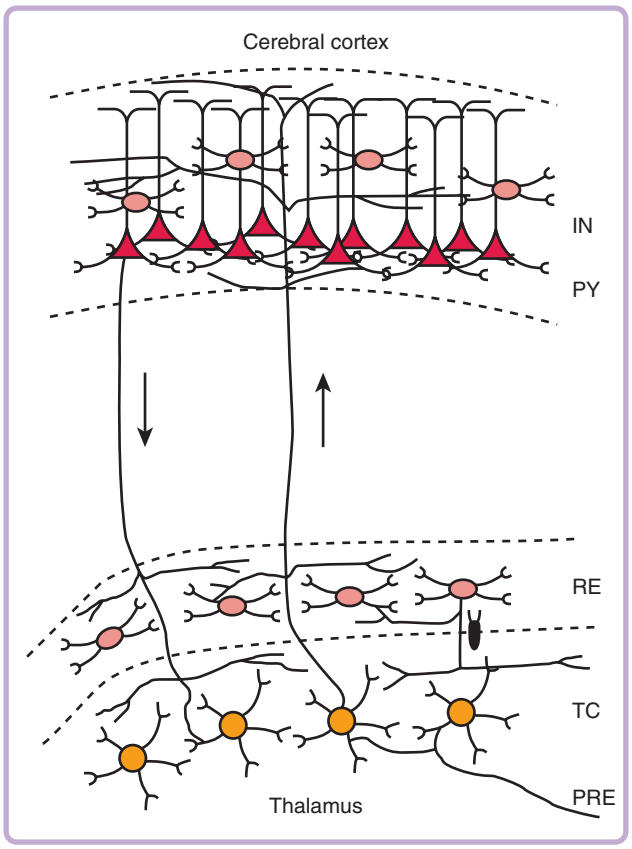


Figure 7-7 Schematic representation of thalamic and cortical cell types involved in the generation of sleep EEG rhythms, and of the synaptic connectivity among the cell types. Four cell types are shown: thalamocortical relay (TC) cells, thalamic reticular (RE) neurons, cortical pyramidal (PY) cells, and cortical interneurons (IN). TC cells receive excitatory inputs from prethalamic afferent fibers (PRE) arising from specific sensory systems and from cholinergic and monoaminergic arousal systems located in the brainstem and posterior hypothalamus. Activity in sensory systems is relayed to the appropriate cortical area by ascending thalamocortical axons *(up arrow)*. TC neurons also send an axon collateral that makes synaptic contact with RE neurons. RE neurons are GABAergic, and they send an inhibitory projection back to TC neurons *(down arrow)*. Corticothalamic feedback is mediated by layer VI, PY neurons that project back to the same relay neurons from which they derive thalamic input, and they send an axon collateral to RE neurons. Corticothalamic projections are excitatory on both RE and TC cells, but cortical stimulation can evoke net inhibitory effects on TC neurons because of activation of GABAergic RE neurons. (From Destexhe A, Sejnowski TJ. Interactions between membrane conductances underlying thalamocortical slow-wave oscillations. *Physiol Rev* 2003;83:1401-53.)

produce strong inhibitory postsynaptic potentials (IPSPs) in relay neurons that are followed by rebound slow Ca^{2+} spikes and a high-frequency burst. This burst of firing in the relay neuron is conveyed back to the RE, evoking an excitatory postsynaptic potential that triggers a calcium spike and a burst in RE neurons. In thalamic slices in which connectivity between the RE and the adjacent relay nuclei is preserved, this disynaptic circuit can generate spontaneous spindle-like oscillations that can propagate across the slice.[81] In the intact brain, the spindle oscillation in the thalamus is conveyed to the cortex by the pattern of burst firing in thalamic relay neurons.[83,84] However, relay of specific sensory information through the thalamus to the cortex is severely compromised during spindle oscillations because of the combination of disfacilitation of relay neurons resulting from loss of excitatory input from arousal systems and the rhythmic IPSPs evoked by RE input. This sensory deafferentation of the cortex is believed to play an important role in maintaining NREM sleep continuity. Although the spindle oscillation originates in the thalamus, cortical feedback to the thalamus and cortico-cortical connections are important in synchronizing the occurrence of spindles over widespread thalamic and cortical areas.[83,84] Phasic activation of layer VI pyramidal neurons excites neurons in the RE and synchronizes IPSP–rebound burst sequences in thalamic relay neurons with cortical activity.

Delta Waves

The delta oscillation of NREM sleep appears to have both cortical and thalamic components. This dual origin is evidenced by the fact that cortical delta activity in the 1- to 4-Hz range persists after complete thalamectomy[4] and by the demonstration that isolated thalamic relay neurons can generate a spontaneous clocklike delta oscillation resulting from the interplay of I_t and a hyperpolarization-activated cation current, known as the h-current.[81] In the intact, sleeping brain, both sources of delta oscillation contribute to the frequency content of the cortical EEG. As discussed previously for spindles, corticothalamic and cortico-cortical connections function to synchronize delta oscillations over widespread cortical areas.

Slow Oscillations

Slow oscillations (with a frequency less than 1 Hz) are a key aspect of the sleep EEG because they function to coordinate the occurrence of other synchronous EEG events (e.g., delta waves, spindles, and K-complexes). Slow oscillations are thought to be primarily of cortical origin. They are absent from the thalamus in chronic decorticate animals and are present in the cortex after thalamectomy as well as in isolated cortical slices.[83] It has been recently reported that cortical deafferentation causes an acute suppression of slow oscillations that recovers over time, indicating a thalamic contribution (as reviewed by David and Schmiedt[85]). Underneath the slow oscillations, fluctuations occur between two states of activity in nearly all cortical neurons.[86,87] "Up" states are characterized by depolarization and generation of trains of action potentials. Up states occur simultaneously in all cell types, including interneurons, and both fast excitatory and inhibitory postsynaptic potentials are characteristic of cortical neuronal activity during this state. Up states are followed by a prolonged period of hyperpolarization and quiescence, referred to as "down" states (see Figure 7-8).

Generation of up states occurs through recurrent excitation in local cortical circuits and depends on excitatory transmission through α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) and NMDA receptors.[88] Transitions from up to down states involve a combination of activation of outward K^+ currents and disfacilitation resulting from depression of excitatory synapses.[88,89] Discharge of layer IV corticothalamic projection neurons during up phases can synchronize IPSPs in thalamic relay neurons through activation of RE neurons, causing the expression of EEG spindles and delta activity of thalamic origin on the background of the slow oscillation.[84] Slow oscillations organize the synchronization and propagation of cortical delta activity through cortico-cortical connections. Frequency spectra of the human NREM sleep EEG reflect this dynamic, with prominent spectral peaks in the delta (1 to 4 Hz) and slow oscillation (less than 1 Hz) frequency ranges.[90]

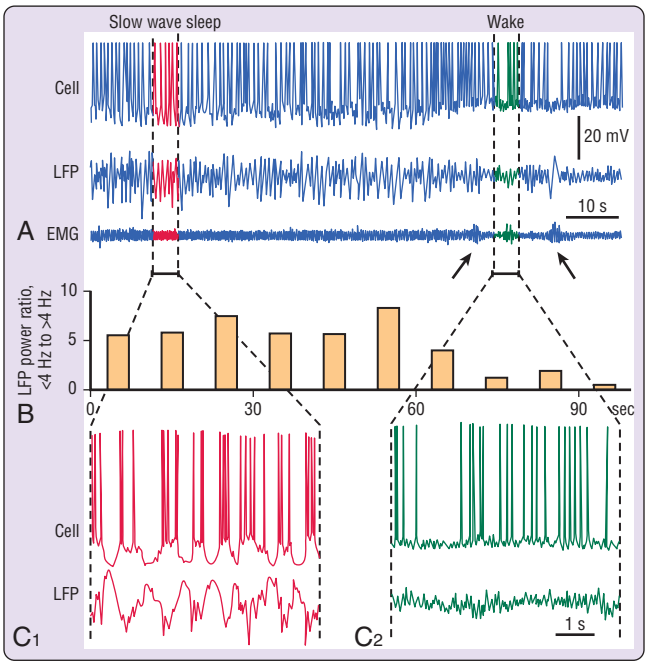


Figure 7-8 Slow oscillations in local cortical field potentials (LFPs) and in the membrane potential of a cortical neuron during NREM sleep. **A,** Simultaneous intracellular, LFP, and EMG recording during sleep and wakefulness. The animal is in NREM sleep at the beginning of the recording, with a transition to waking after approximately 70 seconds (*arrows* indicate EMG activation). Action potentials are truncated in the intracellular recording. **B,** Levels of delta power in the LFP are higher during NREM sleep than during waking. Plotted are 10-second bins of the ratio of spectral power (<4 Hz/>4 Hz) recorded in the LFP. **C,** Intracellular activity and LFP recording from **A** shown at expanded time scale. Note clear fluctuations of the membrane potential between depolarized (up) and hyperpolarized (down) states during slow wave (NREM) sleep **(C1)** in association with the slow oscillation (<1 Hz) in the LFP. During wakefulness **(C2),** cell is tonically depolarized, and no sustained episodes of hyperpolarization are present. EMG, Electromyogram. (From Mukovski M, Chauvette S, Timofeev I, Volgushev M. Detection of active and silent states in neocortical neurons from the field potential signal during slow-wave sleep. *Cereb Cortex* 2007;17:400-14.)

By orchestrating the temporal and spatial coherence of rhythmic oscillations in thalamocortical circuits, slow oscillations are thought to be important in promoting the functional sensory deafferentation of the cortex during sleep, which in turn enhances sleep continuity and sleep depth. The transitions between up and down states in cortical neurons have been hypothesized to underlie changes in synaptic plasticity, or synaptic strength, during sleep, and to contribute to sleep-dependent changes in learning and memory.[91]

Spindle and delta oscillations are blocked by stimulation of the rostral brainstem, which activates cholinergic, monoaminergic, orexinergic, and glutamatergic inputs to the thalamus. Activity of cholinergic and monoaminergic neurons facilitates thalamic *depolarization* through inhibition of potassium channels.[92] Accordingly, *inhibition* of these arousal systems facilitates hyperpolarization of thalamic neurons, permitting the activation of voltage-dependent membrane currents underlying spindles and slow waves.

INTEGRATION OF CIRCADIAN RHYTHMS AND SLEEP

The suprachiasmatic nucleus (SCN) of the POA generates the signals that bring about the circadian patterns of sleep-waking.[93] Direct projections from the SCN to areas of the POA implicated in sleep regulation are sparse, but multisynaptic pathways by which SCN signals can control sleep-active neurons have been described. Introduction of lesions of a primary SCN projection target, the subparaventricular zone, like those of the SCN, itself, eliminates circadian rhythms of sleep-waking.[94] The SPVZ projects directly to the MnPN and indirectly to the VLPO, MnPN, and other POA regions through the dorsomedial hypothalamic nucleus.[95,96] Lesions of the dorsomedial hypothalamic nucleus disrupt the circadian distribution of sleep-waking states in rats. In diurnal animals, activity of SCN neurons could inhibit sleep-promoting neurons during the light phase and facilitate sleep-promoting neurons in the dark phase, with the reciprocal pattern occurring in nocturnal animals.

THE PREOPTIC AREA, THERMOREGULATION, AND CONTROL OF SLEEP

Several types of evidence support the hypothesis that sleep onset is coupled to body cooling. Depending on ambient conditions, sleep onset evokes heat loss–effector processes such as cutaneous vasodilation and sweating.[97] In humans, sleep onset occurs soon after vasodilation of the hands and feet, which increases heat loss. Sleep in humans and animals is modulated by ambient temperature. Mild to moderate ambient temperature elevation increases coincident sleep as well as subsequent sleep (as described in an earlier review[98]). An increase in heat production by selective activation of brown adipose tissue, using a beta$_3$ adrenoreceptor agonist, also increases NREM sleep.[99] It is therefore reasonable to hypothesize that one function of sleep is body cooling, particularly after increased body warming during wakefulness.

A circadian temperature rhythm is well documented. In humans, sleep normally occurs on the descending phase of the circadian temperature cycle, but sleep onset evokes a further decrease in temperature, even during continuous bed rest.[100] Self-selected human bedtimes coincide with the time of the maximal rate of decline of core body temperature.[101] The association of sleep onset and the circadian temperature rhythm could represent two independent outputs of the circadian oscillator. However, under certain experimental conditions, including short sleep-wake cycles, internal desynchronization, and forced desynchronization, the interactions of the temperature rhythm and sleep propensity can be studied and partially isolated from effects of prior waking (see Chapter 35). Although sleep may occur at any circadian phase, sleep propensity is increased on the late descending phase of the circadian temperature rhythm and is highest when temperature is low. Awakenings tend to occur as temperature increases, even if sleep time is short. These findings can be interpreted as evidence that the thermoregulatory mechanism that lowers body temperature also promotes sleep.

The coupling of sleep propensity and body cooling is controlled by the POA. The POA is recognized as a thermoregulatory control site on the basis of the effects of local warming and cooling, lesions, local chemical stimulation, and neuronal unit recording studies (reviewed by Boulant[102]). In vivo and in vitro studies have confirmed that the POA contains populations of warm-sensitive and cold-sensitive neurons (WSNs and CSNs), which are identified by changes in neuronal discharge in response to locally applied mild thermal stimuli.

Activation of WSNs induces heat loss mechanisms. Local POA warming promotes NREM sleep and EEG slow wave activity in cats, rabbits, and rats (as detailed in our earlier review of these experiments[98]). NREM sleep is increased for several hours during sustained POA warming.[103] Local POA warming also increases EEG delta activity in sustained NREM. Augmentation of sleep by ambient warming is prevented by coincident local POA cooling.[103] Because POA cooling prevents activation of WSNs, this finding suggests that POA WSN activation mediates the sleep augmentation induced by ambient warming.

Most POA WSNs exhibit increased discharge during NREM sleep compared with waking; most CSNs are wake-active (described in the earlier review[98]). Increases in WSN discharge and decreases in CSN discharge anticipate EEG changes at sleep onset by several seconds. Thus activity of WSNs and CSNs is likely to modulate spontaneous sleep-wake as well as sleep induced by local POA warming. As summarized earlier, POA sleep-active neurons send afferents to putative arousal systems. Local POA warming suppresses the discharge of arousal-related neurons in the PH, dorsal raphe (5-HT neurons), LH orexin neuronal field, and basal forebrain cholinergic field (Figure 7-9).

These studies demonstrate that functional inhibitory regulation of arousal regulatory neurons originates in WSNs located in the POA. Activation of WSNs has two effects: increasing sleep propensity and inducing heat loss, accounting for the coupling of these two processes at sleep onset.

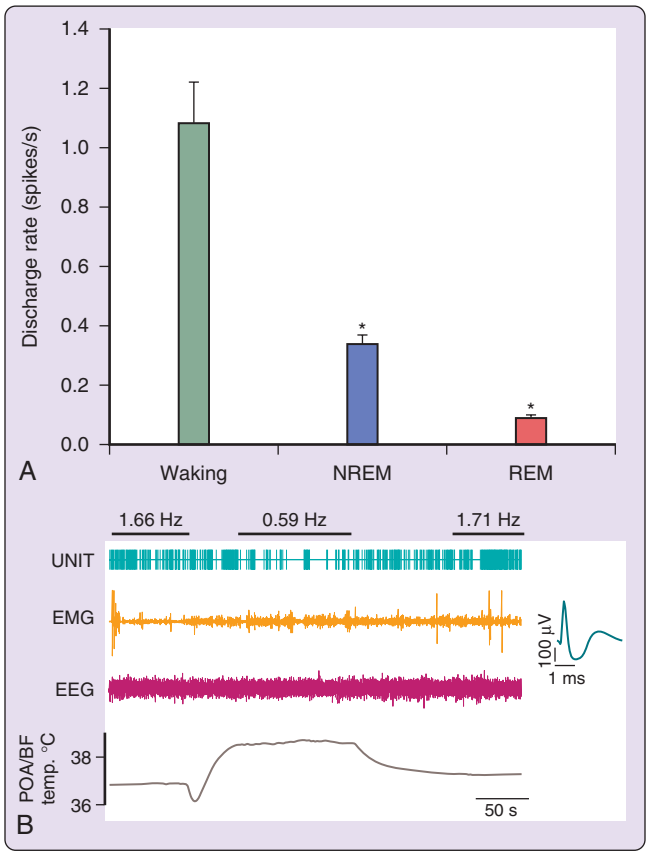


Figure 7-9 Effects of activation of preoptic area (POA) warm-sensitive neurons (WSNs) by local warming on discharge of a putative serotoninergic neuron of the dorsal raphe nucleus (DRN) in the rat. **A,** Pattern of activity of DRN neurons across the sleep-wake cycle. Putative DRN serotoninergic neurons exhibit reduced discharge during NREM compared with waking, and very low discharge in REM sleep. The identification of this pattern of discharge as representing serotoninergic neurons is supported by several types of indirect evidence. **B,** Effects of POA warming on the discharge of a putative serotoninergic REM-off neuron during waking. A POA warming pulse was applied for approximately 100 seconds (*bottom trace*). During POA warming, the discharge was reduced by 64%, to a typical NREM sleep rate. A waking state was maintained during warming, as shown by EEG and EMG recordings. Increased discharge of POA WSNs during spontaneous NREM sleep is thought to contribute to the concurrent reduction of DRN discharge. The central role of temperature-sensitive neurons in sleep control provides a neurologic basis for the coupling of sleep and the circadian temperature rhythm (see text). BF, Basal forebrain; EEG, electroencephalogram; EMG, electromyogram (From Guzman-Marin R, Alam MN, Szymusiak R, et al. Discharge modulation of rat dorsal raphe neurons during sleep and waking: effects of preoptic/basal forebrain warming. *Brain Res* 2000;875:23-34.)

HIERARCHICAL SLEEP CONTROL MODEL

As reviewed earlier, available evidence supports the existence of sleep-promoting mechanisms at all levels of the neuraxis, both in forebrain, including the neocortex, and in the brainstem. Sleep-like behavior may be generated by an isolated lower brainstem, and sleep is facilitated by midbrain, thalamus, and neocortex. Introduction of lesions encompassing the *ventral* pontomesencephalic reticular formation and adjacent ventral structures in cats produced severe sustained sleep reductions.[104] Both NREM and REM sleep were reduced by approximately 50%. Lesioning of the parafacial area in the medulla of rats causes significant reductions in sleep.[9] However, a central role for the POA in the orchestration of sleep is supported by a variety of findings. To rationalize these diverse findings, it is useful to consider a hierarchical control concept, such as that applied in the context of thermoregulation.[105] In this model, a fundamental behavior such as rest-activity or sleep-waking is organized at all levels of the neuraxis. The POA may act as a master controller, integrating sleep-promoting circuitry with sleep homeostatic processes, other behavioral and hormonal regulatory systems, and circadian signals. The limbic system and neocortex are likely sources of additional controls, related to more complex behavioral contingencies including learning processes, instinctive behaviors, and sensory stimuli. These controls may be conveyed to the hypothalamus through projections from neocortex and limbic system.

SLEEP-PROMOTING NEUROCHEMICAL AGENTS

Conceptions of sleep control based on neuronal circuitry, as outlined earlier, are deficient in that they do not provide explanations for quantitative features of sleep or sleep homeostasis. The control of neuronal activity by neurochemical mechanisms, including the release of GABA, occurs in a time frame of seconds. Sleep regulation and homeostasis operate within a time frame of hours to days, not seconds. In addition, a complete description should account for biologic variations in sleep such as the high daily sleep quotas in some species,[106] higher sleep quotas in infants, sleep facilitation after body heating, and increased sleep propensity during acute infection. The investigation of biochemical mechanisms with sustained actions is needed to address these issues. Presented next is a brief overview of this approach, focusing on neurochemical agents with sleep-promoting properties. (Complete details of the biochemical mechanisms of sleep-wake regulation are available in a review by Obal and Krueger.[107])

Adenosine

Adenosine is recognized as an inhibitory neuromodulator in the central nervous system, whose role in sleep is suggested by the potent arousal-producing effects of caffeine, an antagonist of adenosine A_1 and A_{2A} receptors. Adenosine and its analogues were found to promote sleep after systemic administration by intraperitoneal injection, intracerebroventricular administration, and intra-POA microinjection, and after administration by microdialysis in the basal forebrain (reviewed by McCarley[108]). In the basal forebrain and, to a lesser extent, in neocortex, adenosine recovered through microdialysis increases during sustained waking in cats. No increase is found in thalamus or brainstem sites. Application in the basal forebrain of an antisense oligonucleotide to the adenosine A_1 receptor messenger RNA (mRNA), which blocks synthesis of receptor protein, slightly reduced spontaneous sleep but strongly reduced rebound after sleep deprivation.[109] Adenosine A_1 agonists delivered using microdialysis adjacent to basal forebrain neurons inhibited wake-active neurons during both waking and sleep.[110] A current hypothesis is that the effects of adenosine on sleep-waking are mediated by basal forebrain cholinergic neurons through A_1 receptors.[108] Adenosine could act at multiple sites. Adenosine A_{2A} receptors also are present in restricted brain regions and seem to mediate some sleep-promoting effects of adenosine. Administration of A_{2A} agonist by either intracerebroventricular infusion or infusion into the subarachnoid space ventral to the preoptic area increases sleep amounts and increases c-Fos expression in GABAergic neurons in the MnPN and VLPO.[111] Knockdown of A_{2A} receptors in the shell of the nucleus accumbens significantly attenuates the alerting effects of caffeine.[112]

Brain adenosine levels rise when ATP production is reduced; under these conditions, inhibition of neuronal activity is neuroprotective. Astrocytes as well as neurons are potential sources of adenosine. Genetically inhibiting the release of gliotransmitters, including ATP, by astrocytes diminishes homeostatic responses to sleep deprivation.[113] It has been proposed that increased adenosine is a signal of reduced brain energy reserves that develop during waking, and that sleep is induced as an energy-restorative state.[114] With respect to the brain energy restorative concept, some but not all studies show reduced cerebral glycogen, an energy-supply substrate, after sleep deprivation.[115] Functional evidence that brain energy supply is compromised after sleep deprivation is lacking. This absence of evidence is a critical gap in the theory. Of course, adenosine could be a sleep-promoting signal based on functions other than energy supply limitation.

Proinflammatory Cytokines

Several proinflammatory cytokines have sleep-promoting properties. Summarized next is work on a well-studied and prototypical molecule, interleukin (IL)-1. When administered intravenously, intraperitoneally, or into the lateral ventricles, IL-1 increases sleep, particularly NREM sleep (as reviewed by Krueger et al.[116]). Basic findings have been confirmed in several species. REM usually is inhibited by IL-1. Additional evidence supports a hypothesis that IL-1 modulates spontaneous sleep. Administration of agents that block IL-1 reduces sleep. In rats, IL-1 mRNA is increased in the brain during the light phase, when rat sleep is maximal. Sleep deprivation also increases IL-1 mRNA in the brain.

Increased sleep associated with peripheral infection also may be mediated by responses to circulating IL-1, either through vagal afferents or by induction of central IL-1 or other hypnogenic signals. POA sleep-active neurons are activated by local application of IL-1, and wake-active neurons are inhibited.[117] IL-1 fulfills several criteria for a sleep-promoting signal, and it is likely to be important in facilitating sleep during infection.

Prostaglandin D_2

Administration of prostaglandin D_2 (PGD_2) by intracerebroventricular infusion or by microinjection into the POA increases sleep, but the most potent site for administration is the subarachnoid space ventral to the POA and basal forebrain (reviewed by Huang and coworkers[118]). Sleep induced by PGD_2 administration is indistinguishable from normal sleep on the basis of EEG analysis. The enzyme required for PGD_2 synthesis, lipocalin-PGD synthase (L-PGDS) is enriched in the arachnoid membrane and choroid plexus, but the receptor (the D-type prostanoid receptor) is localized more locally in the leptomeninges under the basal forebrain and the TMN. L-PGDS–knockout mice experienced normal baseline sleep, but unlike the control animals, they exhibited no rebound increase in NREM sleep after sleep deprivation. The PGD_2 concentration was higher in the cerebrospinal fluid during NREM sleep than in wakefulness and was higher in the light phase in the rat (when sleep amounts are high). Sleep deprivation increased the PGD_2 concentration in the cerebrospinal fluid. On this basis, it was proposed that PGD_2 plays a central role in sleep homeostasis. The hypnogenic action of PGD_2 is hypothesized to be mediated by adenosine release acting on an adenosine A_{2A} receptor. Administration of A_{2A} antagonists blocked the effects of PGD_2. A functional basis for the role of PGD_2 in sleep homeostasis or its primary localization in the meninges has not been identified.

Growth Hormone–Releasing Hormone

Growth hormone–releasing hormone (GHRH) is known primarily for its role in stimulating the release of growth hormone (GH). A surge in GH release occurs early in the major circadian sleep period in humans, specifically during the earliest stage 3 or 4 NREM sleep episodes (see Chapter 26). GHRH is a peptide with a restricted localization in neurons in the hypothalamic arcuate nucleus and in the adjacent ventromedial and periventricular nuclei. Neurons in the latter locations are thought to be the source of projections to the POA and basal forebrain. GHRH promotes NREM sleep after intraventricular, intravenous, intranasal, or intraperitoneal administration, or after direct microinjection into the POA (as reviewed by Obal and Krueger[119]). Blockade of GHRH by administration of a competitive antagonist reduces baseline sleep and rebound sleep after short-term sleep deprivation. Mutant mice with GHRH signaling abnormalities have lower amounts of NREM sleep. GHRH stimulates cultured GABAergic hypothalamic neurons; these may constitute the GHRH target in the sleep-promoting circuit.[120] Intracerebroventricular infusion of GHRH activates c-Fos expression in GABAergic neurons in the MnPN and VLPO, whereas intracerebroventricular infusion of a competitive GHRH antagonsist suppresses c-Fos in these neurons.[121] It has been suggested that GHRH may elicit sleep onset in conjunction with release of GH as a coordinated process for augmenting protein

synthesis and protecting proteins from degradation during the fasting associated with sleep. In the brain, protein synthesis increases during sleep, and inhibition of protein synthesis augments sleep.[122] GHRH is proposed to be one element of a multiple-element sleep-promoting system.[107]

Endoplasmic Reticular Stress.

Brain protein synthesis is increased during sleep. The endoplasmic reticulum (ER) is a large cellular organelle that houses the machinery that synthesizes and folds new proteins on instructions from messenger RNA, including proteins needed for neuronal growth, repair, and plasticity. The capacity of the ER, particularly with respect to the process of new protein folding, is limited, and overloading can contribute to cell death. The ER generates a set of signals to prevent overloading, called the unfolding protein response (UPR), that temporarily inhibits new protein synthesis. UPR signals are induced by sleep deprivation.[123] Studies in both *Drosophila*[124] and rats[125] suggest that UPR signals also can facilitate sleep. Neuronal damage from long-term sleep deprivation has been linked to the failure of protection by the UPR.[126]

Sleep as Detoxification or Protection from Oxidative Stress

Reactive oxygen species (ROS) include superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radical (OH^-), nitric oxide (NO), and peroxynitrite ($OONO^-$). These molecules are generated during oxidation reactions or reactions between O_2^- and H_2O_2 or NO. ROS normally are reduced by constitutive antioxidants such as oxidized glutathione (GSSR), endogenous glutathione (GSH), and different forms of superoxide dismutase (SOD). Oxidative damage ensues when antioxidant mechanisms fail to adequately scavenge ROS, resulting in generation of oxidized lipids, identified by malondialdehyde (MDA), oxidized proteins (carbonyl proteins), and oxidized nucleic acids, identified by 8-hydroxy-deoxy-guanosine.

GSSR was identified as one of four sleep-promoting substances in brain tissue extracted from sleep-deprived rats.[127] Infusion of GSSR into the lateral ventricle of rats during the dark phase increases both NREM and REM sleep (as reviewed by Ikeda and associates[128]). GSH levels were lower in the hypothalamus of rats after 96 hours of sleep deprivation in studies using the platform-over-water method.[129] Sleep deprivation for 5 to 11 days using the disk-over-water method reduces Cu/Zn SOD (cytosolic SOD) as well as glutathione peroxidase in the hippocampus and brainstem.[130] Also in rats, four days of sleep deprivation using the platform-over-water method reduced GSH and increased MDA in hippocampus,[131] and 72 hours of sleep deprivation using the same method reduced GSH in whole-hippocampus and neocortex samples.[132] In mice, 48 or 72 hours of sleep deprivation using the platform-over-water method increased MDA levels in whole-hippocampus samples.[133] These studies suggest that by reducing antioxidant availability, sleep deprivation can increase the risk of oxidative damage, and that sleep is protective against actions of ROS. The possibility that sleep deprivation could cause neuronal damage was suggested by a finding that supraoptic nucleus neurons exhibited signs of subcellular damage after exposure to sleep deprivation.[134] Supraoptic nucleus neurons may be sensitive to sleep deprivation because of their high rate of protein synthesis.[134] In mice, after only 8 hours of sleep deprivation, locus coeruleus neurons were found to lose antioxidant expression (SOD2, CAT) and expression of a key regulator of responses to metabolic stress, Sirt3.[135] Locus coeruleus neurons were reduced in number by 30%.

What are the signaling molecules related to oxidative stress that increase as a function of sustained wakefulness and can promote sleep? In addition to GSSH, as summarized previously, a probable signal is NO, which can be a response to glutamatergic stimulation, to cytokines and inflammation, and to oxidative stress.[136] Activity of an NO-synthetic enzyme, cytosolic nitric oxide synthase (NOS), was increased during the dark phase in rats, most strongly in the hypothalamus (as reviewed by Gautier-Sauvigne and associates[137]), and NO metabolites are increased during waking and decreased during sleep.[138] Intracerebroventricular or intravenous administration of an NOS inhibitor strongly reduced sleep in rabbits and rats and suppressed NREM sleep response to sleep deprivation. Administration of NO donors increased sleep. NO inhibits oxidative phosphorylation and may stimulate the production of adenosine. NO production could therefore be a mediator of several sleep factors.

ROS may play a role in the underlying pathology of obstructive sleep apnea. Affected patients exhibit signs of increased oxidative stress, including increased O_2^- production by neutrophils, monocytes, and granulocytes derived from patients (see Chapter 20). OSA patients exhibit elevated levels of vascular endothelial growth factor that normally is induced by ROS. These changes were reversed by treatment with continuous positive airway pressure. On the basis of these and several additional findings, it has been proposed that the increased cardiovascular disease in patients with obstructive sleep apnea results from oxidative damage to vascular walls. Oxidative stress is hypothesized to play a central role in other forms of vascular disease and in neurodegenerative disease. It is important to recognize that adenosine, ROS, glutamate, and NO have brief lives in the synaptic space—no more than a few seconds. If these molecules regulate sleep, they must have sustained release, or they may regulate the gene expression to generate sustained downstream effects. These mechanisms are the subject of current investigations.

Sleep-promoting molecules are clearly related to the multiple functions of sleep: adenosine to resupply brain energy reserves, cytokines to facilitate immune functions, GHRH to promote anabolic processes, UPR signals to prevent protein misfolding, and GSSR and NO to prevent oxidative stress–induced cell damage.

CLINICAL PEARLS

- Sleep is reduced in the presence of a wide variety of localized brain lesions; this correlation may account for the association of insomnia with various neuropathologic conditions.
- Degeneration of the hypothalamic wake-promoting cell type that expresses the neuropeptide orexin is known to underlie narcolepsy.
- Identification of certain sleep-promoting molecules strongly suggests that sleep serves to restore brain energy reserves, facilitate immune functions, promote macromolecule synthesis, prevent protein misfolding, and curtail oxidative stress–induced cell damage.

SUMMARY

Rapid progress has been achieved in the elucidation of the neural circuitry underlying the facilitation of sleep and the orchestration of NREM sleep, as well as the facilitation of arousal. Wake and arousal are facilitated by several chemically distinct neuronal groups, including groups synthesizing and releasing acetylcholine, serotonin, norepinephrine, dopamine, histamine, orexin/hypocretin, and glutamate. These neuronal groups distribute axons and axon terminals throughout the brain, providing a basis for concurrent changes in physiology associated with arousal. At the center of the sleep-promoting circuitry is the POA of the hypothalamus. The POA sleep-promoting circuitry has reciprocal inhibitory connections with several arousal-promoting systems. A balance between the activities of sleep-promoting and arousal-promoting neuronal systems determines sleep-waking state. The circadian clock in the SCN has both direct and multisynaptic connections with wake and sleep regulatory neurons, providing a neurologic basis for generation of the daily rhythm of sleep-waking. Both sleep-promoting and arousal-promoting neuronal groups are modulated by a host of processes, including sensory, autonomic, endocrine, metabolic, and behavioral influences, accounting for the sensitivity of sleep to a wide range of centrally acting drugs and behavioral manipulations.

The long-term regulation of sleep—sleep homeostasis—is the subject of competing and still incomplete hypotheses. Long-term sleep homeostasis may reflect the actions of several neurochemical processes that express "sleep factors." Some of these sleep factors, including adenosine, PGD_2, IL-1β, and GHRH, are known to act directly on the POA or adjacent basal forebrain neuronal targets. Sleep factors have been linked to distinct functional models of sleep homeostasis, including brain energy supply (adenosine); control of protein synthesis (GHRH and UPR signaling); local sleep, immune protection, and temperature elevation (IL-1); and protection against oxidative or glutamatergic stress (NO, antioxidant enzymes). All of these factors may be involved in sleep homeostasis, but the relative importance of each factor for daily sleep has yet to be established.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported by Public Health Service (PHS) grants MH 075076, MH 04708, MH 61354, and MH 63323, and by the Medical Research Service of the Department of Veterans Affairs.

Selected Readings

Destexhe A, Sejnowski TJ. Interactions between membrane conductances underlying thalamocortical slow-wave oscillations. *Physiol Rev* 2003;**83**:1401–53.

Haas HL, Sergeeva O, Selbach O. Histamine in the nervous system. *Physiol Rev* 2008;**88**:1183–241.

Huang Z-L, Urade Y, Hayaishi O. Prostaglandins and adenosine in the regulation of sleep and wakefulness. *Curr Opin Pharmacol* 2007;**7**:33–8.

Kilduff T, Cauli B, Gerashchenko D. Activation of cortical interneurons during sleep: an anatomical link to homeostatic sleep regulation? *Trends Neurosci* 2011;**34**:10–19.

Krauchi K. The thermophysiological cascasde leading to sleep initiation in relation to the phase of entrainment. *Sleep Med Rev* 2007;**11**:439–51.

McCarley RW. Neurobiology of REM and NREM sleep. *Sleep Med* 2007;**8**:302–30.

McGinty D, Szymusiak R. Brain structures and mechanisms involved in the generation of NREM sleep: focus on the preoptic hypothalamus. *Sleep Med Rev* 2001;**5**:323–42.

McGinty D, Szymusiak R. Hypothalamic regulation of sleep and arousal. *Front Biosci* 2003;**8**:1074–83.

Obal F Jr, Krueger JM. Biochemical regulation of non-rapid-eye-movement sleep. *Front Biosci* 2003;**8**:520–50.

Obal F Jr, Krueger JM. GHRH and sleep. *Sleep Med Rev* 2004;**8**:367–77.

Saper CB, Chou TC, Scammel TE. The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci* 2001;**24**:726–31.

Shan L, Dauvilliers Y, Siegel JM. Interactions of the histamine and hypocretin systems in CNS disorders. *Nat Rev Neurol* 2015;**1**(7):401–13.

Varin C, Rancillac A, Geoffroy H, et al. Glucose induces slow-wave sleep by exciting the sleep-promoting neurons in the ventrolateral preoptic nucleus: a new link between sleep and metabolism. *J Neurosci* 2015;**35**(27):9900–11.

A complete reference list can be found online at ExpertConsult.com.

Chapter
8

Rapid Eye Movement Sleep

Jerome M. Siegel

Chapter Highlights

- Rapid eye movement (REM) sleep was first identified by its most obvious feature: rapid eye movements occurring during sleep. In most adult mammals the electroencephalogram (EEG) of the neocortex is low in voltage during REM sleep. The hippocampus exhibits regular high-voltage theta waves throughout REM sleep.
- The key brain structure for generating REM sleep is the brainstem, particularly the pons and adjacent portions of the caudal midbrain. The isolated brainstem can generate REM sleep, including rapid eye movements, spike potentials linked to eye movements called ponto-geniculo-occipital (PGO) waves, muscle tone suppression (atonia), and autonomic variability. The structures rostral to the caudal midbrain–pontine brainstem cannot generate the forebrain aspects of REM sleep, such as PGO waves or rapid eye movements. The brainstem and the hypothalamus contain cells that are maximally active in REM sleep, called "REM-on cells," and cells that are minimally active in REM sleep, called "REM-off cells." Subgroups of REM-on cells each use a specific transmitter—gamma-aminobutyric acid (GABA), acetylcholine, glutamate, or glycine. Subgroups of REM-off cells use the transmitters norepinephrine, epinephrine, serotonin, histamine, and GABA.
- Destruction of large regions within the midbrain and pons can prevent the occurrence of REM sleep. More limited damage to portions of the brainstem can cause abnormalities in certain aspects of REM sleep. Of particular interest are manipulations that affect the regulation of muscle tone within REM sleep. Early animal work found that lesions of several regions in the pons and medulla can cause REM sleep to occur without the normal loss of muscle tone. In REM sleep without atonia, animals exhibit locomotor activity, appear to attack imaginary objects, and execute other motor programs during a state that otherwise resembles REM sleep. Subsequent work found a similar syndrome in humans that has been termed the REM sleep behavior disorder. Stimulation of portions of the REM sleep–controlling area of the pons can produce a loss of muscle tone in antigravity and respiratory musculature during waking, without eliciting all aspects of REM sleep.
- Narcolepsy is characterized by abnormalities in the regulation of REM sleep. Most cases of human narcolepsy are caused by a loss of hypocretin (orexin) neurons. Hypocretin neurons, which are located in the hypothalamus, contribute to the regulation of the activity of norepinephrine, serotonin, histamine, acetylcholine, glutamate, and GABA cell groups. Hypocretin neurons have potent effects on alertness and motor control and normally are activated in relation to particular, generally positive, emotions in humans as well as in animals.

OVERVIEW

Rapid eye movement (REM) sleep was discovered by Aserinsky and Kleitman in 1953.[1] These workers reported that REM sleep was characterized by the periodic recurrence of rapid eye movements, linked to a dramatic reduction in the amplitude of the electroencephalogram (EEG) signal from that of the higher-voltage activity of the previous non–rapid eye movement (NREM) sleep period. In their study, the EEG pattern in REM sleep closely resembled that in alert waking; and those subjects awakened from REM sleep reported vivid dreams. Dement identified a similar state of low-voltage EEG with eye movements in cats.[2] Jouvet repeated this observation, finding in addition a loss of muscle tone (i.e., atonia) in REM sleep. He used the term *paradoxical sleep* to refer to this state. The "paradox" was that the EEG resembled that recorded during waking, while behaviorally the animal remained asleep and unresponsive.[3-5] Subsequent authors have described this state as "activated" sleep or "dream" sleep. More recent work in humans has shown that some mental activity can be present in NREM sleep but has supported the original finding linking the most vivid dreaming to the REM sleep state. Lesions of parietal cortex and certain other regions prevent dreaming in humans, even in persons continuing to show normal REM

sleep as judged by cortical EEG activity and suppression of muscle tone and rapid eye movements.[6] Children younger than 6 years of age, in whom the amount of REM sleep is greater than in adults, do not typically report dream mentation, perhaps because the cortical regions involved have not yet fully developed.[7] The physiologic signs of REM sleep in both the platypus, the animal showing the most REM sleep,[8] and a related monotreme, the short-nosed echidna,[9] are largely restricted to the brainstem, in contrast with their propagation to the forebrain in adult placental and marsupial mammals. These findings make it questionable whether all or any nonhuman mammals that apparently experience REM sleep, all of which have cortical regions whose structure differs from that in adult humans, have dream mentation.

Surveyed in this chapter are (1) the defining characteristics of REM sleep, including its physiology and neurochemistry; (2) the techniques used to investigate the mechanisms generating REM sleep; (3) the mechanisms responsible for the suppression of muscle tone during REM sleep and the pathologic effects of disruption of these mechanisms; (4) narcolepsy and its link to mechanisms involved in REM sleep control and especially to the peptide hypocretin; and (5) the functions of REM sleep.

CHARACTERISTICS OF RAPID EYE MOVEMENT SLEEP

The principal electrical signs of REM sleep include a reduction in EEG amplitude, particularly in the power of its lower-frequency components (as shown in Figure 8-1, *top*). REM sleep also is characterized by a suppression of muscle tone, called *atonia*, visible in the electromyogram (EMG). In males, erections are more likely to occur.[10] Thermoregulation (e.g., sweating, shivering) largely ceases in most animals, and body temperatures drift toward environmental temperatures, as in reptiles.[11] Pupils constrict, reflecting a parasympathetic dominance in the control of the iris.[12] These changes, which are present throughout the REM sleep period, have been termed its "tonic" features.

Also visible are electrical potentials that can be most easily recorded in the lateral geniculate nucleus of the cat.[13] These potentials originate in the pons, appear after a few milliseconds in the lateral geniculate nucleus, and can be observed with further delay in the occipital cortex, leading to the designation *ponto-geniculo-occipital* (PGO) *spikes*. They occur as large-amplitude, isolated potentials arising 30 or more seconds before the onset of REM sleep as defined by EEG and EMG criteria. After REM sleep begins, they arrive in bursts of 3 to 10 waves, usually correlated with rapid eye movements. PGO-linked potentials also can be recorded in the motor nuclei of the extraocular muscles, where they trigger the rapid eye movements of REM sleep. They also are present in thalamic nuclei other than the geniculate and in neocortical regions other than the occipital cortex.

In humans, rapid eye movements are loosely correlated with contractions of the middle ear muscles of the sort that accompany speech generation and that are part of the protective response to loud noise.[14] Other muscles also contract during periods of rapid eye movement, briefly breaking through the muscle atonia of REM sleep. Periods of marked irregularity in respiratory and heart rates are characteristic of REM sleep, in contrast with NREM sleep, during which respiration and heart rate are regular. No single pacemaker for all of this irregular activity has been identified. Rather, the signals producing twitches of the peripheral or middle ear muscles may lead or follow PGO spikes and rapid eye movements. Bursts of brainstem neuronal activity may likewise lead or follow the activity of any particular recorded muscle.[15-17]

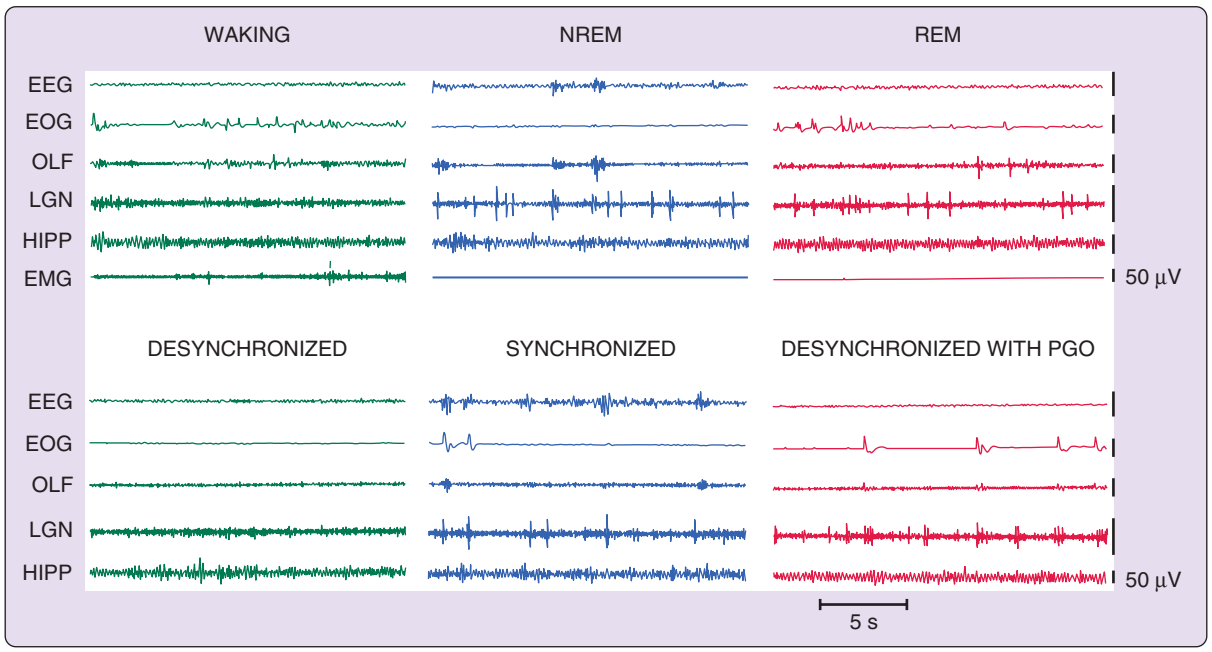


Figure 8-1 *Top*, Polygraph tracings of states seen in the intact cat. *Bottom*, Polygraph tracings of states seen in the forebrain 4 days after transection at the pontomedullary junction. EEG, Sensorimotor electroencephalogram; EMG, dorsal neck electromyogram; EOG, electrooculogram; HIPP, hippocampus; LGN, lateral geniculate nucleus; OLF, olfactory bulb; PGO, ponto-geniculo-occipital.

These changes that occur episodically in REM sleep have been called its "phasic" features.

As described further on, certain manipulations of the brainstem can eliminate only the phasic events of REM sleep, whereas others can cause the phasic events to occur in waking; yet other manipulations can affect tonic components. These tonic and phasic features also are expressed to various extents in different species, and not all of these features are present in all species that have been judged to have REM sleep.[18]

The distribution of REM sleep in the animal kingdom is discussed in Chapter 10.

RAPID EYE MOVEMENT GENERATION MECHANISMS

Technical Considerations

The identification of sleep-generating mechanisms can be achieved by *inactivation* or destruction of particular brain regions or neurons, by the *activation* of populations of neurons, or by *observation* of the activity of neurons or measurement of the release of neurotransmitters. Each approach has its advantages and limitations.

Inactivation of Neurons by Lesions, Inhibition, Antisense Administration, or Genetic Manipulation Including Optogenetic Inhibition

More has been learned about brain function and about sleep control from brain damage caused by stroke, injury, or infection in patients and by experimentally induced brain lesions in animals, than by any other technique. However, some basic principles need to be borne in mind when interpreting such data.

Brain lesions can result from ischemia, pressure, trauma, and degenerative or metabolic changes. In animals, experimental lesions most commonly are induced by aspiration, transection of the neuraxis, electrolysis, local heating by radio frequency currents, or the injection of cytotoxins. These substances include certain amino acids, such as *N*-methyl-D-aspartate (NMDA) and kainate, that cause cell death by excitotoxicity, and targeted cytotoxins, such as saporin coupled to a particular ligand, which will kill only cells containing receptors for that ligand. Cytotoxic techniques have the considerable advantage of sparing axons passing through the region of damage, so that deficits will be attributable to the loss of local neurons, rather than interruption of these axons. Injection of inhibitory neurotransmitters, such as muscimol, allows reversible inactivation of neurons in the injection region. DREADDs (designer receptors exclusively activated by designer drugs) also can be used to inactivate or activate groups of neurons. Viral vectors or transgenic mouse models can be used to express the receptors in the desired populations, which can then be manipulated by the locally or systemically applied "designer drug."

If damage to or inactivation of a brain region causes the loss of a sleep state, that region cannot be assumed to be where a "center" for the state resides. Lesion effects usually are maximal immediately after the lesion is created. Swelling and circulatory disruption make the functional loss larger than will be apparent from standard postmortem histologic techniques. The loss of one brain region also can disrupt functions that are organized elsewhere. For example, so-called spinal shock is a well-known phenomenon in which severing the spinal cord's connection to more rostral brain regions causes a loss of functions known to be mediated by circuits intrinsic to the spinal cord.

On the other hand, this sort of denervation-induced shock dissipates with the passage of time. In addition, adaptive changes occur that allow other regions to take over lost functions. This process is mediated by sprouting of new connections to compensate for the loss. A striking phenomenon seen after placement of lesions aimed at identifying the brain regions responsible for REM and NREM sleep is that with even massive lesions targeted at putative sleep-generating "centers," often only a transient disruption or reduction of sleep occurs, presumably as a result of this compensation.[19]

A particularly useful approach to the understanding of REM sleep generation has been the transection technique. In this approach, the brain is cut at the spinomedullary junction, at various brainstem levels, or at forebrain levels by passing a knife across the coronal plane of the neuraxis. Regions rostral to the cut may be left in situ or may be removed. Such a manipulation might be expected to completely prevent sleep phenomena from appearing on either side of this cut. To a surprising extent, however, this is not the case. As reviewed further on, REM sleep reappears within hours after introduction of some of these lesions. When both parts of the brain remain, signs of REM sleep usually appear on only one side of the cut. This kind of positive evidence is much more easily interpreted than loss of function after lesions, because undoubtedly the removed regions are not essential for the signs of REM sleep that survive.

It is increasingly possible to acquire mutant mice in which any one or several of more than 20,000 genes are inactivated. Investigation of two mutants[20,21] led to major insights into the etiology of human narcolepsy.[22-24] Techniques for the postnatal inactivation of genes permit investigation of gene deletions without the developmental effect of these deletions. They can also be used for investigation of the effects of gene inactivation within particular brain regions. A similar inactivation can be achieved by localized microinjections of antisense. Many if not most such mutants can be expected to have some sleep phenotype, such as increases or decreases in total sleep or REM sleep time, altered sleep rebound, altered responses of sleep to environmental variables, or altered changes in sleep with development and aging. The same interpretive constraints long appreciated in lesion studies apply to the interpretation of manipulations that inactivate genes or prevent gene expression, with the additional possibility of direct effects of genetic manipulation on tissues outside the brain.

Activation of Neurons by Electrical or Chemical Stimulation, Gene Activation, Insertion of Messenger RNAs, or Optogenetic Stimulation

Sites identified by lesion or anatomic studies can be stimulated to identify their roles in sleep control. Older studies used electrical stimulation and were successful in identifying the medial medulla as a region mediating the suppression of muscle tone[25] and basal forebrain as a site capable of triggering sleep.[26] Electrical stimulation is an obviously aphysiologic technique, involving the forced depolarization of neuronal membranes by ion flow at a frequency set by the stimulation device, rather than by the patterned afferent impulses that normally control neuronal discharge. For this reason, it has been supplanted for many purposes by administration of

neurotransmitter agonists, either by direct microinjection or by diffusion from a microdialysis membrane placed in the target area and perfused with high concentrations of agonists, and most recently by optogenetic activation.

One cannot assume that responses produced by such agonist administration demonstrate a normal role for the applied ligand. For example, many transmitter agonists and antagonists have been administered to the pontine regions thought to trigger REM sleep. In some cases this administration has increased REM sleep. The only permissible conclusion from this finding, however, is that cells in the region of infusion have receptors for the ligand and have connections to REM sleep–generating mechanisms. Under normal conditions these receptors may not have a role in triggering the state. Only by showing that the administration duplicates the normal pattern of release of the ligand in this area, and that blockade of the activated receptors prevents normal REM sleep, can a reasonable suspicion be raised that a part of the normal REM sleep control pathway has been identified.

Because it is far easier to inject a substance than to collect and quantify physiologically released ligands, many reported studies have implicated various substances as critical for REM sleep control solely on the basis of microinjection techniques. These results must be interpreted with caution. For example, hypocretin is known to depolarize virtually all neuronal types. It should therefore not be surprising to find that hypocretin microinjection into arousal systems such as the locus coeruleus produces arousal,[27] that microinjection of hypocretin into sites known to control feeding increases food intake,[28] that injection into regions known to contain cells that are waking-active increase waking,[29] that injection into regions known to contain cells selectively active in REM sleep will increase the occurrence of this state,[30,31] that injection into regions known to facilitate muscle tone will increase tone, that identical injections into regions known to suppress tone will decrease tone,[32] and that intracerebroventricular injection of hypocretin can increase water intake[33] and can activate other periventricular systems.[30] Such types of findings do not by themselves demonstrate a role for hypocretin (or any other neurotransmitter) in the observed behavior. It is necessary to obtain data on the effects of inactivation of, for example, hypocretin or hypocretin receptors and recording evidence that indicates activity of hypocretin neurons at the appropriate times before seriously entertaining such conclusions.

Genetic manipulations enable activation of neurons or nonneuronal cells of a particular type. A recent example of a genetic approach is the insertion of a light-sensitive ion channel into hypocretin cells using a lentivirus. Fiberoptic delivery of light could then be used to activate just these cells and determine the effect on sleep-waking transitions.[34] The interpretation of optogenetics responses is discussed elsewhere.[26]

Observation of Neuronal Activity

Recording the activity of single neurons in vivo can provide a powerful insight into the precise time course of neuronal discharge. Unit activity can be combined with other techniques to make it even more useful. For example, electrical stimulation of potential target areas can be used to antidromically identify the axonal projections of the recorded cell. Intracellular or "juxtacellular"[35] labeling of neurons with dyes, with subsequent immunolabeling of their transmitter, can be used to determine the neurotransmitter phenotype of the recorded cell. Combined dialysis and unit recording or iontophoresis of neurotransmitter from multiple barrel recording and stimulating micropipettes can be used to determine the transmitter response of the recorded cell, although whether the effects seen are the direct result of responses in the recorded cell or are mediated by adjacent cells projecting to the recorded cell cannot easily be determined. Such distinctions can be made in in vitro studies of slices of brain tissue by blocking synaptic transmission or by physically dissociating studied cells, but in this case their role in sleep may not be readily ascertained.

Although the role of a neuron in fast, synaptically mediated events happening in just a few milliseconds can be traced by inspection of neuronal discharge and comparison of that discharge with the timing of motor or sensory events, such an approach may be misleading when applied to the analysis of sleep state generation. The sleep cycle consists of a gradual coordinated change in EEG and EMG activity and other phenomena over a period of seconds to minutes, as the awake state turns into NREM sleep and then as NREM sleep is transformed into REM sleep.

Despite this mismatch of time courses, the *tonic latency,* a measure of how long before REM sleep onset the activity in a recorded cell changes, has been calculated in some studies. Neurons purported to show a "significant" change in activity many seconds or even minutes before REM sleep onset have been reported. Such a measure is of little utility, however, because at the neuronal level, the activity of key cell groups can best be seen as curvilinear over the sleep cycle, rather than changing abruptly in the way that activity follows discrete sensory stimulation. A major determinant of the tonic latency, calculated as defined earlier, is the level of "noise" or variability in the cell's discharge, which affects the difficulty of detecting a significant underlying change in rate in a cell population. It is therefore not surprising that cell groups initially designated as "executive neurons" for REM sleep control on the basis of their tonic latencies were later found to have no essential role in the generation of REM sleep.[36-38] The more appropriate method of assessment of the unit activity cycle relative to state control is to compare two different cell types to observe the specific phase relation of the peaks or troughs of their activity under similar conditions. This kind of study is difficult, involving the simultaneous long-term recording of multiple cells, and is rarely performed. Even in this case, a phase lead does not by itself prove that the "lead" neuron is driving activity seen in the "following" neuron, but it does indicate that the reverse is not the case. Awakening, however, is a process that can be studied in this way, because it can be elicited by stimuli and appears to be preceded by abrupt changes in the activity of many neuronal groups.[39] A major advantage of unit recording approaches in the intact animal to investigating sleep and other behavioral processes is their high level of temporal resolution.

Observation of the normal pattern of neurotransmitter release and neuronal activity can help determine the neurochemical correlates of sleep states. The natural release of neurotransmitters can be most easily determined by placing a tubular dialysis membrane 1 to 5 mm long in the area of interest and circulating artificial cerebrospinal fluid through it. Neurotransmitters released outside the membrane will diffuse through the membrane and then can be collected.

Each sample is obtained at intervals typically ranging from 2 to 10 minutes. The collected dialysates can be analyzed by chromatography, radioimmunoassay, mass spectroscopy, or other means. The temporal resolution of this technique typically is on the order of a few minutes for each sample.[40-42]

Unit recording and dialysis approaches require a sharp research focus on a particular neurotransmitter or neuronal group. By contrast, histologic approaches can be used to measure the activity of the entire brain at cellular levels of resolution. The most popular such approach in animal studies labels the activation of immediate early genes. These genes are expressed in the nucleus when a neuron is highly active, and their expression is an early step in the activation of other downstream genes mobilizing the response of the cell to activation. Activation of these genes can be detected by immunohistochemistry techniques, most commonly staining for the production of the Fos protein or the mRNA used to synthesize this protein.[43] Neurons can be double-labeled to identify the transmitter they express, allowing investigators to determine, for example, whether histaminergic neurons in the posterior hypothalamus were activated in a particular sleep or waking state. Metabolic labels such as 2-deoxyglucose also can provide an indication of which neurons are active.[43,44] Similar techniques using radioactive ligands in positron emission tomography studies can be used in living humans or animals. In vivo measurements of blood flow can be made throughout the brain with functional magnetic resonance imaging. All of these techniques have in common their ability to make anatomically driven discoveries of brain regions that are active in particular states, independent of specific hypotheses, thereby leading to major advances in understanding. However, another common feature of these types of "recording" techniques is their very poor temporal and spatial resolutions in comparison to neuronal recording approaches. Fos activation can take 20 minutes or more. Positron emission tomography takes a similar amount of time, and functional magnetic resonance imaging can observe events lasting on the order of 1 to 15 seconds. Accordingly, whether areas active during a particular state caused the state or were activated because of the state cannot be determined with certainty.

Summary of Technical Considerations

Clearly there is no perfect technique for determining the neuronal substrates of sleep states. Ideally, all three approaches should be used in concert to reach conclusions. Explored next are the major findings derived from lesion (transection), stimulation, and recording studies of REM sleep control mechanisms.

Transection Studies

The most radical types of lesion studies are those that slice through the brainstem, severing the connections between regions rostral and caudal to the cut. Sherrington discovered that animals in which the forebrain is removed after transection of the neuraxis in the coronal plane at the rostral border of the superior colliculus showed tonic excitation of the "antigravity muscles" or extensors (Figure 8-2, level A). This decerebrate rigidity was visible as soon as anesthesia was discontinued. Bard and Macht reported in 1958 that animals with decerebrate rigidity would show periodic limb relaxation.[45] It is now known that these researchers were observing the periodic muscle atonia of REM sleep.

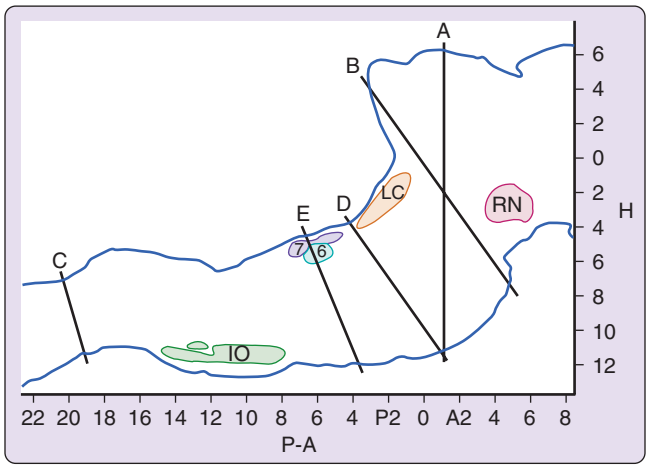


Figure 8-2 Outline of a sagittal section of the brainstem of the cat (level L =1.6) indicating the level of key brainstem transection studies. 6, Abducens nucleus; 7, genu of the facial nerve; H, horizontal (x-axis scale); IO, inferior olive; LC, locus coeruleus; P-A, posterior-anterior (y-axis scale); RN, red nucleus. (Modified from Berman AL. *The brain stem of the cat*. Madison [Wisc.]: University of Wisconsin Press; 1968.)

After the discovery of REM sleep in the cat,[2] Jouvet found that this state of EEG desynchrony normally was accompanied by muscle atonia.[4] Jouvet then examined the decerebrate cat preparation used by Sherrington and by Bard and Macht, with the addition of measures of muscle tone, eye movement, and EEG. One might have expected that REM sleep originates in the forebrain, but Jouvet found something quite different. When he recorded in the forebrain after separating the forebrain from the brainstem at the midbrain level (Figure 8-2, level A or B), he found no clear evidence of REM sleep. In the first few days after transection, the EEG in the forebrain always showed high-voltage activity, but when low-voltage activity appeared, the PGO spikes that help identify REM sleep in the intact animal were absent in traces from the thalamic structures, particularly the lateral geniculate, where they can be most easily recorded. Thus it appeared that activity in the isolated forebrain included slow wave sleep states and possibly waking, but with no clear evidence of REM sleep.

By contrast, the midbrain and brainstem behind the cut showed clear evidence of REM sleep. Muscle atonia appeared with a regular periodicity and duration, similar to that of the intact cat's REM sleep periods. This atonia was accompanied by PGO spikes with a morphology resembling that in the intact animal. The pupils were highly constricted during atonic periods, as in REM sleep in the intact cat.

An interesting feature of REM sleep in the decerebrate animal is that its frequency and duration varied with body temperature. In the decerebrate animal, the forebrain thermoregulatory mechanisms are disconnected from their brainstem effectors. Shivering and panting do not occur at the relatively small temperature shifts that trigger them in the intact animal. For this reason, if the body temperature is not maintained by external heating or cooling, it will tend to drift toward room temperature. Arnulf and colleagues[46] found that if body temperature was maintained at a normal level, little or no REM sleep was seen. But if temperature was allowed to fall, REM sleep amounts increased to levels well above those seen in the intact animal. This observation suggests that REM sleep

facilitatory mechanisms are on balance less impaired by reduced temperature than are REM sleep inhibitory mechanisms. Another way of looking at this phenomenon is that brainstem mechanisms are set to respond to low temperatures by triggering REM sleep, perhaps to stimulate the brainstem, and that high brainstem temperatures inhibit REM sleep. It is unclear whether this mechanism is operative in the intact animal, in which temperature shifts are within a much narrower range.

A further localization of the REM sleep control mechanisms can be achieved by examining the sleep of humans or animals in which the brainstem–spinal cord connection has been severed (Figure 8-2, level C). In this case, normal REM sleep in all its manifestations, except for spinally mediated atonia, is present.[47] These findings support the conclusion that the region between the caudal medulla and rostral midbrain is sufficient to generate REM sleep.

This approach can be continued by separating the caudal pons from the medulla (Figure 8-2, level D or E). In such animals no atonia is present in musculature controlled by the spinal cord, even though electrical or chemical stimulation of the medial medulla in the decerebrate animal suppresses muscle tone.[48] Furthermore, neuronal activity in the medulla does not resemble that seen across the REM-NREM sleep cycle, with neuronal discharge very regular for periods of many hours, in contrast with the periodic rate modulation that is linked to the phasic events of REM sleep in the intact animal[49] (Figure 8-3). This observation demonstrates that the medulla and spinal cord together, although they may contain circuitry whose activation can suppress muscle tone, are not sufficient to generate this aspect of REM sleep when disconnected from the pons and more rostral brainstem structures.

By contrast, the regions *rostral* to this cut show aspects of REM sleep[50] (Figure 8-4; see also Figure 8-1, *bottom*). Within these regions can be seen the progression from isolated to grouped PGO spikes and the accompanying reduction in PGO spike amplitude that occurs in the pre-REM sleep period and the REM sleep periods in the intact animal. Also evident is increased forebrain unit activity, with unit spike bursts in conjunction with PGO spikes, just as in REM sleep.[49,51]

To summarize, this work shows that when pontine regions are connected to the medulla, atonia, rapid eye movements, and the associated unit activity of REM sleep occur, whereas the medulla and spinal cord together, disconnected from the pons, are not sufficient to generate these aspects of REM sleep. When the pons is connected to the forebrain, forebrain aspects of REM sleep are seen, but the forebrain without attached pons does not generate these aspects of REM sleep. Further confirmation of the importance of the pons and caudal midbrain comes from the studies of Matsuzaki and associates.[52] These workers found that when two cuts were placed, one at the junction of the midbrain and pons and the other at the junction of the pons and medulla, periods of PGO spikes were seen in the isolated pons, but no signs of REM sleep were evident in structures rostral or caudal to the pontine "island."

These transection studies demonstrate, by positive evidence, that the pons is sufficient to generate the pontine signs of REM sleep—that is, the periodic pattern of PGO spikes and irregular neuronal activity that characterizes REM sleep. One can conclude that the pons is the crucial region for the generation of REM sleep. The structures within this region that synthesize the core elements of REM sleep are considered in greater detail further on.

Also clear, however, is that the pons alone does not generate all of the phenomena of REM sleep. Atonia requires the activation of motor inhibitory systems in the medulla.[53] In the intact animal, forebrain mechanisms interact with pontine mechanisms to regulate the amplitude and periodicity of PGO spikes,[54] which in turn are linked to the twitches and rapid eye movements of REM sleep. As documented in cases of human REM sleep behavior disorder, the motor activity expressed in dreams is linked to the imagery of the dream.[55]

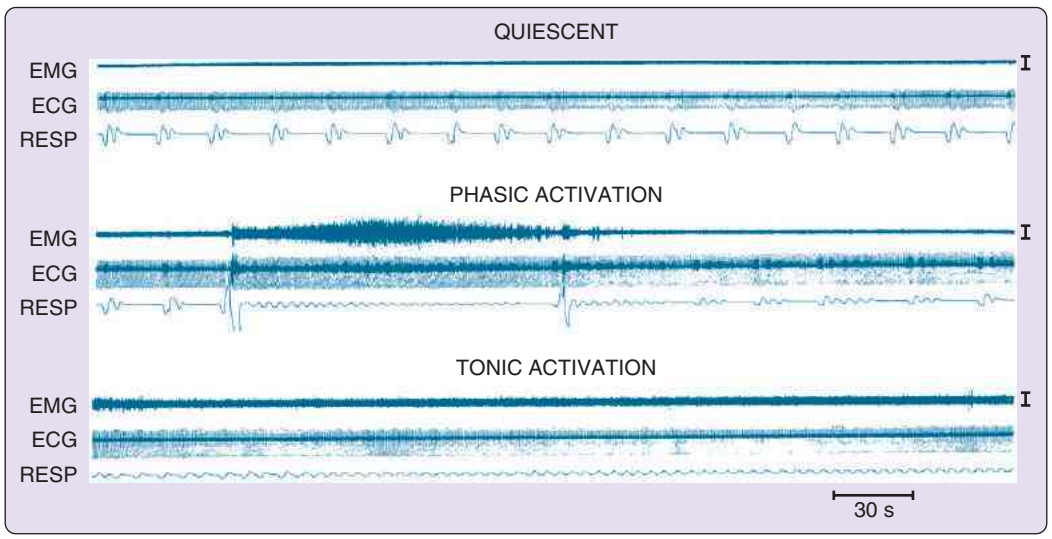


Figure 8-3 States seen caudal to chronic transection at the pontomedullary junction in the cat. Note the absence of periods of atonia. ECG, Electrocardiogram; EMG, electromyogram; RESP, thoracic strain gauge (i.e., respiratory movements). Calibration, 50 μV. (From Siegel JM, Tomaszewski KS, Nienhuis R. Behavioral states in the chronic medullary and mid-pontine cat. *Electroencephalogr Clin Neurophysiol* 1986;63:274-88.)

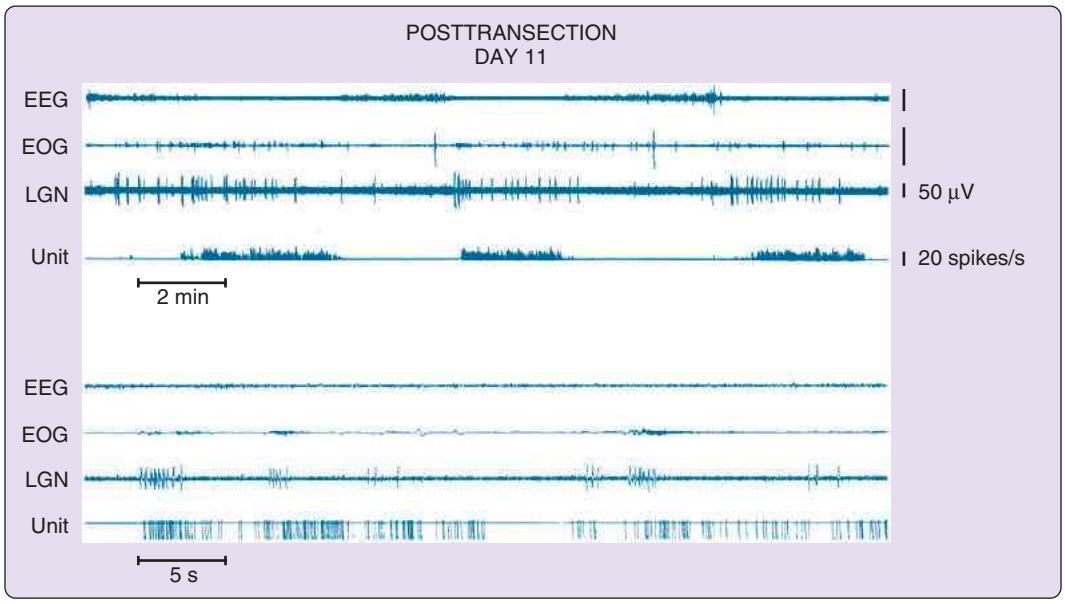


Figure 8-4 States seen rostral to chronic transection at the pontomedullary junction in the cat. Note the presence of ponto-geniculo-occipital (PGO) spikes and associated increases in unit activity triggered by the pons. Midbrain unit: electroencephalogram (EEG), electrooculogram (EOG), and lateral geniculate nucleus (LGN) activity rostral to chronic transections at the pontomedullary junction. *Top traces:* The unit channel displays the output of an integrating digital counter resetting at 1-sec intervals. *Bottom traces:* One pulse is produced for each spike by a window discriminator. (From Siegel JM. Pontomedullary interactions in the generation of REM sleep. In: McGinty DJ, Drucker-Colin R, Morrison A, et al, editors. *Brain mechanisms of sleep.* New York: Raven Press; 1985, p. 157-74.)

With extrapolation to dream imagery in normal humans, it can be hypothesized that because the structure of REM sleep results from an interaction of forebrain and brainstem mechanisms, the dream itself is not just passively driven from the brainstem but rather represents the result of a dynamic interaction between forebrain and brainstem structures.

Localized Lesion Studies

The transection studies point to a relatively small portion of the brainstem, the pons and caudal midbrain, as critical for REM sleep generation. Further specification of the core regions can be achieved by destroying portions of the pons in an otherwise intact animal and then seeing which areas are necessary and which are unnecessary for REM sleep generation. An early systematic study by Carli and Zanchetti in the cat[56] and other subsequent studies emphasized that lesions of locus coeruleus[57] and the dorsal raphe[58] nuclei or of simultaneous lesions of locus coeruleus, forebrain cholinergic neurons, and histamine neurons[19] did not block REM sleep. These investigators concluded that lesions that destroyed the region ventral to the locus coeruleus, called the *nucleus reticularis pontis oralis* or *subcoeruleus region*, produced a massive decrease in the amount of REM sleep. These studies used the electrolytic lesion technique, in which a current is passed through brainstem tissue, depositing metal that kills cells and axons of passage. As cytotoxic techniques that allowed poisoning of cell bodies without the damage to axons of passage came into use, these initial conclusions were confirmed and refined. It was shown that neurons in medial pontine regions including the giant cell region were not important in REM sleep control,[53,59,60] because near-total destruction of these cells was followed by normal amounts of REM sleep as soon as anesthesia dissipated.[37,61] However, lesions of the subcoeruleus and adjacent regions with cytotoxins were associated with a prolonged reduction in the amount of REM sleep. According to one study, the extent of this loss was proportional to the percentage of cholinergic cells lost in subcoeruleus and adjacent regions of the brainstem of the cat.[62] In rats, lesioning or inactivation of the same region below the locus coeruleus (called the sublaterodorsal nucleus in the terminology of Swanson[63]) has been found to reduce REM sleep.[64]

Although large lesions may lead to elimination of all aspects of REM sleep, introduction of small, bilaterally symmetric lesions within the pons can eliminate specific aspects of REM sleep. With lesions of lateral pontine structures, muscle atonia during REM sleep is seen. However, PGO spikes and the associated rapid eye movements are absent when lesions include the region surrounding the superior cerebellar peduncle of the cat[65] (Figure 8-5, *top*). This observation points to a role for this lateral region in the generation of PGO waves and the associated phasic activity of REM sleep.

Small lesions confined to portions of the subcoeruleus regions identified as critical for REM sleep by Carli and Zanchetti, or to the medial medulla,[53] result in a very unusual syndrome. After NREM sleep, affected animals enter REM sleep as indicated by lack of responsiveness to the environment, PGO spikes, EEG desynchrony, and pupil constriction. However, they lack the muscle atonia that normally characterizes this state[5,66] (Figure 8-5, *bottom*). During "REM sleep without atonia" these animals appear to act out dreams, attacking objects that are not visible, exhibiting unusual affective behaviors and ataxic locomotion. When they are awakened, normal behavior resumes. More recent studies have

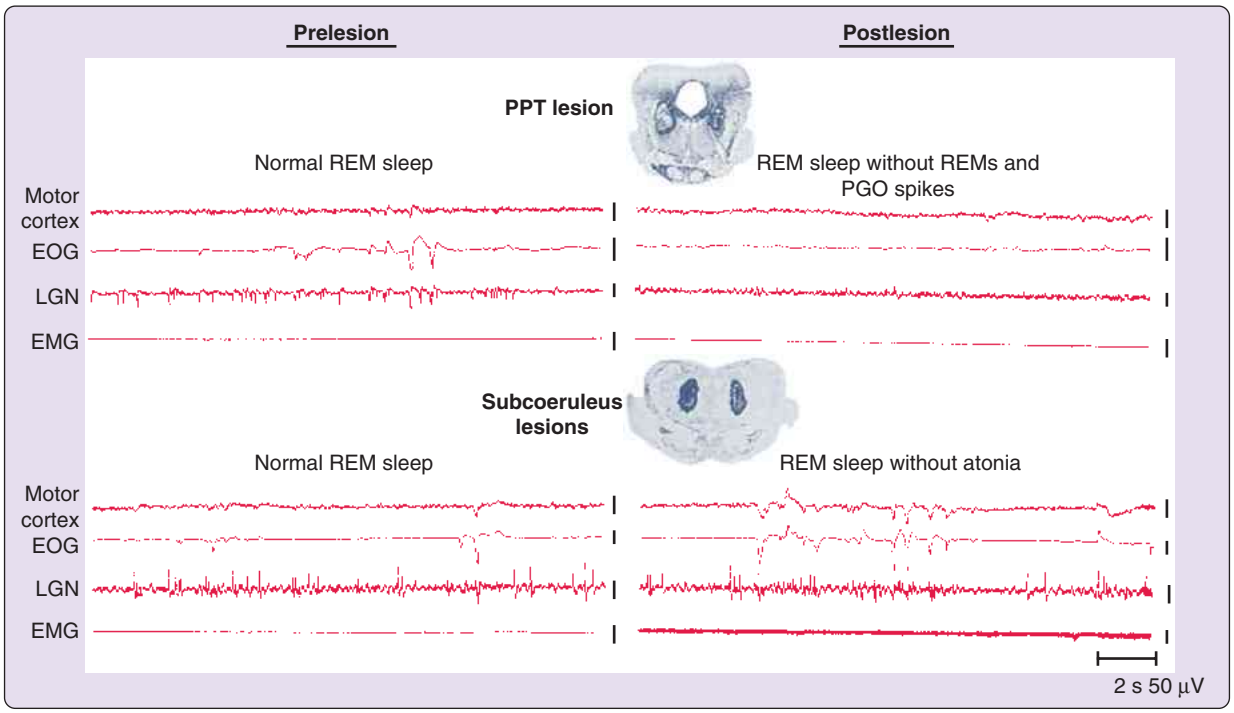


Figure 8-5 Disruption of phasic or tonic aspects of REM sleep by lesions. Twenty-second polygraph tracings during REM sleep before and after introduction of lesions, together with a coronal section through the center of the pontine lesions. Electroencephalographic voltage reduction of REM sleep (recorded from motor cortex) was associated with both lesion locations: pediculopontine tegmental (PPT) and subcoeruleus. *Top traces:* Radiofrequency lesions of the PPT region diminished ponto-geniculo-occipital (PGO) spikes and eye movement bursts during REM sleep. *Bottom traces:* Lesions in the region ventral to the locus coeruleus produced REM sleep without atonia without any diminution of PGO spike or REM frequency. (From Shouse MN, Siegel JM. Pontine regulation of REM sleep components in cats: integrity of the pedunculopontine tegmentum [PPT] is important for phasic events but unnecessary for atonia during REM sleep. *Brain Res* 1992;571:50-63.)

demonstrated that lesions of a system extending from the ventral midbrain to the medial medulla can cause REM sleep without atonia and that activation of this system can suppress muscle tone.[53,67-69]

This subcoeruleus region is under the control of midbrain regions. A midbrain region located just beneath and lateral to the periaqueductal gray (and called the dorsocaudal central tegmental field in the cat) appears to inhibit REM sleep by inhibiting the critical "REM on" subcoeruleus neurons. Muscimol, a $GABA_A$ receptor agonist, injected into this midbrain region silences these cells and increases REM sleep, presumably by blocking the inhibition.[70] The same phenomena have been observed when muscimol is injected into the corresponding region in the guinea pig[71] and the rat[72] (in the rat, this midbrain region has been called the deep mesencephalic nucleus.) The midbrain region of the deep mesencephalic nucleus is the heart of the classic reticular activating system, shown to induce waking when electrically stimulated[73] and coma when lesioned.[74]

Increasing the levels of GABA in the subcoeruleus region (also called the pontine oralis nucleus in the rat and cat) produces an increase in waking, rather than the increase in REM sleep seen with GABA injection into the midbrain regions as previously indicated.[75,76] This finding is another reminder that despite the sleep-inducing effect of systemic administration of GABAergic hypnotic medications (such as benzodiazepines), the effect of GABA on sleep and waking states induced by local manipulation varies across brain regions. Blocking GABA in the subcoeruleus has been reported to increase REM sleep in the cat.[77]

Stimulation Studies

The first study showing that stimulation could elicit REM sleep was carried out by George and colleagues.[78] These investigators found that application of the acetylcholine agonist carbachol to specific regions of the pons ventral to the locus coeruleus could elicit REM sleep in the cat. An impressive proof that a unique REM sleep generation mechanism was being activated was the long duration of the elicited REM sleep periods, which could last hours. Microinjection of acetylcholine into this region in the decerebrate cat produces an immediate suppression of decerebrate rigidity. Later studies showed that, depending on the exact site, either REM sleep or just atonia in a waking state could be triggered by such stimulation.[79-81] When stimulation was applied to the lateral regions with lesions blocking PGO waves, continuous PGO spikes were generated even though the animal was not always behaviorally asleep.

Increased REM sleep has been reported in the rat after microinjection of cholinergic agonists into the subcoeruleus region,[82-84] although this effect is certainly not as robust as it is in the cat.[85]

The first study demonstrating a role for glutamate in the control of REM sleep was done in the cat. The investigators found that a profound suppression of muscle tone could be elicited by the injection of glutamate into the subcoeruleus region or into the ventral medullary region.[48,86,87] Further work has demonstrated that the pontine cells in this inhibitory region receiving cholinergic input use glutamate as their transmitter and project directly to glutamate-responsive regions of the medial medulla.[86,88-90]

Work in the rat has emphasized the strong triggering of REM sleep by glutamatergic excitation of this region.[64,91] Glutamatergic excitation of this region in the cat also increases REM sleep,[92] suggesting that both cholinergic and glutamatergic mechanisms are intimately involved in the triggering of REM sleep, although the evidence points tospecies differences in the relative potency of the effect of microinjection of these two neurotransmitters.

Neuronal Activity, Transmitter Release

The transection, lesion, and stimulation studies all point to the same regions of the pons and caudal midbrain as the critical region for the generation of the REM sleep state as a whole, and to smaller subregions in the brainstem and forebrain in the control of its individual components. The pons contains a complex variety of cells differing in their neurotransmitter, receptors, and axonal projections. Unit recording techniques allow an analysis of the interplay between these cell groups and their targets to further refine dissection of REM sleep mechanisms.

Medial Brainstem Reticular Formation

Most cells within the medial brainstem reticular formation are maximally active in waking, greatly reduce discharge rate in NREM sleep, and increase discharge rate back to waking levels in REM sleep.[15,16,60,93,94] Discharge is most regular in NREM sleep and is relatively irregular in both waking and REM sleep. The similarity of the waking and REM sleep discharge patterns suggests similar roles for these cells in both states. Indeed, most of these cells have been shown to be active in waking in relation to specific lateralized movements of the head, neck, tongue, face, or limbs. For example, a cell may discharge only with extension of the ipsilateral forelimb or abduction of the tongue. The twitches that normally are visible in facial and limb musculature during REM sleep and the phenomenon of REM sleep without atonia suggest that these cells command movements that are blocked by the muscle tone suppression of REM sleep. Lesioning of these cells has little or no effect on REM sleep duration or periodicity[37,38] but does dramatically prevent movements of the head and neck in waking.[95]

Cholinergic Cell Groups

Cholinergic cell groups have an important role in REM sleep control in the cat. As was pointed out above, microinjection of cholinergic agonists into the pons of the cat reliably triggers long REM sleep periods that can last for minutes or hours. Microdialysis studies show that pontine acetylcholine release is greatly increased during natural REM sleep when compared with either NREM sleep or waking.[96] Recordings of neuronal activity within the cholinergic cell population demonstrate the substrates of this release. Certain cholinergic cells are maximally active in REM sleep (REM-on cells). Others are active in both waking and REM sleep.[97,98] Presumably the REM sleep-on cholinergic cells project to the acetylcholine responsive region in the subcoeruleus area.[99]

Cells with Activity Selective for REM Sleep

Cells with activity selective for REM sleep can be identified within the subcoeruleus area in both cats[100] and rats.[72] Anatomic studies using Fos labeling and tract tracing and unit recording studies indicate that these neurons are glutamatergic and GABAergic[98] and that some of them project to the ventral medullary region involved in the triggering of the muscle atonia of REM sleep.[48,64,72,86,88-90]

Monoamine-Containing Cells

Monoamine-containing cells have a very different discharge profile. Most if not all noradrenergic[101,102] and serotoninergic[103] cells of the midbrain and pontine brainstem and histaminergic[104] cells of the posterior hypothalamus are continuously active during waking, decrease their activity during NREM sleep, and further reduce or cease activity during REM sleep (Figure 8-6). As pointed out earlier, these cell groups are not critical for REM sleep generation, but it is likely that they modulate the expression of REM sleep. The cessation of discharge in monoaminergic cells during REM sleep appears to be caused by the release of GABA onto these cells,[105-108] presumably by REM sleep–active GABAergic brainstem neurons.[109,110] Administration of a GABA agonist to target the raphe cell group increases REM sleep duration,[106] demonstrating a modulatory role for this cell group in REM sleep control. Some studies indicate that dopamine cells do not change discharge across sleep states,[42,111,112] other work suggests an increased release of dopamine in REM sleep[113,114] or shows decreased release in REM sleep,[115] and still other work shows selective waking activity in these neurons.[116] These

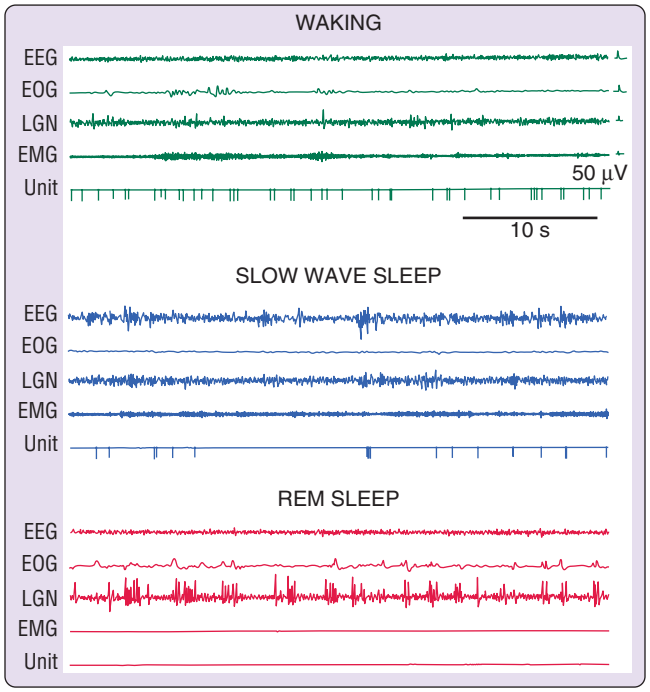


Figure 8-6 Activity of an "REM sleep–off" cell recorded in the locus coeruleus. (Slow wave sleep = NREM sleep.) EEG, Sensorimotor electroencephalogram; EMG, neck electromyogram; EOG, electrooculogram (eye movements); LGN, lateral geniculate activity; Unit, pulses triggered by the locus coeruleus cell.

findings may reflect heterogeneity of firing of different dopamine cell groups and presynaptic control of release in dopamine terminals.

Other Cholinergic Cells in Lateral Pontine Regions

Other cholinergic cells in lateral pontine regions discharge in bursts before each ipsilateral PGO wave.[117,118] These cells may therefore participate in the triggering of these waves. As shown in other studies, PGO waves are tonically inhibited in waking by serotonin input.[119-121] It is likely, therefore, that certain groups of cholinergic cells receive direct or perhaps indirect serotoninergic inhibition in waking and that the decrease of this inhibition in NREM sleep and REM sleep facilitates PGO wave and REM sleep generation.

Fos Labeling

A more global mapping of neurons active in REM sleep can be achieved by using the Fos labeling to identify neurons active within the 20-minute (or longer) period before sacrifice of the experimental animal. Quattrochi and associates demonstrated that microinjections of the cholinergic agonist carbachol that triggered episodes of continuous PGO waves in waking activated neurons within the laterodorsal and pedunculopontine nuclei. Destruction of these nuclei blocks these waves.[121-123]

More extensive Fos mapping has been done to identify neurons activated during REM sleep in the rat. Verret and colleagues[124] found that only a few cholinergic neurons from the laterodorsal and pedunculopontine tegmental nuclei were Fos-labeled after REM sleep. By contrast, a large number of noncholinergic Fos-labeled cells was observed in the laterodorsal tegmental nucleus, subcoeruleus region and lateral, ventrolateral, and dorsal periaqueductal gray of the midbrain. In addition, other regions outside of the brainstem regions critical for REM sleep control were labeled. These included the alpha and ventral gigantocellular reticular nuclei of the medulla, dorsal and lateral paragigantocellular reticular[125] nuclei, and the nucleus raphe obscurus. Half of the cells in the latter nucleus were cholinergic, suggesting that these neurons might be a source of acetylcholine during REM sleep. In a second study, an effort was made to identify the source of the GABAergic input thou ght to cause the cessation of discharge in locus coeruleus cells during REM sleep.[107] Verret and coworkers[87] also showed that the dorsal and lateral paragigantocellular reticular nuclei of the medulla and regions of the periaqueductal gray of the midbrain, regions with large percentages of GABAergic cells, are active in REM sleep. Maloney and associates[109] found GABAergic cells adjacent to the locus coeruleus that expressed Fos during periods of high REM sleep. Because the critical phenomena of REM sleep do not appear to require the medulla, it seems likely that the periaqueductal gray GABAergic neurons and GABAergic neurons adjacent to locus coeruleus and raphe nuclei are sufficient to suppress the activity of noradrenergic and serotoninergic neurons,[106,126] although medullary neurons may participate in the intact animal.

Fos mapping also has been used to identify forebrain regions likely to control REM sleep. The preoptic region, important in NREM sleep control (see Chapter 7), contains neurons that express Fos maximally in REM sleep–deprived animals, suggesting that these neurons may be related to the triggering or duration of REM sleep by brainstem systems.[127] Fos studies also indicate that melanin-concentrating hormone neurons, which are located in the hypothalamus, express Fos during periods with large amounts of REM sleep and that intracerebroventricular administration of melanin-concentrating hormone increases the amount of subsequent REM sleep.[128,129] These results suggest that melanin-concentrating hormone neurons also contribute to forebrain modulation of REM sleep.

Of note, the identity of the cells involved in triggering and controlling REM sleep is not easily determined. The Fos studies do not necessarily identify all cells active during REM sleep; only those of a phenotype that allows them to express Fos during the tested manipulations are so labeled. Certain cell types do not readily express Fos even when very active. In other words, cells not expressing Fos during periods of REM sleep may be involved and may even have a critical role in REM sleep control. Conversely, cells expressing Fos because of their activity during REM sleep may be responding to the motor and autonomic changes characteristic of this state, rather than causing these changes. With neuronal activity recording, identification of the cells responsible for starting the process of REM sleep triggering cannot be easily be achieved without a complete profile of discharge across the sleep cycle and a direct comparison of candidate cell groups, for the reasons just reviewed. Finally, recording from neurons in head-restrained animals, although easier than in freely moving animals, can be misleading, because such restraint can lower the activity of movement-related cells in waking, making them appear to be selectively active in REM sleep.[36] Nevertheless, by comparing the results of multiple recording and stimulation techniques with data on lesions, the evidence thus obtained helps identify the brainstem and forebrain neuronal groups that are the best candidates for controlling the REM sleep state.

CONTROL OF MUSCLE TONE

Abnormalities of muscle tone control underlie many sleep disorders. During REM sleep, central motor systems are highly active, whereas motoneurons are hyperpolarized.[130] The normal suppression of tone in the tongue and laryngeal muscles in REM sleep is a major contributing factor in sleep apnea (see Chapters 109, 110, and 113–118). The failure of muscle tone suppression in REM sleep causes REM sleep behavior disorder[131] (Chapter 103). Triggering of the REM sleep muscle tone control mechanism in waking is responsible for cataplexy.[132]

Early work using intracellular recording and microiontophoresis had shown that motoneuron hyperpolarization during REM sleep was accompanied by the release of glycine onto motoneurons.[130,133] Microdialysis sampling showed that both GABA and glycine are released onto motoneurons during atonia induced by carbachol in the cat.[41] This release occurs in ventral horn motoneurons as well as in hypoglossal motoneurons. The glycinergic inhibition during a carbachol elicited REM sleep–like state was investigated with immunohistochemistry and found to be due to the activation of glycinergic neurons in the nucleus reticularis gigantocellularis and nucleus magnocellularis in the rostroventral medulla and the ventral portion of the nucleus paramedianus reticularis,[133] regions whose activation has been shown to suppress muscle tone in the unanesthetized decerebrate animal.[86] A second population was located in the caudal medulla adjacent to the nucleus ambiguus; these neurons may be responsible for the

REM sleep–related inhibition of motoneurons that innervate the muscles of the larynx and pharynx.

In related work, it has been shown that norepinephrine and serotonin release onto motoneurons is decreased during atonia.[134] Because these monoamines are known to excite motoneurons and because GABA and glycine are known to inhibit motoneurons, it appears that the coordinated activity of these cell groups produces motoneuron hyperpolarization and hence atonia in REM sleep by a combination of inhibition and disfacilitation.

The inhibitory and facilitatory systems are strongly and reciprocally linked. Electrical stimulation of the pontine inhibitory area, located in the subcoeruleus region,[86] produces muscle tone suppression. Even though the pontine inhibitory area is situated within a few millimeters of the noradrenergic locus coeruleus, electrical stimulation in the pontine inhibitory area that suppresses muscle tone will always cause a *cessation* of activity in the noradrenergic neurons of the locus coeruleus and other facilitatory cell groups.[135] Cells that are maximally active in REM sleep ("REM-on" cells) are present in the pontine inhibitory area and also in the region of the medial medulla which receives pontine inhibitory area projections (Figure 8-7).

The release of GABA and glycine onto motoneurons during REM sleep atonia is most likely to be mediated by a pathway from the pontine inhibitory area to the medial medulla.[89,90] The pontine region triggering this release is not only sensitive to acetylcholine but also responds to glutamate[88] (Figure 8-8).[86] The medullary region with descending projections to motoneurons can be subdivided into a rostral portion responding to glutamate and a caudal portion responding to acetylcholine[48,136] (Figure 8-8). The medullary interaction with pontine structures evidently is critical for muscle tone suppression, because inactivation of pontine regions greatly reduces the suppressive effects of medullary stimulation on muscle tone.[137,138] This ascending pathway from the medulla to the pons may mediate the inhibition of locus coeruleus during atonia and also may help recruit other active inhibitory mechanisms. Thus damage anywhere in the medial pontomedullary region can block muscle atonia by interrupting ascending and descending portions of the pontomedullary inhibitory system, as can muscimol injection into the pons,[137]

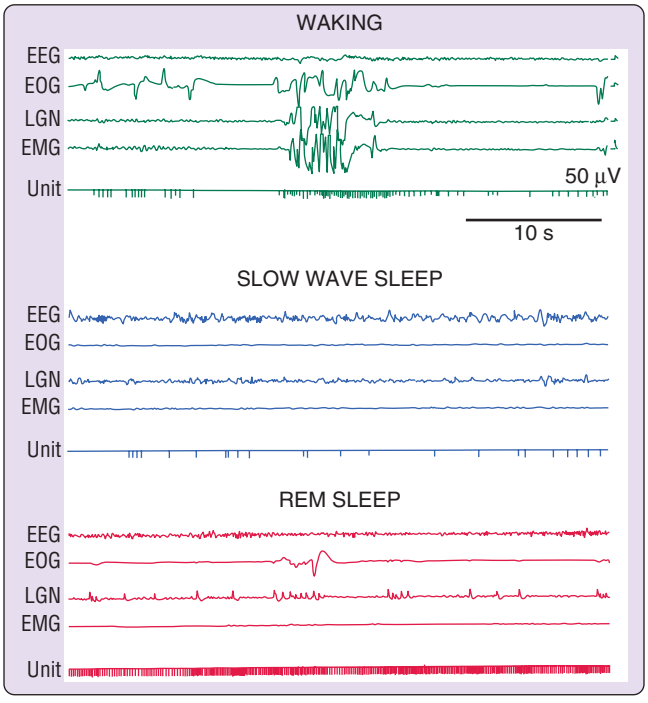


Figure 8-7 Activity of medullary "REM sleep–on" cell. Note the tonic activity during REM sleep. (Slow wave sleep = NREM sleep.) In waking, activity generally is absent even during vigorous movement. Some activity, however, is seen during movements involving head lowering and postural relaxation. EEG, Sensorimotor electroencephalogram; EMG, neck electromyogram; EOG, electrooculogram (eye movements); LGN, lateral geniculate activity; Unit, pulses triggered by the locus coeruleus cell.

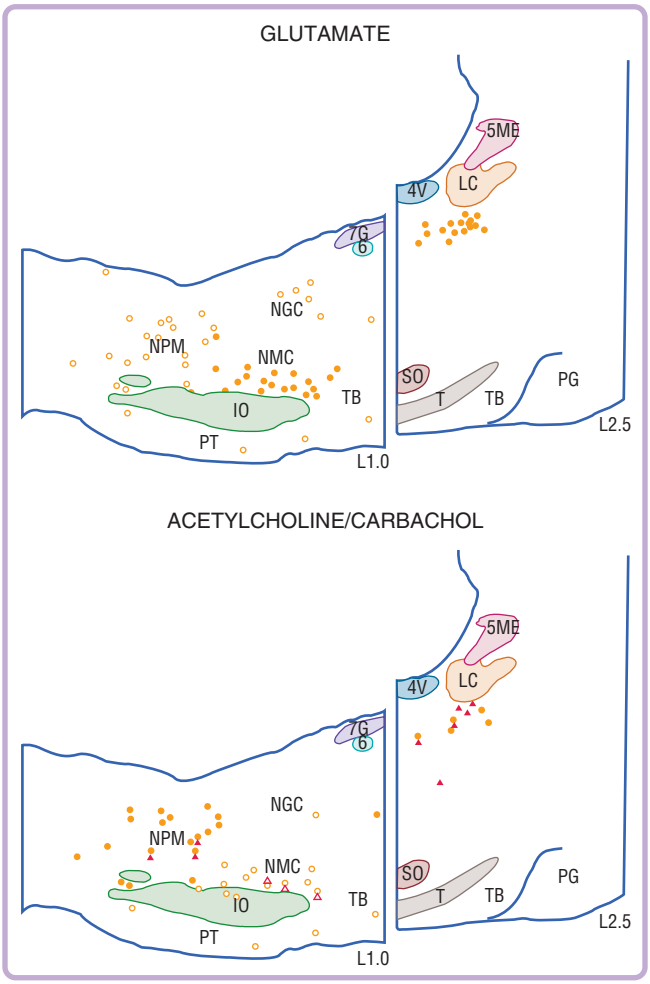


Figure 8-8 Sagittal map of pontomedullary inhibitory areas. Electrical stimulation produced atonia at all points mapped. All electrically defined inhibitory sites were microinjected with glutamate or cholinergic agonists. *Filled symbols* represent points at which microinjections decreased muscle tone (to less than 30% of baseline values or to complete atonia). *Open circles* indicate points at which injections increased or produced no change in baseline values. Data for glutamate injections are shown at the top; data for acetylcholine (ACh) and carbachol (Carb) injections at the bottom, with *circles* and *triangles* representing ACh and Carb injections, respectively. 4V, Fourth ventricle; 5ME, mesencephalic trigeminal tract; 6, abducens nucleus; 7G, genu of the facial nerve; IO, inferior olivary nucleus; LC, locus coeruleus nucleus; NGC, nucleus gigantocellularis; NMC, nucleus magnocellularis; NPM, nucleus paramedianus; PG, pontine gray; PT, pyramid tract; SO, superior olivary nucleus; T, nucleus of the trapezoid body; TB, trapezoid body. (From Lai YY, Siegel JM. Medullary regions mediating atonia. *J Neurosci* 1988;8:4790–6.)

again indicating that the pons is a key component of the circuit producing motor inhibition.

The studies just reviewed focused largely on ventral horn and hypoglossal motoneurons. However, the control of jaw muscles also is a critical clinical issue. The success of jaw appliances indicates that reduced jaw muscle activity can contribute to closure of the airway in sleep apnea (see Chapters 109, 110, and 113–118). Jaw muscle relaxation is a common initial sign of cataplexy, and tonic muscle activation underlies bruxism. Investigation of the control of masseter motor neurons allows analysis of the regulation of muscle tone on one side of the face, with use of the other side as a control for changes in behavioral state caused by application of neurotransmitter agonist and antagonists.[139] Using this model, it was determined that tonic glycine release reduces muscle tone in both waking and NREM sleep. Blockade of glycine receptors, however, did not prevent the suppression of muscle tone in REM sleep. In a similar manner, blockade of GABA receptors alone or in combination with glycine receptors increased tone in waking and NREM sleep but did not prevent the suppression of masseter tone[140] or of genioglossus tone in REM sleep.[141] Both of these manipulations, however, increased phasic masseter muscle activity in REM sleep.

Further studies showed that a blockade of glutamate receptors reduces the normal enhancement of muscle tone in waking relative to the level in NREM sleep. Glutamate also contributes to the phasic motor activity during REM sleep. Reduction in glutamate alone, however, is not sufficient to account for the suppression of muscle tone in REM sleep, because stimulation of NMDA and non-NMDA glutamate receptors does not appear to restore muscle tone in REM sleep.[142]

A study in the anesthetized rat suggested that activation of norepinephrine receptors, in combination with the activation of glutamate receptors, was sufficient to potently increase muscle tone in the masseter muscles.[143] A study of the hypoglossal motor nucleus in the unanesthetized rat concluded that the suppression of muscle tone in REM sleep was mediated to a large extent by a reduction in norepinephrine release, but not by reduced serotonin release.[144] In the context of previous microdialysis analysis of transmitter release, these studies suggest that the reduction of norepinephrine release may be a key factor regulating muscle tone, along with the aforementioned changes in amino acid release. These conclusions are consistent with earlier work indicating that cataplexy was linked to a reduction in the activity of noradrenergic neurons (see further on).[145] Although the current literature suggests that trigeminal, hypoglossal, and ventral horn motoneurons are subjected to similar neurochemical control across the sleep cycle, direct comparison of these systems has not been made, and it is likely that some aspects of control may differ across systems as well as species.

The role of reduced serotonin release in the suppression of muscle tone has been investigated in the hypoglossal nucleus of the rat. It was found that the modulation of genioglossus activity across natural sleep-wake states was not greatly affected by endogenous input from serotoninergic neurons, although earlier studies in vagotomized and anesthetized rats had shown an effect of serotonin on muscle tone under these aphysiologic conditions.[146-148]

Subsequent work suggested that inhibition of motor output is accompanied by a neurochemically similar inhibition of sensory relays during REM sleep.[149] Such sensory inhibition may be important in preserving sleep and, in particular, in blocking the sensory input produced by twitches breaking through the motor inhibition of REM sleep. The failure of this inhibition may contribute to sleep disruption and increased motor activity in sleep in pathologic states.

In contrast with the norepinephrine, serotonin, and histamine cell groups, it was reported that mesencephalic dopaminergic neurons do not appear to alter their discharge rate across the sleep cycle.[111] Dopamine release in the amygdala measured by dialysis does not significantly vary across the sleep cycle.[150] In disagreement with this finding, a Fos study indicted that dopaminergic neurons within the ventral portion of the mesencephalic tegmentum were activated during periods of increased REM sleep.[151] A unit recording study indicated that dopaminergic neurons in the ventral tegmental area of the midbrain show maximal burst firing in both waking and REM sleep.[113] Other work using the Fos labeling technique identified a wake-active dopaminergic cell population in the ventral periaqueductal gray.[116] In dialysis measurements of dopamine release, dopamine release was reduced in the dorsal horn of the spinal cord during the REM sleep–like state triggered by carbachol. Such a decrease was not seen in the ventral horn or hypoglossal nucleus.[134] These data suggest either heterogeneity in the behavior of sleep cycle activity of dopaminergic neurons or presynaptic control of dopamine release independent of action potentials in the cell somas.

Figure 8-9 illustrates some of the anatomic and neurochemical substrates of the brainstem generation of REM sleep.

NARCOLEPSY AND HYPOCRETIN

Narcolepsy has long been characterized as a disease of the REM sleep mechanism. Narcoleptic patients often enter REM sleep within 5 minutes of sleep onset, in contrast with normal persons, who rarely show such "sleep-onset REM sleep." Most narcoleptics experience cataplexy,[152] a sudden loss of muscle tone with the same reflex suppression that is seen in REM sleep. High-amplitude theta activity in the hippocampus, characteristic of REM sleep, is also prominent in cataplexy as observed in dogs.[145] Further evidence for links between narcolepsy and REM sleep comes from studies of neuronal activity during cataplexy. Many of the same cell populations in the pons and medulla that are tonically active only during REM sleep in normals become active during cataplexy in narcoleptics, including cells in the medial medullary inhibitory region that are selectively active in relation to the atonia of REM sleep.[17,132] Likewise, cells in the locus coeruleus, which cease discharge only in REM sleep in normal animals, invariably cease discharge in cataplexy.[153] However, just as cataplexy differs behaviorally from REM sleep in its maintenance of consciousness, not all neuronal aspects of REM sleep are present during cataplexy. As noted previously, in the normal animal, noradrenergic, serotoninergic, and histaminergic cells are all tonically active in waking, reduce discharge in NREM sleep, and cease discharge in REM sleep.[145,153] Unlike noradrenergic cells, however, serotoninergic cells do not cease discharge during cataplexy, only reducing discharge to quiet waking levels. Histaminergic cells actually increase discharge in cataplexy relative to quiet waking levels[154] (Figure 8-10). These findings allow identification of some of

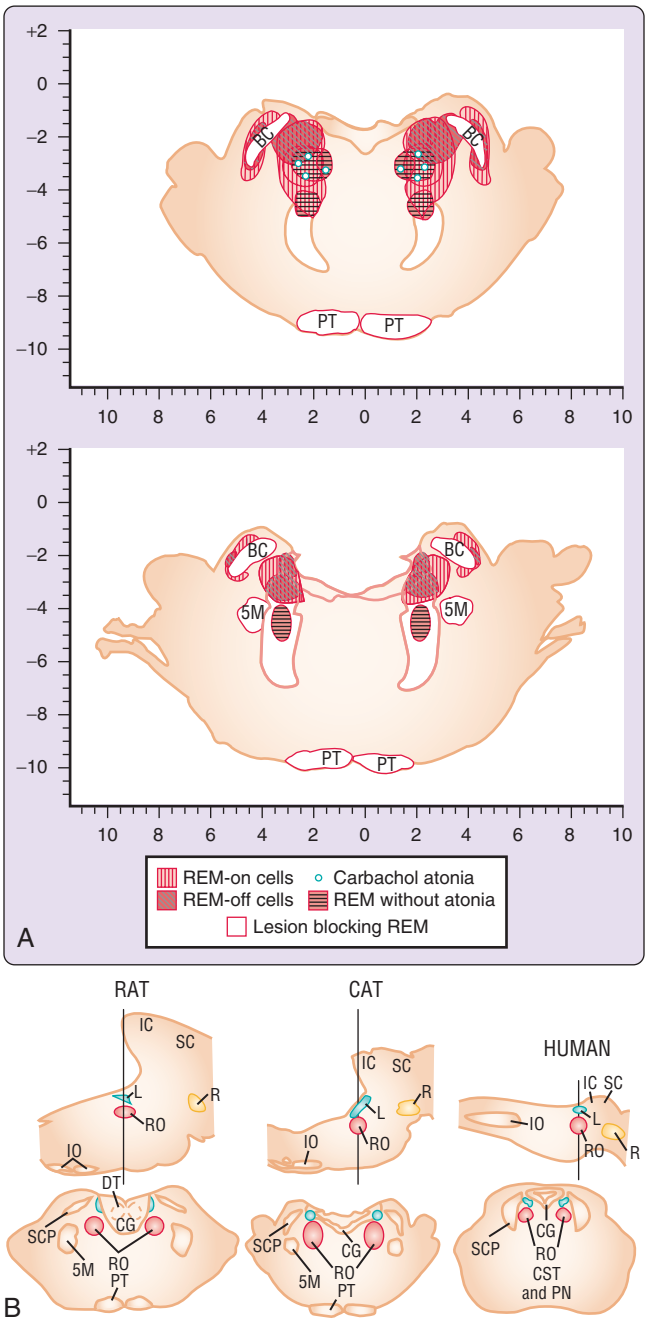


Figure 8-9 **A** and **B,** Anatomic relation of "REM sleep–on" and "REM sleep–off" cells, carbachol-induced atonia sites, lesions blocking atonia but not preventing REM sleep, and lesions completely blocking REM sleep. **B** shows anatomic locations of REM on areas in cat and rat brains and projected location in the human in sagittal and coronal views. 5M, Motor nucleus of the trigeminal nerve; BC, brachium conjunctivum; CG, central gray; CST, corticospinal tract; DT, dorsal tegmental; IC, inferior colliculus; IO, inferior olive; L, locus coeruleus; PN, pontine nuclei, PT, pyramidal tract; R, red nucleus; RO, reticularis oralis nucleus; SC, superior colliculus; SCP, superior cerebellar peduncle (brachium conjunctivum). (**A,** From Siegel JM, Rogawski MA. A function for REM sleep: regulation of noradrenergic receptor sensitivity. *Brain Res* 1988;13:213-33. **B,** From Siegel JM. The stuff dreams are made of: anatomical substrates of REM sleep. *Nat Neurosci* 2006;9:721–2.)

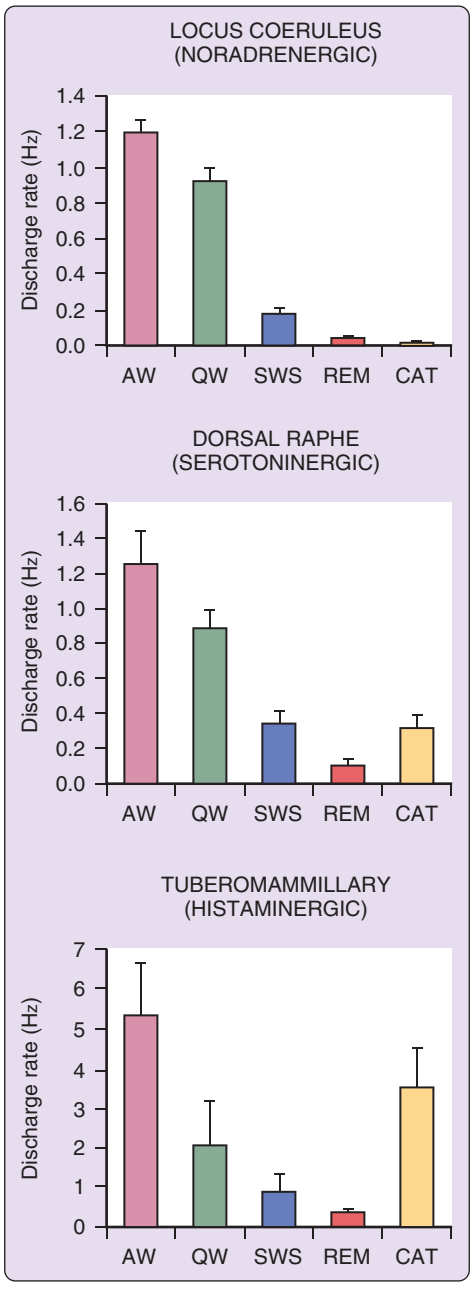


Figure 8-10 Comparison of mean discharge rates in sleep-waking states and cataplexy for REM–off cells recorded from three brain regions. Posterior hypothalamic histaminergic neurons remain active, whereas dorsal raphe serotoninergic neurons reduced discharge, and locus coeruleus noradrenergic neurons ceased discharge during cataplexy. All of these cell types were active in waking, reduced discharge in NREM sleep, and were silent or nearly silent in REM sleep. AW, Active waking; CAT, cataplexy; QW, quiet waking; REM, REM sleep; SWS, slow wave (NREM) sleep. (From John J, Wu MF, Boehmer LB, SiegelJM. Cataplexy-active neurons in the posterior hypothalamus: implications for the role of histamine in sleep and waking behavior. *Neuron* 2004;42:619–34.).

the cellular substrates of cataplexy. Medullary inhibition and noradrenergic disfacilitation are linked to cataplexy's loss of muscle tone. By contrast, the maintained activity of histamine neurons is a likely substrate for the maintenance of consciousness during cataplexy that distinguishes cataplexy from REM sleep. Thus the study of neuronal activity in the narcoleptic animal provides insight into both narcolepsy and the normal role of these cell groups in maintaining consciousness and muscle tone.

In 2001 it was discovered that most human narcolepsy was caused by a loss of hypothalamic cells containing the peptide hypocretin[23,24] (Figure 8-11). On average, 90% of these cells are lost in narcolepsy. Subsequently it was discovered that a lesser reduction in the number of hypocretin cells was seen in Parkinson disease, with a loss of up to 60% of hypocretin cells.[155,156] It was found that administration of the peptide to genetically narcoleptic dogs reversed symptoms of the disorder[157] and that nasal administration reversed sleepiness in monkeys,[158] suggesting that similar treatment could be uniquely effective for narcolepsy and perhaps for other disorders characterized by sleepiness.[159-161] Recently it also has been found that in human narcoleptics, the number of detectable histamine cells is increased more than 65%.[162,163] It has been speculated that because this change is not seen in any of four different animal genetic models of narcolepsy, the increase may be related to the presumed immune activation that causes human narcolepsy.[162]

In further work in normal animals, it was determined that identified hypocretin neurons discharge at their highest rates during active waking[35,164] (Figure 8-12). This discharge was reduced or absent during aversive waking situations, even if the EEG indicated high levels of alertness.[35] The hypocretin level in normal dogs is nearly doubled when they are let out into a yard to play with other dogs. By contrast, when these same dogs run at maximal speed on a treadmill, hypocretin levels are unchanged, demonstrating that motor activity and associated changes in respiratory rate, heart rate, and body temperature do not by themselves determine the release of hypocretin. Findings in studies of hypocretin release in the cat[165] also are consistent with this hypothesis. Hypocretin cells send ascending projections to cortical and basal forebrain regions, in addition to their descending projection to locus coeruleus and other brainstem regions. In the absence of hypocretin-mediated facilitation of forebrain arousal centers, waking periods are truncated, resulting in the sleepiness of narcolepsy.[166]

The functions of hypocretin have been investigated in genetic knockout animals lacking the peptide and in their wild-type littermates, using operant reinforcement tasks. Hypocretin-knockout mice are deficient in the performance of bar presses to secure food or water reinforcement. However, they do not differ from their normal littermates in their performance when trained to bar press to avoid foot shock. Periods of poor performance on the positive reinforcement tasks are characterized by EEG deactivation.[167] This deficit is restricted to the light phase, suggesting that hypocretin neurons mediate the arousing and mood-elevating effects of light,[167] effects that are central to the current understanding of depression. Fos labeling studies in normal littermates showed that the positive reinforcement task used in this study is characterized by activation of hypocretin neurons. However, hypocretin neurons are not activated in the negative reinforcement task or during the same positively motivated task in the dark phase, despite high levels of EEG activation, indicating that nonhypocretin systems mediate arousal during these behaviors.

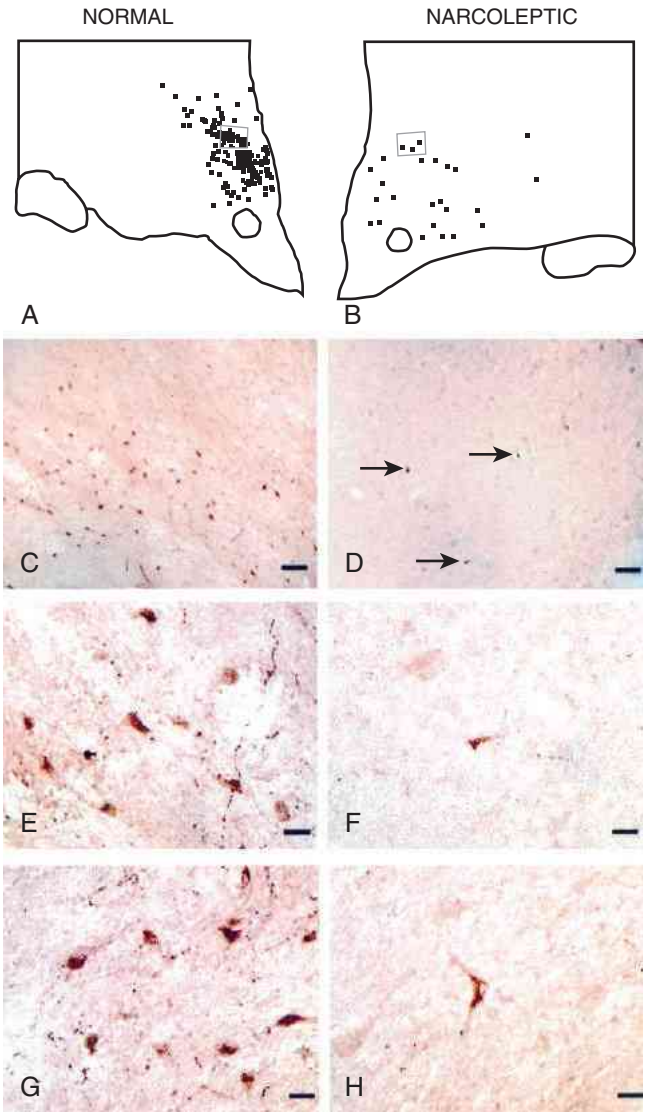


Figure 8-11 Loss of hypocretin cells in human narcolepsy. Distribution of cells in perifornical and dorsomedial hypothalamic regions of normal (**A, C, E, G**) and narcoleptic (**B, D, F, H**) humans. (From Thannickal TC, Moore RY, Nienhuis R, et al. Reduced number of hypocretin neurons in human narcolepsy. *Neuron* 2000;27:469-74.)

The conclusions of these animal studies were extended in the first study of hypocretin release within the human brain. Hypocretin levels were shown to be maximal during positive emotion, social interaction, and anger, behaviors that induce cataplexy in human narcoleptics. This finding is consistent with the hypothesis that release of hypocretin facilitates motor activity during emotionally charged activities of the sort that trigger cataplexy in narcoleptics.[166,168,169] Even normal persons experience weakness at these times, seen in the "doubling over" that often accompanies laughter or the weakness that can result from other strong emotions of sudden onset. In the absence of the hypocretin-mediated motor facilitation of locus coeruleus and other brainstem regions, muscle tone is lost at

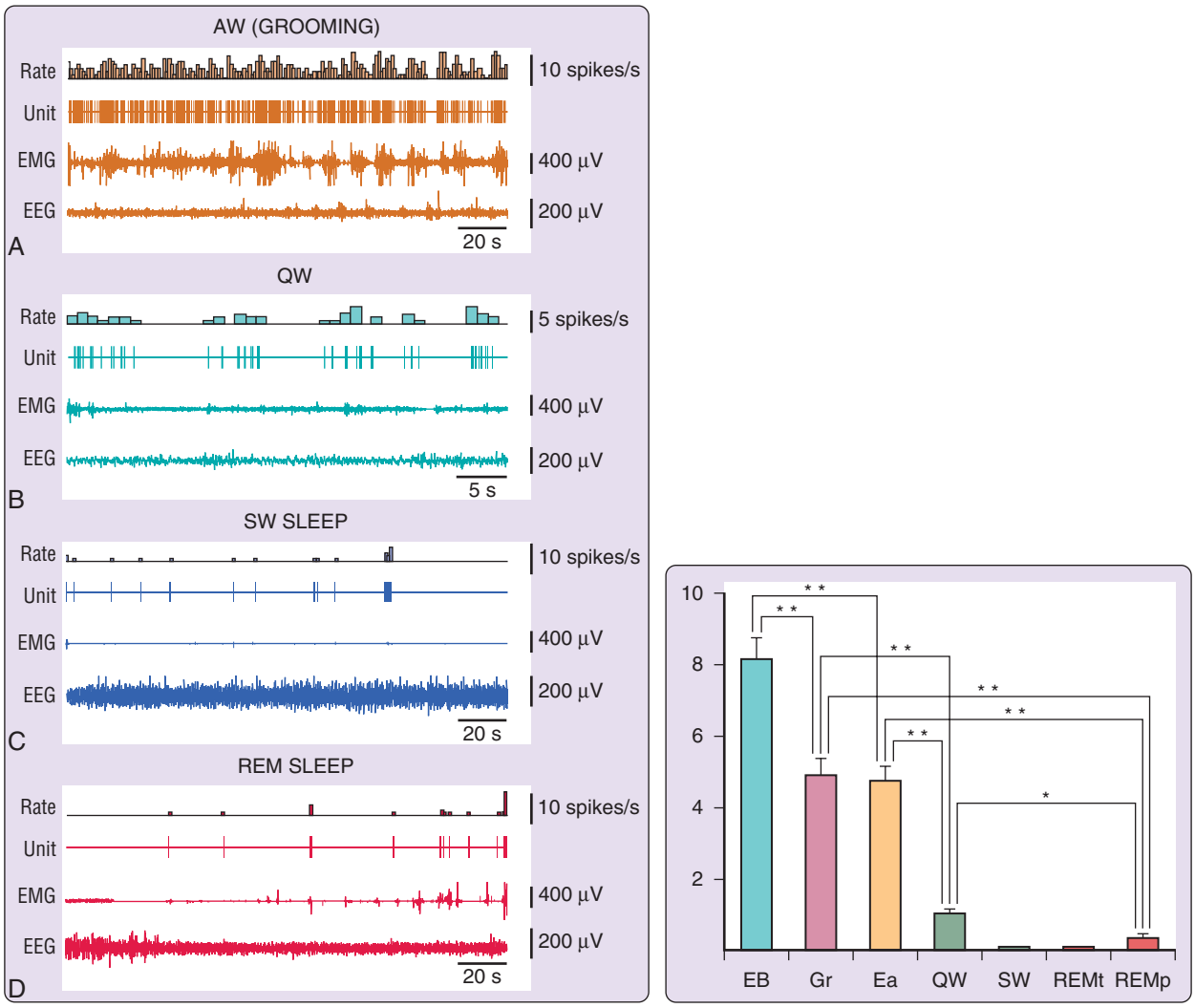


Figure 8-12 Firing rate of hypocretin cells in waking and sleep behaviors in freely moving rats. *Left:* The discharge pattern of a representative hypocretin neuron across the sleep-waking cycle in the freely moving rat. **A,** High firing rates are seen during active waking (AW) with grooming. **B,** Reduced firing rate or cessation of activity is seen in quiet waking (QW) and drowsiness. **C,** A further decrease or cessation of firing is seen during slow wave (NREM) sleep. **D,** Minimal firing rate is seen during the tonic phase of REM sleep. Brief hypocretin cell discharge bursts are correlated with muscle twitches during the phasic events of REM sleep. *Right:* Summary data from identified hypocretin cells: exploratory behavior (EB), grooming (Gr), eating (Ea), QW, SW sleep (SW), and tonic (REMt) and phasic (REMp) sleep. Maximal discharge is seen during exploration-approach behavior. (From Mileykovskiy BY, Kiyashchenko LI, Siegel JM. Behavioral correlates of activity in identified hypocretin (orexin) neurons. *Neuron* 2005;46:787-98.)

these times. By contrast, the release in humans of melanin-concentrating hormone, a peptide produced by neurons intermixed in the hypothalamus with the hypocretin neurons, is minimal during social interaction but is increased after eating. Both peptides are at minimal levels during periods of postoperative pain despite high levels of arousal. Melanin-concentrating hormone levels increase at sleep onset, consistent with a role in sleep induction,[170] whereas hypocretin-1 levels increase at wake onset, consistent with a role in wake induction. Levels of these two peptides in humans are not simply linked to arousal but rather are correlated with specific emotions and state transitions[171] (Figure 8-13).

The findings that hypocretin is released and hypocretin neurons are active only during arousal linked to certain emotions suggests a new approach to the understanding of arousal systems. Hypocretin is clearly related to arousal linked to certain generally positive emotions. Other arousal systems must mediate arousal during aversive situations. An analysis of the differential activation of arousal systems as a function of emotion, light level, and other variables might provide important clinical and basic science insights into the unique roles of each arousal system.

Hypocretin appears to act largely by modulating the release of amino acid neurotransmitters.[172] Systemic injection of hypocretin causes a release of glutamate in certain hypocretin-innervated regions, producing a potent postsynaptic excitation.[139,173] In other regions it facilitates GABA release, producing postsynaptic inhibition.[165,174] The loss of these competing inhibitory and facilitatory influences in narcolepsy appears to leave brain motor regulatory and arousal systems

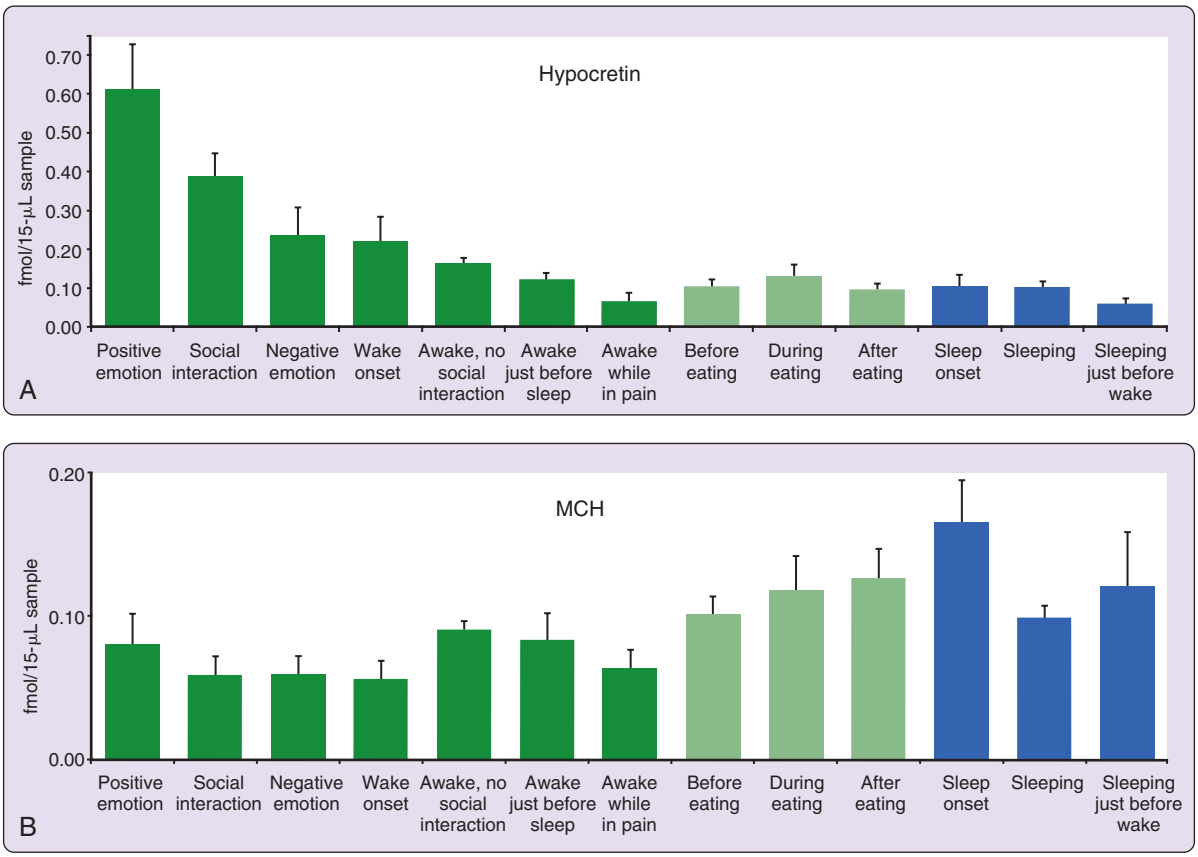


Figure 8-13 Hypocretin and melanin-concentrating hormone (MCH) levels across waking and sleep activities in humans. **A,** Maximal hypocretin levels in waking are seen during positive emotions and social interactions and on awakening; minimal levels are seen before sleep and in alert waking associated with reported pain. Changes during and after eating are smaller than those during monitored non–eating-related activities. Waking values are shown in *shades of green*; sleep values, in *blue*. For all awake samples, subjects were awake but were not exhibiting social interaction or reporting emotion. **B,** Maximal MCH levels are seen at sleep onset and after eating. Minimal levels are seen during wake onset, social interaction, and pain. Error bars represent ±S.E.M. (From Blouin AM, Fried I, Wilson CL, et al. Human hypocretin and melanin-concentrating hormone levels are linked to emotion and social interaction. *Nat Commun* 2013;4:1547.)

less stable than the tightly regulated balance that can be maintained in the presence of hypocretin (Figure 8-14). According to this hypothesis, this loss of stability is the underlying cause of narcolepsy, with the result being inappropriate loss of muscle tone in waking and inappropriate increases in muscle tone during sleep, resulting in a striking *increased* incidence of REM sleep behavior disorder in narcoleptics (see Chapters 89, 90, and 103). In the same manner, although a principal symptom of narcolepsy is intrusions of sleep into the waking period, narcoleptic persons sleep poorly at night, with frequent awakenings.[175-177] In other words, narcoleptics are not simply weaker and sleepier than normal subjects. Rather, their muscle tone and sleep-waking state regulation is less stable than in normal persons as a result of the loss of hypocretin function.

THE FUNCTIONS OF RAPID EYE MOVEMENT SLEEP

Research into the control of REM sleep turns into a seemingly infinite regression, with REM-on cells inhibited by REM-off cells, which in turn may be inhibited by other REM-on cells. It is very difficult to identify the sequence in which these cell groups normally are activated because the axonal condition and synaptic delays could not be more than a few milliseconds between these cell groups, yet REM sleep onset occurs over a period of minutes in the human and cat and at least 30 or more seconds in the rat. It also does not completely enlighten researchers with respect to the ultimate functional question: What is REM sleep for? To answer this question requires determining what if any physiologic process is altered over REM sleep periods. Is some toxin excreted or some protein synthesized? If so, how can the widely varying durations of the typical REM sleep be accounted for? In the human, REM sleep typically lasts from 5 to 30 minutes, whereas in the mouse, it typically lasts 90 seconds.[178] What can be accomplished in 90 seconds in the mouse but requires an average of approximately 15 minutes in humans? If a vital process is accomplished, why do drug treatments that abolish REM sleep have no discernable effect on any vital process, even when such drugs are taken continuously for many years? The biologic need that initiates REM sleep remains unknown, as well as the source of the REM sleep "debt" that accumulates during REM sleep deprivation.[179] Why do some marine mammals have no apparent REM sleep (see Chapter 10)?

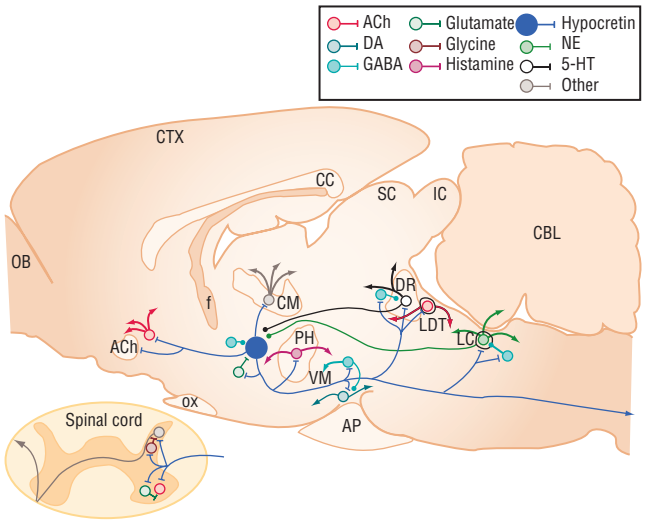


Figure 8-14 Major identified synaptic interactions of hypocretin neurons. Lines terminated by *perpendicular* lines denote excitation; *circular* terminations indicate inhibition. Acb, Nucleus accumbens; ACh, acetylcholine; AP, anterior pituitary; CBL, cerebellum; CC, corpus callosum; CM, centromedian nucleus of the thalamus; CTX, cortex; DA, dopamine; DR, dorsal raphe; f, fornix; 5-HT, 5-hydroxytryptamine (serotonin); IC, inferior colliculus; LC, locus coeruleus; LDT, laterodorsal tegmental; NE, norepinephrine; OB, olfactory bulb; OX, optic chiasm; PH, posterior hypothalamus; SC, superior colliculus; VM, ventral midbrain.

Why is REM sleep present in homeotherms (i.e., birds and mammals) but apparently absent in the reptilian ancestors of homeotherms?

Great progress has been made in localizing the mechanisms that generate REM sleep. As described previously, many of the key neurotransmitters and neurons involved have been identified. The discovery of the role of hypocretin in narcolepsy serves as a reminder that key cell groups may still need to be identified before fundamental insights into the generation mechanism and functions of REM sleep can be gained. Yet despite this caveat, a substantial amount of information about what goes on in the brain during REM sleep has already been accumulated.

Clearly, increased brain activity in REM sleep consumes considerable amounts of metabolic energy. The intense neuronal activity shown by most brain neurons, similar to or even more intense than that seen during waking, exacts a price in terms of energy consumption and "wear and tear" on the brain. Such a state would be unlikely to have produced a Darwinian advantage and remained so ubiquitous among mammals if it did not have benefits compensating for its obvious costs. But what might these benefits be?

One idea that has received much media attention is that REM sleep has an important role in memory consolidation. However, the evidence for such a role is poor.[180] Although early animal work suggested that REM sleep deprivation interfered with learning, subsequent studies showed that it was the stress of the REM sleep deprivation procedure, rather than the REM sleep loss itself, that was critical.[181] A leading proponent of a sleep and memory consolidation relationship has concluded that sleep has no role in the consolidation of declarative memory,[182] which would exclude a role for sleep in rote memory, language memory, and conceptual memory, leaving only the possibility of a role in procedural memory—the sort of memory required for learning to ride a bicycle or play a musical instrument. However, studies supporting a role for sleep in the consolidation of human procedural learning have made contradictory claims about similar learning tasks, with some concluding that REM but not NREM sleep is important and others stating just the reverse, and still others claiming that both sleep states are essential.[180] Millions of people have taken monoamine oxidase inhibitors or tricyclic antidepressants, often for 10 to 20 years. These drugs profoundly depress or in many cases completely eliminate all detectable aspects of REM sleep. Of note, however, not a single report of memory deficits attributable to such treatment has emerged. Likewise, well-studied patients with permanent loss of REM sleep resulting from pontine damage show normal learning abilities; the best-studied of these patientscompleted law school after his injury[183] and was last reported to be the puzzle editor of his city newspaper. People with multiple systems atrophy can have a complete loss of slow wave sleep and disruption of REM sleep without manifesting any substantial memory deficit.[184] A recent well-controlled study showed that REM sleep suppression with selective serotonin reuptake inhibitors or serotonin-norepinephrine reuptake inhibitors was associated with no significant decrement in memory consolidation on any task and even produced a small but significant improvement in a motor learning (i.e., procedural) task.[185]

Another idea that has been suggested repeatedly is that REM sleep serves to stimulate the brain.[186-188] According to this theory, the inactivity of NREM sleep causes metabolic processes to slow down to an extent that the animal would be unable to respond to a predator, capture prey, or meet other challenges upon awakening. Such alterations would leave mammals functioning like reptiles, with slow response after periods of inactivity. This hypothesis explains the appearance of REM sleep after NREM sleep under most conditions. It also explains the well-documented increased proportion of sleep time in REM sleep as the sleep period nears its end in humans and other animals. Humans are more alert when aroused from REM sleep than from NREM sleep, as are rats[189]—findings consistent with this idea. The very low amounts or absence of REM sleep in dolphins, in which the brainstem is continuously active and which never show bilateral EEG synchrony, can be explained by this hypothesis. If one hemisphere is always active, there is no need for the periodic stimulation of REM sleep to maintain the ability to respond rapidly. However, the brain stimulation hypothesis of REM sleep function does not explain why waking cannot substitute for REM sleep in terrestrial mammals. REM sleep–deprived persons experience a REM sleep rebound even if they are kept in an active waking state for extended periods, although this effect may be a result of stress, rather than REM sleep loss.[181]

A phenomenon that may explain REM sleep rebound is the cessation of activity of histamine, norepinephrine, and serotonin neurons during REM sleep. This cessation does not occur during the awake state, so waking would not be expected to substitute for this aspect of REM sleep.[190] REM sleep rebound may therefore be due to an accumulation of a need to inactivate these aminergic cell groups. Several cellular processes might benefit from the cessation of activity in aminergic cells. Synthesis of these monoamines and their receptors might be facilitated during this period of reduced release. The

receptors for these substances might be resensitized in the absence of their agonist. The metabolic pathways involved in the reuptake and inactivation of these transmitters also may potentially benefit from periods of inactivity. Some but not all studies have supported this hypothesis.[191-195]

Further investigation at the cellular level may lead to an "inside-out" explanation of REM sleep function, deriving a functional explanation from a better understanding of the neuronal basis of REM sleep control.

CLINICAL PEARL

The loss of hypocretin neurons is responsible for most cases of human narcolepsy. It is thought that this cell loss may be the result of an immune system attack on these neurons, but convincing evidence for this explanation is lacking. Administration of hypocretin is a promising future avenue for the treatment of narcolepsy. Because the hypocretin system has potent effects on arousal systems including the norepinephrine, serotonin, acetylcholine, and histamine systems, manipulation of the hypocretin system with agonists and antagonists is likely to be important in further pharmacotherapies for narcolepsy, insomnia, and other sleep disorders, as well as for depression.

SUMMARY

REM sleep was first identified by its most obvious behavior: rapid eye movements during sleep. In most adult mammals the EEG of the neocortex is low in voltage during REM sleep. The hippocampus has regular high-voltage theta waves throughout REM sleep. The tone of the postural muscles is greatly reduced or abolished during this state.

The key brain structure for generating REM sleep is the brainstem, particularly the pons and adjacent portions of the midbrain. Considerable progress has been made in identifying the neurons most closely linked to REM sleep within these regions and the transmitters that they employ. Massive damage to the REM-generating region can abolish REM sleep. Small lesions can cause REM sleep without atonia in animals or REM sleep behavior disorder in humans

Narcolepsy is characterized by abnormalities in the regulation of REM sleep. Most cases of human narcolepsy are caused by a loss of hypocretin (orexin) neurons, a cell group whose somas are localized to the hypothalamus. Hypocretin neurons have potent effects on alertness and motor control and normally are activated in relation to particular, generally positive emotions in humans as well as in animals. In the absence of this cell group, cataplexy, a REM sleep–like loss of muscle tone, occurs.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported by the Medical Research Service of the Department of Veterans Affairs and by National Institutes of Health (NIH) grants NS14610, MH64109, and DA034748.

Selected Readings

Howell MJ, Schenck CH. Rapid eye movement sleep behavior disorder and neurodegenerative disease. *JAMA Neurol* 2015;**72**:707–12.

Rasch B, Pommer J, Diekelmann S, Born J. Pharmacological REM sleep suppression paradoxically improves rather than impairs skill memory. *Nat Neurosci* 2009;**12**:396–7.

Schenck CH, Montplaisir JY, Frauscher B, et al. REM sleep behavior disorder (RBD): devising controlled active treatment studies for symptomatic and neuroprotective therapy—a consensus statement by the International RBD Study Group. *Sleep Med* 2013;**14**:795–806.

Siegel JM. The REM sleep-memory consolidation hypothesis. *Science* 2001;**294**(5544):1058–63.

Siegel JM. Clues to the functions of mammalian sleep. *Nature* 2005;**437**:1264–71.

A complete reference list can be found online at ExpertConsult.com.

Chapter
9

Novel Techniques for Identifying Sleep Mechanisms and Disorders

John H. Peever; Priyattam J. Shiromani

Chapter Highlights

- In the past decade, new genetically engineered tools have been developed that enable identification of the neural circuits that control sleep-wake behavior.
- Candidate cell circuits can be genetically targeted to determine how they regulate sleep and wakefulness, and specific genes can be transferred into these circuits to determine if they can correct abnormal sleep behaviors (e.g., narcolepsy).
- Novel genetic tools can be used to advance the current understanding of sleep control and sleep disorders such as narcolepsy.

Considerable effort has been directed at identifying the brain circuits that generate wakefulness, non–rapid eye movement (NREM) and rapid eye movement (REM) sleep because their dysfunction contributes to sleep disorders such as narcolepsy and REM sleep behavior disorder.[1,2] The brain structures responsible for sleep control were first identified by transecting or lesioning specific brain areas and determining how these manipulations affect sleep-wake behavior. For example, transection studies were used to show that structures in the pons are responsible for controlling REM sleep, and lesion studies were used to show that regions of the hypothalamus are critical for regulating NREM sleep. Researchers then began to monitor the electrophysiologic activity patterns of individual neurons and discovered that in regions identified by the lesion studies, neurons were active during either wakefulness or NREM or REM sleep. This discovery strengthened the idea that the sleep-wake states were controlled by activity of specific cell groups. However, neurons in these groups have been found to be intermingled with neurons controlling other behaviors. Moreover, astroglial cells are now considered to be important partners with neurons in regulating behavior. In view of the growing complexity of the circuit, the challenge faced by a new generation of sleep researchers is how to disentangle the network to identify the culprit neurons regulating behavior. This chapter surveys some of the new methodologies used to investigate sleep control at whole-organism, regional, and cellular levels.

NOVEL TOOLS FOR STUDYING SLEEP CIRCUITRY

Emerging technologies are changing the way in which sleep researchers ask questions about the functions and mechanisms of sleep. How some of these new tools have been used to dissect the neural circuits that control sleep and wake behaviors in both health and disease is a main focus of this chapter. Although cutting-edge tools have been applied to study rest-activity behavior in fruit flies and worms, this work is not reviewed here; rather, the focus is on advances in mechanisms of mammalian sleep and how such discoveries have led to better understanding of causes and treatments of narcolepsy.

Figure 9-1 outlines some of the new additions to the neuroscience toolkit that have been (or can be expected to be) influential in furthering the current understanding of sleep-wake behaviors. Optogenetics, pharmacogenetics, and gene therapy are three of the most powerful additions to this toolkit and are described in detail in this chapter. Specific aspects include how sleep researchers are using these technologies to determine (1) how individual neuronal populations generate sleep states, (2) the specific neurotransmitter identities of each of these cell groups, and (3) how these "sleep circuits" communicate with one another to facilitate the switching between different sleep states.

Understanding Sleep Control Using Conventional Genetics

By way of background, presented next is a summary of insights gained from more traditional transgenic approaches in which a specific gene of interest has been either deleted (by so-called knockout strategies) or added (using "knockin" methods) to the genome. The animal models that sleep researchers are most familiar with are ones in which the hypocretin (orexin) system has been manipulated. At the outset it is important to acknowledge the enormous impact that William Dement had on the sleep science field by developing a canine model of narcolepsy. He was the first to recognize canine narcolepsy and believed that this model system would unlock the mystery of the disease. He was right because the "dog model" eventually led to the identification of one of the causes of narcolepsy.

In 1999, researchers found that a genetic mutation in the hypocretin-2 receptor causes canine narcolepsy,[3] and targeted deletion of the gene encoding hypocretin induced narcolepsy in mice[4] (Jackson Labs stock 008867). Shortly thereafter, two independent reports showed selective loss of hypocretin neurons in human narcolepsy.[5,6] A transgenic mouse line was created that mimics the degeneration of the hypocretin neurons in human narcolepsy.[7] This was accomplished by fusing ataxin to the hypocretin promoter (called a fusion gene hypocretin:ataxin) so that a polyglutamine cytoxicity accumulates in hypocretin neurons, resulting in loss of 90% of

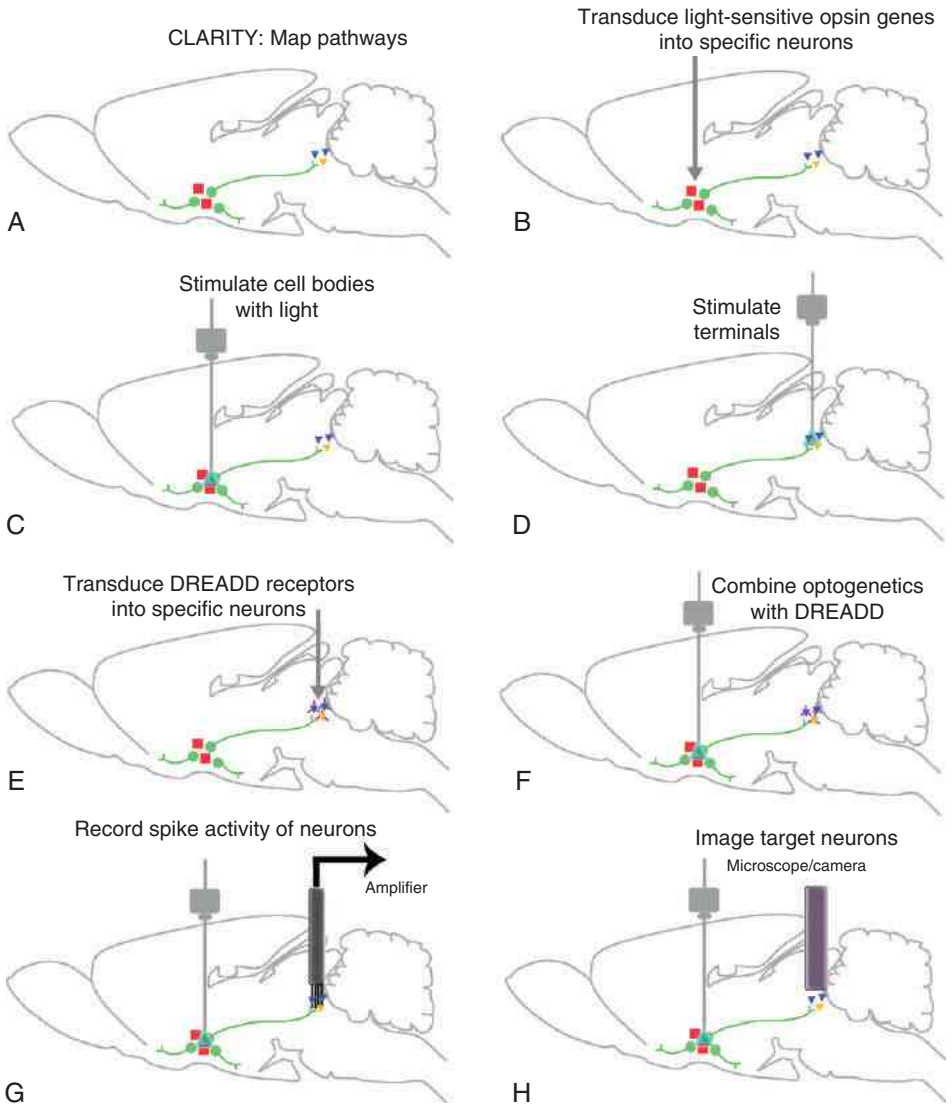


Figure 9-1 New genetically engineered tools to identify the neural circuitry responsible for waking, NREM, and REM sleep. Two areas in the brain, one in the hypothalamus (*green-* and *red-labeled* neurons) and the other in the pons (*triangles*), have been implicated in regulating sleep. These areas were first identified by transection and lesion studies, but now it is necessary to pin down the phenotype of the culprit neurons regulating sleep. To map the circuit, researchers can use CLARITY, a neuroanatomic method that renders tissue such as the brain transparent. The transparent brain is scanned with a confocal microscope allowing visualization of the underlying neurons and pathways **(A).** The activity of specific neurons can be manipulated by introducing light-sensitive genes into these cells **(B)** and using a light source such as a laser or light-emitting diode of a specific frequency to either activate them (via channelrhodopsins) or inhibit them (via halorhodopsin) **(C).** The terminals of these neurons can be manipulated with light to determine whether the target areas are involved in the behavior **(D).** Pharmacogenetic methods such as DREADD also can be used to selectively manipulate a specific phenotype of neurons **(E).** Both optogenetics and DREADD can be applied in the same experimental plan to manipulate specific nodes in the circuit **(F).** Electrophysiologic activity of the neurons can be monitored **(G).** However, to provide definitive evidence about the phenotype of the recorded neuron, a camera attached to a microendoscope can image the behavior of the specific neurons **(H).**

hypocretin cells within 1 month after birth[7] and virtually 100% by 6 months.

In typical gene knockout mice, such as the hypocretin-knockout model, gene deletion is constitutive and starts from conception. Such alterations may produce compensatory responses that could have unexpected effects on behavior, so a potentially better approach is to control the exact timing of gene deletion or cell death. This approach has now been realized by expressing the diphtheria toxin gene in hypocretin neurons and placing it under the control of the Tet-off system[8] (Tet from tetracycline). The use of antibiotics such as doxycycline to control the DNA is referred to as the Tet-off/Tet-on system. This method provides control over gene expression or repression simply through access to tetracycline or doxycycline in the water. In the presence of doxycycline, hypocretin neurons remain intact, but once doxycycline administration (in drinking water) is stopped, diphtheria toxin accumulates in hypocretin neurons, and within 7 days, 80% of hypocretin

neurons die. This approach can be used to determine the time course of development of narcoleptic symptoms as a function of hypocretin neuron loss. In fact, it has been used to show that mice develop cataplexy only when approximately 95% of hypocretin cells are lost.[8]

The Tet-driven system also has been used to study the functional role of neurons containing melanin-concentrating hormone (MCH).[9] Selective loss of MCH neurons produces a selective decrease in NREM sleep without changing either wakefulness or REM sleep.[9]

Cre Recombinase and the Advent of New Genetic Tools

The doxycycline and Tet-off/Tet-on system was made possible by the discovery of an enzyme called Cre recombinase, which was isolated from the P1 bacteriophage. The Cre (cycle recombinase) gene encodes a protein of 343 amino acids, which recognizes two sites called loxP (locus of X-over P1) sites, and the enzyme will very efficiently recombine essentially any DNA substrates that contain these sites. What is remarkable is that the orientation of the loxP sites determines whether the DNA between the loxP sites is deleted, translocated, or inverted. When two loxP sites are oriented in the same direction, Cre excises the DNA flanked by the loxP sites. When the loxP sites are oriented in the opposite direction, Cre flips the flanked DNA into the antisense orientation. Expression of Cre recombinase occurs only in specific cells, which ensures that the gene manipulation (deletion or insertion) is restricted only to cells expressing Cre. Crossing Cre mice with a loxP-flanked (also referred to as floxed sequence) mouse line enables Cre-mediated recombination in the Cre cells of the offspring.

It is no longer necessary to establish breeding colonies to use the Cre-transgenic animals. A new tool has been designed to insert specific genes of interest only in those cells expressing Cre recombinase. This system, developed by Karl Diesseroth's group at Stanford, involves a double-floxed inverse open reading frame (DIO) construct that causes serial recombination in Cre-expressing cells, resulting in the expression of the transgene. The DIO system was designed specifically to express light-sensitive opsins (discussed in a later section) in certain cells in a particular location. The DIO construct comprises a promoter, elongation factor 1 alpha (EF1α) (Table 9-1); DIO, the light-sensitive opsin to be activated (hChR2 or eArch3.0); and a reporter gene, enhanced yellow fluorescent protein (eYFP). These are linked to a viral vector—recombinant adenoassociated virus (AAV), in Table 9-1)—and microinjected into the desired target area. The virus will subsequently infect all of the cells in the area, but only the cells containing Cre recombination will express the transgene (for example, ChR2-eYFP).

It is common to include a reporter gene in the construct because its detection will indicate that the gene insertion procedure was successful. The reporter gene also serves as a proxy for proteins, such as ion channels (ChR2), pumps (halorhodopsin), or receptors that are difficult to detect using immunohistochemical methods. The use of reporter genes as markers of gene expression has been so transformative that Martin Chalfie, Osamu Shimomura, and Roger Y. Tsien were awarded a Nobel Prize for their discovery.

Ready-made viruses that target Cre-expressing cells constitute the most commonly used tool to deliver candidate effector genes (e.g., ChR2). A number of viruses can be used (e.g., adeno-associated virus) for this technology, and a review article[10] has highlighted strengths and weaknesses of each. The choice of a virus depends on the size of the total package of genes, called the cassette, that are going to be transferred into the surrogate cell. The cassette will include the promoter elements (so that the gene is expressed in a specific phenotype of neurons), the gene itself, and the reporter gene (to identify success of the gene transfer).

Table 9-1 Viral Vectors that Can Be Administered by Microinjection into Specific Brain Regions in Cre-transgenic Mice or Rats

Application	Excitatory	Inhibitory
Experimental	AAV-EF1a-DIO-hChR2 (H134R)-eYFP (optogenetics)	AAV-EF1a-DIO-eArch3.0-eYFP AAV-EF1a-DIO-eNpHR3-eYFP (optogenetics)
Experimental	AAV-hSyn-DIO-hM3D(Gq)-mCherry (DREADD)	AAV-hSyn-DIO-hM4D(Gi)-mCherry (DREADD)
Control	AAV- EF1a-DIO-eYFP (optogenetics or DREADD)	
Live-imaging of cells	AAV-EF1a-DIO-GCaMP6m	

AAV, Adenoassociated virus infects cells inserting the genes linked to it into the cells; DDO, double-floxed inverse open reading frame; DREADD, designer receptor exclusively activated by designer drug; eArch3.0, enhanced Archerhodopin3.0, inhibitory opsin; EF1a, elongation factor 1 alpha, a promoter; eNpHR, enhanced Natronomonas pharaonis3.0, an inhibitory opsin; eYFP, enhanced yellow fluorescent protein, reporter gene; GCaMP6m, genetically encoded calcium indicator; hChR2(H134R), humanized version of channelrhodopsin-2 that has been mutated to produce larger photocurrents compared to wild-type hChR2; hM3D and hM4D, human M3 muscarinic DREADD receptors; hSyn, human synapsin-1 promoter directs expression of genes into neurons.

THE OPTOGENETIC METHOD

Optogenetics ("opto" for optical stimulation and "genetics" for genetically targeted cell types) uses the light-sensitive proteins channelrhodopsin-2 (ChR2) and halorhodopsin (NpHR) to either activate or inhibit (respectively) the activity of neurons in which they are targeted.[11] ChR2 is a cation channel cloned from green algae, *Chlamydomonas reinhardtii,* that opens in response to blue light (with a wavelength of approximately 473 nm) and allows Na^+ to flow into a cell. Cells expressing ChR2 are depolarized when exposed to blue light (i.e., of the same wavelength). Halorhodopsin is a light-gated chloride pump present in *Natronomonas pharaonis,* a single-celled organism that is abundant in brackish water. Cells expressing NpHR are hyperpolarized when exposed to yellow light (wavelength of approximately 590 nm). The fast temporal kinetics of light-sensitive opsins makes it possible to reliably drive trains of high-frequency action potentials in ChR2-expressing cells or suppress action potentials in NpHR-expressing neurons. These two types of light-sensitive proteins can be inserted into cells, conferring the ability to manipulate the activity of specific cells with millisecond precision. Thus, depending on the wavelength of light, it is possible to control

specific circuits in a neural network. Because cells controlling specific behaviors are intermingled with cells controlling other behaviors, optogenetics empowers researchers to use light to control specific neurons in a neural network.

Optogenetics in Studying Sleep-Wake Control

Optogenetics is currently being used to determine how brain circuitry promotes sleep-wake behaviors. For example, the hypocretin and noradrenergic neurons have been optogenetically manipulated to determine how they control arousal[12-14] (Figure 9-1). In the first use of optogenetics in sleep neurobiology, a lentivirus delivered the gene for channelrhodopsin-2 into hypocretin neurons in mice. Light pulses at a frequency of 1 Hz did not significantly induce waking episodes from either NREM or REM sleep, but stimulation at higher frequencies (5 to 30 Hz) increased transitions from both NREM and REM sleep into wakefulness.[12] Optically induced arousals were blocked with a hypocretin receptor antagonist (almorexant), indicating that hypocretin neurons promote arousal by a hypocretin-dependent mechanism.

Optogenetic methods also have been used to show that hypocretin neurons promote arousal by activating noradrenergic cells in the locus coeruleus (LC). LC neurons were manipulated using channelrhodopsin-2 (activating) or halorhodopsin (inhibiting).[14] The two genes were inserted into LC neurons by injecting a Cre-recombinase-dependent AAV into knockin mice that selectively expressed Cre in tyrosine hydroxylase–expressing neurons. Photostimulation of LC neurons caused rapid transitions from sleep into wakefulness, whereas photoinhibition of these same cells reduced in the length of wake bouts, suggesting that one function of noradrenergic cells is to promote arousal. Of interest, very-high-frequency (5 to 20 Hz) stimulation of LC cells caused "behavioral arrests" that at face value appear similar to cataplexy bouts in narcoleptic mice.[4] One interpretation of this observation is that high-frequency LC stimulation depletes noradrenergic release onto motoneurons, which in turn triggers loss of muscle tone and hence cataplexy-like behavior. This observation is congruent with experiments showing that inactivity of LC neurons and reduced noradrenergic release onto motoneurons are associated with cataplexy.[15,16]

A newer inhibitory opsin, archaerhodopsin, which is more potent in its effects (compared with halorhodopsin), has been used to inhibit hypocretin neurons.[17] One hour of photoinhibition of the hypocretin neurons during the active period (i.e., dark phase) increased total time spent in NREM sleep but had no effect on REM sleep. Of interest, such inhibition did not induce more sleep during the normal inactive period (i.e., light phase) in these same mice, suggesting that optogenetic manipulation of hypocretin neurons has predictable effects on waking behaviors.

Optogenetics also has been used to determine how "sleep-active" neurons participate in sleep control. Although gamma-aminobutyric acid (GABA)/galanin neurons in the median and ventral lateral preoptic area[18,19,20] are hypothesized to induce NREM sleep, their specific role in sleep promotion has not been studied using optogenetic approaches. However, the role of hypothalamic MCH neurons in sleep promotion has been well documented.

MCH neurons are located in the zona incerta, dorsomedial hypothalamus, and lateral hypothalamus.[21] They are intermingled with hypocretin neurons and project to many of the same target regions.[22-25] Electrophysiology studies have shown that MCH neurons are inactive during waking, begin firing during NREM sleep, and become most active during REM sleep.[26] C-Fos studies support these findings.[27,28] Consistent with their sleep activity pattern, MCH release is correlated with sleep onset in both humans and rats,[29,29a] suggesting that MCH could function to initiate sleep. Pharmacologic studies in rodents showed that intracerebroventricular injection of MCH during the dark period increases both NREM and REM sleep in dose-dependent fashion.[29] MCH injection into sleep-promoting areas such as the ventral lateral preoptic area significantly increases NREM sleep,[30] whereas injection into REM sleep-promoting areas such as the nucleus pontis oralis or dorsal raphe increases REM sleep amounts.[31,32]

Because MCH neurons coexpress a variety of transmitters, including GABA,[25] CART (cocaine- and amphetamine-regulated transcript),[33] and nesfatin,[34] it is difficult to determine which neurochemical is responsible for their sleep-inducing effects. In a recent study, however, selectively killing MCH neurons by expressing the diphtheria toxin gene resulted in more wakefulness and less NREM sleep.[9] This finding is supported by observations in MCH-knockout mice, which experience more wakefulness and less NREM sleep,[35,36] suggesting that MCH per se is a sleep-promoting agent.

Optogenetic stimulation provides a direct test of the role of MCH neurons in sleep. In the very first study, Konadhode and coworkers[37] inserted the gene for channelrhodopsin-2 into the MCH neurons in wild-type mice and photostimulated infected neurons with blue-light pulses of 10 milliseconds (for 1 minute on every 5 minutes). These workers found that optogenetic activation of MCH neurons at night reduced the length of waking bouts and increased both NREM and REM sleep amounts. The increase in sleep was most robust when the light pulses were given at 10 Hz than at 5 Hz. The 10-Hz stimulation also increased delta power, which is a marker of sleep intensity. The increase in sleep occurred during the animal's active period, indicating that MCH neurons inhibit the activity of arousal-promoting neurons.

Jego and associates[38] also assessed the role for MCH neurons in sleep control by inserting channelrhodopsin-2 into MCH neurons. These investigators stimulated only during the second half of the normal 24-hour light-cycle and only when the mice entered into REM sleep, and found that such stimulation prolonged REM sleep bouts. They also used light-sensitive halorhodopsin (eNpHR3.0) to inhibit MCH neurons but found no affect on the duration of REM sleep bouts. Their finding broadly supports the hypothesis that MCH function to promote sleep.

Mechanisms generating REM sleep are controversial. One body of data suggests that REM sleep is generated by a cholinergic mechanism originating within the pedunculopontine tegmentum (PPT) and laterodorsal tegmentum (LDT) nuclei, whereas other data suggest that it originates from glutamatergic cells in the subcoeruleus area (synonymous with the SLD). Recently, Van Dort and colleagues aimed to clarify the role of PPT and LDT cholinergic neurons in REM sleep generation.[39] On optogenetically activating cholinergic neurons in the PPT or LDT during NREM sleep, these researchers found that 33% of the time, this intervention triggered REM sleep bouts, but without changing the length of REM sleep episodes. This observation suggests that PPT or

LDT is capable of inducing REM sleep. These workers did not find that optogenetic stimulation of the LDT/PPT cholinergic neurons triggered waking, even though these neurons are active in both waking and REM sleep. Thus NREM sleep may bias these neurons toward REM sleep.

The only other study to assess the functional role of sleep-active neurons using optogenetics is that of Anaclet and colleagues.[40] These investigators showed that optogenetic stimulation of the parafacial zone (PZ) induced NREM sleep, regardless of circadian time. They found that PZ neurons monosynaptically innervate and release GABA onto parabrachial neurons, which in turn project to and release glutamate onto magnocellular basal forebrain neurons. It is therefore hypothesized that GABAergic PZ neurons trigger NREM and slow wave activity by way of this newly identified pathway. Identification of this pathway was made possible by the use of optogenetics, underscoring the important contribution this new technology has made in elucidation of the circuit control of sleep.

THE PHARMACOGENETIC METHOD

Pharmacogenetics, also called chemogenetics, is a complementary technology to optogenetics, but instead of using light to control cell activity, it employs a drug-based approach. The particular branch of pharmacogenetics discussed here is DREADDs (designer receptors exclusively activated by designer drugs). This technology uses two separate types of engineered G protein–coupled receptors to genetically target cells and thereby manipulate their activity. Neurons that are forced to express the hM3Dq muscarinic receptor are excited by a Gαq signaling pathway, whereas neurons expressing the hM4Di muscarinic receptor are inhibited by a Gαi pathway. Both hM3Dq and hM4Di receptors are under exclusive control of the biologically inert compound clozapine-*N*-oxide (CNO) and are unresponsive to endogenous acetylcholine. Because CNO crosses the blood-brain barrier, it can be delivered by intraperitoneal injection. Unlike in optogenetics, which allows rapid, reversible control of neuron activity, manipulation of hM3Dq and hM4Di receptors affects neuronal activity over many minutes (approximately 60 to 120 minutes). However, the drug penetrates a wider area than can be achieved with optogenetics, where neurons only in the immediate arc of light are activated. DREADDs is advantageous for studying phenotypes of neurons that are diffusely scattered.

Pharmacogenetics in Studying Sleep-Wake Control

DREADD technology has been used to determine how manipulating hypocretin neurons affects sleep and wakefulness.[41] In a recent study, CNO-induced activation of hM3Dq receptors on hypocretin neurons increased wakefulness, whereas CNO-induced activation of hM4Di receptors on hypocretin neurons increased sleep.[41] The investigators also have combined both optogenetics and DREADDs to define the brain region in which hypocretin deficiency produces narcoleptic symptoms. They used mice in which both hypocretin receptors were deleted, and then they reinserted either hypocretin receptor 1 or 2 in phenotype-specific neurons in the pons.[42] Cataplexy was suppressed when the hypocretin-2 receptors were restored in the serotoninergic neurons of the dorsal raphe, and waking behavior and sleep fragmentation were rescued only when the hypocretin-1 receptors were restored in the LC. To confirm their findings, the authors used the hypocretin-ataxin-3 mice and selectively activated either the dorsal raphe serotonin neurons or the LC neurons with DREADD technology. DREADDs induced activation of serotonin neurons rescued cataplexy while activation of LC neurons increased waking. This finding suggests that cataplexy results from loss of hypocretin-induced activation of raphe serotonin neurons, and that sleepiness is caused by reduced LC activity.

GENE THERAPY METHODS

Gene therapy is being used to treat some intractable diseases including neurodegenerative diseases (http://www.wiley.com/legacy/wileychi/genmed/clinical/). For instance, in a phase 1 study, eight patients with mild Alzheimer disease were given ex vivo NGF gene therapy, and their rate of cognitive decline was slowed.[43] In another study, motor functions were improved in six patients with Parkinson disease who were given a gene for synthesis of dopamine and serotonin.[44] Because narcolepsy is considered a neurodegenerative disease, it should be feasible to restore network function and hence behavior by transferring the hypocretin gene into surrogate neurons.[5,6]

Mouse models of narcolepsy (see earlier under Understanding Sleep Control Using Conventional Genetics) provide an ideal platform for testing the effects of hypocretin gene therapy on narcolepsy symptoms, namely, sleepiness and cataplexy. One study used the herpes simplex virus to insert the hypocretin gene into neurons in the lateral hypothalamus of hypocretin-knockout mice. This particular mouse model was used because the underlying neuronal network regulating sleep remains unchanged. Liu and associates found that increasing hypocretin expression in the lateral hypothalamus markedly decreased cataplexy.[45]

A second study[46] used the hypocretin–ataxin-3 mouse model,[7] which more closely resembles human narcolepsy in that the hypocretin neurons die. The hypocretin gene (*HCRT*) was inserted into surviving neurons in the lateral hypothalamus, which led to a significant decrease in cataplexy in both number and duration of attacks.[46] Diluted virus produced intermediate effects, and transfer of the hypocretin gene into a control region that is not linked to sleep had no effect on cataplexy. Levels of sleep-wake were not changed by gene transfer in this study.[46]

Another group of investigators injected the hypocretin gene into more ventral regions of the lateral hypothalamus in hypocretin–ataxin-3 mice, and this approach modestly improved waking but not cataplexy.[47] In a fourth study,[48] the hypocretin gene was transferred into pontine neurons on the basis of evidence that they are linked to maintaining muscle tone.[49] Ths strategy produced a significant decrease in cataplexy in hypocretin-knockout mice. Another study combined optogenetics and DREADD to provide a better cellular resolution of the phenotype of neurons in the pons involved in cataplexy.[42] The researchers showed that the hypocretin input onto serotonin neurons reduces cataplexy, whereas hypocretin input onto LC neurons maintains waking.[42] Other sites also regulate waking, however, and hypocretin may exert its action at one of these. One group of workers[50] used mice that could not maintain long waking bouts as a result of a mutation in the hypocretin-2 receptor. They normalized the hypocretin-2

receptor (using Cre recombinase) only in neurons of the posterior hypothalamus including the tuberomammillary nucleus (TMN) and found that the mice could sustain long wake bouts, similar to wild-type mice.[50] The hypocretin-2 receptor is dense on histaminergic TMN neurons, and because the hypocretin ligand is already present in the mice, the normalized hypocretin-2 receptor on the histamine neurons corrected the waking defect. Thus two gene transfer studies show that waking can be sustained by hypocretin action onto TMN or LC neurons.

The conclusions from these gene transfer studies is that hypocretin gene transfer is most successful in decreasing cataplexy and maintaining long waking bouts provided that (1) the surrogate neurons carrying the hypocretin gene or the hypocretin-2 receptor are part of the circuit responsible for the behavior, and (2) the surrogate neurons carrying the ligand in the circuit are active during waking because then they can release hypocretin at target sites to normalize behavior.

CLARITY: Whole-Brain Imaging of Anatomic Pathways

CLARITY is a new neuroanatomic method that makes brain (and other body) tissues transparent. This process allows cells and their neural pathways to be visualized and reconstructed in three dimensions within the intact brain.[51-53] CLARITY has been used to examine cellular and pathway connections within the brain and spinal cord as well as peripheral organs such as the heart and kidney.

CLARITY is achieved by impregnating brain tissue with hydrogel that maintains tissue cellular structure and matrix, including antigenicity of the proteins. After the hydrogel has secured the cellular structure, the lipids in the tissue are dissolved and removed. A third set of chemicals adjusts the refractive index of the tissue to allow light to pass through, rendering it transparent. The entire block of tissue can now be scanned with a light-sheet microscope, and the network visualized in three-dimensional format. Whole-brain clearing methods can be combined with neuronal activity markers, such as c-Fos or Arc, to determine functional activation of the network.[54] CLARITY will be useful in reconstructing visual maps of sleep-wake circuitry to help delineate how and where these circuits are located in the central nervous system. Because CLARITY is largely descriptive in nature, however, functional methods are needed to assess exactly when these pathways become active during specific behaviors. This methodology reconstructs an intact brain in three-dimensional format, so foreseeably, it will prove to be crucial in deconstructing pathways in animal models of narcolepy, restless legs syndrome, or REM sleep behavior disorder.

Optical Live-Cell Imaging in Vivo

One of the newest and still developing technologies in neuroscience is live-cell imaging of neurons in the intact brain during behavior. Initially, only neurons in the cortex or hippocampus could be imaged. Today, however, it is feasible to image neurons deep (5 mm from cortical surface) within brain tissue in freely behaving mice.[55,56]

The target neurons are reached through a microendoscope (only 0.5 mm in diameter) and the optical signals imaged with a miniature fluorescence microscope.[57] Live imaging monitors changes in calcium influx associated with action potentials, and fluorescent calcium sensors, including those contained in the cells of genetically engineered mice, are now available.[58] To image a specific phenotype of neurons, AAVDJ-EF1a-DIO-GCaMP6m can be injected into the target neurons in specific Cre mice, and GCaMP6m fluorescence detected with a fluorescence microscope.

The advantage of this approach is that it allows for monitoring of the activity of individual genetically identified cells (e.g., bone fide hypocretin cells) in real time in the context of specific sleep-wake behaviors. Electrophysiology studies can record activity of single neurons, but this technique is extremely labor-intensive and slow, and it cannot identify the phenotype of the recorded neuron in freely behaving animals. By contrast, with live-cell imaging, a researcher can now "watch" the activity of individual and populations of neurons of known phenotypic origin and rapidly determine their role in behavior. Indeed, the activity of a single neuron can be assessed in this manner to determine whether it is a participant in the behavior. This methodology was recently used to dissect the role of GABAergic neurons in the lateral hypothalamus in appetitive versus consummatory behavior. The investigators found that even within a homogeneous neuronal population, the activity of individual neurons depends on the behavior.[57]

Live-cell imaging also confers the ability to either add or replace genes and to determine how this intervention affects their activity in the context of the intact nervous system during identified behaviors. This approach would be very useful to deconstruct the activity of specific neurons during cataplexy. Previously, electrophysiologic activity of single neurons was monitored during cataplexy in narcoleptic mice,[59] but live-cell imaging could definitively identify the phenotype of the recorded neurons.

CLINICAL PEARL

Technological advances have made it possible for scientists to peer into events at the subatomic level and also allowed humans to gaze out into the universe. The complexity of the human brain also is yielding to technology. In the past decade, new genetically engineered tools have been developed that enable identification of the neural circuits that control sleep-wake behavior. For example, cell circuits can be genetically targeted to determine how they regulate sleep and wakefulness, and specific genes can be experimentally transferred into these circuits to correct abnormal behavior such as in narcolepsy. Clinicians should be aware of these tools because they facilitate developing more rational treatments for sleep disorders.

SUMMARY

Conventional neuroscience tools such as lesions, cell recordings, c-Fos, and track-tracing methodologies have been instrumental in identifying the complex and intermingled populations of sleep- and arousal-promoting neurons that orchestrate and generate wakefulness and NREM and REM sleep. The recent development of new genetic technologies such as optogenetics, DREADDs, and gene therapy, however, has begun to transform the current understanding of sleep-wake regulation to uncover more specific details. These strategies have helped identify how discrete populations of genetically targeted cells (e.g., MCH, hypocretin) function to trigger and shape the timing and transition into and out of

different sleep-wake states. One of the greatest potentials for these new strategies, however, lies in explaining how pathologic activation and inactivation of discrete elements of the sleep-wake circuitry contribute to sleep disorders. For example, optogenetics can be expected to elucidate how (or if) recruitment of the REM sleep circuit contributes to muscle paralysis in cataplexy, and to determine how inactivity in this circuitry contributes to REM sleep without atonia in REM sleep behavior disorder.

ACKNOWLEDGMENTS

The work of PJS on which this chapter is based was supported in part by the Veterans Health Administration,Office of Research Development of the Department of Veterans Affairs, and by National Institutes of Health (NIH) grants NS052287, MH55772, and NS079940. The work of JHP was supported by grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada.

Selected Readings

Adamantidis AR, Zhang F, Aravanis AM, et al. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* 2007;**450**:420–4.

Apergis-Schoute J, Iordanidou P, Faure C, et Al. Optogenetic evidence for inhibitory signaling from orexin to MCH neurons via local microcircuits. *J Neurosci* 2015;**35(14)**:5435–41.

Brooks PL, Peever JH. Identification of the transmitter and receptor mechanisms responsible for REM sleep paralysis. *J Neurosci* 2012;**32**:9785–95.

Burgess CR, Peever JH. A noradrenergic mechanism functions to couple motor behavior with arousal state. *Curr Biol* 2013;**23**:1719–25.

Carter ME, Yizhar O, Chikahisa S, et al. Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat Neurosci* 2010;**13**:1526–33.

Konadhode RR, Pelluru D, Blanco-Centurion C, et al. Optogenetic stimulation of MCH neurons increases sleep. *J Neurosci* 2013;**33**:10257–63.

Liu M, Thankachan S, Kaur S, et al. Orexin (hypocretin) gene transfer diminishes narcoleptic sleep behavior in mice. *Eur J Neurosci* 2008;**28**: 1382–93.

Mochizuki T, Arrigoni E, Marcus JN, et al. Orexin receptor 2 expression in the posterior hypothalamus rescues sleepiness in narcoleptic mice. *Proc Natl Acad Sci U S A* 2011;**108**:4471–6.

A complete reference list can be found online at ExpertConsult.com.

Sleep in Animals: A State of Adaptive Inactivity

Jerome M. Siegel

Chapter 10

Chapter Highlights

- In most adult animals sleep is incompatible with mating and feeding. Many researchers view animals as being highly vulnerable to predation during sleep. Why do animals devote from 2 to 20 hours of the day to sleep, in what appears to be a nonproductive state? Why has evolution preserved this state?The presence of sleep in nearly all animals and the enormous variation in sleep time across species are best explained as adaptations to ecologic and energy demands.
- Sleep is not a maladaptive state that needs to be explained by undiscovered functions (which nevertheless undoubtedly exist). Rather, the major function of sleep is to increase behavioral efficiency. Greater waking activity does not necessarily lead to increased numbers of viable offspring and, hence, genetic success. Rather, genetic success is closely linked to the efficient use of resources and to the avoidance of risk. Thus, inactivity can reduce predation and injury. It also reduces brain and body energy consumption. As often stated, energy conservation is not a sufficient explanation for sleep, because the energy saved in a night's sleep in humans is only equivalent to that contained in a slice of bread. In the wild, however, most animals are hungry and are seeking food most of the time they are awake. If ample food is available, the population of a species quickly expands until faced again with food scarcity, a phenomenon that is illustrated by the great increase in the human population.
- The ability of sleep to conserve energy when food is scarce constitutes a major survival benefit even if only a small amount of energy is saved. Conversely, if food is available but is time-consuming to acquire, it is highly advantageous for animals to be able to reduce sleep time without behavioral impairment. Similarly, it is highly advantageous to reduce or eliminate sleep to allow migration and to respond to certain other needs. Several examples of extended periods of elimination or substantial sleep reduction without rebound have recently been documented and appear to be strongly linked to species success.
- Many researchers have assumed that predation risk is increased during sleep—that more animals are killed per hour during sleep than during waking. However there is scant evidence to support this contention. Most animals seek safe sleeping sites, often underground, in trees or in groups that provide communal protection. Those large herbivores that cannot find safe sleeping sites appear to require smaller amounts of sleep and to sleep less deeply. Large animals that are not at risk for predation, such as big cats and bears, can sleep for long periods, often in unprotected sites, and appear to sleep deeply.

ADAPTIVE INACTIVITY

Sleep should be viewed in the context of other forms of so-called *adaptive inactivity*. Most forms of life have evolved mechanisms that permit the reduction of metabolic activity for long periods of time when conditions are not optimal. In animals, this usually includes a reduction or cessation of movement and sensory response. The development of dormant states was an essential step in the evolution of life, and dormancy continues to be essential for the preservation of many organisms. Many species have evolved seasonal dormancy patterns that allow them to anticipate periods that are not optimal for survival and propagation (predictive dormancy, including hibernation). In other species, dormancy is triggered by environmental conditions (consequential dormancy). Many organisms spend most of their lifespan in dormancy, becoming active only when conditions are optimal. A continuum of states of adaptive inactivity can be seen in living organisms including plants, unicellular and multicellular animals, and animals with and without nervous systems.[1]

In the plant kingdom, seeds often are dormant until the correct season, heat, moisture, and pH conditions are present. A documented example of markedly delayed emergence was production of a healthy tree from a lotus seed after a 1300-year period of dormancy.[2] In another, more recent report, a 2000-year-old date palm seed produced a viable sapling.[3] Some forms of vegetation can germinate only after fires that may come decades apart. These include the giant sequoias

native to the U.S. Southwest, as well as many other species of trees and grass. Most deciduous trees and plants have seasonal periods of dormancy during which they cease photosynthesis, a process called abscission. These periods of dormancy enhance plant survival by synchronizing growth to optimal conditions. Clearly this mechanism has evolved to time germination to optimal conditions. Energy savings is not the only reason for dormancy in plants or animals.

Many unicellular organisms (protozoans) have evolved to live in environments that can sustain them for only portions of the year, because of changes in temperature, water availability, or other factors. Their survival requires that they enter dormant states that can be reversed when optimal conditions reappear.

Rotifers, a group of small multicellular organisms of microscopic or submicroscopic size (up to 0.5 mm long), have extended dormant periods lasting from days to months in response to environmental stresses, including lack of water or food.[5,6]

Parasites can become dormant within the host animal's tissues for years, emerging during periods when the immune system is compromised.[7] Some invertebrate parasites have extended dormant periods, defending themselves by forming a protective cyst.[8] In some cases the cyst can be dissolved and the parasite activated only by digestive juices. Many sponges have a similar dormant state that allows them to survive suboptimal conditions by being encased in "gemmules."

Insect dormancy or diapause can be seasonal, lasting several months; anecdotal reports indicate that under some conditions, diapause can last for several years to as long as a century.[9] This can occur in an embryologic, larval, pupal, or adult stage. During diapause, insects are potentially vulnerable to predation, as are some sleeping animals. Passive defense strategies are employed, such as entering dormancy underground or in hidden recesses, having hard shells, and tenacious attachment to substrates. In a few cases, insects have evolved a vibrational defensive response that is elicited when pupae are disturbed. Land snails and slugs can secrete a mucous membrane for protection and enter a dormant state when conditions are not optimal.[10]

Reptiles and amphibia that live in lakes that either freeze or dry seasonally and snakes that live in environments with periods of cold or extreme heat have the ability to enter dormant states (called *brumination* in reptiles). These dormant periods may occur just during the cool portion of the circadian cycle or may extend for months in winter.[11] *Estivation* is a form of dormancy that occurs during warm periods, typically during summer. It allows reptiles, amphibia, fish, and insects[12-16] to emerge with the first rains from what had been a barren, apparently lifeless lakebed.

In the mammalian class, a continuum of states ranging from dormancy to continuous activity can be seen. Small animals that cannot migrate long distances and live in temperate or frigid environments often survive the winter by hibernating. Some bats and many species of rodents, marsupials, and insectivores hibernate. This condition is entered from, and generally terminates in, non–rapid eye movement (NREM) sleep periods. During *hibernation,* body temperature can be reduced to below 10° C, down to as low as –3° C, with greatly reduced energy consumption.[17,18] Animals are quite difficult to arouse during hibernation, with full arousal taking many minutes. Consequently, hibernators are vulnerable to predation and survive hibernation by seeking protected sites. *Torpor*[17] is another form of dormancy that can be entered by mammals and birds daily. Torpor is entered and exited through sleep and can recur in a circadian rhythm or can last for weeks or months. Animals in shallow torpor are less difficult to arouse than hibernating animals but are still unable to respond quickly when stimulated. Some other mammals such as bears enter extended periods of sleep in the winter during which their metabolic rate and body temperature are reduced by 4° to 5° C,[19] but they remain more responsive than animals in torpor.

Sleep can be seen as a form of adaptive inactivity lying on this continuum. What is most remarkable about sleep is not the unresponsiveness or vulnerability it creates but rather its ability to reduce activity and body and brain metabolism but still allow a high level of responsiveness relative to the states of dormancy just described. The often-cited example of a parent's arousing at a baby's whimper but sleeping through a thunderstorm illustrates the ability of the sleeping human brain to continuously process sensory signals during the sleep period and trigger complete awakening to significant stimuli within a few hundred milliseconds. This capacity is retained despite the great reduction in brain energy consumption achieved in sleep relative to quiet waking.[20,21]

Adolescent humans are less responsive than adults to stimuli presented during sleep, as anyone who has lived with teenagers can attest. This trait may have been selected for by evolution, because protection from predators normally is provided by older members of the family group who also tend to the nocturnal needs of infants. The inactivity of children benefits the group by reducing their relatively large portion of the family's food needs and diverting food energy to growth, rather than activity.

Some animals that live in climates with a pronounced seasonal reduction in food or light availability or a periodic increase in threat from predators may need to migrate to survive. Many species of birds do this, as do certain species of marine mammals (see later). Although some may maintain circadian rhythms of activity during migration, others remain continuously active for weeks or months. Some vertebrate species do not ever appear to meet the behavioral criteria for sleep, remaining responsive, or responsive and active, throughout their lifetime.[22]

Humans with insomnia typically are not sleepy during the day despite a reduced (or in many cases normal) duration of nighttime sleep. With respect to their sleep pattern, they may be viewed as falling closer to migrating animals or short-sleeping animals, in contrast with humans with sleep disturbed by sleep deprivation, sleep apnea, or pain, who are sleepy during the day.[23] Patients with restless legs syndrome are similarly unlikely to be sleepy during the day despite low levels of nightly sleep. Conversely, many people with hypersomnia appear to need more sleep and in fact sleep more deeply, rather than being the victims of a shallow or disrupted sleep that is compensated for by extended sleep time. Perhaps such persons may be expressing genes and behaviors that were highly adaptive in distant phylogenetic ancestors needing to reduce energy consumption.

To summarize, evolution has produced a wide range of forms of diurnal or seasonal "adaptive inactivity," some of which are accompanied by a virtual cessation of metabolism and responsiveness. Clearly, evolution rewards judicious

activity, not continuous activity. Sleep often is viewed as a liability because of the associated reduced alertness in comparison with quiet waking. However, seen in the context of adaptive inactivity shown by most species, what is most notable about sleep in humans is its intermediate status, between the highly inactive unresponsive states seen in rotifers, insects, and hibernating mammals (which show little neuronal activity during hibernation) and the virtually continuous periods of activity and wakefulness that have been seen in migrating birds and cetaceans.

QUANTITATIVE ANALYSES OF THE CORRELATES OF SLEEP DURATION IN MAMMALS

Many studies have attempted to correlate the data that have been collected on sleep duration in mammals with physiologic and behavioral variables to develop hypotheses regarding the function of sleep. The data on which these studies are based are not ideal. Only approximately 70 mammalian species have been studied with sufficient measurements to determine the amounts of rapid eye movement (REM) and NREM sleep over the 24-hour period. These are by no means a random sample of the more than 5000 mammalian species. Rather, they are species that are viable and available for study in laboratories or, in some instances, for noninvasive (and less accurate) studies in zoos.

In laboratories, animal subjects for sleep studies typically are fed ad libitum and are housed at relatively invariant, thermoneutral temperatures and on artificial light cycles. These environments differ greatly from those in which they evolved. Digital recording and storage technologies now exist that will enable the collection of polygraphic data on animals in their natural environments,[24] but they have not yet been widely used. Such observations are necessary to determine the variation in sleep times caused by hunger, response to temperature changes, predation, and the other variables that have driven evolution. Very few of these animals have been tested for arousal threshold, the nature and extent of sleep rebound, and other aspects of sleep whose variation across species may potentially contribute to an understanding of sleep evolution and function. An important issue in comparing sleep times in animals is determining sleep depth. In humans, sleep depth, as assessed by either arousal threshold or EEG amplitude, increases after sleep deprivation and often is greater during early stages of development when total sleep time is greatest. Can sleep time be profitably compared across animals without incorporating information on sleep depth? Can it be assumed that animals that sleep for longer periods also sleep more deeply, as is true across human development, or is the reverse more likely—that short-sleeping animals sleep more deeply as has been hypothesized?[25] It would be best not to make either assumption; any conclusions should be based on hard evidence about these dimensions of sleep.

One of the earliest studies comparing REM and NREM sleep durations with physiologic variables found that sleep duration was inversely correlated with body mass.[26,27] A subsequent analysis found that this relationship applied only to herbivores, not to carnivores or omnivores.[28] This study also showed that, as a group, carnivores slept more than omnivores, who in turn slept more than herbivores (Figure 10-1). In an early study, a significant negative correlation was found between brain weight and REM sleep time (but not total sleep time). A point worthy of emphasis is that this latter correlation was extremely small, accounting for only 4% of the variance in REM sleep time (Figure 10-2). The largest correlation emerging from these early studies was that between body or brain mass and the *duration* of the sleep cycle—that is, the time from the start of one REM sleep period to the start of the next, excluding interposed waking. This correlation accounted for as much as 80% of the variance in sleep cycle time between animals and has held up in subsequent studies in mammals. Sleep cycle duration is approximately 10 minutes in mice, 90 minutes in humans, and 120 minutes in elephants. Because sleep is linked to a reduction in body temperature[29] and reduces energy usage, it has been hypothesized that energy conservation may be a function of sleep.[30]

Several studies have reanalyzed the phylogenetic data set with the addition of data on the few more recently studied animals. These studies took a variety of strategies to extract relations from this data set. Lesku and colleagues[31] used a method of "independent contrasts" in an attempt to control for the relatedness of species being compared. Inclusion of many rodent species in earlier analyses would give the data for those animals a disproportionate effect on conclusions. These workers confirmed previous findings of a negative relationship between basal metabolic rate (which is correlated with body mass) and sleep time. In contrast with earlier and subsequent studies of the same data set, they reported a positive correlation between REM sleep and relative brain mass and a negative relationship between REM sleep time and predation risk.

Another study, confining its analysis to studies that met the investigators' more rigorous criteria, found that metabolic rate correlates negatively rather than positively with sleep quotas,[32] in contrast with earlier studies.[27] This result is not inconsistent with some earlier work.[28] They also reported that neither adult nor neonatal brain mass correlates positively with adult REM or NREM sleep times, differing from earlier studies.[27,32] In agreement with earlier analyses, animals with high predation risk were found to sleep less.[28,33] In keeping with the concept of some fixed need for an unknown function performed only during sleep, the researchers proposed that short-sleeping species sleep more intensely to achieve this function in less time, but they presented no experimental evidence for this hypothesis.

A notable feature of the studies by Lesku and Capellini and their coworkers is that both excluded animals that the investigators concluded had unusual sleep patterns. So the echidna, which combines REM and NREM features in its sleep,[34] was eliminated from the analysis. The platypus, with the largest amount of REM sleep of any animal yet studied,[35] also was excluded from this analysis, as it was from another study focusing on brain size relations.[36] The dolphin and three other cetacean species and two species of manatee were excluded from the Lesku et al. study because of their low levels of REM sleep and presence of unihemispheric slow waves. Including these species in such analyses would undoubtedly negate or reverse the positive relationship reported between brain size and REM sleep, because the platypus exhibits the largest amount of REM sleep time of any studied animal and one of the smallest brain sizes and the dolphin, which appears to have little or no REM sleep, has a larger brain size than humans.[37,38] As discussed further on, these "unusual" species that have been excluded from previous

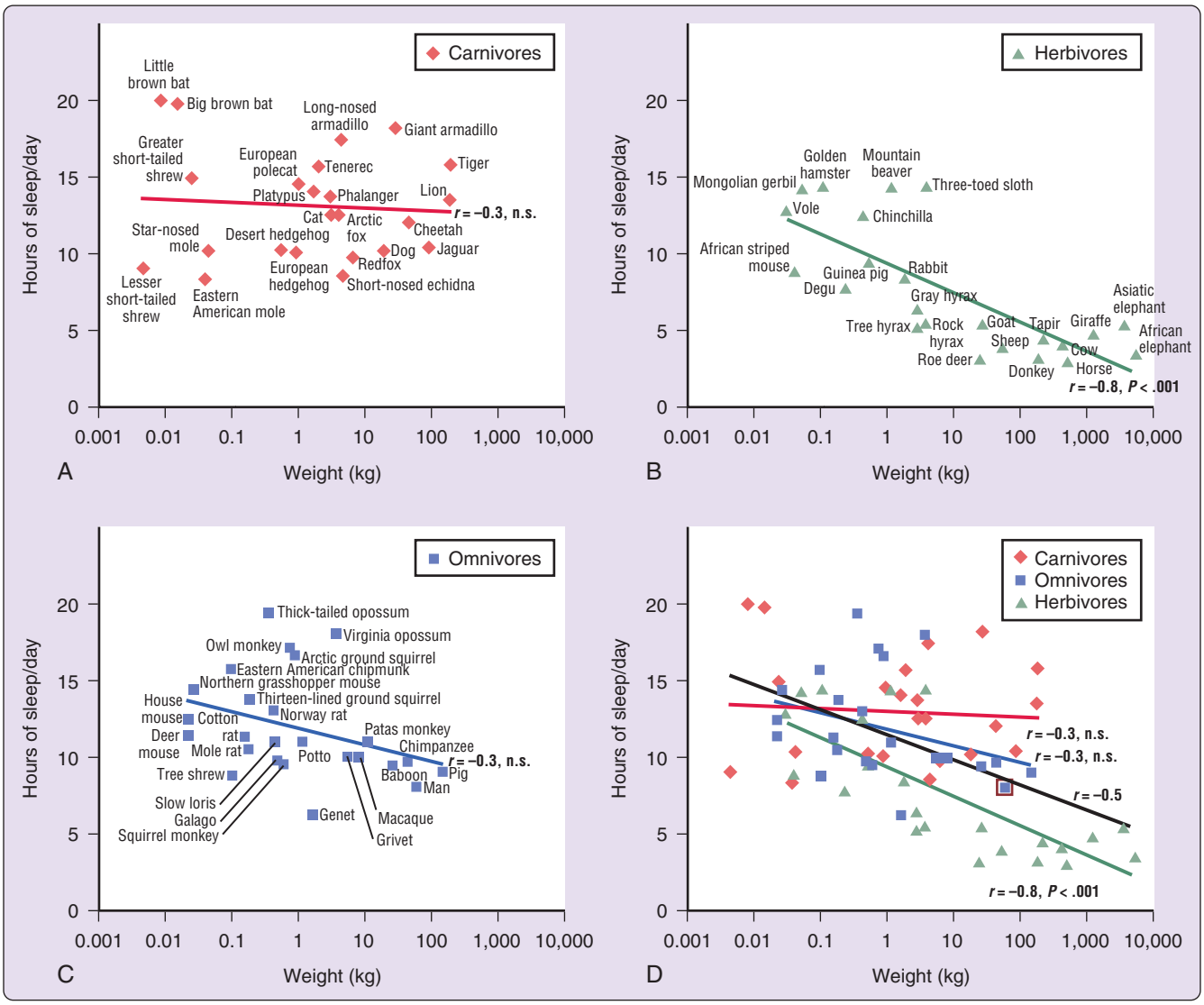


Figure 10-1 Sleep Times in Mammals. Sample data plotted by color: **A,** carnivores, *dark red*; **B,** herbivores, *green*; **C,** omnivores, *blue*; **D,** all data (*red box* indicates human). Sleep times in carnivores, omnivores, and herbivores differ significantly, with carnivore sleep amounts significantly greater than those for herbivores. Sleep amount is an inverse function of body mass over all terrestrial mammals (*black line*). This function accounts for approximately 25% of the interspecies variance (**D**) in reported sleep amounts. Herbivores are responsible for this relation, because body mass and sleep time were significantly and inversely correlated in herbivores but not in carnivores or omnivores. (From Siegel JM. Clues to the functions of mammalian sleep. *Nature* 2005;437:1264–71.)

analyses may in fact hold important clues to the function of sleep across species.

In considering the possibility of universal functions of sleep across species, from humans to *Drosophila*, it is important to appreciate the presence of REM and NREM sleep in birds. A correlational analysis of sleep parameters in birds paralleling the studies done in mammals found no relationship between brain mass, metabolic rate, relative metabolic rate, maturity at birth, and total sleep time or REM sleep time.[39] All values for these parameters were found to be "markedly nonsignificant." The only significant relation found was a negative correlation between predation risk and NREM sleep time (but not REM sleep time), in contrast with the relation reported earlier in mammals between predation risk and REM sleep time (but not NREM sleep time). This lone significant relation explained only 27% of the variance in avian NREM sleep time.

To summarize, a variety of correlation studies reach disparate and often opposite conclusions about the physiologic and functional correlates of sleep time. Of note, with the exception of the strong relationship between sleep cycle length and brain and body mass, all of the "significant" correlations reported explain only a small portion of the variance in sleep parameters, throwing into question whether the correlational approach as currently used is getting at the core issues of sleep function. Despite similar genetics, anatomy, cognitive abilities, and physiologic functioning, closely related mammalian species can have very different sleep parameters, and distantly related species can have very similar sleep parameters. Many such examples exist despite the relatively small number of species in which REM and NREM sleep times have been determined (Figure 10-3). For example, the guinea pig and baboon have the same daily amounts of REM and NREM sleep.[40]

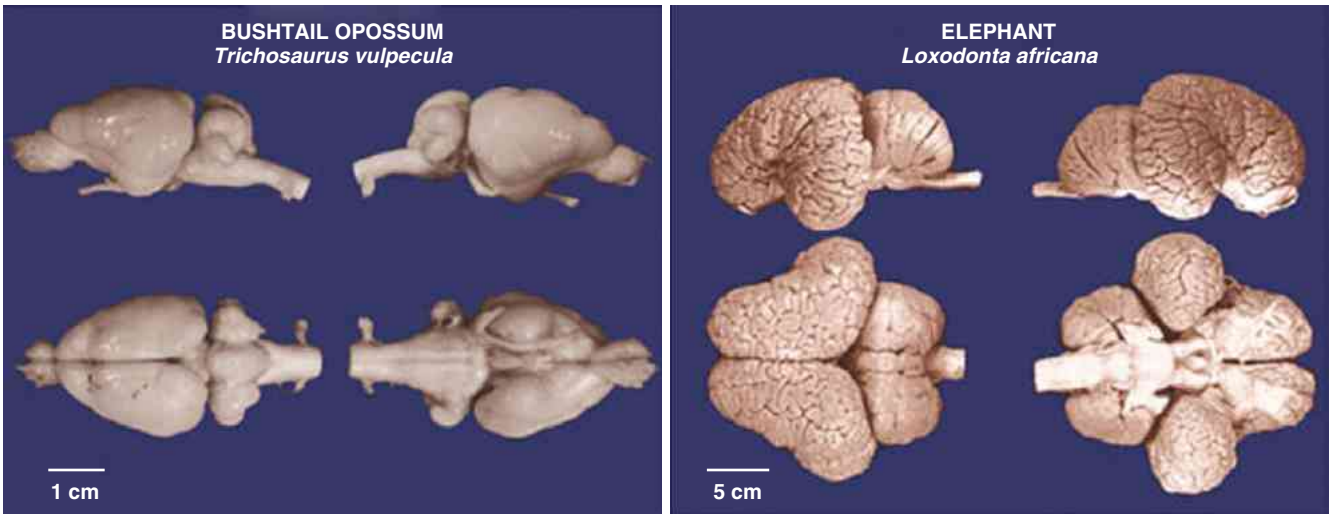


Figure 10-2 Sleep amount is not proportional to the relative size of the cerebral cortex or to the degree of encephalization, as illustrated by these two examples. (From Siegel JM. Clues to the functions of mammalian sleep. *Nature* 2005;437:1264–71.)

THE DIVERSITY OF SLEEP

Overview

On the assumption that sleep satisfies an unknown yet universal function in all animals, some work has been carried out in animals whose genetics and neuroanatomy are better understood and more easily manipulated than in mammals. Much of this work has focused on the fruit fly, *Drosophila melanogaster.* These animals appear to meet the behavioral definition of sleep. Their response threshold is elevated during periods of immobility, but they will rapidly "awaken" when sufficiently intense stimuli are applied. They make up for "sleep" deprivation with a partial rebound of inactivity when left undisturbed. However, major differences between the physiology and anatomy of these organisms and those of mammals make it difficult to transfer insights gleaned from studies of *Drosophila* sleep to human sleep. The *Drosophila* brain does not resemble the vertebrate brain. Octopamine, a major sleep-regulating transmitter in *Drosophila,* does not exist in mammals. Hypocretin, a major sleep-regulating transmitter in mammals, is not produced by *Drosophila.*[40] *Drosophila* flies are not homeotherms, whereas thermoregulation has been closely linked to fundamental aspects of mammalian sleep.[28,29,41] There is no evidence for the occurrence of a state resembling REM sleep in *Drosophila.* Thus the neurochemistry, neuroanatomy, and neurophysiology of sleep must necessarily differ between *Drosophila* and humans and other mammals. Any commonality of sleep phenomena would have to be restricted to cellular level processes. Two studies have shown that *Drosophila* sleep and sleep rebound are markedly impaired by genetic alteration of a potassium current that regulates neuronal membrane excitability.[42,43] Regulation of potassium currents may be a core function of sleep, or it may instead affect the excitability of circuits regulating activity and quiescence, much as such currents affect seizure susceptibility.[44,45]

Caenorhabditis elegans, a roundworm with a nervous system much simpler than that of *Drosophila,* has also been investigated for sleep like behavior.[46] *C. elegans* reaches adulthood in 60 hours and has periods of inactivity during this maturation, called "lethargus," occurring before each of the four molts it undergoes before reaching maturity. Stimulation of *C. elegans* during the lethargus period produced a small but significant decrease in activity during the remainder of the lethargus period but did not delay the subsequent period of activity or increase quiescence overall, phenomena that differ from the effects of sleep deprivation in mammals. It is not clear if adult *C. elegans* shows any aspect of sleep behavior.[47]

Fundamental species differences in the physiology and neurochemistry of sleep have been identified even within the mammalian line. Although many similarities have been recognized, the EEG aspects of sleep also differ considerably among humans, rats, and cats, the most-studied species.[48-50] Human stage 4 NREM sleep is linked to growth hormone secretion. Disruption of stage 4 sleep in children is thought to be a factor in the pathogenesis of short stature. In dogs, however, growth hormone secretion normally occurs in waking, not sleep.[51] Melatonin release is maximal during sleep in diurnal animals, but is maximal in waking in nocturnal animals.[52] Erections have been shown to be present during REM sleep in humans and rats[53]; the armadillo, however, has erections only in NREM sleep.[54] Blood flow and metabolism differ dramatically between neocortical regions in adult human REM sleep,[55] although most animal sleep deprivation and sleep metabolism studies treat the neocortex as a unit. Lesions of parietal cortex and certain other regions prevent dreaming in humans, even in subjects who continue to show normal REM sleep as judged by cortical EEG activity, rapid eye movements, and suppression of muscle tone.[56] Children younger than 6 years of age do not generally report dream mentation, perhaps because these cortical regions have not yet developed.[57] These findings make it questionable whether nonhuman mammals that exhibit REM sleep, all of which have cortical regions whose structure differs from that in adult humans, have dream mentation.

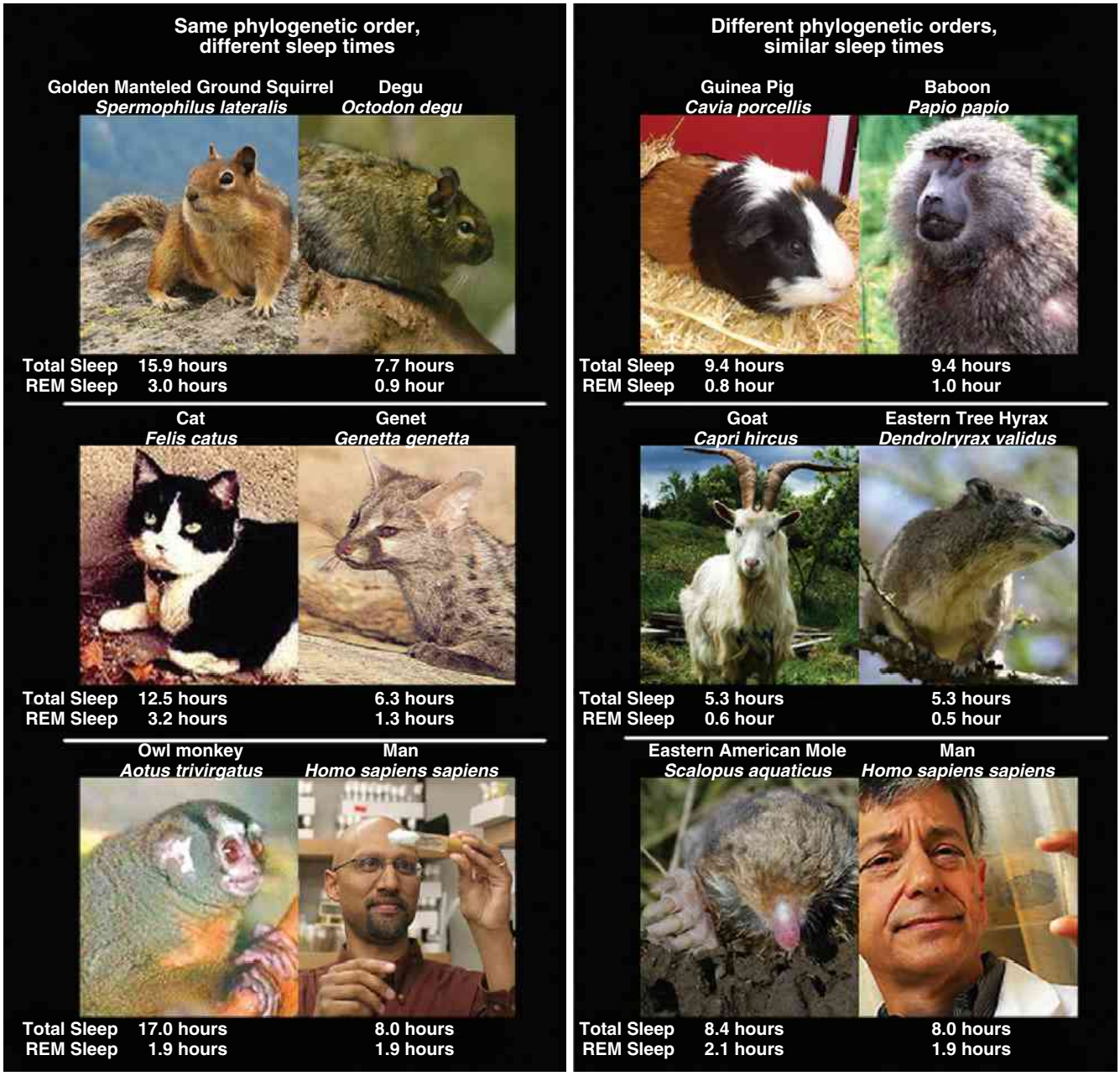


Figure 10-3 Mammalian phylogenetic order is not strongly correlated with sleep parameters. **Left,** Three pairs of animals that are in the same order but have very different sleep parameters. **Right,** Three pairs of animals from different orders with similar sleep amounts. Mammalian sleep times are not strongly correlated with phylogenetic order. (From Allada R, Siegel JM. Unearthing the phylogenetic roots of sleep. *Curr Biol* 2008;18:R670–9.)

Reindeer

Reindeer are ruminants. Like some other ruminants, they appear to remain active over the entire circadian cycle to an extent not seen in most carnivores and omnivores. A study examined the activity of two species of reindeer living in polar regions where they experience periods of continuous darkness in the winter and continuous light in the summer. Activity was monitored for an entire year. It was found that the circadian rhythm of melatonin and circadian rhythms of behavioral activity dissipated in winter and summer. Activity time ranged from 22% to 43% greater in summer than in winter (calculated from data obtained in studies by van Oort and Tyler).[58,59] EEG recording and arousal threshold tests were not done; however, the activity changes suggest that major changes in sleep duration occur seasonally.

Walrus

A study of the walrus revealed that these animals frequently become continuously active for periods of several days even when fed ad libitum and under no apparent stress.[60] Animals living in marine environments may not be as strongly affected by circadian variables because their evolution has been shaped by tidal and weather features that do not adhere to 24-hour cycles.

Sleep in Cetaceans: Dolphins and Whales

REM sleep is present in all terrestrial animals that have been studied, but signs of this state have not been seen in cetaceans, which are placental mammals. These animals show only unihemispheric slow waves (USWs), which can be confined to one hemisphere for 2 hours or longer. The eye contralateral to the hemisphere with slow waves typically is closed, although covering the eye is not sufficient to produce slow waves.[35,61] These animals never show persistent high-voltage waves bilaterally. Sometimes they float at the surface during USW activity. Often, however, they swim while USWs are being produced (Figure 10-4). When they swim during USW activity, no asymmetry in their motor activity is observed, in contrast with the behavior seen in the fur seal (see further on). Regardless of which hemisphere is showing slow wave activity, they tend to circle in a counterclockwise direction (in the northern hemisphere[62]). No evidence has been presented for elevated sensory response thresholds contralateral to the hemisphere that produces slow waves. Indeed, a substantial elevation of sensory thresholds on one side of the body presumably would be quite maladaptive, in light of the danger of collisions while moving. Similarly, brain motor systems must be bilaterally active to maintain the bilaterally coordinated movement that they exhibit during USW activity. As indicated by these findings, forebrain and brainstem sensory and motor activity must differ radically during USW from that seen in terrestrial mammals during sleep (see Chapter 8)[63,64] The one study of USW rebound after USW deprivation in dolphins produced very variable results, with little or no relation between the amount of slow waves lost in each hemisphere and the amount of slow waves recovered when the animals were subsequently left undisturbed.[65] In two other studies it was shown that dolphins are able to maintain continuous vigilance 24 hours/day, responding at 30-second intervals, for 5 and for 15 days with no decline in accuracy. At the end of this period, no detectable decrease of activity or evidence of inattention or sleep rebound, such as would be expected of a sleep-deprived animal, was seen.[22,66,67]

USWs would be expected to save nearly one half of the energy consumed by the brain that is saved during bihemispheric slow wave activity.[20,21] USWs are well suited to the dolphin's group activity patterns. Because dolphins and other cetaceans swim in pods, the visual world can be monitored by dolphins on each side of the pod, and the remaining dolphins merely have to maintain contact with the pod. In routine "cruising" behavior, this can be done with only one eye, allowing the other eye and connected portions of the brain to reduce activity, as occurs in USWs. This hypothesis needs to be explored by electroencephalographic observations of groups of cetaceans in the wild.

In some smaller cetaceans, such as the harbor porpoise[68] and Commerson's dolphin,[69] motor activity is essentially continuous from birth to death—that is, they never float or sink to the bottom and remain still. These animals move rapidly, and it is evident that they must have accurate sensory and motor performance and associated brain activation to avoid collisions. It is difficult to accept this behavior as "sleep" without discarding all aspects of the behavioral definition of sleep.[22]

All studied land mammals have been reported to show maximal sleep and maximal immobility at birth, leading to the conclusion that sleep is required for brain and body development. Newborn killer whales and dolphins, however, are continuously active. In captivity, they swim in tight formation and turn several times a minute to avoid conspecifics in the pool and pool walls. During this period the calves learn to nurse, breathe, and swim efficiently. Although some USWs might be present at these times, the eyes are open bilaterally when they surface, at average intervals of less than 1 minute, indicating that any slow wave pattern could not last longer than this period.[70] Sleep interruption at such intervals can be lethal to rats,[71] and human sleep is not restorative if interrupted on such a schedule.[72] The cetacean mothers also cease eye closure at the surface and during floating behavior and are continuously active during the postpartum period. No loss of alertness is apparent during the "migratory" period. In the wild, mother and calf migrate together, typically for thousands of miles, from calving to feeding grounds. Sharks, killer whales, and other predatory animals target the migrating calves, and a high level of continuous alertness is necessary for both mother and calf during migration. The maternal and neonatal pattern could be described as "sleep" with well-coordinated motor activity, accurate sensory processing, and effective response to threats in the environment, and without the likelihood of any EEG slow waves or prolonged eyelid closure. This designation, however, does not comport with the accepted behavioral definition of sleep.[73] Thus both cetaceans and migrating birds (see later) greatly reduce sleep time during migrations without any sign of degradation of physiologic functions, sluggishness, loss of alertness, or impairment of cognitive function.

Sleep in Otariids: Eared Seals

On land, sleep in the fur seal generally resembles that in most terrestrial mammals. The EEG is bilaterally synchronized, and the animal closes both eyes, appears unresponsive, and cycles between REM and NREM sleep. By contrast, when the fur seal is in the water, it usually shows an asymmetric pattern of behavior, with one of the flippers being active in maintaining body position, while the other flipper is inactive. During these periods, the fur seal has a high-voltage EEG, with slow waves in one hemisphere with the contralateral eye generally closed. The other eye generally is open or partially open, with an activated, waking-like EEG pattern (Figure 10-5). Therefore, unlike in the dolphin, it appears that one half of the brain and body may in some sense be "asleep" and the other half "awake." Microdialysis studies showed that during asymmetric sleep, the waking hemisphere has significantly higher levels of acetylcholine release than the sleeping hemisphere.[74] By contrast, levels of serotonin,[75] histamine, and norepinephrine,[76] transmitters traditionally considered to be linked to arousal, do not differ between the hemisphere with high-voltage EEG activity and the hemisphere with a low-voltage EEG pattern. This work indicates that acetylcholine has a unique role in mediating the waking EEG in this species, and probably in other mammals, including humans.

Sleep in Monotremes

The mammalian class can be subdivided into three subclasses: placentals, marsupials, and monotremes. There are just three extant monotreme species: the short-beaked and long-beaked echidna and the platypus. Fossil and genetic evidence indicates that the monotreme line diverged from the other mammalian lines approximately 150 million years ago and that

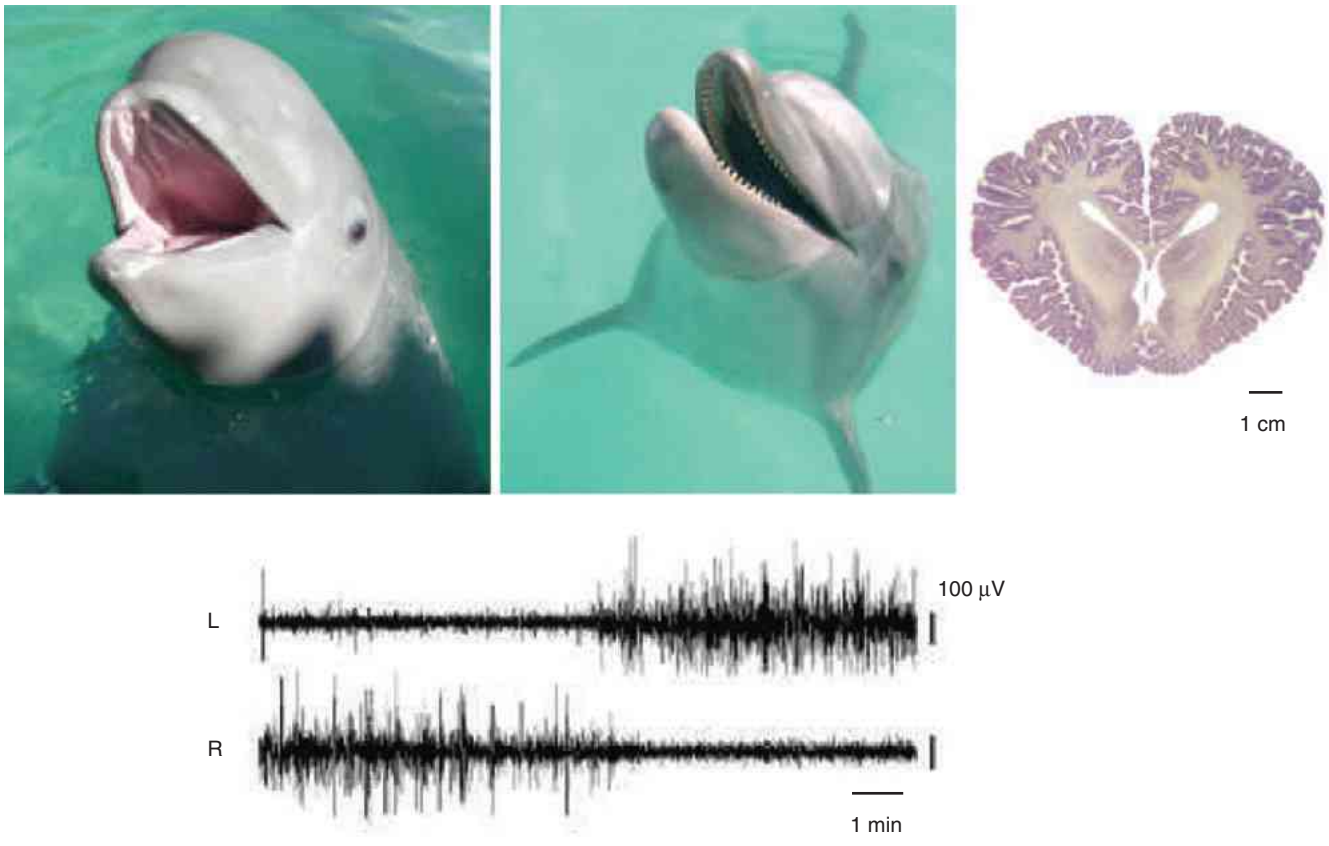


Figure 10-4 Cetacean Sleep: Unihemispheric Slow Waves in Cetaceans. **Top,** Photos of immature beluga (*left*), adult dolphin (*center*), and section of adult dolphin brain (*right*). Electroencephalogram (EEG) of adult cetaceans, represented here by the beluga, during sleep. All species of cetacean for which sleep EEGs have been recorded so far have demonstrated unihemispheric slow waves. *Top traces* show left and right EEG activity. The spectral plots (*bottom traces*) show 1- to 3 Hz power in the two hemispheres over a 12-hour period. The pattern in the cetaceans contrasts with the bilateral pattern of slow waves seen under normal conditions in all terrestrial mammals, represented here by the rat. (From Siegel JM. Clues to the functions of mammalian sleep. *Nature* 2005;437:1264–1271.)

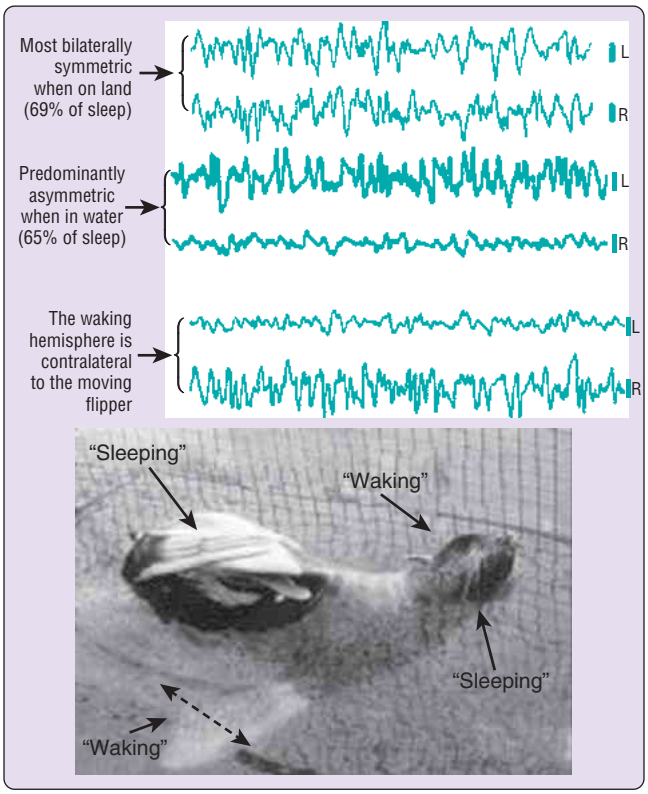


Figure 10-5 **Fur Seal Sleep.** On land, fur seals usually sleep like terrestrial mammals, with bilateral EEG synchrony and REM sleep (*not shown*). When in water, however, they typically show asymmetric slow wave sleep, with a sleep-like EEG pattern in one hemisphere and a waking-like EEG in the other hemisphere. Unlike in the dolphin, the asymmetric EEG of the fur seal is accompanied by asymmetric posture and motor activity, with the flipper contralateral to the hemisphere showing low-voltage activity used to maintain the animal's position in the water and the other flipper and its controlling hemisphere showing "sleep." EEG, Electroencephalogram.

both echidna species are derived from a platypus-like ancestor.[77-80] The monotremes have shown a remarkably conservative evolutionary course since their divergence from the two other mammalian lines. For example, fossil teeth from *Steropodon galmani* dated at 110 million years ago show many similarities to the vestigial teeth of the present-day platypus, *Ornithorhyncus anatinus*.[81] Analyses of fossilized skull remains indicate remarkably little change in platypus morphology over at least 60 million years.[81,82] The low level of speciation throughout the fossil record is another indicator of the uniquely conservative lineage of monotremes. The 150 million years of platypus evolution has produced no species radiation, apart from the echidna line, and only two living and one extinct species of echidna have been documented. Although monotremes are distinctly mammalian, they do display a number of reptilian features, making study of their physiology a unique opportunity to determine the commonalties and divergences in mammalian evolution.[78,83,84]

This phylogenetic history led to an early study of the echidna to test the hypothesis that REM sleep was a more recently evolved sleep state. No clear evidence of the forebrain low-voltage EEG acctivity that characterizes sleep was seen in this study, leading to the tentative conclusion that REM sleep evolved in placentals and marsupials after the divergence of the monotreme line from the other mammals.[85] This issue was subsequently reexamined using single-neuron recording techniques, in addition to the EEG measures employed in the earlier studies. REM sleep is generated in the mesopontine brainstem (see Chapter 8) and is characterized by a highly variable burst-pause activity of brainstem neurons. This activity is responsible for driving the rapid eye movements, twitches, and other aspects of REM sleep. The investigators recorded from these brainstem regions in unrestrained echidnas to see if this activation was absent throughout sleep. Instead of the slow, regular activity that characterizes brainstem neurons in many nuclei during NREM sleep in placental mammals,[63,64] the echidna showed the irregular activity pattern of REM sleep throughout most of the sleep period[34,86] (Figure 10-6). It appeared that the brainstem was in a REM sleep–like state while the forebrain was in an NREM sleep state.

These interesting findings led to the performance of additional electrophysiologic studies of sleep in the platypus. In this work, the platypus was found to have pronounced phasic motor activity typical of that seen in REM sleep (see Video 10-1).[87] This intense motor activity could occur while the forebrain EEG exhibited high-voltage activity,[35] similar to the phenomenon seen in the echidna. Not only was the motor activity during sleep of intensity equal to or greater than than that seen in REM sleep in other animals, but the daily amount of this REM sleep state was greater than that in any other animal. However, unlike in adult placental and marsupial mammals, the signs of REM sleep were largely confined to the brainstem (Figure 10-7). This observation indicates some resemblance to the conditions in most mammals that are born in an immature (altricial) state, which do not show marked forebrain EEG activation during REM sleep early in life. The tentative conclusion reached in the initial studies of the echidna, that the monotremes experienced no REM sleep and that REM sleep was a recently evolved state, had to be reversed. Apparently, a brainstem manifestation of REM sleep probably was present in the earliest mammals, perhaps in very large amounts. It may be the brainstem quiescence of NREM sleep, along with the cortical EEG desynchrony of REM sleep, that are the most recently evolved aspects of sleep in the mammalian line.

Birds

Birds have REM sleep that appears physiologically very similar to that seen in mammals, although REM sleep values tend to be lower than total sleep values in mammals.[39] Many bird species migrate over long distances. The effect of this migratory behavior on sleep has been studied in the white-crowned sparrow (*Zonotrichia leucophrys gambelii*). These birds, even when confined in the laboratory, decrease sleep time by two thirds during the periods when they would normally be migrating.[88] Of note, this is a common feature of cycles of adaptive inactivity. A ground squirrel that normally hibernates in the winter will enter a state of torpor at the appropriate season even when maintained in a laboratory under constant conditions.[89]

During the migratory period, the sparrow's learning and responding were unimpaired or improved. In these birds, sleep was not deeper by EEG criteria than that seen when they were not migrating, despite its greatly reduced duration. Their sleep latency did not differ from that during nonmigrating periods.[88]

A recent study of sleep in birds presented an novel and important example of "adaptive" sleep suppression. It was

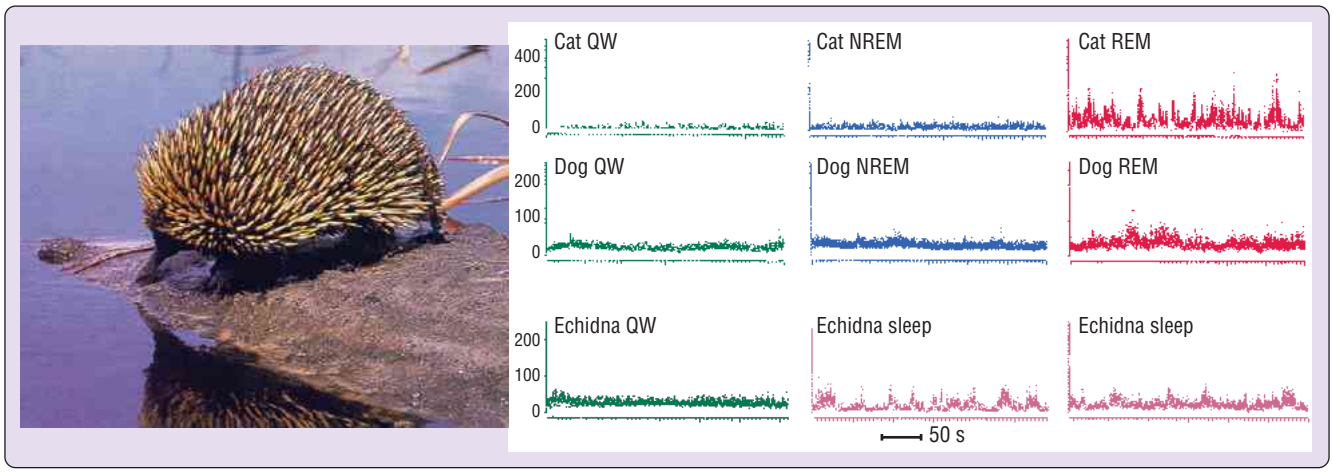


Figure 10-6 **Brainstem Activation during Sleep in the Echidna.** Instantaneous compressed rate plots of representative units recorded in nucleus reticularis pontis oralis of the cat, dog, and echidna. Each point represents the discharge rate for the previous interspike interval. In cat quiet waking (QW) and NREM sleep, the discharge rate is low and relatively regular. The rate increases and becomes highly variable during REM sleep. A similar pattern can be seen in a unit recorded in the dog. In the echidna, sleep is characterized by variable unit discharge rates, as is seen in REM sleep, but this occurs while the cortex is showing high-voltage activity. (From Siegel JM, Manger P, Nienhuis R, et al. The echidna Tachyglossus aculeatus combines REM and nonREM aspects in a single sleep state: implications for the evolution of sleep. *J Neurosci* 1996;16:3500–6.)

found that many polygynous pectoral sandpipers, which breed during a period of continuous summer light, cease sleeping or greatly reduce sleep during breeding. Furthermore, the males with the greatest reduction in sleep sired the most offspring—a unique and dramatic example of adaptive sleep loss increasing genetic propagation.[90] What was most surprising, in view of the strength of this selective benefit, was that *any* males remained sleeping during the breeding period. It was speculated that the continuously active males would be at a competitive disadvantage if their food was scarce after the breeding season relative to those that saved energy by sleeping. In periods of reduced food availability, it is the second group of birds that would survive to mate the next year, leading to a dynamic balance between birds with these two behaviors.[91]

Studies in the ostrich, considered to be in many respects a "primitive" bird, provided novel support for the link between bird and mammalian sleep. It was found that sleep in the ostrich resembles that in the platypus and echidna, with rapid eye movements and muscle tone suppression, brainstem aspects of REM sleep occurring with high-voltage EEG activity, resembling the weak or nonexistent EEG voltage reduction seen in monotremes, whereas their brainstem shows the neuronal activity activation and rapid eye movements that characterize REM sleep.[92]

The observations in monotremes and birds suggest that the reptilian common ancestor of both mammals and birds exhibited REM sleep or a closely related precursor state, rather than the previously advanced speculation that REM sleep must have evolved twice, based on the conclusion that monotremes did not have REM sleep. Although scattered early reports claimed to have identified REM sleep in reptiles, these findings have not been replicated.[28] In the experience of my own research group, when the same recording techniques used in the echidna were appplied in the turtle in a search for evidence of REM sleep, no evidence of phasic brainstem neuronal activity during quiescent states in this reptile was found.[93]

SLEEP REBOUND

The phenomenon of sleep rebound[94] is not always seen. When fur seals go in the water for extended periods, as they do in winter, REM sleep time is greatly reduced. Little or no rebound of lost REM sleep occurs when the animals return to land, even after several weeks in the water.[95] In the cases of the dolphins and killer whales mentioned previously, a near-total abolition of "sleep-like behavior" for periods of several weeks during migration is followed by a slow increase back to baseline levels, with no rebound. The same phenomenon is seen in migrating white sparrows, a migratory species that has been carefully studied under laboratory conditions.[88] In human studies, persons with bipolar disorder in a "manic" phase greatly reduce sleep time for extended periods, and persuasive evidence for progressive degradation of performance, physiologic function, or sleep rebound during this period is lacking. Zebra fish can be completely deprived of sleep for more than three days by placing them in continuous light but show no rebound when returned to a "12-12" light-dark cycle.[96] By contrast, when they are deprived by repetitive tactile stimulation, they do show rebound, suggesting that the deprivation procedure rather than the sleep loss underlies the rebound.

Typically, 30% or less of sleep time lost during deprivation is recovered in the human and rodent, in which the phenomenon has been most extensively studied. A similar percentage of rebound is seen in other species including some invertebrates.[97] One may ask why, if sleep is essentially a maladaptive state, animals that have the ability to regain lost sleep in 30% of the time it would normally have taken have not evolved shorter sleep times to take advantage of the adaptive benefits of increased waking. If sleep is viewed as a form of adaptive inactivity, however, this paradox vanishes. A small sleep rebound may be necessary to compensate for processes that can occur only, or optimally, in sleep, but for the most part,

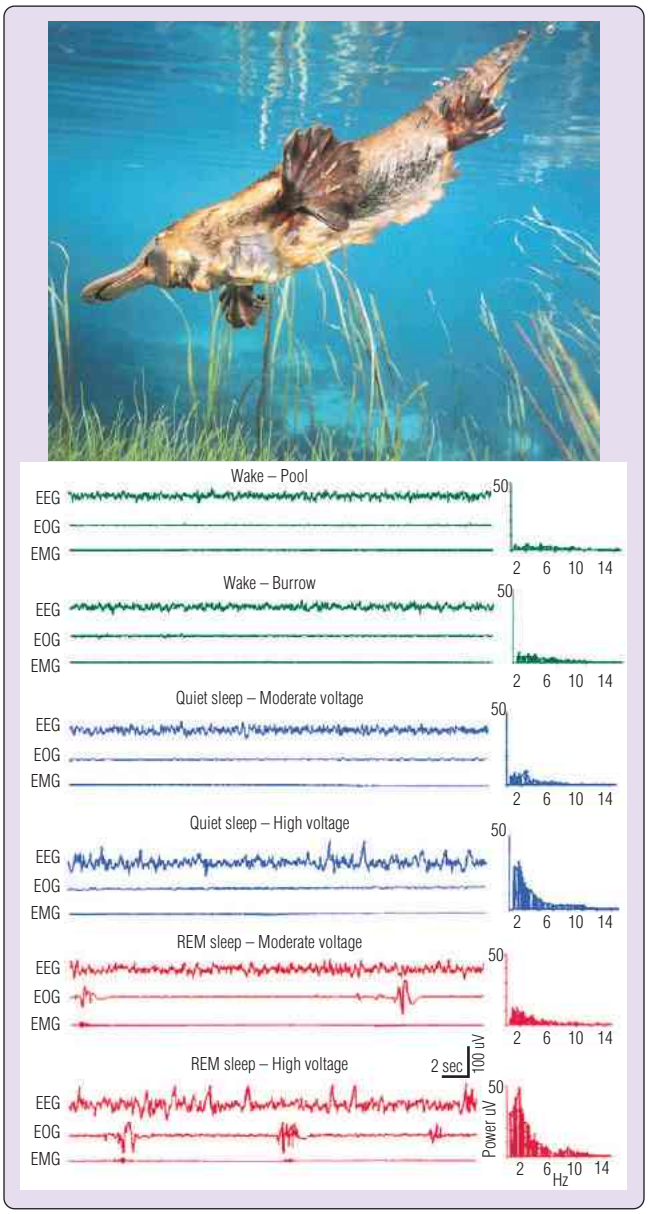


Figure 10-7 **Brainstem REM Sleep State in the Platypus.** Rapid eye movements and twitches can occur while the forebrain is showing a slow wave activity pattern. EEG, EOG, EMG, and EEG power spectra of samples shown of sleep-wake states in the platypus. EEG, Electroencephalogram; EMG, electromyogram; EOG, electrooculogram. (From Siegel JM, Manger PR, Nienhuis R, et al. Sleep in the platypus. *Neuroscience* 1999;91:391–400.)

sleep time is determined in each species by the evolved tradeoffs between active waking and adaptive inactivity.

The variation in rebound within and across species needs to be more carefully studied. Some aspects of rebound have been shown to be due to the deprivation procedure, rather than to the sleep loss itself. For example, stressing rats by restraint can produce increased REM sleep even when no sleep has been lost. This effect is mediated by the release of pituitary hormones.[98,99] It is possible that in some species, other aspects of rebound are driven by changes in hormonal release linked to sleep deprivation,[1] rather than by some intrinsic property of sleep.

CLINICAL PEARL

Although sleep and sleep stages differ in amount between species, human sleep does not appear to be qualitatively unique. This factor makes animal models suitable for the investigation of many aspects of pharmacology and pathology in sleep science.

SUMMARY

Sleep can be seen as an adaptive state, benefiting animals by increasing the efficiency of their activity. Sleep does this by suppressing activity at times associated with maximal predator risk and permitting activity at times of maximal food and prey availability and minimal predator risk. It also increases efficiency by decreasing brain and body metabolism. However, unlike the dormant states employed in plants, simple multicellular organisms, and ectothermic organisms, and the hibernation and torpor employed in some mammals and birds, sleep allows rapid arousal for tending to infants, dealing with predators, and responding to environmental changes. A major function of REM sleep may be to allow rapid awakening with alertness, by means of periodic brainstem activation. Many organisms can reduce sleep for long periods of time without rebound during periods of migration or other periods in which a selective advantage can be obtained by continuous waking.

The big brown bat specializes in eating mosquitoes and moths that are active from dusk to early evening. The big brown bat typically is awake only approximately 4 hours a day.[27] Not surprisingly, this waking is synchronized to the period when its insect prey species are active. It is not likely that this short waking period, one of the shortest yet observed, can be explained by the need for some time-consuming unknown process that occurs only during sleep and requires 20 hours to complete. This extremely brief period of wakefulness can be more easily explained by the ecologicl specializations of this bat. Similarly, "sleep" in ectothermic animals is most likely to be determined by temperature and other environmental variables, rather than any information processing or physiologic maintenance requirement. An approach that takes the environmental conditions in which each species evolved into account can better explain the variance in sleep time among mammals.

Many vital processes occur in both waking and sleep, including recovery of muscles from exertion, control of blood flow, respiration, growth of various organs, and digestion. Some processes may occur more efficiently in sleep but can also occur in waking. It has been claimed that sleep has an essential role in learning, but further investigations have disputed such claims.[100-105] It is highly probable that some functions have migrated into or out of sleep in various animals. Neurogenesis,[106] synaptic downscaling,[107] immune system activation,[108] and reversal of oxidative stress[109,110] may be accomplished in sleep in mammals. It remains to be seen if these or any other vital functions can be performed only in sleep. As suggested by the available evidence, however, such functions cannot explain the variation of sleep amounts and the apparent flexibility of sleep physiology within and between animals. Viewing sleep as a period of well-timed adaptive inactivity that regulates behavior may better explain this variation.

ACKNOWLEDGMENT

Work on which this chapter is based was supported by the Medical Research Service of the Department of Veterans Affairs, the National Institutes of Health, and the National Science Foundation.

Selected Readings

Lesku JA, Rattenborg NC. Avian sleep. *Curr Biol* 2014;**24**:R12–14.

Schmidt MH. The energy allocation function of sleep: a unifying theory of sleep, torpor, and continuous wakefulness. *Neurosci Biobehav Rev* 2014;**47**:122–53.

Siegel JM. Clues to the functions of mammalian sleep. *Nature* 2005;**437**:1264–71.

Siegel JM. Sleep viewed as a state of adaptive inactivity. *Nat Rev Neurosci* 2009;**10**:747–53.

Yetish G, Kaplan H, Gurven M, et al. Natural sleep and its seasonal variations in three pre-industrial societies. *Curr Biol* 2015;**25**:1–7.

A complete reference list can be found online at ExpertConsult.com.

Section 3

Physiology in Sleep

Chapter 11

Relevance of Sleep Physiology for Sleep Medicine Clinicians

Gilles Lavigne; Meir Kryger

Chapter Highlights

- The physiology section of this book covers a wide spectrum of very specialized concepts, from molecular and behavioral genetics to system physiology (temperature control, cardiovascular and respiratory physiology, immune and endocrine functions, sensory motor neurophysiology), and integrative functions (e.g., mental performance, memory, mood, and daily physical functioning).
- Some chapters are entirely new or have been revised together with new collaborators. For example, new content has been added on the role of opioids in sleep and breathing and on the neurobiology of brain trauma, a topic of major interest owing to its relevance for sleep and wake disorders.
- A deeper understanding of physiologic dysfunctions will help clinicians explain to their patients how to cope with sleep disorders, a process that is instrumental for patient satisfaction and well-being.
- Broader knowledge of physiology will assist sleep science and health stakeholders to clarify new and relevant priorities for basic and clinical research and public health investigations and policy making.
- Bidirectional effective communication and knowledge transfer between clinicians and researchers will benefit patients and, at the same time, will keep clinicians abreast of the latest developments in the field of sleep medicine and better prepare them to cope with the growing challenges in health care systems.

In this new edition of *Principles and Practice of Sleep Medicine,* the chapters on physiology (Section 3) contain a substantial amount of updated knowledge that can be expected to greatly expand the current understanding of sleep disturbances and to help identify and treat persons affected by them. This section includes new, in-depth information that will raise students' awareness of the relevance of the basic sciences and their application to sleep medicine.

New content has been included on how brain imaging has contributed to the understanding of sleep disruption and memory loss, how the cardiac control system regulates sleep and reacts to sleep disruptions, how respiration is related to the heart function as well as the relevance for health, and how the heart and the brain rest and regenerate, as well as the role of altitude in sleep breathing problems; the latest evidence on the interactive effects of insulin and other hormones in normal and deprived or disordered sleep is reviewed as well. Changes in sensory perception and pain during sleep due to the potential deleterious influence of opioids on sleep architecture and breathing also are covered. Finally, sleep-wake disturbances associated with traumatic brain injury are addressed, highlighting the importance of sleep for cognitive functioning and well-being. These chapters contain a wealth of background knowledge and are meant as primers for understanding the broad clinical content of this textbook.

One comment frequently encountered at our lectures on sleep physiology relates to its relevance for the practicing physician or researcher: "Why do doctors need to understand sleep physiology?" First, for doctors, whether in basic or clinical science, a primary objective is to improve public health by promoting and restoring healthy sleep patterns. A thorough knowledge of sleep physiology is essential for developing and validating more accurate diagnostic tools, for developing innovative patient management systems, and for keeping abreast of the latest advancements in sleep medicine. As they read these chapters on sleep physiology, some clinicians may be reminded of their early training days. Nevertheless, physiology is central to many current advances in clinical practice (e.g., breathing and cardiovascular measurements, brain imaging), and it forms the basis for translating genetics and proteomics into innovative diagnostic and therapeutic methods.

Because the adverse impact of poor sleep and a number of sleep disorders as well as overall societal and economic health are interrelated, government health agencies tend to seek discoveries that can improve the public's quality of life and reduce health care costs and mortality. A more thorough understanding of sleep physiology has resulted in innovative pharmaceutical therapies and improved designs for diagnostic and therapeutic devices. These include new recording and scoring systems, continuous positive airway pressure devices and mandibular advancement appliances with intelligent chips to monitor treatment adherence and assess effectiveness, and nerve and brain stimulation techniques to manage sleep-disordered breathing and sleep disturbances. Government agencies generally grant permission to market new products on the basis of epidemiologic data, recognition of the physiologic and pathologic mechanisms involved, and the findings of randomized controlled trials to determine product efficacy and safety. Research agencies are more prone to support pathophysiology-based projects that test hypothesized causes of diseases and disorders to rapidly develop innovative therapies (e.g., to improve mental health in North America). The lack of objective and valid measures to assess sleep improvement or therapeutic efficacy and effectiveness can retard the introduction of innovative developments into sleep medicine practices.

Although questionnaires are used to screen patients for many sleep disorders, physiologic (e.g., oxygen and breathing rate, hormone and endocrine release, electrocardiogram heart rate, electroencephalography brain activity) and psychophysiologic (e.g., reaction time, multiple sleep latency test, sensory perception) measures are used to confirm the accuracy and validity of the responses. Are these physiologic tools perfect? The answer is no. Some need further verification in terms of ecologic validity and the public health consequences of a given physiologic outcome (e.g., morbidity, mortality, cost for payers). Furthermore, it is essential to conduct randomized clinical trials and effectiveness studies with limited sample sizes but with stricter selection criteria, in which questionnaire-based information is combined with clinical findings and genetic findings or physiologic outcomes, and using more powerful analysis methods. These methods would go far beyond the usual Student's *t*-test and analysis of variance (ANOVA). They will require massive calculation capacity to enable finer discrimination of disease states and the most influential variables.

The identification of polymorphic genes associated with specific sleep disorders is another area of intense interest (covered in various chapters in Section 4). With the advent of proteomics, sleep medicine has entered the postgenomic era. *Proteomics* refers to large-scale protein characterization in relation to biologic processes, disorders, and diseases. Most sleep disorders, including sleep breathing disorders (e.g., sleep apnea), parasomnias (e.g., sleepwalking, enuresis, REM behavior disorder [RBD]), sleep-related movement disorders (e.g., restless legs syndrome/periodic limb movements), and circadian rhythm sleep disorders, are associated with genetic and/or molecular deficits that may define potential therapeutic avenues. Genetic epidemiology is a growing field that integrates all aspects of physiology in the aim of identifying further targets (gene loci) along with their individual and environmental variance.

Physiology provides the tools for phenotyping sleep disorders, and genetics advances the clinical domain by identifying risk and vulnerability factors as well as endophenotypes (e.g., interactions with environmental influences). Interindividual trait differences can be identified and phenotypic determinants of vulnerability recognized. Integrated analyses of a disorder's characteristics and outcome variables can form the basis for innovative approaches to sleep therapy, lead to a better understanding of why they work, and explain why some patients are nonresponders.

Large sample size population and genetic study findings using limited physiologic outcomes need to be correlated with the findings of smaller sample size studies, and the variable sets liable to best predict the outcomes of relevance in a given population presenting with a specific condition (e.g., chronic fatigue syndrome) need to be identified.[1,2] Quantitative models also can be applied to assess the strength of sleep medicine hypotheses.[3] Cell to system-wide integration and the reverse, system to cell-wide integration, are among the most promising approaches to track emerging factors and to integrate known factors into the body of knowledge,

leading to innovative diagnostic and therapeutic methods. Research and public health stakeholders have led the movement toward patient-oriented research, which is no longer a debatable issue; it has become a reality and is instrumental in preventing the transition of acute and simple conditions to complex, severe, or chronic conditions that require specialized tertiary care.

A complete reference list can be found online at ExpertConsult.com.

Chapter
12

What Brain Imaging Reveals About Sleep Generation and Maintenance

Eric A. Nofzinger; Pierre Maquet

Chapter Highlights

- The development of neuroimaging techniques has made it possible to characterize regional cerebral function in humans under a variety of sleep-related conditions. These techniques were first used to characterize brain activity throughout the sleep-wake cycle in normal human subjects.
- Regional brain activity during sleep is segregated and integrated within cortical and subcortical areas differently than during wakefulness.
- Regional brain activity is influenced by incoming stimuli as well as by previous waking experience.
- Functional neuroimaging identified the neural correlates of sleep-wake regulation by homeostatic sleep pressure and the nonvisual effects of light.
- Finally, functional imaging of patients with sleep disorders and in response to treatment interventions indicated reliable changes in neural systems across the sleep-wake cycle.

Functional neuroimaging consists of all techniques that can generate images of brain activity. In humans, such techniques usually include single photon emission computed tomography (SPECT), positron emission tomography (PET), functional magnetic resonance imaging (fMRI), optical imaging, diffusion tensor imaging (DTI), multichannel electroencephalography (EEG), and magnetoencephalography (MEG). Each technique has its own advantages and drawbacks in terms of spatial and temporal resolutions, accessibility, safety, and cost. For instance, EEG and MEG record brain oscillations with an excellent temporal resolution (usually on the order of a millisecond), but their localizing capacity is limited because a method of accurately modeling the electric or magnetic sources of the signal remains unrealized. By contrast, PET and MRI provide excellent spatial resolution (a few millimeters), but they are based on the measure of hemodynamic or metabolic parameters, which reduces their temporal resolution from a few seconds to many minutes. Accordingly, a comprehensive understanding of brain function probably requires human brain function to be characterized using as many techniques as possible.

This chapter summarizes the main advances made using functional neuroimaging in the current understanding of human sleep and its intimate relationships with waking performance and cognition, both in normal healthy sleepers and in patients with sleep disorders. These advances are considered in the context of four main topics: the characterization of regional brain activity during normal human sleep, the neural correlates of the regulation of sleep-wake cycle by circadian influences and nonclassical photoreception, the regional brain function in conditions of increased sleep pressure, and some clinical applications to the management of sleep disorders.

FUNCTIONAL SEGREGATION AND INTEGRATION DURING NORMAL HUMAN SLEEP

Preclinical research has identified the basic circuits in the brain that are responsible for promoting arousal. In general, reduction in activity in these systems is essential for generating and maintaining sleep. A major component of this network is the brainstem reticular core, a diffuse network of predominantly glutamatergic long-projecting neurons and a smaller collection of presumed local circuit gamma-aminobutyric acid (GABA) neurons. Additional components include a collection of nuclei in the brainstem tegmentum dorsal to the reticular formation that include cholinergic and monoaminergic (serotoninergic, noradrenergic, and dopaminergic) neurons. These nuclei send rostral projections that parallel to and are interconnected with those of the brainstem reticular core. Cholinergic nuclei also are clustered rostrally in the basal forebrain, including septal nuclei and the diagonal band of Broca.

Research attention has focused on a significant role for the hypothalamus in arousal and in regulating transitions between sleep and waking states. Specifically, the tuberomamillary histaminergic neurons and the perifornical hypocretin neurons in the posterior hypothalamus have extensive interconnections and interactions with the basic arousal systems (for more information, see Chapters 7, 8, 23, 34, and 39). Activity changes in these primary regulating areas result in profound modifications in activity patterns in thalamocortical circuits and associated structures, such as basal ganglia or cerebellum. A primary aim of functional imaging studies has been to characterize this reorganization of regional brain function during normal human sleep as well as the responses to external stimuli and the influence of previous waking experience on

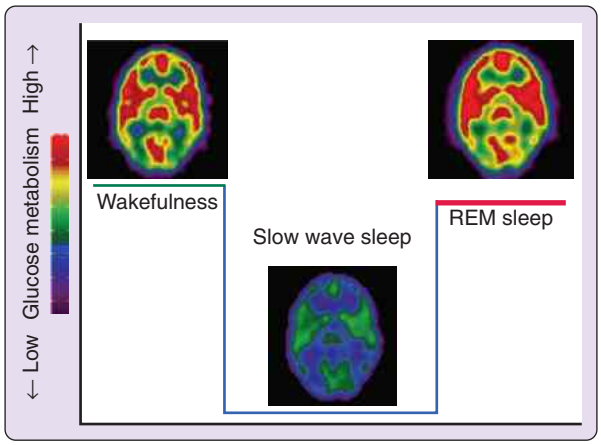


Figure 12-1 Schematic representation of the variations in global cerebral glucose metabolism in resting wakefulness, slow wave sleep, and REM sleep. The images represent the cerebral glucose metabolism measured in a single subject during three different sessions with fluorodeoxyglucose F-18 positron emission tomography (^{18}F-FDG PET). Functional images are displayed at the same brain level and using the same color scale. Similar rates of brain glucose metabolism are measured during wakefulness and REM sleep. Brain glucose metabolism is significantly decreased during slow wave sleep relative to that in both wakefulness and REM sleep. (Modified from Maquet P, Dive D, Salmon E, et al. Cerebral glucose utilization during sleep-wake cycle in man determined by positron emission tomography and [$_{18}$F]2-fluoro-2-deoxy-D-glucose method. *Brain Res* 1990;513:136-43.)

regional brain activity during sleep. The results of these studies, summarized in Table 12-1, are detailed next (Figure 12-1).

Non–Rapid Eye Movement Sleep

Neurophysiologic recordings in sleeping animals indicate that during non–rapid eye movement (NREM) sleep, the neural activity of the brain is shaped by a slow rhythm (<1 Hz), characterized by a fundamental oscillation of membrane potential made up of a depolarizing phase, associated with important neuronal firing ("up" state), followed by a hyperpolarizing phase, during which cortical neurons remain silent for a few hundred milliseconds ("down" state).[1,2] The slow oscillation occurs synchronously in large neuronal populations in such a way that it can be reflected on EEG recordings as high-amplitude, low-frequency waves.[1,3-5] The slow rhythm entrains other sleep oscillations in a coalescence of multiple rhythms.[6] Among the latter, spindles are associated with burst firing in thalamocortical populations. They arise from a cyclic inhibition of thalamocortical neurons by reticular thalamic neurons, which elicits postinhibitory rebound spike bursts in thalamocortical cells, which in turn entrain cortical populations in spindle oscillations.[7]

At the macroscopic systems level, the measure of brain metabolism or hemodynamics by PET usually requires the integration of the brain activity over extended time periods (from tens of seconds for $H_2{}^{15}O$-PET to 45 minutes for fluorodeoxyglucose F-18 (^{18}F-FDG) PET, resulting in the averaging of brain activity over the up and down states described earlier. Consequently, NREM sleep has systematically been associated with lower brain energy metabolism than that seen in wakefulness.[8] Relative to wakefulness, cerebral glucose and oxygen metabolism, as well as cerebral blood flow, is decreased by 5% to 10% during stage 2 sleep[9,10] and by 25% to 40% during slow wave sleep (SWS)—that is, stage 3 to 4 NREM sleep (see also Chapter 13).[11-13]

These decreases are not homogeneous but show a reproducible regional distribution. It is thought that brain areas with a high proportion of neurons committed in synchronous sleep oscillations are likely to have the lowest regional activity.[8] During light NREM sleep, cerebral blood flow decreases in the pons and thalamic nuclei, as well as in frontal and parietal areas but is maintained in the midbrain.[14] Consistent with the implication of thalamic nuclei in the generation of spindles, thalamic blood flow during stage 2 sleep decreases in proportion to the power density within the sigma frequency range (12 to 15 Hz).[15] During SWS, the most consistent decreases were observed in areas playing a permissive or active role in generating NREM sleep and its characteristic oscillations. These areas are the dorsal pons and mesencephalon, thalami, basal forebrain, and hypothalamus. The topography of the decreases in cortical blood flow during NREM sleep also is very reproducible and encompasses the prefrontal cortex, anterior cingulate cortex, the precuneus, and the mesial aspect of the temporal lobe.

The mechanisms that produce these diminutions in regional blood flow are not completely understood. The frontal and parietal polymodal associative cortices are among the most active brain structures during wakefulness.[8] Consequently, homeostatic sleep pressure locally accrued during a normal waking day generally is thought to be particularly important in these cortical areas, resulting during NREM sleep in a considerable amount of local slow waves, with a substantial decrease in local energy metabolism.[8] Significant decreases in blood flow also were unexpectedly observed in the cerebellum and basal ganglia. It is currently not known whether these changes in blood flow indicate that these two areas have a role in generating and maintaining cortical oscillations or if they are merely entrained by the cortical slow rhythm.

Advances in event-related EEG and fMRI have allowed a finer-grained characterization of brain activities associated with transient events, such as spindles[16] or slow waves. In humans, some evidence suggests the occurrence of two different types of spindles during sleep, slow and fast EEG spindles, which differ by their scalp topography and some aspects of their regulation.[17] Both spindle types trigger significant activity in the thalami, the anterior cingulate and insular cortices, and superior temporal gyri. Beyond the common activation pattern, slow spindles (frequencies of 11 to 13 Hz) are associated with increased activity in the superior frontal gyrus. By contrast, fast spindles (13 to 15 Hz) recruit a set of cortical regions involved in sensorimotor processing and recruit the mesial frontal cortex and hippocampus.

During SWS, human EEG recordings are characterized by high-voltage low-frequency oscillations whose classification is not always clear. The power density in the 0.75- to 4-Hz frequency band, usually referred to as *slow wave activity* (SWA), has proved to be a very useful and popular parameter because it quantifies the dissipation of homeostatic sleep pressure during NREM sleep.[18] However, its frequency bounds do not respect the dichotomy between slow (<1 Hz) and delta rhythms (1 to 4 Hz), which is based on differences in the respective cellular correlates of these oscillations in animals.[6] In the temporal domain, the amplitude of SWS waves is classically larger than 75 μV.[19] However, the largest waves (>140 μV) have been taken as realizations of the slow oscillation (<1 Hz).[20,21] Transient increases in brain activity associated with slow (>140 μV) and delta EEG waves (75 to

Table 12-1 Summary of the Modifications in Global and Regional Hemodynamic or Metabolic Parameters during NREM and REM Sleep

	Deep NREM Sleep					REM Sleep			
	Kety-Schmidt*		PET		fMRI	Kety-Schmidt		PET	
Area	Oxygen Metabolism	Blood Flow	Glucose Metabolism	Blood Flow	BOLD Signal	Oxygen Metabolism	Blood Flow	Glucose Metabolism	Blood Flow
Global Changes									
	↓ relative to W and REM sleep	↓ relative to W and REM sleep	↓ relative to W and REM sleep	↓ relative to W and REM sleep		≅ W	≅ W	≅ W	≅ W
Regional Changes									
Lateral frontal				↓ relative to W	↑ in response to slow waves				↓ relative to W
Medial prefrontal (including orbito-frontal)					↑ in response to slow waves			↑ relative to W	
Lateral parietal				↓ relative to W					↓ relative to W
Medial parietal (including precuneus)				↓relative to W	↑ in response to slow waves				↓ relative to W
Temporo-occipital			↑ relative to W					↑ relative to W	
Anterior cingulate cortex									↑ relative to W or SWS
Medial temporal lobe (hippocampus, parahippocampus)					↑ in response to slow waves				↑ relative to W or SWS
Medial temporal lobe (amygdala)									↑ relative to W or SWS
Thalamus				↓ relative to W					↑ relative to W or SWS
Basal ganglia				↓ relative to W					
Pons, midbrain				↓ relative to W	↑ in response to slow waves				
Cerebellum				↓ relative to W	↑ in response to slow waves				

↑, Increase; ↓, decrease; ≅, approximately the same activity; BOLD, blood oxygen level–dependent; fMRI, functional magnetic resonance imaging; PET, positron emission tomography; SWS, slow wave sleep; W, wakefulness.

*Kety-Schmidt: A method for measuring organ blood flow, first applied to the brain in 1944 by CF Schmidt and SS Kety. A chemically inert indicator gas is equilibrated with the tissue of the organ of interest, and the rate of disappearance from the organ is measured. Blood flow is calculated on the assumption that the tissue and venous blood concentrations of the indicator gas are in diffusion equilibrium at all blood flow rates and that the rate of disappearance of the indicator from the tissue is a function of how much of the indicator is in the tissue at any time; the rate of disappearance is assumed to be exponential.

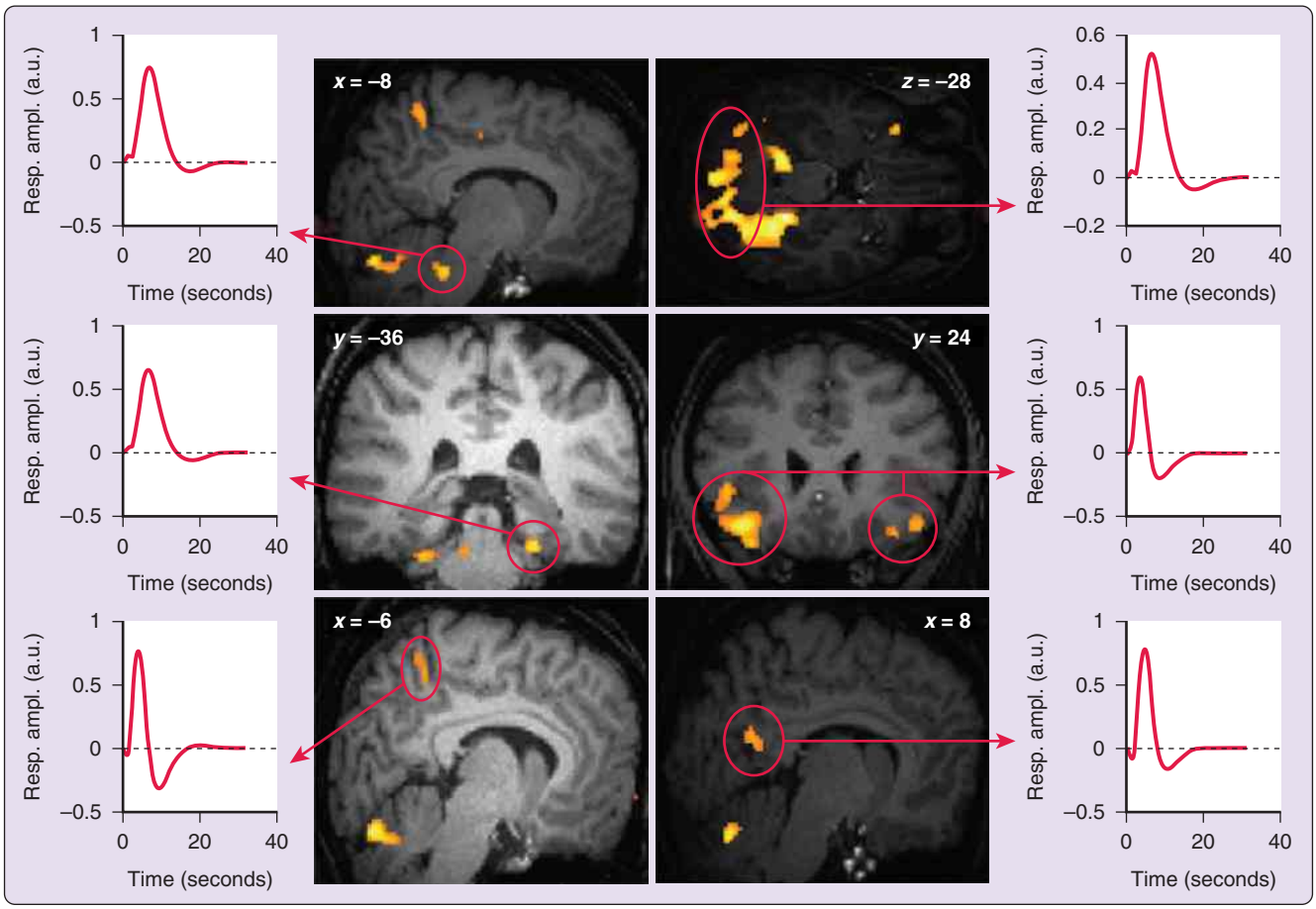


Figure 12-2 Brain regions activated in relation to slow oscillation (both high-amplitude slow waves and delta waves), as observed with combined EEG and functional MRI. *Structural images—center panels,* Significant responses are associated with both slow and delta waves. Functional results are displayed on an individual image (display at $P < .001$, uncorrected) at different levels of the *x, y,* and *z* axes as indicated for each section. *Side panels (left and right),* Time course (in seconds) of fitted response amplitudes (in arbitrary units [a.u.]) during slow waves or delta waves in the corresponding circled brain area. All responses consisted in regional increases of brain activity. From *left to right* and then *top to bottom*: pontine tegmentum, cerebellum; right parahippocampal gyrus, inferior frontal gyrus; precuneus, posterior cingulate cortex. EEG, Electroencephalogram; MRI, magnetic resonance imaging. (Modified from Dang-Vu TT, Schabus M, Desseilles M, et al. Spontaneous neural activity during human slow wave sleep. *Proc Natl Acad Sci U S A* 2008;105:15160-5.)

140 μV) can be detected in the pontine tegmentum (in an area encompassing the locus coeruleus), midbrain, and cerebellum in several cortical areas including inferior frontal, medial prefrontal, precuneus, and posterior cingulate parahippocampal gyrus,[22] areas that have been shown to be current sources underpinning human slow waves.[23]

As compared with baseline activity, slow waves are associated with significant activity in the parahippocampal gyrus, cerebellum, and brainstem, whereas delta waves are related to frontal responses. These findings show that NREM sleep is not a state of brain quiescence but is an active state during which neural activity is consistently synchronized by sleep oscillations (spindles, slow rhythm) in specific cerebral regions. Electrophysiologic and computational evidence from the cortex and thalamus indicates that the tonic firing pattern and fluctuations of the membrane potential during slow oscillation up states are similar to those characteristic of the waking state, suggesting that the up state is a ubiquitous feature of neuronal dynamics in corticothalamic networks reproducing a "micro" wake-like state that facilitates neuronal interactions.[24] The partial overlap between the regional activity pattern related to SWS waves and the waking default mode network is consistent with the hypothesis that the brain responses synchronized by the slow oscillation restore micro wake-like activity patterns[22] (Figure 12-2).

Processing of External Stimuli during NREM Sleep

To date, only a few neuroimaging studies have investigated how the human brain processes external stimuli during sleep, and their results are controversial. In sedated young children, examined with fMRI, visual stimulation elicited a paradoxical decrease in response in the anterior aspect of the medial occipital cortex.[25] Although the neurobiologic significance of this enigmatic finding remains elusive, it has been replicated in naturally sleeping adults during SWS, using fMRI and PET.[26] This response pattern does not seem to be specific to visual stimulations because it also is observed when auditory stimuli are delivered during NREM sleep.[27]

In sharp contrast with these results, other data suggest that the brain still processes auditory stimuli up to cortical areas during NREM sleep.[28] Significant responses to auditory stimuli were detected in bilateral auditory cortex, thalamus,

and caudate nuclei during wakefulness and light NREM sleep. In addition, the left amygdala and the left prefrontal cortex are recruited by stimuli of particular affective significance for the individual subject. (More information on how the brain process sensory input during sleep can be found in Chapter 23.)

REM Sleep

Reorganization of Regional Brain Function during REM Sleep: Relation with Dream Characteristics

The level of energy metabolism recorded during REM sleep is similar to waking levels.[12,13] The distribution of regional brain activity, however, significantly differs from that in wakefulness. In addition to areas known to participate in generation and maintenance of REM sleep (e.g., pontine tegmentum, thalamic nuclei), significant activations of limbic and paralimbic areas (e.g., amygdaloid complexes, hippocampal formation, anterior cingulate cortex, orbitofrontal cortex) are reported during REM sleep[11,29,30] (Figure 12-3).

Although this observation is not reported in all studies, posterior cortices in temporal occipital areas typically are activated during REM sleep.[11] By contrast, the dorsolateral prefrontal cortex, parietal cortex, posterior cingulate cortex, and precuneus are the least active brain regions.[11,29] Although early animal studies had already mentioned the high level of limbic activity during REM sleep, functional neuroimaging in humans highlighted the contrast between the activation of limbic, paralimbic, and posterior cortical areas on the one hand and the relative quiescence of the associative frontal and parietal cortices on the other.

Regional functional integration also is modified during REM sleep relative to wakefulness. For instance, the functional interactions between striate and extra-striate cortices, which are positive during wakefulness, become negative during REM sleep.[31] Likewise, the functional connectivity between the amygdala and temporal occipital areas is tighter during REM sleep than during resting wakefulness.[32]

The organization of human brain function during REM sleep somehow relates to some of the characteristics of dreaming activity.[29,33,34] The perceptual aspects of dreams would be related to the activation of posterior (occipital and temporal) cortices, whereas emotional features in dreams would be related to the activation of amygdalar complexes, orbitofrontal cortex, and anterior cingulate cortex. The recruitment of mesiotemporal areas would account for the memory content commonly observed in dreams. The relative hypoactivation of

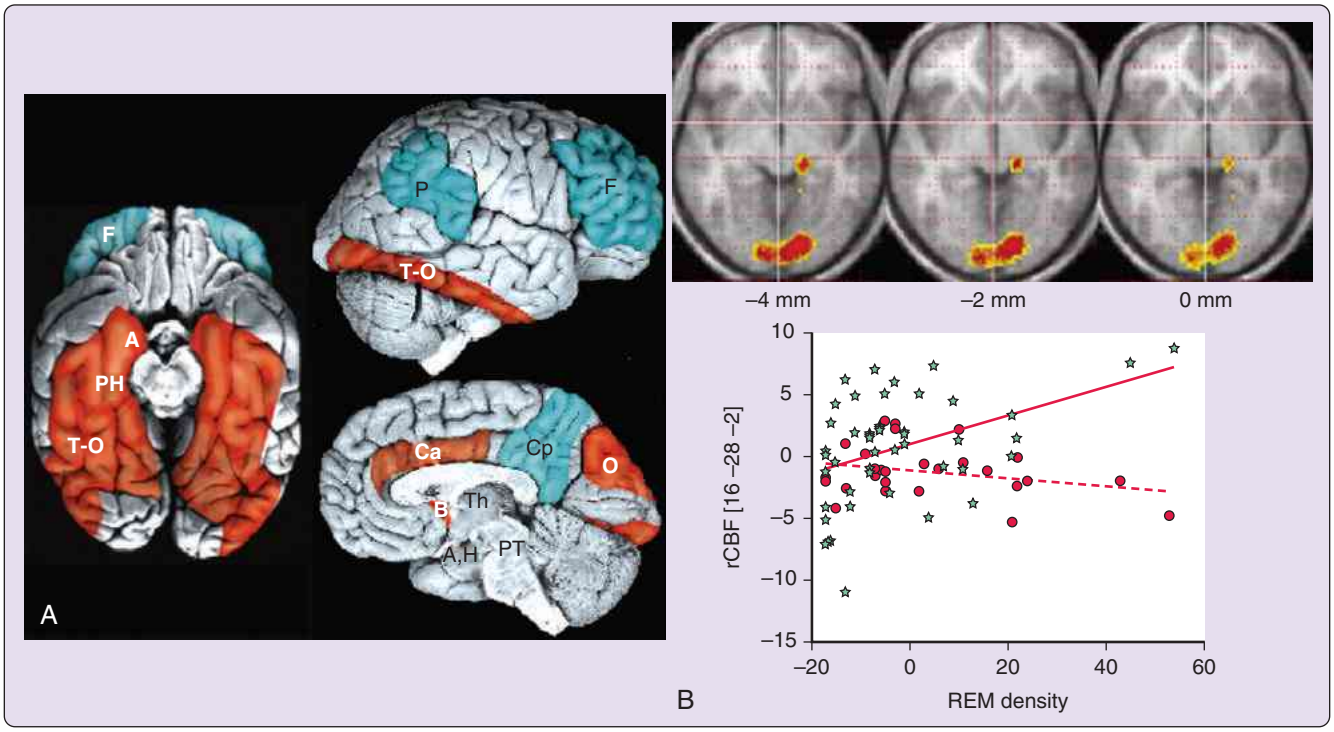


Figure 12-3 **A,** Schematic representation of the relative increases and decreases in neural activity associated with REM sleep, as observed with positron emission tomography. **B,** *Top*, Cerebral areas more active in relation to rapid eye movements during paradoxical sleep than during awake states. Transverse sections are from 24 to 0 mm from the bicommissural plane. The functional data are displayed at $P < .001$ uncorrected, superimposed on the average magnetic resonance imaging scan of the brain of sleeping subjects, coregistered to the same reference space. *Bottom,* Plot of the adjusted regional cerebral blood flow (rCBF), in arbitrary units, in the right geniculate body in relation to the rapid eye movement counts. The geniculate cerebral blood flow is correlated more strongly with the rapid eye movement counts during REM sleep (*red circles*) than during wake (*green stars*). A,H, Amygdala and hyppocampus; B, basal forebrain; Ca, anterior cingulate gyrus; Cp, posterior cingulate gyrus and precuneus; F, prefrontal cortex; H, hypothalamus; M, motor cortex; P, parietal supramarginal cortex; PH, parahippocampal gyrus; PT, pontine tegmentum; O, occipital-lateral cortex; Th, thalamus; T-O, temporooccipital extrastriate cortex. (**A**, Modified from Schwartz S, Maquet P. Sleep imaging and the neuro-psychological assessment of dreams. *Trends Cogn Sci* 2002;6:23-30. **B**, Modified from Peigneux P, Laureys S, Fuchs S, et al. Generation of rapid eye movements during paradoxical sleep in humans. *Neuroimage* 2001;14:701-8.)

the prefrontal cortex might help to explain the alteration in logical reasoning, working memory, episodic memory, and executive functions characterizing dream reports collected from experimentally induced REM sleep awakenings. Activation of the anterior cingulate cortex and surrounding mesial prefrontal cortex has been described, in studies on waking cognitive neuroscience, to be related to self-referential cognition and to the monitoring of performance. Activation of these structures within REM sleep might represent a role for REM sleep in the internal monitoring of aspects of the self, especially those having emotional significance, in keeping with the activation of other related limbic and paralimbic structures.[33]

Brain Imaging and Other Characteristic Features of REM Sleep

Rapid eye movements constitute a prominent feature of REM sleep. Cerebral mechanisms underpinning the generation of spontaneous ocular movements differ between REM sleep and wakefulness in humans. Regional cerebral blood flow changes in the lateral geniculate bodies and in the striate cortex are significantly more correlated with ocular movement density during REM sleep than during wakefulness,[35] a pattern that was later confirmed using fMRI.[36] This pattern of activity is reminiscent of ponto-geniculo-occipital (PGO) waves, prominent phasic bioelectrical potentials associated with eye movements, which occur in isolation or in bursts just before and during REM sleep and are most easily recorded in cats and rats in the mesopontine tegmentum, the lateral geniculate bodies, and the occipital cortex.[37]

Another important feature in REM sleep is the instability in autonomic regulation and especially in cardiovascular regulation. During awake states, the right insula is involved in cardiovascular regulation,[38] but during REM sleep, the variability in heart rate is related to the activity in the right amygdaloid complex.[39] The functional connectivity between the amygdala and the insular cortex, two brain areas involved in cardiovascular regulation, differ significantly in REM sleep as compared with wakefulness.[39] These results suggest a functional reorganization of central cardiovascular regulation during REM sleep (see also Chapter 13).

Experience-Dependent Modifications of Regional Brain Function during NREM and REM Sleep

Waking experience substantially influences regional brain activity during subsequent sleep. For instance, the blood flow of the hippocampus and parahippocampal gyrus during NREM sleep is increased in subjects who were navigating a virtual town (by means of a computer software program) during the previous waking period, as compared with naive participants.[40] The level of hippocampal activity expressed during SWS positively correlates with the improvement of performance in route retrieval the next day, suggesting that hippocampal activity during NREM sleep is related to subsequent "offline" processing of spatial memory.[40] Similarly, several brain areas activated during the execution of a serial reaction time task during wakefulness (brainstem, thalamus, and occipital, parietal and premotor areas) are significantly more active during REM sleep in subjects previously trained on the task than in nontrained subjects.[41] This enhancement of regional brain activity during posttraining REM sleep is observed only when the learning material is presented in a structured manner, versus random presentation.[42] It is then suggested to be associated with significant changes in brain functional connectivity.[43] Collectively, these results support the hypothesis that the memory of a motor sequence is further processed during REM sleep in humans (Figure 12-4). (For more information on memory processing in relation to sleep, see Chapter 23.)

BRAIN IMAGING AND NEURAL CORRELATES OF SLEEP-WAKE CYCLE REGULATION

The timing of sleep and wake episodes is thought to be regulated by the interaction between the homeostatic sleep pressure and an intrinsic circadian oscillation.[44] One study compared regional relative brain glucose metabolism rates for morning and evening wakefulness in healthy humans to define the mechanisms that maintain wakefulness across the day, in relation to the increasing sleep drive that accumulates over the wake period (see Chapter 36 for more information on sleep and wake process). Brain scans, using ^{18}F-FDG PET, were conducted during quiet wakefulness in the morning and in the evening in 13 healthy adults (10 women, 3 men) (mean age, 37 years). As expected, subjective ratings of alertness were lower in the evening than in the morning. Conversely, relative regional glucose metabolism was significantly higher in the evening than in the morning in a large cluster of midline and brainstem structures. More specifically, changes were localized in the pontine and midbrain reticular formation, midbrain raphe, locus coeruleus, and posterior hypothalamus. Of note, evening wakefulness is associated with increased relative metabolism in brainstem and hypothalamic arousal systems and decreased relative metabolism in posterior cortical regions. These patterns might reflect input from the circadian timing system(s) to promote wakefulness, or they might reflect the effects of increasing homeostatic sleep drive (see Chapter 36 and other chronobiology chapters in Section 5).

Using fMRI, two studies capitalized on interindividual differences in sleep-wakefulness regulation to assess the impact of sleep homeostasis and circadian factors on regional brain function. A first study investigated executive responses to a working memory task during a normal sleep-wake cycle and during sleep loss in a population of young healthy volunteers stratified according to homozygosity for a variable-number (4 or 5) tandem-repeat polymorphism in the coding region of the clock gene *PERIOD3 (PER3)*.[45] Homozygosity to the long allele (5/5) is associated with faster buildup of sleep homeostasis and increased vulnerability to sleep-loss.[46] In the less vulnerable genotype (4/4), no changes were observed in brain responses during the normal sleep-wake cycle. During sleep loss, these subjects recruited supplemental anterior frontal, temporal, and subcortical regions, while executive function was maintained. By contrast, in the vulnerable genotype (5/5), activation in a posterior prefrontal area was already reduced in comparison of evening to morning responses during a normal sleep-wake cycle. Furthermore, in the morning after a night of sleep loss, widespread reductions in activation in prefrontal, temporal, parietal, and occipital areas were observed in this genotype. Such differences occurred in the absence of genotype-dependent differences in circadian phase. These findings clearly showed that the allocation of prefrontal resources is constrained by sleep pressure and circadian phase.

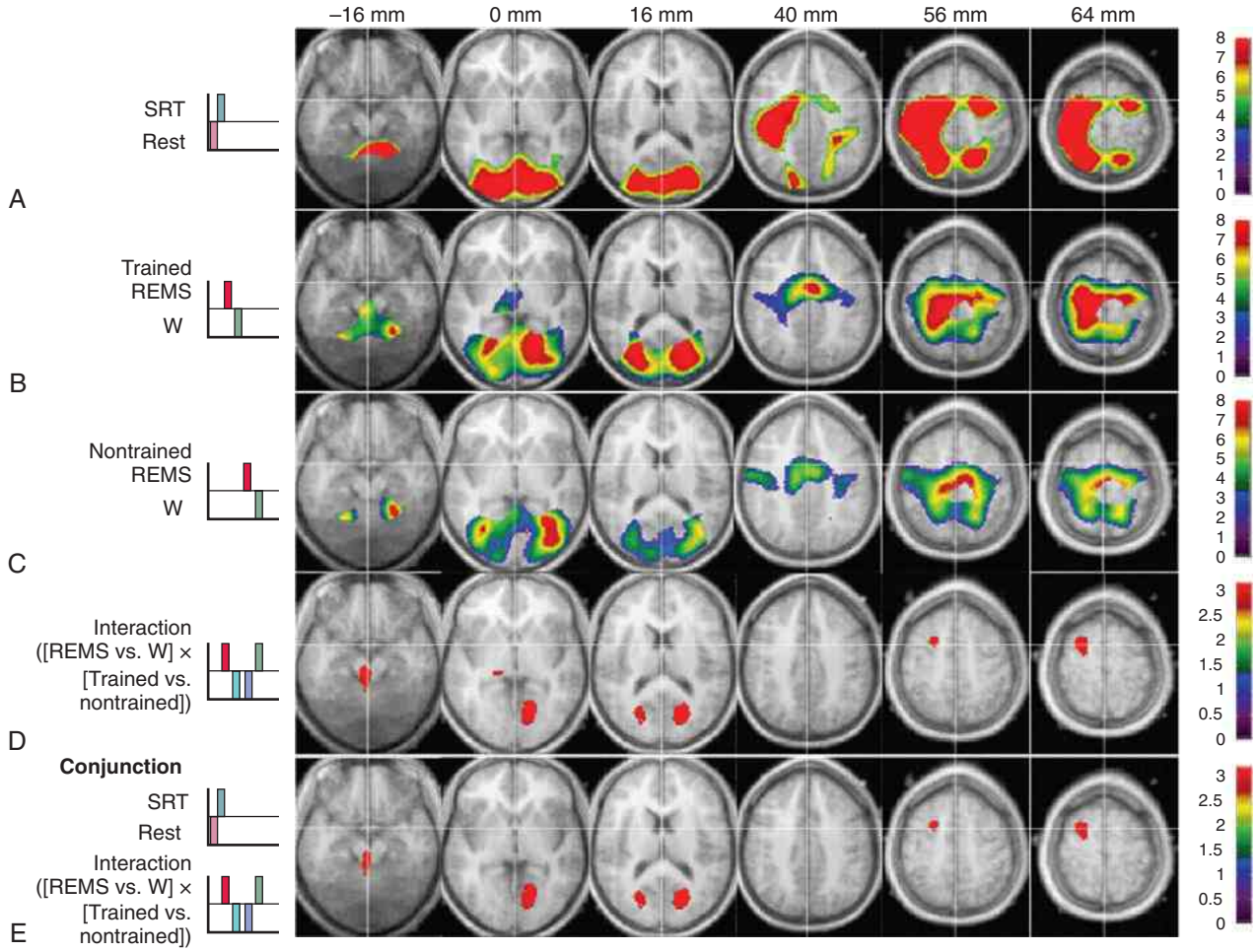


Figure 12-4 Influence of previous waking experience, in this case a procedural motor learning sequence, on the distribution of regional brain activity during subsequent REM sleep (REMS). Statistical parametric maps of different contrasts are displayed at six different brain levels (from 16 mm below to 64 mm above the bicommissural plane), superimposed on the average (coregistered and normalized) magnetic resonance imaging scan of the brain in sleeping subjects. All maps were thresholded at $P < .001$ (uncorrected), except for **A,** which was thresholded at voxel level–corrected ($P < .05$). **A,** Brain regions activated during performance of a serial reaction time (SRT) task during wakefulness (SRT-rest). **B,** Brain regions activated during REM sleep (REM sleep–wakefulness) in subjects previously trained to the SRT task. **C,** Brain regions activated during REM sleep (REM sleep–wakefulness) in nontrained subjects. **D,** Brain regions activated more in trained subjects than in nontrained subjects during REM sleep—that is, the intersection of the condition (REM sleep versus wakefulness) by group (trained versus nontrained). **E,** Brain regions that were both recruited during the execution of motor tasks and more activated in trained than in nontrained subjects scanned during REM sleep—that is, the conjunction of (SRT-rest) with the condition (REM sleep versus wakefulness) by group (trained versus nontrained). (Modified from Maquet P, Laureys S, Peigneux P, et al. Experience-dependent changes in cerebral activation during human REM sleep. *Nat Neurosci* 2000;3:831-6.)

The second fMRI study assessed the cerebral correlates of sustained attention in persons categorized as extreme early and late chronotype individuals, in the morning and in the evening (1.5 and 10.5 hours, respectively, after preferred waking time).[47] Early and late chronotypes are known to differ in terms of circadian phase, but early-morning types also are known to accrue a larger homeostatic sleep pressure buildup during wakefulness than is seen in evening types.[48]

Brain responses differed between chronotypes in the thalamus and areas compatible with the locus coeruleus and suprachiasmatic area. Furthermore, estimated fMRI blood oxygen level–dependent (BOLD) activity in the suprachiasmatic region was negatively related to the slow wave activity recorded during the subsequent first sleep cycle. This finding suggested that the activity profile of the suprachiasmatic nucleus (SCN) master clock acting on the cerebral correlates of sustained attention inherently depends on the status of the sleep homeostat. In other words, whereas evening types can benefit from the increasing circadian alertness signal to achieve optimal performance levels during the evening hours, morning types have to fight against a disproportionally increasing homeostatic sleep pressure at this time of the day and are therefore less able to profit from the beneficial circadian alertness signal to accomplish sustained attention performance.

The activity of the SCN, the master circadian clock, is influenced by external temporal markers (zeitgebers), the most important of which is light. In addition to vision, light profoundly affects human physiology and modulates sleep-wake cycles, body temperature, endocrine functions, alertness, and performance.[44] Animal and human studies demonstrated that

a *nonvisual* photoreception system mediates these effects, which include the synchronization of the circadian system, suppression of melatonin, regulation of sleep, and improvements in alertness and cognition.[49-53] This photoreception system recruits the retinal photoreceptors (rods and cones) and intrinsically photosensitive retinal ganglion cells expressing melanopsin.[54,55] These retinal ganglion cells project to numerous nuclei of the brainstem, hypothalamus, thalamus, and cortical structures; such anatomic connectivity suggests that the nonvisual system can influence many brain functions.

However, light has been shown to enhance cortical and subcortical responses induced by various cognitive challenges. Polychromatic bright white light (>7000 lux) was first shown to enhance responses to attention tasks in subcortical (hypothalamus and thalamus) and cortical areas during the night.[56] Similar results were observed during daytime.[57] These early experiments did not make it possible to specify the relative contribution of the different retinal photoreceptors involved. Monochromatic blue (470 nm) light was then shown to enhance brain responses to a working memory task in areas such as the thalamus and association cortices; it prevented the decline otherwise observed with green (550 nm) light exposure.[58] These results supported the potential involvement of melanopsin in eliciting these nonvisual modifications in brain responses. However, no definite claims could be made concerning the contribution of different photoreceptor classes. Subsequently, the relative contribution of S-cones, melanopsin-expressing ganglion cells, and M-cones to nonvisual brain responses to light was assessed using short-duration monochromatic light exposures of 430 nm (violet), 473 nm (blue), and 527 nm (green) wavelength, respectively.[59] Because light exposures were brief, this protocol allowed the detection of subcortical brain structures involved in early nonvisual responses to light, such as the thalamus and the brainstem.

It also appears that these immediate effects of ambient light on cognition in turn depend on circadian phase, homeostatic sleep pressure, and genotype.[60] An fMRI study showed that in a genotype reported to be more vulnerable to sleep loss (*PER3*$^{5/5}$), exposure to as little as 1 minute of low-intensity ambient blue light in the morning following sleep deprivation increases responses in a left thalamo-fronto-parietal circuit to a larger extent than does green light. By contrast, no impact of ambient light was observed during sleep loss in the genotype reported to be less vulnerable to sleep (*PER3*$^{4/4}$). These results support the view that the impact of light on cognitive brain function is especially present in challenging conditions when endogenous alerting mechanisms are not already active.

Collectively, these findings show that light, and especially blue light, not only can influence the timing of sleep and wake cycles but can also profoundly and swiftly influence regional brain function. The underlying mechanism is more likely to be a modulation of the activity in subcortical structures promoting alertness (e.g., anterior hypothalamus, mesopontine tegmentum, thalamus)[49-53] (Figure 12-5).

BRAIN IMAGING AND NEURAL CORRELATES OF HUMAN SLEEP DEPRIVATION

Sleep generation and maintenance are strongly affected by the homeostatic sleep drive. Accordingly, an understanding of the neural correlates of sleep deprivation, which increases the homeostatic sleep drive, can be expected to provide additional clues from functional neuroimaging about the mechanisms of sleep generation and sleep maintenance.

One study[61] assessed regional cerebral metabolism using the ^{18}F-FDG PET method in healthy subjects before and after 32 hours of sleep deprivation. The investigators noted prominent decreases in metabolism in the thalamus, basal ganglia, temporal lobes, and cerebellum and increases in visual cortex. Whole-brain absolute metabolic rate was not different. Another study[62,63] described the effects of 24, 48, and 72 hours of sleep deprivation on waking regional cerebral metabolism assessed using ^{18}F-FDG PET, as well as alertness and cognitive performance. Sleep deprivation was associated with global declines in absolute cerebral metabolism. Regionally, these declines were most notable in frontoparietal cortex and in the thalamus. These findings are consistent with studies showing that the effects of sleep deprivation on SWS are greatest in frontal EEG leads. Alertness and cognitive performance on a sleep deprivation–sensitive serial addition and subtraction test declined in association with the sleep deprivation–associated regional deactivations.

In other work, blood flow in the thalamus and pontomesencephalic tegmentum, as assessed by $H_2{}^{15}O$ PET, was found to positively correlate with frequency of arousals in sleep,[64] with performance on vigilance tasks,[65] and with loss of consciousness associated with anesthesia.[66] In some instances, this arousal network also included the basal forebrain and anterior cingulate cortex.[65] These findings regarding sleep deprivation support the role for sleep in restoring brain function in thalamocortical networks associated with higher-order cognition.

Another study[67] assessed changes in regional brain function during sleep recovery after sleep deprivation to define the specific neural correlates of the sleep recovery process. Homeostatic sleep need was modulated in a within-subjects design based on sleep deprivation. In four young adult healthy male subjects (mean age, 24.9 years ± 1.2 years), NREM sleep was assessed using ^{18}F-FDG PET after a normal night of sleep and again after 36 hours of sleep deprivation. Both absolute and relative regional cerebral glucose metabolic data were obtained and analyzed. In relation to baseline NREM sleep, the subjects' recovery NREM sleep was associated with increased slow wave activity, global reductions in whole-brain metabolism, and relative reductions in glucose metabolism in broad regions of frontal, parietal, and temporal cortex. These findings demonstrate that the homeostatic recovery function of sleep is associated with global reductions in whole-brain metabolism during NREM as well as greater relative reductions in broad regions of frontal, parietal, and temporal cortex. As indicated by these results, the homeostatic function of sleep in humans involves a reduction of glucose metabolism throughout the cortex. Neurobiologic models of sleep homeostasis, therefore, need to account for the inverse relationship between slow wave activity and cerebral glucose utilization and may suggest that the increased slow wave activity associated with recovery from sleep deprivation is a marker of an increased need for cerebral metabolic restoration.

FUNCTIONAL NEUROIMAGING IN SLEEP DISORDERS

Functional neuroimaging studies in patients with sleep disorders provide additional clues to the role of various brain

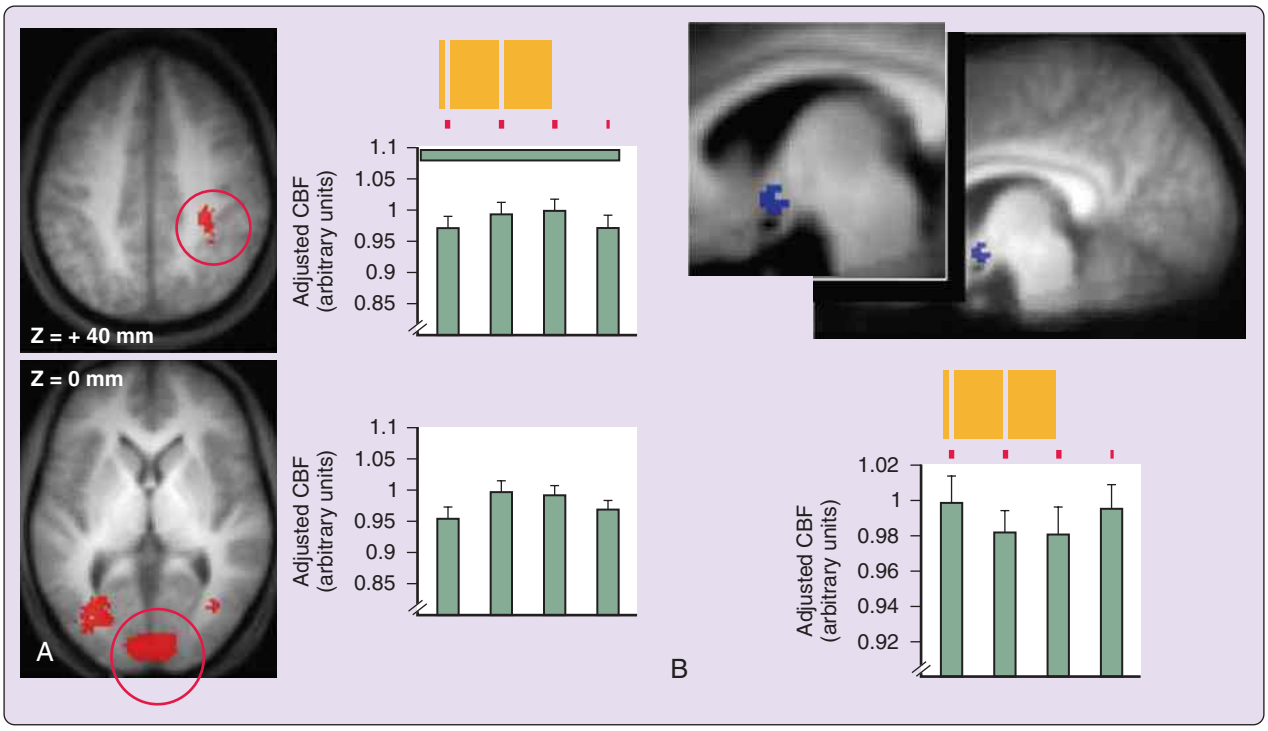


Figure 12-5 Effects on regional brain activity of exposure to bright white light during the night, as assessed with positron emission tomography. Regional cerebral blood flow (rCBF) was measured in subjects attending to auditory stimuli in near darkness after light exposures (>8000 lux) of different durations (0.5, 17, 16.5, and 0 min) during the biologic night. **A,** Parietal and occipital areas in which the regional brain blood flow is significantly increased in proportion to the duration of the previous exposure to light. In the images at *left,* functional data are displayed at $P < .05$ (voxel level), superimposed on the mean normalized MRI scan. *Right,* Plots of the adjusted regional cerebral blood flow in these areas. The duration of the light exposure preceding each scan is indicated by the *yellow boxes* (inset) above the activity estimates. **B,** Suprachiasmatic area in which the blood flow is significantly decreased in proportion to the duration of the previous exposure to light. In the images at *top,* functional data are displayed at $P < .001$ (uncorrected), superimposed on a parasagittal view of the mean normalized MRI scan (*x* coordinate, 2 mm). *Inset,* Enlargement of the hypothalamic area. The plot *below,* shows the corresponding adjusted blood flow for the four scans of the blocks. (Modified from Perrin F, Peigneux P, Fuchs S, et al. Nonvisual responses to light exposure in the human brain during the circadian night. *Curr Biol* 2004;14:1842-6.)

structures in generating and maintaining sleep. If the previously stated hypotheses regarding the role of various brain structures in these processes are correct, then these brain structures should function abnormally in some manner in patients who do not sleep well. Insomnia, for example, is a disorder in which patients have difficulty falling asleep or staying asleep or who have nonrestorative sleep along with daytime dysfunction. A survey of the functional neuroimaging findings in this disorder can be expected to provide additional clues, because new data are emerging on a regular basis about the neural mechanisms of sleep generation and maintenance in humans.

Although an overview of all neuroimaging studies in the full spectrum of all sleep disorders would be beyond the scope of this chapter, it bears mention here that extensive studies exist in circadian rhythm disorders, narcolepsy and the hypersomnia disorders, sleep-related breathing disorders, parasomnias, restless legs syndrome (RLS), and periodic limb movement disorder (PLMD) as well as neuroimaging studies reporting the effects of various treatments for sleep disorders on brain function. These findings have been reviewed in more extensive detail in the textbook *Neuroimaging of Sleep and Sleep Disorders,* by Nofzinger, Maquet, and Thorpy (2013), listed in the Selected Readings section of this chapter.

Insomnia and Brainstem and Hypothalamic Arousal Networks

Insomnia is the most prevalent of all sleep problems. *Insomnia* is the subjective complaint of difficulty falling asleep, difficulty staying asleep, poor quality sleep, or inadequate sleep duration despite having an adequate opportunity for sleep. Of importance, EEG sleep alterations in insomnia are not always found in persons with subjective complaints of insomnia. The leading neurobiologic model of insomnia is that of hyperarousal. For instance, people with insomnia have been shown to have elevated temperature and muscle tone at sleep onset, elevated heart rate and elevated sympathovagal tone in heart rate variability, and positive correlations between wake times after sleep onset and urinary norepinephrine and dopamine metabolites.[68-71] Studies of whole-body metabolic rate, assessed by oxygen consumption, show elevated rates for patients with insomnia compared with healthy control subjects, a difference that persists 24 hours per day.[72] The hyperarousal of insomnia is supported by higher rates of self-reported ruminations and intrusive thoughts among insomniac patients.

Perhaps more than for all other sleep disorders, brain imaging studies have been most helpful in elucidating the neural substrates of hyperarousal. An emerging body of evidence exists showing specific regional brain alterations despite

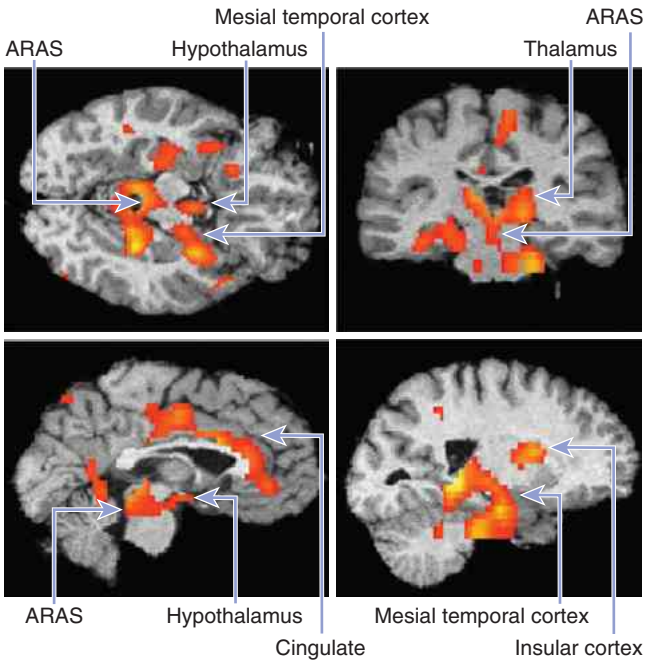


Figure 12-6 Brain structures that do not show decreased metabolic rate from waking to sleep in insomniacs. All regions shown reach statistical significance at the $P < .05$, corrected, level of significance in relation to healthy sleeper control subjects. ARAS, Ascending reticular activating system. (From Nofzinger EA, Buysse DJ, Germain A, et al. Functional neuroimaging evidence for hyperarousal in insomnia. *Am J Psychiatry* 2004;161:2126-31.)

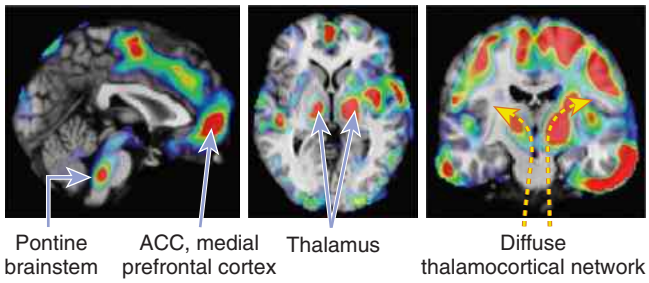


Figure 12-7 Brain structures in which metabolism during NREM sleep correlates with wakefulness after sleep onset in insomniac patients. ACC, Anterior cingulate cortex. (From Nofzinger EA, Nissen C, Germain A, et al. Regional cerebral metabolic correlates of WASO during NREM sleep in insomnia. *J Clin Sleep Med* 2006;2:316-22.)

the relative absence of EEG sleep signs of the disorder. These findings hold promise that the use of brain imaging in this disorder can considerably advance the current understanding of the disorder and lead to new mechanisms for treatment in a manner previously unattainable through traditional EEG sleep assessments.

Human sleep neuroimaging studies in insomniac subjects support the involvement of basic arousal networks in disturbances in NREM sleep (Video 12-1). Nofzinger and colleagues[73] investigated the neurobiologic basis of poor sleep in insomnia. Insomniac patients and healthy subjects underwent regional cerebral glucose metabolic assessments during both waking and NREM sleep using ^{18}F-FDG PET. Healthy subjects reported better sleep quality than did those with insomnia. The two groups did not differ on any measure of visually scored or automated measure of sleep. Grouping by state interaction analysis confirmed that insomniac subjects showed a smaller decrease than for healthy subjects in relative glucose metabolism from waking to NREM sleep in the brainstem reticular core and in the hypothalamus. This study supports the concept that persistent activity in this basic arousal network may be responsible for the impaired objective and subjective sleep in insomniac patients (Figures 12-6 and 12-7).

In terms of interventions, the action of sedative-hypnotics may be primarily on these basic arousal systems. For example, one study[74] assessed regional cerebral blood flow, as a correlate of neuronal activity, during NREM sleep in response to triazolam, a short-acting benzodiazepine sedative-hypnotic. They found that blood flow in the basal forebrain was lower during NREM sleep after administration of triazolam than after administration of placebo.

One study aimed to determine if eszopiclone, a nonbenzodiazepine cyclopyrrolone, reversed the pattern of brainstem, hypothalamus, and basal forebrain abnormalities found in insomniac patients.[75] In this study, eight subjects (four women and four men) (mean age, 35 years ± 13 years) completed 2 weeks of open-label eszopiclone treatment, 3 mg at bedtime. Pre- and posttreatment assessments included a sleep diary, three nights of polysomnography, and waking and NREM sleep ^{18}F-FDG PET scans. From pre- to posttreatment, insomniac patients showed improvements in all subjective measures of sleep, sleep quality, mood, and next-morning alertness. The Pittsburgh Sleep Quality Index total was 11.9 ± 2.5 pretreatment and 7.5 ± 2.3 posttreatment; paired $t = 3.86$, $P = .006$. Brain imaging analyses showed that the reduction in relative metabolism in an arousal network from waking to NREM sleep was greater after eszopiclone treatment than before. Specific arousal network regions associated with such reduction included the pontine reticular formation and ascended into the midbrain, subthalamic nucleus, culmen of the cerebellum, and thalamus. Related neocortical areas showing this interaction included the orbitofrontal cortex, superior temporal lobe, right paracentral lobule of the posterior medial frontal lobe, right precuneus, dorsal cingulate gyrus, and portions of the frontal lobe. Comparisons involving only sleep, but not waking states, revealed similar regions of posttreatment reductions in relative metabolism.

These results demonstrate that eszopiclone reverses a pattern of central nervous system hyperarousal in insomniac patients. This effect is most pronounced during NREM sleep, at a time when the concentration of eszopiclone in the brain, when given before sleep, should be highest. The inhibitory actions of eszopiclone, and probably similar nonbenzodiazepine sedative-hypnotics and potentially the benzodiazepine sedative-hypnotics, then, that are likely to be responsible for the sedating properties of these medications appear to be largely on an arousal neural network within sleep that includes the pontine and midbrain reticular activating system. Such studies further corroborate the essential role of these structures in generating and maintaining sleep in humans (Figure 12-8).

Insomnia, Disorders of Emotion, and Limbic and Paralimbic Arousal Networks

Importantly, the basic biology of arousal can be modified by neural systems that regulate emotional and goal-directed behavior. These systems may play an important role in modulating or perpetuating the increased frequency of arousal of insomniac patients. Demonstration of this regulatory effect

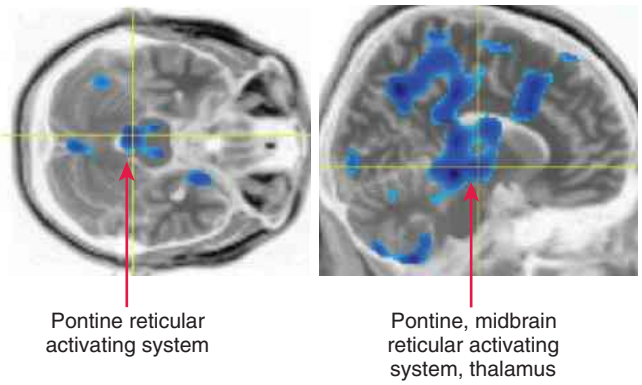


Figure 12-8 Brain structures that show a greater decline in metabolism from waking to NREM sleep in insomniac patients after 2 weeks of medication management with eszopiclone. (From Nofzinger EA, Buysse D, Moul D, et al. Eszopiclone reverses brain hyperarousal in insomnia: evidence from ^{18}F-FDG PET. *Sleep* 2008;31:A232.)

provides significant support for an essential role for these systems in generating and maintaining sleep. This is especially true given the significant epidemiologic and neurobiologic overlaps between insomnia and mental disorders.

The results of preclinical neuroimaging studies of healthy subjects and depressed persons support the importance of two neural systems in emotional behavior. A more ventrally located system, with important contributions from the amygdala, has been shown to be fundamental to the initial experience of emotions and to the automatic generation of emotional responses. The function of this system is a reactive one in response to emotional stimuli. Other structures related to this system include the anterior insula, ventral striatum, and ventral regions of the anterior cingulate cortex and ventral prefrontal cortex. A more dorsally located system, with important contributions from the dorsolateral prefrontal cortex, has been shown to be fundamental to the conscious, planned regulation of emotional behavior in light of future behavior. The function of this system is one of planning behavior in response to emotional stimuli. Other structures related to this system include the hippocampus and the dorsal regions of the anterior cingulate cortex. A primary structure in the ventral system is the amygdala. It has been shown to participate in the sensory component of emotional behavior and in the initial organization of a reactive emotional response. In humans, the amygdala shows increased activation in response to a variety of emotional stimuli including fearful faces, sad faces, threatening words, and fearful vocalizations. A reactive motor role for the amygdala includes the recruitment and coordinating of cortical arousal and vigilant attention for optimizing sensory and perceptual processing of stimuli associated with underdetermined contingencies.

Recent work shows that the amygdala is anatomically connected with and functionally modulates effects on the brainstem centers involved in arousal and sleep regulation. Similarly, other components of the ventral emotional system such as the ventral striatum, the subgenual anterior cingulate cortex, and the ventromedial prefrontal cortex are known to have anatomic and functional relationships with brainstem centers that are thought to play a role in behavioral state regulation in addition to the primary roles they each play in cortical arousal.

Human sleep neuroimaging studies support the role for components of the *ventral emotional system* in pathologic sleep associated with both depression and insomnia. Nofzinger and colleagues[76] used ^{18}F-FDG PET to define regional cerebral correlates of arousal in NREM sleep in 9 healthy and 12 depressed patients. These workers assessed EEG power in the beta high-frequency spectrum as a measure of cortical arousal. They then correlated beta power with metabolism in NREM sleep. They found that beta power negatively correlated with sleep quality. Further, beta power positively correlated with ventromedial prefrontal cortex metabolism in a group of depressed subjects and a group of healthy subjects. These investigators concluded that elevated function in the ventromedial prefrontal cortex, an area associated with obsessive behavior and anatomically linked with brainstem and hypothalamic arousal centers, might contribute to dysfunctional arousal.

Nofzinger and coworkers[73] also investigated the neurobiologic basis of poor sleep in insomnia. Insomniac patients and healthy subjects completed regional cerebral glucose metabolic assessments during waking and during NREM sleep using ^{18}F-FDG PET. A group-by-state interaction analysis confirmed that insomniac persons showed a smaller decrease than in healthy subjects in relative metabolism from waking to NREM sleep in the insular cortex, amygdala, hippocampus, anterior cingulate, and medial prefrontal cortices. This study supports the notion that persistent overactivity in a limbic or paralimbic level of the arousal system contributes to the nonrestorative sleep in insomnia patients.

Pharmacotherapy for insomnia may entail direction of some component of its mechanism of action to these limbic and paralimbic structures, especially with use of the antidepressant medications. In a study[77] assessing regional cerebral blood flow during NREM sleep in subjects receiving triazolam, blood flow in the amygdaloid complexes, in addition to basal forebrain, was lower than after administration of placebo. Although not always consistent, several neuroimaging studies show that serotoninergic antidepressants tend to lower, or inhibit, brain function (blood flow or metabolism) in ventral limbic and paralimbic cortex including the amygdala,[78,79] perhaps through actions on 5-hydroxytryptamine type 1A (5-HT$_{1A}$) postsynaptic receptors, in keeping with their high density in these areas.[80] Inhibiting hyperarousal in this limbic-paralimbic emotional neural network could be of benefit in insomniac patients.

Recent studies have focused on the role of sleep in memory consolidation utilizing both cognitive neuroscience and brain imaging of the hippocampus. One study[81] compared sleep-related consolidation of procedural memory in 7 patients with primary insomnia and 7 healthy control subjects. Performance on a mirror-tracing task was assessed before and after sleep. Performance measures in the mirror-tracing task before sleep did not differ between the groups. Both groups performed significantly better in the recall condition after sleep. Healthy control subjects showed an improvement of 42.8% ± 5.8% in the mirror tracing draw time, whereas patients with insomnia showed an improvement of 20.4% ± 14.8% (multivariate analysis of variance [ANOVA] test session × group interaction: $F = 10.9$, $P = .002$). These preliminary findings support the view that sleep-associated consolidation of procedural memories may be impaired in patients with primary insomnia.

Another study focused on morphology of the hippocampus in insomniac patients.[82] Morphometric analysis of MRI brain scans was used to investigate possible neuroanatomic differences between patients with primary insomnia and with good

sleepers. MR images (1.5T) of the brain were obtained from both groups of subjects. MRI scans were analyzed bilaterally by manual morphometry for different brain areas including hippocampus, amygdala, anterior cingulate, and orbitofrontal and dorsolateral prefrontal cortex. Patients with primary insomnia demonstrated significantly reduced hippocampal volumes bilaterally in comparison with the good sleepers. None of the other regions of interest analyzed revealed differences between the two groups. Although replication of the findings in larger samples is needed to confirm the validity of the data, these pilot data raise the possibility that chronic insomnia is associated with alterations in brain structure. The integration of structural, neuropsychological, neuroendocrine, and polysomnographic studies is necessary to further assess the relationships between insomnia and brain function and structure.

One study suggested a role for the basal ganglia in the neurobiology of insomnia.[83] Patients with insomnia and good-sleeper control subjects were studied polysomnographically for three nights with a whole-brain SPECT scan of NREM sleep on night 3 of the series. Tomographs of regional cerebral blood flow during the first NREM sleep cycle were successfully obtained. Contrary to expectations, patients with insomnia showed a consistent pattern of hypoperfusion across all eight preselected regions of interest, with particular deactivation in the basal ganglia ($P = .006$). The frontal medial, occipital, and parietal cortices also showed significant decreases in blood flow compared with those in good sleepers ($P < .05$). Subjects with insomnia had decreased activity in the basal ganglia relative to the frontal lateral cortex, frontal medial cortex, thalamus, and occipital and parietal cortices ($P < .05$). These preliminary results suggest that primary insomnia may be associated with abnormal central nervous system activity during NREM sleep that is particularly linked to basal ganglia dysfunction.

Insomnia, Disorders of Emotion, and Neocortical Arousal Networks

Beyond the subcortical brainstem and the hypothalamic and limbic and paralimbic neural systems, the prefrontal cortex might play a role in insomnia in several respects. As noted, significant epidemiologic links have been documented between insomnia and disorders of emotion. The dorsolateral prefrontal cortex is a primary structure in the dorsal emotional neural system. This structure has been shown to play an important role in executive function, including selective attention, planning, and effortful regulation of affective states. The dorsolateral prefrontal cortex maintains the representation of goals and the means to achieve them. It sends bias signals to other areas of the brain to facilitate the expression of task-appropriate responses in the face of competition with other responses.

Left-lateralized prefrontal cortex regions are important in approach-related appetitive goals and the right, in behavioral inhibition and vigilant attention. The dorsolateral prefrontal cortex not only is responsible for recruiting or inhibiting limbic regions as appropriate to performing tasks[84] but also appears to be modulated by early limbic processing. A related component of the dorsal emotional system is the dorsal anterior cingulate cortex. This region has been associated with conflict monitoring (e.g., it is particularly active during conditions requiring swift arbitration quickly between two likely responses). In view of its role in the executive aspects of emotional behavior, an abnormal increase in vigilant functions of the prefrontal cortex could lead to insomnia, whereas deficient executive behavior could be a consequence of inadequate sleep resulting from insomnia.

Nofzinger's group[85] again investigated regional cerebral glucose metabolism in insomniac patients and healthy subjects during both waking and NREM sleep using ^{18}F-FDG PET. Insomniac subjects scored worse on measures of daytime concentration and fatigue, consistent with prefrontal cortex impairment. Insomniac patients showed increased global cerebral glucose metabolism during sleep and wake states, suggesting an increased vigilant or attentive function of the neocortex consistent with hyperarousal. A group-by-state interaction analysis confirmed that insomniac subjects showed a smaller decrease than that in healthy subjects in relative metabolism from waking to NREM sleep in the thalamus, the anterior cingulate, and medial prefrontal cortices, suggesting a persistence of thalamocortical arousal even within sleep in insomniac patients. In comparison with healthy subjects, insomniac subjects, while awake, showed relative hypometabolism in a broad region of the frontal cortex bilaterally; left hemispheric superior temporal, parietal, and occipital cortices; and the thalamus. Their daytime fatigue might reflect decreased activity in prefrontal cortex that results from inefficient sleep.

Several interventions may alter activity in the prefrontal cortex in a beneficial manner for insomniac patients. For example, one study[86] assessed regional brain function associated with yoga nidra, a meditative state characterized by a loss of conscious control and an increased awareness of sensory experience. In this study, the investigators found reduced blood flow during meditation in an attentional network that included the dorsolateral prefrontal cortex and anterior cingulate cortex, as well as increased blood flow in posterior sensory and associative cortex associated with visual imagery. Cognitive approaches to the treatment of insomnia may operate by means of similar mechanisms of action in prefrontal areas.

Serotonergically active antidepressants also increase brain function (blood flow or metabolism) in dorsal paralimbic and dorsolateral prefrontal cortex. Increasing activity in prefrontal cortex might reverse prefrontal deficits in insomnia patients, leading to improved daytime cognitive function. Alternatively, further increase of an already metabolically overactive prefrontal cortex might increase attentive and vigilant functions, thereby producing further insomnia, a not uncommon side effect of selective serotonin reuptake inhibitor (SSRI) therapy in depressed or insomnia patients.

One study has documented prefrontal hypoactivation in insomnia as predicted from the background review previously mentioned. This study investigated functional brain activation differences as a possible result of chronic insomnia, and the reversibility of these differences after nonmedicated sleep therapy. Twenty-one persons with insomnia and 12 carefully matched control subjects underwent fMRI scanning during the performance of a category task and a letter-fluency task. Insomniac subjects were randomly assigned to either a 6-week period of nonpharmacologic sleep therapy or a wait-list period, after which fMRI scanning was repeated using parallel tasks. Task-related brain activation and number of generated words were considered as outcome measures. Compared with control subjects, insomniac patients showed hypoactivation of the medial and inferior prefrontal cortical areas (Brodmann areas 9, 44, and 45); activation function recovered after sleep

therapy but not after a wait-list period. These studies support the hypothesis that insomnia interferes in a reversible fashion with activation of the prefrontal cortical system during daytime task performance.

Another study attempted to explore frontal cortex function in subsets of insomnia patients subtyped as either morning- or evening chronotypes.[87] Although insomnia is a well-established risk factor for the initial onset, recurrence, or relapse of affective disorders, the specific characteristics of insomnia that connote risk remain unclear. Insomniac patients with an evening chronotype may constitute one particularly high-risk group, perhaps owing to alterations in positive affect and its related affective circuitry. This study explored this possibility by comparing diurnal patterns of positive affect and the activity of positive affect–related brain regions in morning and evening types of insomnia. The investigators assessed diurnal variation in brain activity by means of the relative regional cerebral metabolic rate of glucose uptake by using [^{18}F]fluorodeoxyglucose PET during morning and evening wakefulness. They focused on regions in the medial prefrontal cortex and striatum, which have been consistently linked with positive affect and reward processing. As predicted, chronotypes differed in their daily patterns in both self-reported positive affect and associated brain regions. Evening types displayed diurnal patterns of positive affect characterized by phase delay and smaller amplitude compared to those of morning-types with insomnia. In parallel, evening types showed a reduced degree of diurnal variation in the metabolism of the medial prefrontal cortex and the striatum, as well as lower overall metabolism in these regions across both morning and evening wakefulness. Taken together, these preliminary findings suggest that alterations in the diurnal activity of positive affect–related neural structures may underlie differences in the phase and amplitude of self-reported positive affect between morning and evening chronotypes and may constitute one mechanism for increased risk of mood disorders among evening-type insomniac patients.

REM Sleep in Depression

Because REM sleep activates the limbic and anterior paralimbic cortex in healthy subjects, the increased REM sleep in depressed patients may reflect a greater reactivation of these structures in REM sleep. Nofzinger and coworkers[88] tested this hypothesis in 24 depressed patients and 14 healthy subjects, who underwent EEG sleep studies and regional cerebral glucose metabolism assessments during both waking and REM sleep using ^{18}F-FDG PET. Depressed patients showed greater REM sleep percentage. Consistent with the hypothesis that depressed patients would show increased activation in limbic and anterior limbic structures from waking to REM, depressed patients showed greater increases in relative metabolism from waking to REM sleep than those seen in healthy subjects in the midbrain reticular formation, including the pretectal area, and in a larger region of anterior paralimbic cortex. Additionally, depressed patients showed greater increases in relative metabolism from waking to REM sleep than for healthy subjects in a broadly distributed region of predominantly left hemispheric dorsolateral prefrontal, parietal, and temporal cortex. This area included the frontal and parietal eye fields.

Increased activation of the brainstem reticular formation from waking to REM sleep in depressed patients is consistent with the model of an altered balance in brainstem monoaminergic (norepinephrine and serotonin) systems and brainstem acetylcholine neuronal systems in depressed patients. A second important finding in this study was the increased activation of limbic and anterior paralimbic (hippocampus, basal forebrain/ventral pallidum, anterior cingulate, and medial prefrontal) cortex from waking to REM sleep in the depressed patients. The highest density of cholinergic axons is in core limbic structures such as the hippocampus and amygdala. Limbic and anterior paralimbic cortices also exhibit high densities of inhibitory 5-HT$_{1A}$ postsynaptic receptors in relation to other areas of cortex. In terms of behavior, increased activation of limbic and paralimbic cortex in depressed patients may reflect a susceptibility to experiencing stimuli in a more affectively intense, negative context, in keeping with the increased activation of these structures in response to negatively valenced stimuli or increased affective states. A third major finding in this study is the relatively greater activation of executive cortex from waking to REM sleep in depressed patients. This differential effect may reflect a change in modulation of cortical function from monoaminergic during waking to cholinergic in REM sleep, coupled with a monoaminergic or cholinergic imbalance in depressed patients. Behaviorally, this finding also might reflect a greater involvement of executive function during REM sleep in depressed patients, perhaps in response to the increased affective state produced by the abnormal reactivation of limbic and paralimbic cortex during REM sleep in these patients.

Fatal Familial Insomnia

One study[89] by Perani and colleagues assessed cerebral metabolism in four patients with fatal familial insomnia, a prion disease with a mutation at codon 178 of the prion protein gene. Thalamic hypometabolism was found in all cases, and more widespread nonspecific cortical hypometabolism was noted in some. These workers suggested that the thalamic dysfunction is consistent with the neuropathologic findings in the disorder and is a hallmark of the disease.

Another study[90] reported the results of a [^{123}I]β-CIT SPECT study in two cases of fatal familial insomnia. The researchers showed a 57% and 73% reduced availability of serotonin transporters in a thalamus-hypothalamus region in the two patients in comparison with age-expected control values. Although the interpretation is not entirely clear, they suggested that this decrease might reflect altered serotoninergic function in regions of the brain thought to be important in sleep-wake regulation in this patient group.

CLINICAL PEARLS

- Brain imaging technology has helped the field of sleep medicine to explore more precisely the relation between wake-sleep and sleep disturbances and their relation to cognitive function.
- Neural systems related to sleep-wake regulation and the function of sleep overlap extensively with neural systems involved in essential aspects of waking cognitive and emotional behavior.
- Disruptions in sleep in patients with sleep disorders can therefore be associated with alterations in these neural systems, and in turn, altered sleep leads to fundamental changes in these neural systems that impair waking behavior.

SUMMARY

The application of functional neuroimaging methods to the study of sleep in health and disease in human subjects has provided unique insights into the neural mechanisms of sleep generation and maintenance. In many instances, these studies provide secondary support for those mechanisms that have been discovered in preclinical research. They also have shed light on the involvement and interaction of broad neural networks operating at subcortical and cortical levels in a defined and regular manner to produce the final experience of sleep in humans. Brain imaging studies are contributing to the understanding of how wake and sleep networks can behave pathologically to produce various sleep disorders, and to the identification of those cases in which specific treatments, either pharmacologic or behavioral, can reverse these abnormalities.

ACKNOWLEDGMENTS

The research data reported in this chapter were collected with the help of grants from the Belgian Fonds National de la Recherche Scientifique (FNRS), Fondation Médicale Reine Elisabeth (FMRE), Research Fund of the University of Liège, and the Interuniversity Attraction Poles Programme—Belgian State—Belgian Science Policy, as well as U. S. National Institutes of Health (NIH) grants MH24652, AG00972, AG20677, RR00056, RR024153, and MH30915.

Selected Readings

Engström M, Hallböök T, Szakacs A, et al. Functional magnetic resonance imaging in narcolepsy and the Kleine-Levin syndrome. *Front Neurol* 2014;**5**:105.

Gaggioni G, Maquet P, Schmidt C, et al. Neuroimaging, cognition, light and circadian rhythms. *Front Syst Neurosci* 2014;**8**:126.

Ma N, Dinges DF, Basner M, Rao H. How acute total sleep loss affects the attending brain: a meta-analysis of Neuroimaging studies. *Sleep* 2015;**38**: 233–40.

Nofzinger EA, Maquet P, Thorpy M. *Neuroimaging of sleep and sleep disorders.* Cambridge (U.K.): Cambridge University Press; 2013.

O'Byrne JN, Berman Rosa M, Gouin JP, Dang-Vu TT. Neuroimaging findings in primary insomnia. *Pathol Biol (Paris)* 2014;**62**(5):262–9.

A complete reference list can be found online at ExpertConsult.com.

Chapter
13

Cardiovascular Physiology and Coupling with Respiration: Central and Autonomic Regulation

Ronald M. Harper; Richard L. Verrier

Chapter Highlights

- Sleep states contribute to substantial variability in heart rate and coronary artery flow, mediated by pronounced activation and deactivation of parasympathetic and sympathetic components of the autonomic nervous system. Surges in both components of the autonomic nervous system concurrent with electroencephalographic phasic activity during rapid eye movement sleep can significantly alter heart rate and coronary flow, potentially compromising individuals with heart disease or sleep-disordered breathing.
- Sleep-disordered breathing results in significant changes to brain tissue, including cell and axonal loss, and glial alterations indicative of injury. The changes appear in neural structures that serve essential autonomic, hormonal, cognitive, affective, and motor control functions; involve nuclei that are origins of neurotransmitter systems, as well as important medullary cardiovascular sites; and are often lateralized.
- The specificity and asymmetry of injury to autonomic regulatory areas, which include autonomic cortical sites as well as hypothalamic and medullary regions, result in marked functional consequences expressed as pathologic hypertension, potential for cardiac arrhythmia, and altered cerebral perfusion.
- Several syndromes, including those of sudden infant death (SIDS), sudden unexplained death in epilepsy (SUDEP), sudden unexplained nocturnal death syndrome (SUNDS), and sudden death in diabetes, and multiple cardiovascular conditions appear to have sleep-related disturbed autonomic regulation as their basis.

OVERVIEW

Because of the close neurohumoral coupling between central structures and cardiorespiratory function, a dynamic coordination exists between heart rhythm, arterial blood pressure, coronary artery blood flow, and ventilation. Non–rapid eye movement (NREM) sleep is associated with relative autonomic stability and functional coordination between respiration, pumping action of the heart, and maintenance of arterial blood pressure. During rapid eye movement (REM) sleep, surges in cardiac-bound sympathetic and parasympathetic activity provoke accelerations and pauses in heart rhythm, respectively. These surges occur in association with alterations on the electroencephalogram (EEG) indicative of phasic central nervous system activation in REM sleep, so-called phasic REM sleep. Whereas autonomic nervous system perturbations are well tolerated in normal persons, patients with heart disease or sleep-disordered breathing may be at heightened risk for cardiac events, especially due to the exaggerated sympathetic tone of REM sleep. Autonomic regulation often is compromised from injury to central autonomic regulatory sites in obstructive sleep apnea (OSA) and heart failure, resulting in preferentially lateralized high sympathetic nerve output and inappropriate responses to momentary challenges. In patients with severely compromised hearts, NREM sleep is associated with the potential for onset of hypotension, which can in turn impair blood flow through stenotic coronary vessels. During a typical night's sleep, a broad spectrum of autonomic patterns unfolds that provides both respite and stress to the cardiovascular system. These effects result from natural carefully orchestrated changes in central nervous system physiology, as the brain periodically reexcites during REM sleep, from the relative tranquility of NREM sleep.

This chapter surveys central and autonomic nervous system mechanisms that regulate cardiovascular function during sleep and identifies some dysfunctions that can arise in patients with underlying pathologic conditions. Particular attention is focused on cardiac electrical stability and coronary artery blood flow, because disturbances in these factors of cardiovascular function can trigger life-threatening cardiac arrhythmias and myocardial ischemia and infarction in patients with heart disease. Attention also is directed toward central mechanisms underlying the high sympathetic tone found in conditions associated with sleep-disordered breathing, including OSA and heart failure, and with cardiovascular function in infants, particularly because nocturnal regulatory perturbations of these systems may be an important factor in sudden infant

death syndrome (SIDS). The importance of these issues to public health is underscored by the annual toll of nocturnal, sleep-related cardiac events, which account for an estimated 20% of myocardial infarctions (approximately 250,000/year) and for 15% of sudden cardiac deaths (47,500/year) in the United States.[1] For detailed reports and discussion of clinical findings, see Chapters 124 and 125.

SLEEP STATE CONTROL OF CARDIOVASCULAR FUNCTION

The first NREM sleep cycle, from sleep onset, is characterized by a period of relative autonomic stability, with vagus nerve dominance and heightened baroreceptor gain. During NREM sleep, a near-sinusoidal modulation of heart rate variation occurs from coupling of respiratory and cardiorespiratory brain centers, resulting in what is termed "normal respiratory sinus arrhythmia" (Figure 13-1). During inspiration, heart rate accelerates to accommodate increased venous return, increasing cardiac output, followed by progressive rate slowing during expiration. This normal sinus variability in heart rate, particularly during NREM sleep, generally is indicative of cardiac health, whereas absence of intrinsic variability has been associated with cardiac pathology and advancing age.[2] The reflexive cardiovascular changes during breathing, manifested as cyclic heart rate variation, also have a converse relationship seen as transient elevation of arterial blood pressure, which results in slowing, cessation, or diminution of breathing efforts. This cascade of reflex adaptation is enhanced during sleep,[3] when even small reductions in arterial blood pressure increase respiratory rates.[4,5] These breathing pauses and increased rates serve as compensatory mechanisms to normalize arterial blood pressure. Reduced breathing variation and absence of normal breathing pauses, as well as declines in respiration-induced heart rate variation, are characteristic of infants who later succumb to SIDS[6] and may hint at failing compensatory mechanisms in the syndrome. Reduced heart rate variability appears in infants afflicted with congenital central hypoventilation syndrome, a condition characterized by reduced drive to breathe during sleep and compromised autonomic regulation, with high sympathetic tone.[7] Exaggerated heart rate variation appears in children with OSA, coincident with the enhanced bradycardia/tachycardia of apnea, but with lower sinus arrhythmia.[8] Patients with heart failure, who typically exhibit high levels of sympathetic tone, also have diminished respiratory function–related heart rate variation.[9] Thus the common denominator of cardiac risk associated with decreased heart rate variability appears to be enhanced sympathetic activity, together with reduced vagus nerve function.

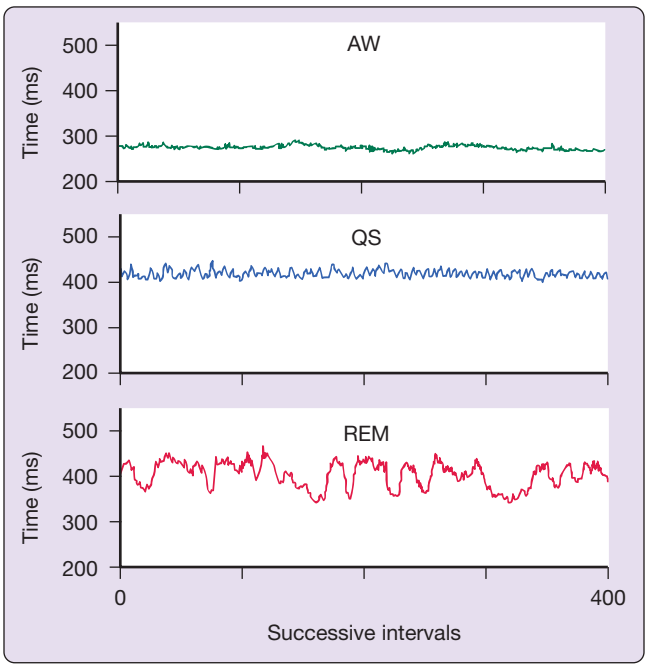


Figure 13-1 The *x*-axis represents successive heartbeats and the intervals between heartbeats from a healthy 4-month-old infant during quiet sleep (QS), REM sleep, and wakefulness (AW). The *y*-axis represents time (in milliseconds) between those heartbeats. Note the rapid modulation of intervals during quiet sleep contributed by respiratory variation. Also evident are lower-frequency modulation during REM sleep and epochs of sustained rapid rate during wakefulness.

Sympathetic nerve activity appears to be relatively stable during NREM sleep, and its cardiovascular input is reduced by more than half from wakefulness to stage N3 of NREM sleep.[10] In general, the autonomic stability of NREM sleep, with hypotension, bradycardia, and reduced cardiac output and systemic vascular resistance, provides a relatively salutary neurohumoral background during which the heart has an opportunity for metabolic restoration.[11] The bradycardias appear to result mainly from increased vagus nerve activity, whereas the hypotension is primarily attributable to reduced sympathetic vasomotor tone.[12] During NREM-to-REM sleep transitions, vagus nerve activity bursts may result in pauses in heart rhythm and frank asystole.[13]

REM sleep is initiated at approximately 90-minute intervals and, in subserving brain neurochemical functions and behavioral adaptations, can disrupt cardiorespiratory homeostasis.[14] The brain's increased excitability during REM sleep can result in major surges in cardiac sympathetic nerve activity to the coronary vessels. Baroreceptor gain is reduced. Heart rate fluctuates strikingly, with marked tachycardia and bradycardia episodes.[15,16] Cardiac efferent vagus nerve tone generally is suppressed during REM sleep,[11] and the highly irregular breathing patterns can lead to lower oxygen levels, particularly in patients with pulmonary or cardiac disease.[14] Neurons serving the principal diaphragmatic respiratory muscles are spared the generalized inhibition (see also Chapter 15),[17] although accessory and upper airway muscles diminish activity.[18] The REM atonia is especially marked in infant thoracic and abdominal muscles,[19] leading to the potential for dangerously low residual tidal volumes from the resulting increased thoracic wall compliance when the rib cage is not yet calcified. During sleep apnea, loss of central respiratory activity and/or airway obstruction may occcur several hundred times each night, with dire consequences for neural structures from the resulting intermittent hypoxia.

CARDIORESPIRATORY INTERACTIONS

Central Mechanisms

Integration of cardiorespiratory function during sleep is achieved at several neuraxis levels. Several pontine and suprapontine, as well as cerebellar mechanisms, can alter cardiorespiratory patterns during sleep and waking, and cerebral cortex sites play major roles, especially in modulating sympathetic and parasympathetic outflow and breathing effort.

Regional pontine roles in REM sleep activation have been documented by imaging studies of REM sleep, which also show preferential activation of limbic and paralimbic regions in REM sleep, compared with waking or with NREM sleep.[20-22] The pontine midline raphe contains serotoninergic neurons exerting significant roles in vascular control[23]; the raphe is damaged in OSA and heart failure (Figure 13-2), probably resulting from altered perfusion and intermittent hypoxia accompanying sleep-impaired breathing in these conditions.[24-27]

The classic sympathetic medullary areas include the rostral and caudal ventrolateral nuclei (RVLM and CVLM) and the nucleus of the solitary tract (NTS); the latter integrates baroreceptor and other sensory signals, relaying information to the CVLM, which in turn projects to the RVLM and then to the spinal cord intermediolateral column for sympathetic outflow. The human RVLM and CVLM are dorsally displaced relative to usual anatomic descriptions in rodents, and their signal changes with muscle sympathetic nerve discharge have been well described.[28] The RVLM shows significant

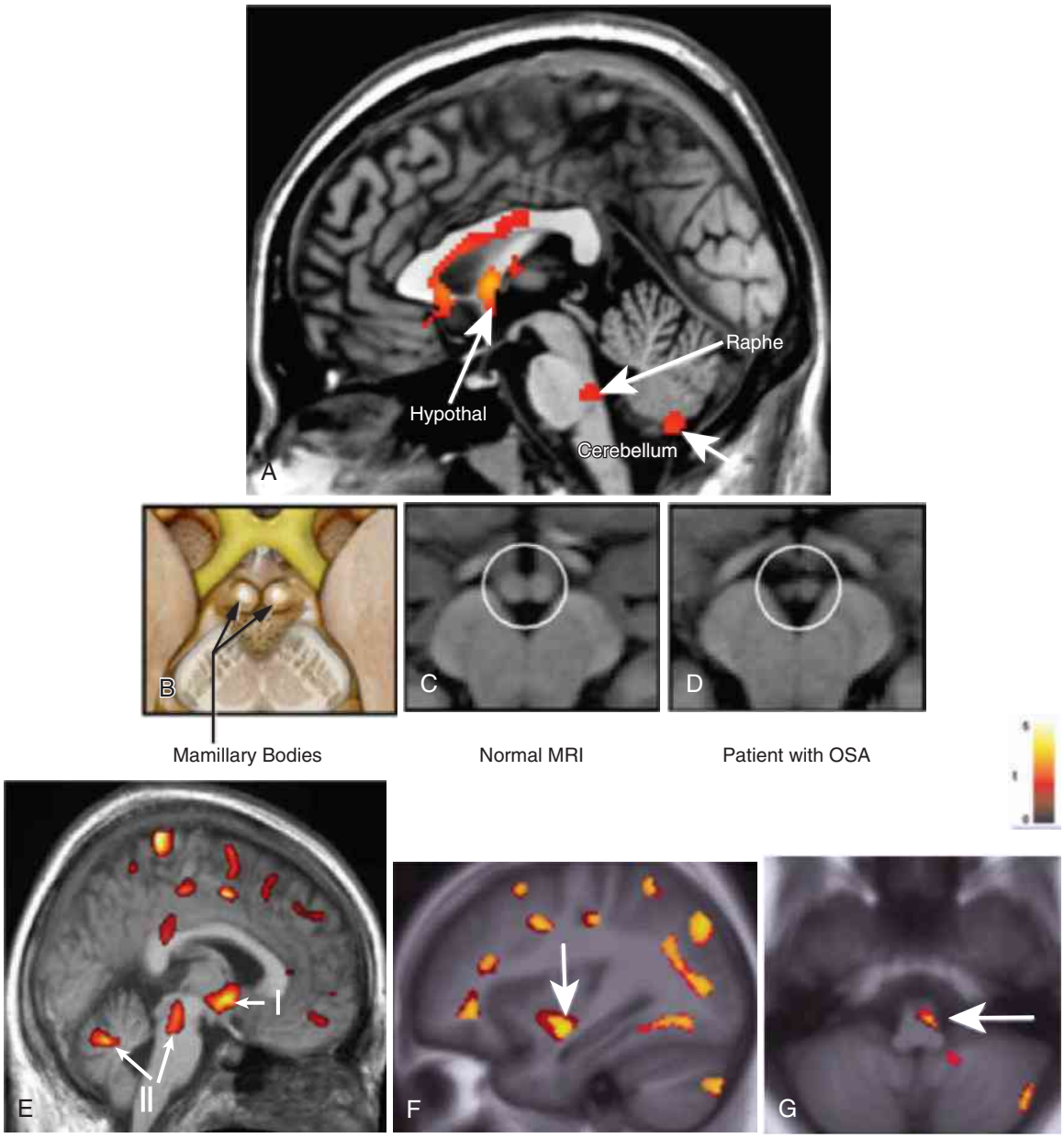


Figure 13-2 **A,** Midline view demonstrating injury in pontine raphe, cerebellum, and hypothalamus of patients with heart failure, as detected by T2 relaxometry procedures. Pontine raphe (*arrow*), fibers of the fornix, hypothalamus, and cerebellum show injury. **B-D,** Mammillary body volume loss in obstructive sleep apnea (OSA). **B,** Cartoon of mammillary bodies. **C-D,** T1-weighted MRI images. **C,** Control subject mammillary bodies. **D,** Mammillary bodies in patient with OSA. **E,** Mean diffusivity scans, hypothalamic (I) and cerebellar and midbrain (II) injury in OSA. **F,** Mean diffusivity scans, insula (*arrow*) injury in OSA. **G,** Mean diffusivity scans, ventrolateral medullary injury (*arrow*) in OSA. (Data from Woo MA, Kumar R, Macey PM, et al. Brain injury in autonomic, emotional, and cognitive regulatory areas in patients with heart failure. *J Cardiac Fail* 2009;15:214-23; Kumar R, Birrer BV, Macey PM, et al. Reduced mammillary body volume in patients with obstructive sleep apnea. *Neurosci Lett* 2008;438:330-4; and Kumar R, Chavez AS, Macey PM, et al. Altered global and regional brain mean diffusivity in patients with obstructive sleep apnea. *J Neurosci Res* 2012;90:2043-52. Drawing by Acerland International.)

injury, measured by diffusion tensor imaging, in both OSA and heart failure (see Figure 13-2).[25,26]

The ventral medial frontal, cingulate, and insular cortices, along with portions of the hippocampal formation, mammillary bodies, and hypothalamic structures, are suprapontine structures participating in cardiorespiratory patterning and autonomic control, and all of these structures (see Figure 13-2) are injured in OSA.[26,29-32] The amygdala central nucleus projects extensively to the parabrachial pons, the NTS, the dorsal motor nucleus, and the periaqueductal gray, all areas exerting significant influences on cardiac action.[33] Portions of the amygdala, hippocampus, and frontal and insular cortices help mediate transient arterial blood pressure changes elicited by cold pressor or Valsalva challenges.[34,35] Injury to those areas may contribute to the impaired dynamic blood pressure control and chronic hypertension in OSA.

The insular cortices deserve special attention in cardiovascular regulation during sleep and waking. Both animal and human studies show that these areas modulate sympathetic and parasympathetic action, with the right insula principally affecting sympathetic outflow and the left, parasympathetic action[36] (although both sides interact).[37] The right anterior insula modulates baroreflex action,[37] and a lateralized and anterior-posterior topography of fMRI signals emerges to different autonomic (Valsalva, hand grip, or cold pressor) challenges.[38] The insular topographic organization affects sleep, heart failure, and stroke fields, because the right insula is damaged in OSA and heart failure[24,26,29,39,40] (Figure 13-3, *A*) and can be vascularly compromised from middle cerebral artery stroke. Heart failure and OSA are accompanied by enhanced sympathetic nerve discharge and hypertension, possibly from injury to the anterior insula and its baroreflex modulation roles. That insular damage is accompanied by impaired amplitude and timing of fMRI signals to autonomic challenges and an inability to mount appropriate heart rate responses.[41-44]

Some types of seizure discharge in epilepsy occur during sleep states. A unilateral seizure focus in the insula could exert profound influences on arterial blood pressure and heart rate,[45] which might include exaggerated sympathetic drive from right-sided foci, leading to circumstances underlying sudden unexplained death during sleep in epilepsy (SUDEP) and diabetes.[46,47] The ventral medial frontal and cingulate cortices also exert major roles in vagal and sympathetic control, respectively.[48-51] The cortical influences on subcortical sites carry significant import for cardiorespiratory control.

The cerebellum is especially important for regulating cardiovascular and respiratory control in both sleep and waking states. Although it is not classically considered a component of either breathing or cardiac regulation, that cerebellar role has been known for over half a century[52] and is mediated partially through vestibular/cerebellar mechanisms in blood pressure coordination.[53] Vestibular mechanisms modify arterial blood pressure responses to rapid postural changes, a process familiar to orthostatic hypotensive individuals who suffer syncope on rising rapidly from the horizontal position. Lesions of feline cerebellar fastigial nuclei result in ineffective compensatory responses to hypotension,[54] with ensuing death. Significant gray matter loss occurs in the cerebellar cortex and deep nuclei in heart failure[39] (Figure 13-3, *B*) and in OSA[40] and probably contributes to aberrant cardiovascular control in these syndromes. Abnormal cerebellar development or cerebellar insult contributions to cardiorespiratory disturbances are well described.[55-57]

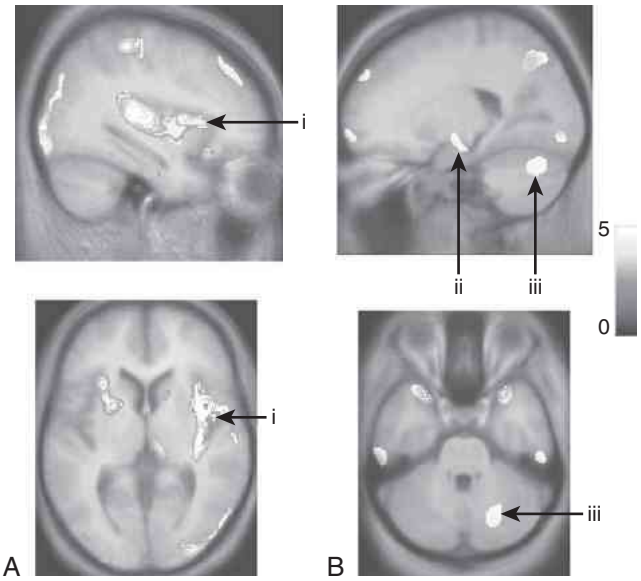


Figure 13-3 Areas of gray matter loss (*arrows*) within the insula (i) of patients with heart failure ($n = 9$) **(A),** and in the hippocampal region (ii) and cerebellum (iii) of patients with obstructive sleep apnea (OSA) ($n = 21$) **(B).** Gray matter loss was calculated from structural magnetic resonance imaging scans relative to those obtained in control subjects. The 0 to 5 scale represents *t* values; all *light areas* are significant ($P < .05$). (**A**, From Woo MA, Macey PM, Fonarow GC, et al. Regional brain gray matter loss in heart failure. *J Appl Physiol* 2003;95:677-84. **B**, Reprinted with permission of the American Thoracic Society. Copyright 2012 American Thoracic Society. Macey PM, Henderson LA, Macey KE, et al. Brain morphology associated with obstructive sleep apnea. *Am J Resp Crit Care Med* 2002;166:1382-7.[40] The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

Cardiorespiratory Homeostasis

An important consideration in preserving circulatory homeostasis during sleep is coordination of control over two systems: the respiratory system, essential for oxygen exchange, and the cardiovascular system, for blood transport. The coordination of two motor systems, one for somatic musculature (i.e., diaphragmatic, intercostal, abdominal, and upper airway musculature) and the other for autonomic regulation (to the heart and vasculature), is a formidable task during sleep and is particularly challenging in patients with diseased respiratory or cardiovascular systems, especially in apnea or heart failure, or in infants, whose developing control systems may become compromised. Respiratory neuron activity varies greatly between sleep states, as does heart rhythm regularity. Tachycardia, polypnea, sweating, and dramatic elevations in arterial blood pressure secondary to intense autonomic activity occur primarily during REM sleep.

Maintaining perfusion of vital organs through adequate arterial blood pressure control is essential for cardiorespiratory homeostasis. Respiratory mechanisms are recruited to support cardiovascular action by assisting venous return and by reflexly altering cardiac rate. REM sleep induces a near-paralysis of accessory respiratory muscles, including upper airway musculature, and diminishes descending forebrain influences on brainstem vasculature and motor control regions.[58,59] This reorganization during REM sleep may interfere with compensatory breathing mechanisms that assist arterial blood pressure management and may lead to removal of protective forebrain influences on hypotension or hypertension. The significant interaction between breathing and arterial blood

pressure appears in the partial normalization of blood pressure by delivery of continuous positive airway pressure in patients with apnea-induced hypertension.[60]

Arterial blood pressure control during sleep is of interest in examining mechanisms of failure in SIDS. Several reports indicate that the fatal sequence in SIDS may originate with a cardiac rhythm failure.[61] Specifically, an arrhythmia, or bradycardia and hypotension, rather than an initial breathing cessation, characterizes the final event.[62,63] Antecedent tachycardia may be present for up to 3 days. The terminal events in SIDS appear to parallel the two stages of shock—namely, an initial sympathoexcitation followed by a sudden, centrally triggered sympathoinhibition and bradycardia, leading to a life-threatening fall in arterial blood pressure. Some monitored SIDS cases show a near-total loss of arterial blood pressure within a minute of onset of the fatal event. Because SIDS deaths occur largely during sleep, some interaction of state and compensatory mechanisms is suspected. The prone sleeping position enhances the risk for SIDS, which may derive from the vestibular and cerebellar contributions to arterial blood pressure control[53] described earlier. Because vestibular and cerebellar mechanisms assist mediation of arterial blood pressure to postural changes, static stimuli, such as those from the prone position, can directly modify cardiovascular responses to blood pressure challenges.[64-67] Sleep effects on vestibular systems must be considered in examination of arterial blood pressure control mechanisms.

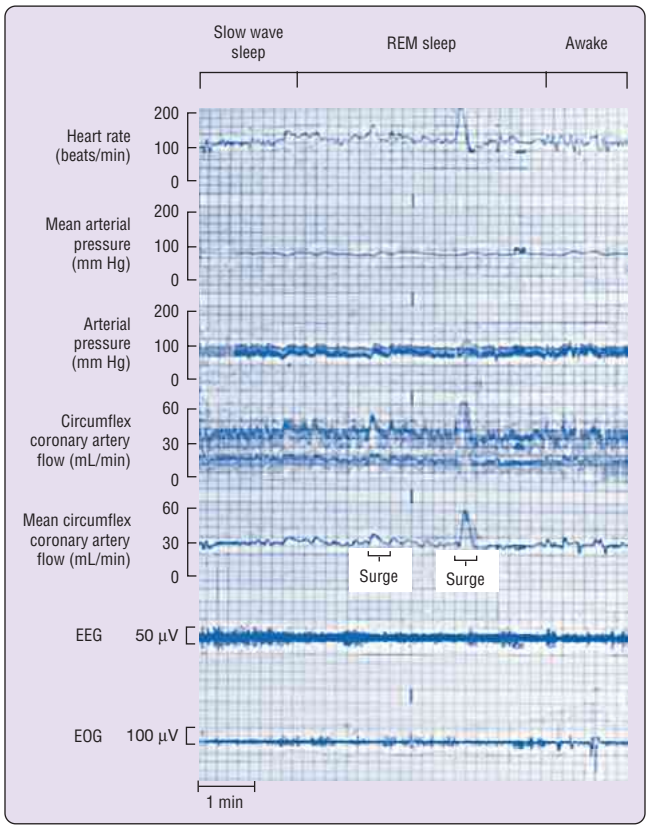


Figure 13-4 Effects of NREM sleep (slow wave sleep), REM sleep, and quiet wakefulness on heart rate, phasic and mean arterial blood pressure, phasic and mean left circumflex coronary flow, electroencephalogram (EEG), and electrooculogram (EOG) in the dog. Sleep spindles are evident during NREM sleep, eye movements during REM sleep, and gross eye movements on awakening. Surges in heart rate and coronary flow occur during REM sleep. (From Kirby DA, Verrier RL. Differential effects of sleep stage on coronary hemodynamic function. *Am J Physiol* 1989;256:H1378-83.)

SLEEP STATE–DEPENDENT CHANGES IN HEART RHYTHM

Recent evidence indicates that the pronounced changes in heart rate occurring during REM sleep and transitions between sleep states are attributable to distinct mechanisms associated with specific brain sites, rather than representing a continuum of autonomic change.

Heart Rate Surges

Several investigators have reported REM sleep–induced increases in heart rate in experimental animals.[12,15,68-71] Accelerations consisting of an abrupt, although transitory, 35% to 37% increase in rate that are concentrated during phasic REM sleep were observed in canines (Figure 13-4). These marked heart rate surges are accompanied by a rise in mean arterial blood pressure and are followed by a rate deceleration that apparently is baroreceptor-mediated. Because the sequence is completely abolished by interruption of sympathetic neural input to the heart,[68-70] the acceleration does not appear to be dependent on withdrawal of parasympathetic nerve activity.[15,70]

REM sleep state–dependent heart rate surges also have been observed in felines. The rate accelerations are linked to central nervous system activation as reflected in a concomitant increase in hippocampal theta frequency, ponto-geniculo-occipital (PGO) activity, and eye movements.[71] In cats, the appearance of theta waves is characteristic of arousal, orienting activity, alertness, and REM sleep.[71-75] The surges are abolished by cardioselective beta-adrenergic blockade with atenolol, suggesting, as in canines, that the peripheral effect is attributable to bursting of cardiac sympathetic efferent fiber activity, which directly affects heart rate. The main difference in rate responses between the two species is that in dogs, the rate acceleration is accompanied within seconds by a baroreflex-mediated deceleration. The precise basis for these differences in the pattern of heart rate responses is unclear, but a plausible explanation is that the canine studies were performed in beagles—dogs that are bred for intense physical activity, a factor known to augment baroreceptor responsiveness.

Heart Rhythm Pauses

A complementary finding to centrally mediated heart rate surges is the observation in cats of an abrupt heart rhythm deceleration that occurs predominantly during tonic REM sleep and is not associated with any preceding or subsequent heart rate or arterial blood pressure change[76] (Figure 13-5). The vagus nerve involvement appears to be directly initiated by central influences, as no antecedent or subsequent change in resting heart rate or arterial blood pressure occurs. The primary involvement of central nervous system activation is shown by the consistent, antecedent, abrupt cessation of PGO activity and the concomitant interruption of hippocampal theta rhythm. In normal human volunteers, Taylor and colleagues[77] observed heart rate decelerations during REM sleep that preceded eye movement bursts by 3 seconds and suggested that the phenomenon reflects an orienting response at dreaming onset. The processes underlying the central nervous

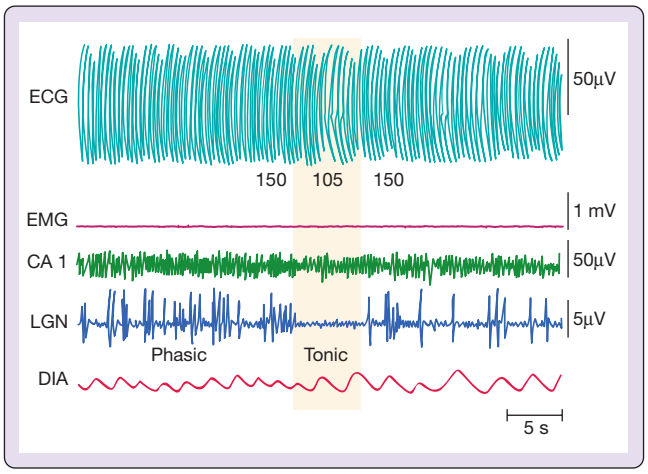


Figure 13-5 Representative polygraphic recording of a primary heart rate deceleration during tonic REM sleep. During this deceleration, heart rate decreased from 150 to 105 beats/minute, or 30%. The deceleration occurred during a period devoid of ponto-geniculo-occipital (PGO) spikes in the lateral geniculate nucleus (LGN) or theta rhythm in the hippocampal (CA 1) leads. The abrupt decreases in amplitude of hippocampal theta waves (CA 1), PGO waves (LGN), and respiratory rate (DIA) are typical of transitions from phasic to tonic REM sleep. ECG, Electrocardiogram; EMG, electromyogram. (From Verrier RL, Lau RT, Wallooppillai U, et al. Primary vagally mediated decelerations in heart rate during tonic rapid eye movement sleep in cats. *Am J Physiol* 1998;43:R1136-41.)

system changes accompanying the tonic REM sleep–induced increase in vagus nerve tone, suppressing sinus node activity, remain unknown. Notwithstanding extensive studies of the physiologic and anatomic bases for PGO activity, little is known about its conductivity and functional relationship to heart rhythm control during sleep.

The most likely basis for the abrupt deceleration in heart rate during tonic REM sleep is a centrally induced decline in sympathetic nerve activity or enhanced vagus nerve tone, or both in combination. In felines, cardioselective beta$_1$-adrenergic blockade with atenolol did not affect the incidence or magnitude of decelerations, but muscarinic blockade with glycopyrrolate completely abolished the phenomenon. These observations suggest that the tonic REM sleep–induced decelerations are primarily mediated by cardiac vagus nerve efferent fiber activity. It is well known that enhanced vagal activity can abruptly and markedly affect the sinus node firing rate.[78] Because beta-adrenergic blockade exerted no effect on the frequency or magnitude of decelerations, it does not appear that withdrawal of cardiac sympathetic tone is an important factor in the observed rate changes. Respiratory interplay is not an essential component of the deceleration, inasmuch as the phenomenon often occurs in the absence of a temporal association with inspiratory effort.

This primary heart rate pause phenomenon appears to be distinct from baroreceptor-mediated reductions in heart rate that almost invariably follow accelerations in rate and elevation of arterial blood pressure[13] (Figure 13-6). This second

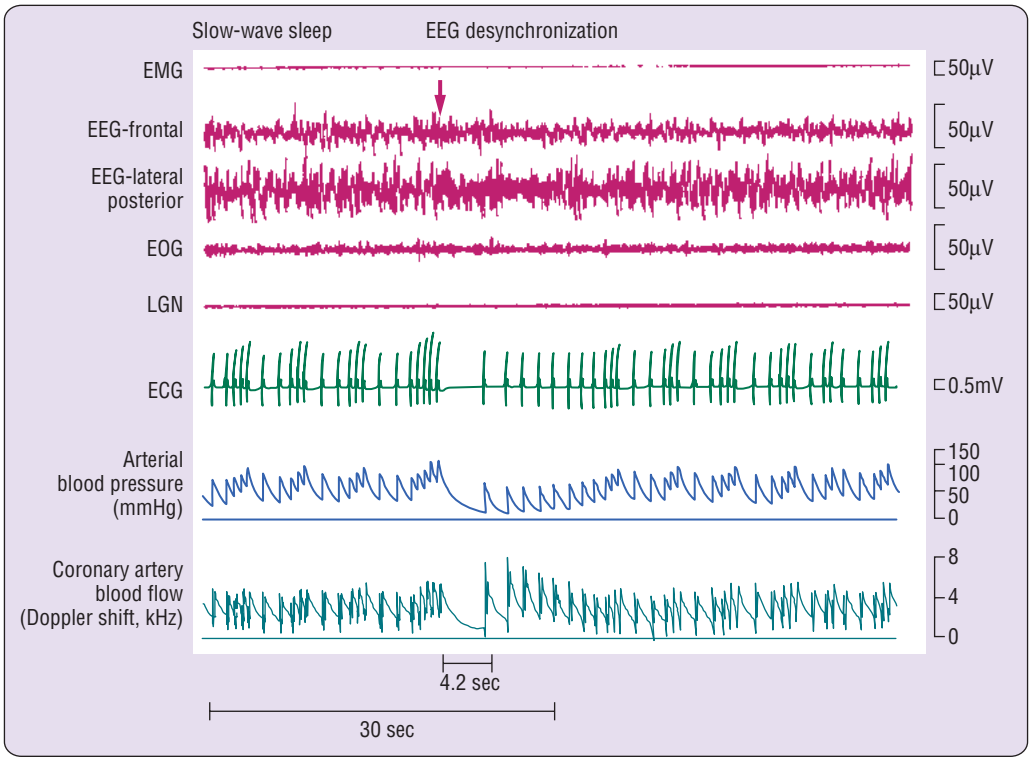


Figure 13-6 Coronary blood flow (CBF) surge during deep NREM sleep interrupted by electroencephalographic desynchronization. This response pattern is common and appears to represent a brief, low-grade arousal. The 4.2-second pause in heart rhythm was followed by a brief increase of 46% in average peak CBF and a decrease of 49% in the heart rate–systolic blood pressure product. ECG, Electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram; LGN, lateral geniculate nucleus field potential recordings; SWS, slow wave sleep. (From Dickerson LW, Huang AH, Nearing BD, et al. Primary coronary vasodilation associated with pauses in heart rhythm during sleep. *Am J Physiol* 1993;264:R186-96.)

group of heart rhythm pauses was observed in canines and occurs mainly during the transition from slow wave sleep to desynchronized sleep and more frequently during phasic than tonic REM sleep. These pauses persist for 1 to 8 seconds and are followed by dramatic increases in coronary blood flow averaging 30% and ranging up to 84%, which are independent of metabolic activity of the heart as reflected in the heart rate–blood pressure product. An intense burst of vagus nerve activity appears to produce the phenomenon, because the pauses develop against a background of marked respiratory sinus arrhythmia with varying degrees of heart block (with nonconducted P waves) and with low heart rate. Moreover, the heart rhythm pauses could be reproduced by electrical stimulation of the vagus nerve. Guilleminault and colleagues[79] documented similar pauses in healthy young adults.

CORONARY ARTERY BLOOD FLOW REGULATION DURING SLEEP

Striking changes in coronary blood flow occur during REM and sleep state transitions.[13,68-70,80] Vatner and coworkers[80] studied the effects of the sleep-wake cycle on coronary artery function in baboons. During the nocturnal period, when the animals were judged to be asleep by behavioral indicators, coronary blood flow increased sporadically by as much as 100%. The periodic oscillations in blood flow were not associated with alterations in heart rate or arterial blood pressure and occurred while the animals remained motionless with eyes closed. Because the baboons were not instrumented for electroencephalographic recordings, no information was obtained regarding sleep stage, nor was the mechanism for the coronary blood flow surge defined.

Concomitant with the heart rate surges of REM sleep found in canines[68-70] as described earlier, were remarkable episodic surges in coronary blood flow, with corresponding declines in coronary vascular resistance. These phenomena occurred predominantly during periods of REM sleep marked by intense eye movement phasic activity.[70] No significant changes in mean arterial blood pressure were seen. Heart rate was elevated during the coronary flow surges, suggesting increased cardiac metabolic activity as the basis for the coronary vasodilation. In fact, the close coupling between rate-pressure product, an index of metabolic demand, and the magnitude of the flow surges indicates that the surges do not constitute a state of myocardial hyperperfusion. These surges in coronary blood flow appear to result from enhanced adrenergic discharge, because they were abolished by bilateral stellectomy, and not from nonspecific effects of somatic activity or respiratory fluctuations.

During severe experimental coronary artery stenosis (with baseline flow reduced by 60%) in canines, phasic decreases in coronary arterial blood flow, rather than increases, were observed during REM sleep coincident with these heart rate surges[69] (Figure 13-7). An increase in adrenergic discharge could lead to a coronary blood flow decrement by at least two possible mechanisms. The first is by stimulation of alpha-adrenergic receptors on the coronary vascular smooth muscle. Such an effect, however, could be only transitory, because alpha-adrenergic stimulation results in brief (10 to 15 seconds) coronary constriction even during sympathetic nerve stimulation in anesthetized animals[81] or during intense arousal associated with aversive behavioral conditioning.[82] The second possible mechanism is mechanical: a decrease in diastolic coronary perfusion time caused by the surges in heart rate. In support of this explanation, a strong correlation ($r^2 = 0.96$) was found between the magnitude of the increase in heart rate and the decrease in coronary blood flow.[69] The link between REM-induced changes in heart rate and the occurrence of myocardial ischemia in patients with advanced coronary artery disease was reported by Nowlin and coworkers.[83]

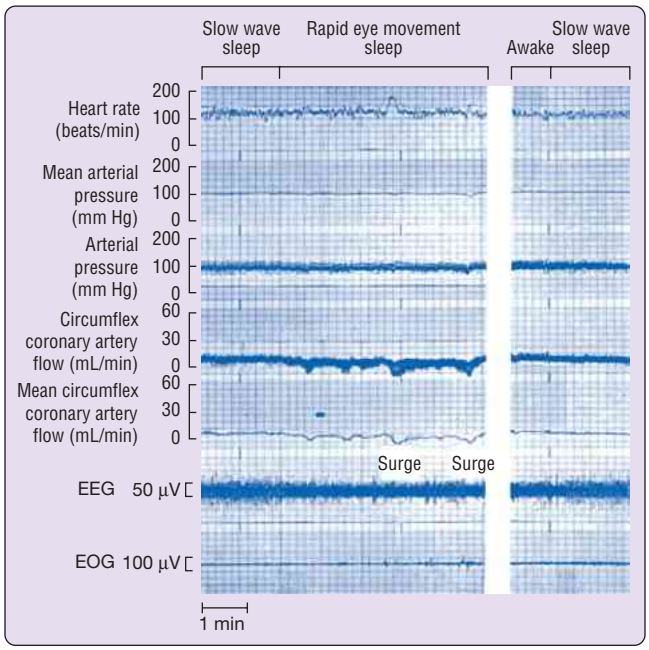


Figure 13-7 Effects of sleep stage on heart rate, mean and phasic arterial blood pressure, and mean and phasic left circumflex coronary artery blood flow in a typical dog during coronary artery stenosis. Note phasic decreases in coronary flow occurring during heart rate surges while the dog is in REM sleep. During REM sleep, the electroencephalogram (EEG) reveals a characteristic lower-amplitude, higher-frequency pattern than in slow wave sleep. The electrooculogram (EOG) tracing indicates the presence of eye movements during REM but not slow wave sleep. (From Kirby DA, Verrier RL. Differential effects of sleep stage on coronary hemodynamic function during stenosis. *Physiol Behav* 1989;45:1017-20.)

Furthermore, nocturnal myocardial ischemia carries significant potential to affect coronary function and cardiac electrical stability in patients with ischemic heart disease, because it has been linked to infarctions, 20% of which occur at night, and arrhythmias (see Chapter 124).

IMPACT OF SLEEP ON ARRHYTHMOGENESIS

Central Nervous System Sites Influencing Cardiac Electrical Stability

Extensive investigation of central nervous system–induced cardiac arrhythmias has provided evidence that triggering of arrhythmias by the central nervous system not only is the consequence of intense activation of the autonomic nervous system but is also a function of the specific neural pattern elicited. Regulation of cardiac neural activity is highly integrated and is achieved by circuitry at multiple levels[84] (Figure 13-8). Higher brain centers operate through elaborate pathways within the hypothalamus, modulated by cortical influences and by medullary cardiovascular regulatory sites. Baroreceptor mechanisms have long been recognized as

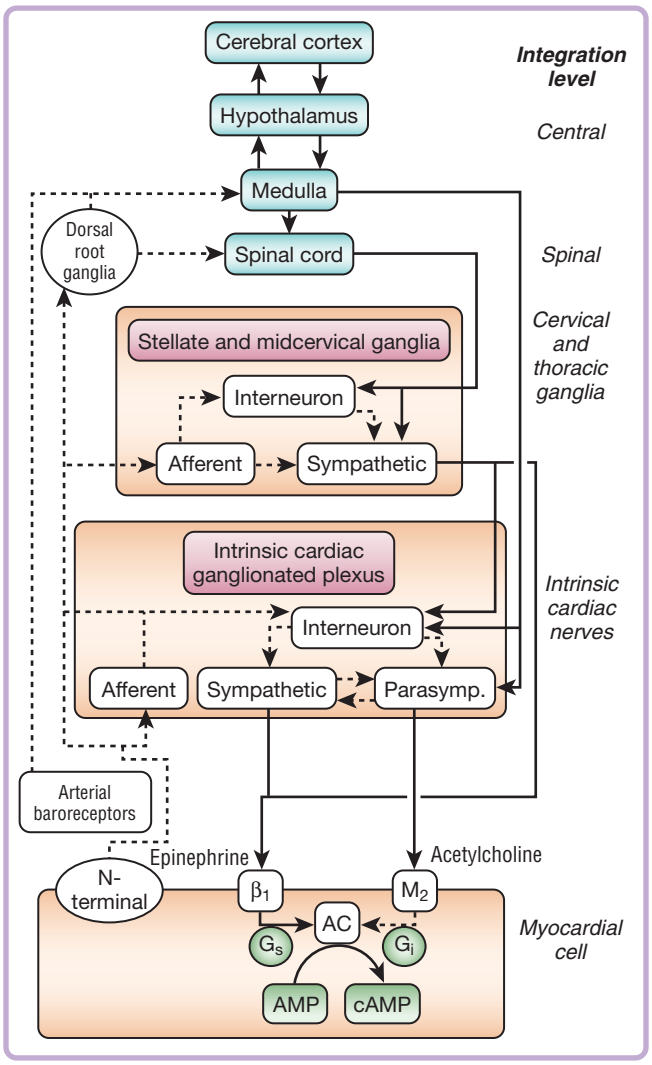


Figure 13-8 Synthesis of new and present views on levels of integration important in neural control of cardiac electrical activity during sleep. More traditional concepts focused on afferent tracts (*dashed lines*) arising from myocardial nerve terminals and reflex receptors (e.g., baroreceptors) that are integrated centrally within hypothalamic and medullary cardiostimulatory and cardioinhibitory brain centers and on central modulation of sympathetic and parasympathetic outflow (*solid lines*) with little intermediary processing at the level of the spinal cord and within cervical and thoracic ganglia. More recent views incorporate additional levels of intricate processing within the extraspinal cervical and thoracic ganglia and within the cardiac ganglionic plexus, where recently described interneurons are envisioned to provide new levels of noncentral integration. Release of neurotransmitters from postganglionic sympathetic neurons is believed to enhance excitation in the sinoatrial node and myocardial cells through norepinephrine binding to beta$_1$ receptors, which enhances adenyl cyclase (AC) activity through intermediary stimulatory G proteins (G_s). Increased parasympathectomy outflow enhances postganglionic release and binding of acetylcholine to muscarinic (M_2) receptors and, through coupled inhibitory G proteins (G_i), inhibits cyclic AMP (cAMP) production. Cyclic AMP alters electrogenesis and pacemaking activity by affecting the activity of specific membrane sodium, potassium, and calcium channels. New levels of integration are shown superimposed on previous views and are emphasized here to highlight new possibilities for intervention. (From Lathrop DA, Spooner PM. On the neural connection. *J Cardiovasc Electrophysiol* 2001;12:841-4, with permission.)

integral to autonomic control of the cardiovascular system, as evidenced by heart rate variability and baroreceptor sensitivity testing in both cardiac patients and normal subjects. The baroreflex is partially modulated by the insular cortex,[37] but intrinsic cardiac nerves and fat pads provide local neural coordination independent of higher brain centers. The phenomenon of electrical remodeling is attributable to nerve growth and degeneration. At the level of the myocardial cell, autonomic receptors influence G proteins to control ionic channels, pumps, and exchangers. The influence of the vagus nerve on ventricular electrical properties is contingent on the level of sympathetic tone, a phenomenon referred to as accentuated antagonism. The underlying mechanism is that acetylcholine, released by vagal activation, exerts its opposing effects by presynaptic inhibition of norepinephrine release from sympathetic nerve endings and through an antagonism of second messenger formation at the cardiac receptor level.[85] Thus the balance in cardiac input from either limb of the autonomic nervous system and their interactions must be considered. Another important concept is that triggering of arrhythmias by central nervous system activity also may depend on several intermediary mechanisms. These include direct effects of neurotransmitters on the myocardium and its specialized conducting system and changes in myocardial perfusion due to alterations in coronary vasomotor tone, enhanced platelet aggregability, or both. The net influence on the heart thus depends on a complex interplay between the specific neural pattern elicited and the underlying cardiac pathology.

More than 100 years ago, Levy[86] demonstrated that ventricular tachyarrhythmias can be elicited in normal animals by stimulating specific areas in the brain, a finding subsequently confirmed in several species. Hockman and colleagues,[87] using stereotactic techniques, demonstrated that cerebral stimulation and hypothalamic activation evoked a spectrum of ventricular arrhythmias. Stimulation of the posterior hypothalamus causes a 10-fold increase in the incidence of ventricular fibrillation elicited by experimental occlusion of the coronary artery.[88] This enhanced vulnerability was linked to increased sympathetic nerve activity, because beta-adrenergic receptor blockade, but not vagotomy, prevented it. These findings are consistent with clinical reports that cerebrovascular disease (particularly intracranial hemorrhage) can elicit significant cardiac repolarization abnormalities and life-threatening arrhythmias.[89,90] Cryogenic blockade of the thalamic gating mechanism or its output from the frontal cortex to the brainstem[91] and of the amygdala[92] delayed or prevented the occurrence of ventricular fibrillation during stress in pigs.

Ventricular arrhythmias also ensue immediately on cessation of diencephalic or hypothalamic stimulation, but the appearance of these arrhythmias requires intact vagi and stellate ganglia.[93,94] The likely electrophysiologic basis for such post–central nervous system stimulation arrhythmias is the loss of rate overdrive suppression of ectopic activity. This phenomenon occurs when the vagus nerve regains its activity after cessation of centrally induced adrenergic stimulation. Accordingly, the enhanced automaticity induced by adrenergic stimulation of ventricular pacemakers is exposed when vagus nerve tone is restored and slows the sinus rate.[94] Although these arrhythmias, including ventricular tachycardia, may be dramatic in appearance, they rarely degenerate into ventricular fibrillation.[95] This proarrhythmic effect of dual autonomic activation has been erroneously interpreted as profibrillatory.

The antiarrhythmic influence of beta-adrenergic receptor blockade may result in part from blockade of central beta-adrenergic receptors. Parker and coworkers[96] determined that intracerebroventricular administration of subsystemic doses of L-propranolol (but not D-propranolol) significantly reduced the incidence of ventricular fibrillation during combined left anterior descending coronary artery occlusion and behavioral stress in pigs. In a surprising turn, intravenous administration of even a relatively high dose of L-propranolol was ineffectual. The latter result may relate in part to a species dependence; unlike canines, pigs do not show a suppression of ischemia-induced arrhythmias in response to beta blockade.[97] It has been proposed that the centrally mediated protective effect of beta blockade results from a decrease in sympathetic nerve activity and in plasma norepinephrine concentration.[96,98,99] Of importance, whereas central actions of beta-adrenergic receptor blockers may play an important role in reducing susceptibility to ventricular fibrillation during acute myocardial ischemia, they are unlikely to constitute the sole mechanism because beta blockers prevent the profibrillatory effect of direct stimulation of peripheral sympathetic structures such as the stellate ganglia.[100] Of note, all three of the beta blockers that have long-term effects on mortality in cardiac patients (propranolol, metoprolol, and carvedilol) are lipophilic[101] and therefore cross the blood-brain barrier readily to affect sleep structure, with significant perturbations of sleep continuity.[102]

Autonomic Factors in Arrhythmogenesis during Sleep

NREM sleep generally is salutary with respect to ventricular arrhythmogenesis, as indicated both by extensive studies of neurocardiac interactions and by clinical experience. Activation of the vagus nerve reduces heart rate, increases cardiac electrical stability, and reduces rate-pressure product, an indicator of cardiac metabolic activity, to improve the supply-demand relationship in stenotic coronary artery segments. In the setting of severe coronary disease or acute myocardial infarction, however, hypotension during NREM can lead to myocardial ischemia as a consequence of inadequate coronary perfusion pressure, thereby provoking arrhythmias and myocardial infarction.[11,103] The abrupt increases in vagus nerve tone that can occur during periods of REM or sleep state transitions can result in significant pauses in heart rhythm, bradyarrhythmias, and, potentially, triggered activity, a mechanism of the lethal cardiac arrhythmia torsades de pointes. Patients with the long QT syndrome who have the type 3 phenotype are more prone to experience torsades de pointes at night rather than during stress or exercise.[64] Tonic control of the vagus nerves over the caliber of the epicardial coronary vessels[104] could be an important factor in dynamic regulation of coronary resistance as a function of the sleep-wake cycle. An important question is whether tonic vagus nerve activity exerts a protective or a deleterious influence on myocardial perfusion and arrhythmogenesis in persons with atherosclerotic disease. In these patients, nocturnal surges in vagus nerve activity could precipitate myocardial ischemia and arrhythmias as a result of coronary vasoconstriction rather than dilation in atherosclerotic segments, because of impaired release of endothelium-derived relaxing factor.[105]

Because of the attendant surges in sympathetic nerve activity and in heart rate, REM sleep has the potential to trigger ventricular arrhythmias.[106,107] The striking variability of heart rate and breathing pattern can exert a significant impact on cardiovascular functioning, as is evident in the development of ischemia and arrhythmias in patients whose myocardium is compromised. Indeed, the only clinical studies in which sleep staging has been employed have identified REM as the state in which arrhythmias occurred.[11,108,109] The increased sympathetic nerve activity that occurs at REM sleep onset[10] provides a potent stimulus for ventricular tachyarrhythmias because of the arrhythmogenic influence of neurally released catecholamines. Sympathetic nerve activation by stimulation of central[86-88,93,94] or peripheral adrenergic structures,[100,110] infusion of catecholamines,[111] or imposition of behavioral stress[112,113] can increase cardiac vulnerability in the normal and the ischemic heart. These profibrillatory influences are substantially blunted by beta-adrenergic receptor blockade.[112] A wide variety of supraventricular arrhythmias also can be induced by autonomic activation.[95]

Enhanced sympathetic nerve activity increases cardiac vulnerability in the normal and in the ischemic heart by complex mechanisms. The major indirect effects include an impaired oxygen supply-demand ratio resulting from increased cardiac metabolic activity and coronary vasoconstriction, particularly in vessels with injured endothelium and in the context of altered preload and afterload. The direct profibrillatory effects on cardiac electrophysiologic function are attributable to derangements in impulse formation or conduction, or both.[95] Increased levels of catecholamines activate beta-adrenergic receptors, which in turn alter adenylate cyclase activity and intracellular calcium flux. These actions probably are mediated by the cyclic nucleotide and protein kinase regulatory cascade, which can alter spatial heterogeneity of calcium transients and consequently increase dispersion of repolarization. The net influence is an increased susceptibility to ventricular fibrillation.[82,114] Conversely, reduction of cardiac sympathetic drive by stellectomy has proved to be antifibrillatory.

Notwithstanding the evidence that autonomic factors can significantly alter susceptibility to arrhythmias, the observation that the heart rate surges of REM sleep are conducive to myocardial ischemia, and the epidemiologic data in humans on the extent of sleep-induced cardiac events,[1] very little information is available on the effects of myocardial infarction on the cardiovascular system during sleep. Ventricular ectopic activity, but not ventricular fibrillation, emerges during NREM sleep in pigs after myocardial infarction.[115] This pattern may be attributable to slowing of heart rate and increased vagus nerve activity during NREM sleep, conditions that can inhibit the normal overdrive suppression of ventricular rhythms by sinoatrial node pacemaker activity and result in firing of latent ventricular pacemakers and triggered activity. Snisarenko[116] found significant elevations in heart rate in both the acute (4 to 10 days) and subacute (3 to 12 months) periods after myocardial infarction in a feline model. In the acute period, these effects were accompanied by increased wakefulness, decreased heart rate variability, and severely disordered sleep. In the intervening weeks, sleep quality recovered fully until, in the subacute period, beta blockade with propranolol led to renewed, pronounced disturbances in sleep structure, with increased wakefulness, reduction in REM sleep, and prolongation of stages N1 and N2 of NREM sleep. Snisarenko attributed these results to reflex activation of adrenergic, noradrenergic, and dopaminergic nerves in several brain structures after coronary artery ligation.[117]

CLINICAL PEARLS

- REM sleep is characterized by surges in sympathetic and vagus nerve activity, which are well tolerated in normal individuals but may result in cardiac arrhythmias, myocardial ischemia, and myocardial infarction in those with heart disease.
- During NREM sleep, systemic blood pressure may fall, potentially reducing flow through stenotic coronary vessels, which may precipitate myocardial ischemia or infarction.
- Sleep-disordered breathing induces significant injury in autonomic regulatory areas.
- In essence, sleep constitutes an autonomic stress test for the heart, and nighttime monitoring of cardiorespiratory function is of considerable diagnostic value.[104]

SUMMARY

Sleep states exert a major impact on cardiorespiratory function, as a direct consequence of the significant variations in brain states that occur in the normal cycling between NREM and REM sleep. Dynamic fluctuations in central nervous system variables influence heart rhythm, arterial blood pressure, coronary artery blood flow, and ventilation. Whereas REM sleep–induced surges in sympathetic and parasympathetic nerve activity, with accompanying significant surges and pauses in heart rhythm, are well tolerated in normal people, patients with heart disease may be at heightened risk for life-threatening arrhythmias and myocardial ischemia and infarction.[83,107] During NREM sleep, in the severely compromised heart, a potential for hypotension exists that can impair blood flow through stenotic coronary vessels to trigger myocardial ischemia or infarction.[11] Cumulative damage incurred as a consequence of sleep-disordered breathing, heart failure, or stroke to the central brain areas that regulate autonomic activity and coordinate upper airway and diaphragmatic action can lead to enhanced sympathetic outflow, increasing risk in heart failure and contributing to hypertension in obstructive sleep apnea. Coordination of cardiorespiratory control is especially pivotal in infancy, when developmental immaturity can compromise function and pose special risks. Throughout sleep, the coexistence of coronary disease and apnea is associated with heightened risk of cardiovascular events[113,118] resulting from the challenge of dual control of the respiratory and cardiovascular systems.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported by the following grants from the National Institutes of Health (NIH): MH13923 from the National Institute of Mental Health, HD22695 from the National Institute of Child Health and Human Development, HL22418 and HL60296 from the National Heart, Lung, and Blood Institute, and NR012810 from the National Institute of Nursing Research.

We thank Sandra S. Verrier for her editorial contributions.

Selected Readings

Harper RM, Kinney HC. Potential mechanisms of failure in the sudden infant death syndrome. *Curr Pediatr Rev* 2010:**6**:39–47.

Harper RM, Kumar R, Macey PM, et al. Functional neuroanatomy and sleep-disordered breathing: implications for autonomic regulation. *Anat Rec (Hoboken)* 2012;**295**:1385–95.

Harper RM, Kumar R, Macey PM, et al. Affective brain areas and sleep-disordered breathing. *Prog Brain Res* 2014;**209**:275–93.

Harper RM, Kumar R, Ogren JA, et al. Sleep-disordered breathing: effects on brain structure and function. *Respir Physiol Neurobiol* 2013;**188**: 383–91.

Harper RM, Macey PM, Henderson LA, et al. fMRI responses to cold pressor challenges in control and obstructive sleep apnea subjects. *J Appl Physiol* 2003;**94**:1583–95.

Harper RM, Macey PM, Woo MA, et al. Hypercapnic exposure in congenital central hypoventilation syndrome reveals CNS control mechanisms. *J Neurophysiol* 2005;**93**:1647–58.

Macey PM, Kumar R, Woo MA, et al. Brain structural changes in obstructive sleep apnea. *Sleep* 2008;**31**:967–77.

Mansukhani MP, Wang S, Somers VK. Chemoreflex physiology and implications for sleep apnoea: insights from studies in humans. *Exp Physiol* 2015;**100**:130–5.

Ogren JA, Macey PM, Kumar R, et al. Impaired cerebellar and limbic responses to the Valsalva maneuver in heart failure. *Cerebellum* 2012;**11**: 931–8.

Woo MA, Palomares JA, Macey PM, et al. Global and regional brain mean diffusivity changes in patients with heart failure. *J Neurosci Res* 2015;**93**: 678–85.

A complete reference list can be found online at ExpertConsult.com.

Chapter
14

Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders

Paola A. Lanfranchi; Jean-Louis Pépin; Virend K. Somers

Chapter Highlights

- Autonomic control of the circulation is pivotal in ensuring an adequate cardiac output to the vital organs through continuous and rapid adjustments of heart rate, arterial blood pressure, and redistribution of blood flow. In the longer term, neural circulatory control appears to be coupled with the circadian rhythm, the sleep-wake cycle, and ultradian rhythms, including rapid eye movement (REM) and non–REM (NREM) sleep processes.
- The cardiovascular autonomic nervous system's primary role is to ensure an adequate cardiac output to the vital organs through continuous and rapid adjustments of heart rate, arterial blood pressure, and redistribution of blood flow.
- Heart rate and blood pressure have a diurnal rhythm characterized by a significant reduction during nighttime hours. This physiologic pattern can be altered by sleep insufficiency and sleep disorders, with important implications for cardiovascular health.
- During NREM sleep, there is an increase in cardiovagal drive and a reduction in cardiac and peripheral sympathetic activity. Baroreflex gain is heightened in response to blood pressure increments rather than decrements during NREM sleep to ensure the maintenance of stable low blood pressure and heart rate during NREM sleep. By contrast, REM sleep is a state of autonomic instability, dominated by remarkable fluctuations between parasympathetic and sympathetic influences.
- Sleep loss, alterations in sleep quality, and sleep disorders are associated with persistence of high sympathetic activity during night and reduction in physiologic nocturnal blood pressure dipping. These effects lead to sustained sympathetic activation with increased blood pressure during the succeeding days.

OVERVIEW

Autonomic circulatory control operates via parasympathetic neurons to the heart and by sympathetic neuronal efferents to the heart, blood vessels, kidneys, and adrenal medulla. Parasympathetic stimulation of the heart, through the activation of cardiac muscarinic receptors, results in bradycardia, whereas sympathetic stimulation of the heart, through activation of $beta_1$-adrenergic receptors, results in tachycardia and increased contractility. Sympathetic activation in the vascular bed induces both vasoconstriction, by stimulating $alpha_1$ adrenoreceptors, and vasodilation, by stimulating $beta_2$ adrenoreceptors. Several reflexes, including the arterial baroreflex, cardiopulmonary reflexes, and chemoreflexes, also are important in the rapid adjustments of circulation that occur in association with postural changes, hypoxemia, temperature changes, and sleep. Heart rate (HR) and blood pressure (BP) have a diurnal rhythm characterized by a significant reduction during nighttime hours, secondary to changes in activity and posture, as well as sleep and circadian influences. This physiologic pattern can be altered by sleep insufficiency and sleep disorders, with important implications for cardiovascular health.

The cardiovascular autonomic nervous system seeks to maintain homeostasis through precise control of numerous hemodynamic variables, including HR, arterial BP, and peripheral blood flow, on a beat-by-beat basis. Neural circulatory control is intimately linked to sleep and circadian physiology, as demonstrated by the disrupted autonomic control that accompanies sleep disruption, for example, with sleep loss and sleep apnea. On the other hand, whether primary alterations in autonomic function may translate into sleep disturbances also needs to be considered.

This chapter begins with a general overview of autonomic cardiovascular regulation and of its central and peripheral controllers, followed by a description of the methods used to explore cardiovascular neural control during sleep in humans and a review of their advantages and limitations. Presented next is an outline of some of the current knowledge of neural circulatory control during normal sleep. Also included is a survey of changes that may occur as a consequence of sleep deprivation, alterations in sleep quality, sleep apnea, and autonomic dysfunction such as may occur in diabetes.

THE CARDIOVASCULAR AUTONOMIC NERVOUS SYSTEM: DEFINITION AND FUNCTIONS

The cardiovascular autonomic nervous system is a highly integrated network that controls visceral functions, which, on a

short timescale (seconds to hours), adjusts the circulation in keeping with behavior, the environment, and emotions. Its primary role is to ensure an adequate cardiac output to the vital organs through continuous and rapid adjustments of HR and arterial BP and redistribution of blood flow. In the longer term, this neural circulatory regulation appears to be coupled with the circadian rhythm, the sleep-wake cycle, and some ultradian rhythms, including rapid eye movement (REM) and non-REM (NREM) sleep processes, as well as hormones implicated in long-term BP regulation.

Neural control of the circulation operates via parasympathetic neurons to the heart and by sympathetic neuronal efferents to the heart, blood vessels, kidneys, and adrenal medulla. Parasympathetic stimulation of the cardiovascular system is mediated primarily by the vagus nerve through the activation of muscarinic receptors and results in bradycardia. Sympathetic stimulation of the heart acts through activation of $beta_1$ adrenoreceptors at the sinoatrial node (the cardiac pacemaker) and in the myocardium (the cardiac muscle) and results in tachycardia and increased contractility. Sympathetic activation in the vascular bed induces vasoconstriction by stimulating $alpha_1$ adrenoreceptors (in the skin and splanchnic regions) and vasodilation by stimulating $beta_2$ adrenoreceptors (in the heart and skeletal muscles). Parasympathetic and sympathetic efferent activity to the heart may also modulate cardiac electrophysiologic properties that can be relevant to the pathogenesis of several types of arrhythmias, particularly in the presence of a pro-arrhythmic substrate.

Central organization of the autonomic nervous system and its relationship with sleep-modulating mechanisms are detailed in Chapter 13. Briefly stated, autonomic impulses to the vasculature and heart originate from the vasomotor center in the brainstem, located bilaterally in the reticular substance of the medulla and pons. The vasomotor center is in turn modulated by higher nervous system regions in the pons, mesencephalon, and diencephalon, including the hypothalamus and many portions of the cerebral cortex. Several cardiovascular reflexes also are important in the rapid adjustments of blood pressure occurring in association with postural changes, hypoxemia, exercise, and moderate temperature changes and, in addition, may be implicated in cardiovascular changes observed during sleep. These mechanisms include the arterial baroreflex, cardiopulmonary reflexes, and chemoreflexes. The renin-angiotensin-aldosterone system, vasopressin, and other vasoactive mechanisms also may contribute to cardiovascular regulation during sleep.

Arterial Baroreflex

The arterial baroreflex is an important regulator of BP in the short term.[1] The baroreceptors are sensory receptors in the aortic arch and carotid sinuses that relay in the medullary regions of the brain. Changes in arterial baroreceptor afferent discharge trigger reflex adjustments that buffer or oppose the changes in BP. For instance, increments in BP stretch the receptors, resulting in heightened afferent traffic to the brainstem neuronal network. This increased trafficking inhibits efferent sympathetic outflow to cardiac and vascular smooth muscle and increases parasympathetic cardiac tone, resulting in slowing of HR, reduction in contractility, and reduced peripheral vasoconstriction, with subsequent compensatory decreases in BP. A decrease in BP has opposite effects: It elicits reflex tachycardia, increased contractility, and peripheral vasoconstriction, with subsequent compensatory increases in BP.

Cardiopulmonary Reflexes

Cardiopulmonary reflexes are triggered by the stimulation of low-pressure receptors located in the atria, ventricles, and pulmonary arteries. The cardiopulmonary receptors are volume receptors that serve to mitigate changes in BP in response to changes in blood volume. The firing pattern of these receptors parallels the pressure changes within the cardiac chambers or vessels and help to regulate blood volume. Cardiopulmonary reflex activation results in peripheral vasodilation, reduction in sympathetic outflow to the kidney, and activation of the posterior pituitary gland to inhibit the release of antidiuretic hormone, resulting in increased urine excretion.

Arterial and cardiopulmonary reflexes are implicated in blood pressure regulation during postural changes. Assumption of the upright position produces a caudal shift in blood volume and acutely reduces stroke volume and blood pressure. The circulatory adjustment to this orthostatic stress is rapid and is characterized by reflex increases in HR and peripheral vascular resistance, followed by enhanced secretion of antidiuretic hormone and activation of the renin-angiotensin system. Recumbency, as during sleep, leads to production of an increased volume load in the cardiac chambers and elicits the opposite effects.

The Chemoreflexes

The chemoreflexes mediate the ventilatory response to hypoxia and hypercapnia and also exert important cardiovascular effects.[2] The peripheral arterial chemoreceptors, the most important of which are located in the carotid bodies, respond primarily to changes in the partial pressure of oxygen. Hypoxemic stimulation elicits an increase in respiratory muscle output, inducing hyperventilation, and an increase in sympathetic outflow to peripheral blood vessels, resulting in vasoconstriction. Hyperventilation in turn activates pulmonary stretch receptors, which buffer the increases in sympathetic and vagal outflow, thereby maintaining homeostasis under normal conditions. During apnea, when hyperventilation is absent or prevented, vasoconstriction is potentiated and occurs simultaneously with activation of cardiac vagal drive resulting in bradycardia, collectively termed the "diving reflex," a protective mechanism that helps preserve blood flow to the heart and brain while limiting cardiac oxygen demand.[3,4]

The central chemoreceptors are located in the brainstem and respond to changes in pH mediated primarily by carbon dioxide tension. Stimulation of central chemoreceptors by hypercapnia also elicits sympathetic and respiratory activation, but without the cardiovagal effects seen with hypoxemia.[2]

MEASURES TO EXPLORE AUTONOMIC CHANGES DURING SLEEP AND THEIR PHYSIOLOGIC SIGNIFICANCE

Heart Rate and Arterial Blood Pressure and Their Variability

The RR interval, the time elapsed between two successive R waves of the QRS signal on the electrocardiogram (and its reciprocal, the HR), is a function of intrinsic properties of the sinus node as well as autonomic influences. BP is a function of vascular resistance (an expression of arterial constriction or

dilation) and cardiac output (the blood volume being pumped by the heart in 1 minute), which is a function of HR, cardiac contractility, and diastolic blood volume, all components controlled in part by the autonomic nervous system.

Autonomic cardiovascular regulation can be investigated through the quantification of average HR and BP as assessed in steady state conditions (wakefulness and sleep, for instance) or in their responses to endogenous or exogenous challenges (e.g., changes in posture, response to respiratory changes, and arousal from sleep).

Both HR and BP exhibit spontaneous fluctuations that can be described by the standard deviation around the mean, or by their rhythmic and nonrhythmic characteristics. When described by the standard deviation over 24-hour ambulatory recordings, high variability of the RR interval is a recognized index of the ability of the cardiovascular system to cope with environmental challenges.[5] By contrast, heightened BP variability is found to accompany aging and hypertension.[6] Among the various cyclic components that characterize 24-hour HR and BP variability, those occurring from daytime wakefulness to nighttime sleep have received great attention. Specifically, HR and BP physiologically decrease during night hours.[7] This diurnal pattern is evident in ambulatory subjects, and even in recumbent subjects maintaining the sleep-wake cycle.[8] The contribution of circadian versus noncircadian factors and how these may be modified in sleep disorders are discussed later in the chapter.

RR and BP also exhibit short-term oscillations, in a frequency range between 0 and 0.5 Hz, which appear to be under the influence of intrinsic autonomic rhythms and of respiratory inputs. Spectral analysis of RR and BP variability provides an estimate regarding how power (i.e., variance) of the signal is distributed as a function of frequency. Indeed, RR and BP variability appear to be organized in three major components: the high-frequency (HF) (>0.15 Hz) respiratory band, the low-frequency band (LF) (around 0.1 Hz), and the very-low-frequency (VLF) band (0.003 to 0.039 Hz) (Figure 14-1 and Table 14-1).[9] The HF components of RR variability primarily reflect the respiration-driven modulation of sinus rhythm, evident as sinus arrhythmia, and have been used as an index of tonic vagal drive. Nonneural mechanical mechanisms, linked to respiratory fluctuations in cardiac transmural pressure, atrial stretch, and venous return, also are determinants of HF power and may become especially important after cardiac denervation such as with heart transplantation.[10] The LF rhythm, which appears to have a widespread neural genesis,[11] reflects in part the sympathetic modulation of the heart,[12] as well as the baroreflex responsiveness to the beat-to-beat variations in arterial BP,[13] but also can be modulated by LF or irregular breathing patterns. Of importance, LF components in respiration confound the interpretation of the LF component of cardiovascular variability in attempts to identify the autonomic characteristics of cardiovascular control. Therefore, in any assessment of the relative contributions of the LF and HF components to any particular physiologic state or disease condition, it is crucial to ensure that the respiratory pattern is limited to the HF component. The LF/HF ratio is used to provide an index of the balance of the

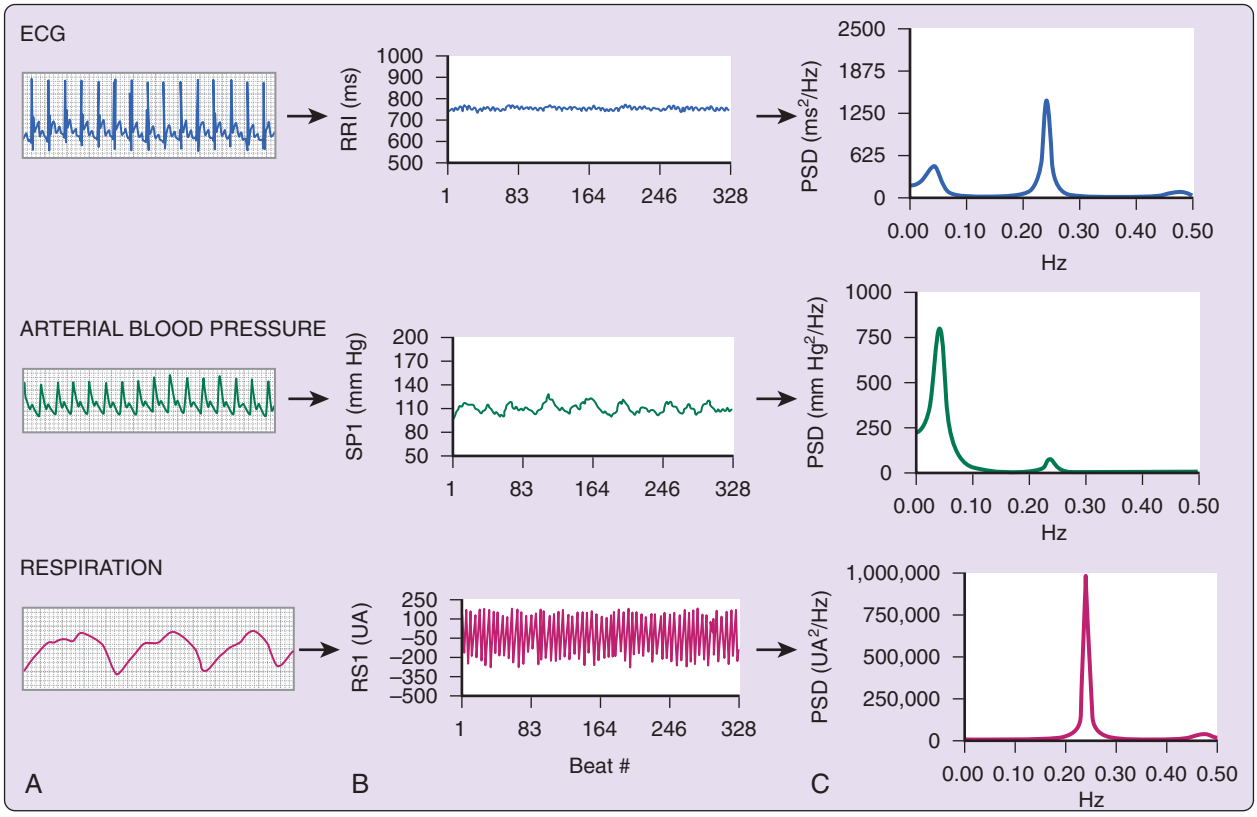


Figure 14-1 **A,** ECG, beat-to-beat blood pressure (BP), and respiration recordings. **B,** Temporal series of RR intervals, BP, and respiration; **C,** Power spectra of RR, BP, and respiration variability **(C)** in a single healthy subject. PSD, Power spectral density; RRI, RR intervals; RS1, Respiratory signal; SP1, Systolic pressure; UA, Arbitrary units.

Table 14-1 Spectral Components of RR Interval Variability in the Short Term*

Variable	Units	Description Analysis of Short-term Recordings (5 min)	Frequency Range
Total power	ms^2	The variance of RR intervals over the temporal series analyzed	Approximately ≤0.4 Hz
VLF	ms^2	Power in the VLF range	≤0.04 Hz
LF	ms^2	Power in the LF range	0.04–0.15 Hz
LF norm	%	LF power in normalized units: LF/(total power – VLF) × 100	
HF	ms^2	Power in HF range	0.15–0.4 Hz
HF norm	%	HF power in normalized units: HF/(total power – VLF) × 100	
LF/HF		Ratio: LF [ms^2]/HF [ms^2]	

HF, High frequency; LF, low frequency; VLF, very low frequency.
*Approximately 5 minutes.

sympathovagal influence on the sinus node,[14] provided that measurements are obtained in strictly controlled conditions. Finally, the VLF component has been hypothesized to reflect thermoregulation and the renin-angiotensin system.[15] Regarding BP variability, LF components in systolic BP variability are considered an index of efferent sympathetic vascular modulation, whereas the HF components reflect mechanical effects of respiration on blood pressure changes.[12] Measurements of HF, LF, and VLF usually are made in absolute (millisecond) values, but LF and HF often are presented in normalized units (nu), which represent the relative value of each power component in proportion to the total power minus the VLF components (see Table 14-1). Normalization allows minimizing the effect of changes in total power on LF and HF components.

Traditional spectral analysis techniques include fast Fourier transform algorithms and autoregressive modeling, which in most instances provide comparable results.[16] These techniques require stationarity of the signal being processed and therefore cannot be applied to processes embodying significant transient activity (e.g., sleep onset, arousals, sleep stage transition and awakening). In addition, such methods have to be used with caution in association with respiratory or motor events (e.g., periodic limb movements, bruxism). More advanced algorithms of signal processing can be used to overcome this limitation and permit the assessment of dynamic changes in autonomic cardiovascular control during transient events (e.g., sleep onset, arousal, bruxism)[17] and help define the temporal relationship between dynamic changes occurring in different systems, such as between the electroencephalogram (EEG) and the electrocardiogram (ECG).[18,19] The most commonly used algorithms include short time Fourier transform, Wigner-Ville distribution, time variant autoregressive models, wavelets, and wavelet-packets.[17]

Finally, in addition to the periodic oscillatory behavior observed in RR interval and arterial blood pressure, a less specific variability occurs with nonperiodic behavior, which can be described by methods based on nonlinear system theory ("chaos theory and fractal analysis").[20] The physiologic basis for this nonharmonic beat-to-beat behavior, which extends over a wide time range (seconds to hours), is still unsettled, although some investigators have proposed that it is under higher central modulation.[21] The application of this type of analysis to sleep cardiovascular physiology is still limited.

Baroreflex Sensitivity

The arterial baroreflex is important in buffering short term changes in BP. The gain of the arterial baroreflex, or baroreflex sensitivity, is measured by the degree of change in heart rate or sympathetic traffic for a given unit change in blood pressure.[22] Two techniques have been mainly used in sleep research to assess spontaneous baroreflex modulation of heart rate: the sequence technique and the spectral analysis technique. The first technique identifies sequences of consecutive beats in which progressive increases in systolic BP are followed by a progressive lengthening in RR (or vice versa). The slope of the regression line between RR intervals and systolic BP within these sequences is taken as the magnitude of the reflex gain. The second technique is based on cross-spectral analysis of short segments of systolic BP and RR and relies on the assumption that a certain frequency band of RR variability, between 0.04 and 0.35 Hz, is modulated by the baroreflex. Baroreflex sensitivity is expressed by the gain of the transfer function relating changes in blood pressure to coherent changes in RR or muscle sympathetic nerve activity (MSNA).

Preejection Period

The *preejection period* (PEP) is the time elapsed between the electrical depolarization of the left ventricle (QRS on the ECG) and the beginning of ventricular ejection and represents the period of left ventricular contraction with the cardiac valves closed. PEP is influenced by sympathetic activity by way of $beta_1$ adrenoreceptors and shortens under stimulation. PEP can be derived noninvasively from impedance cardiography, which converts changes in thoracic impedance (as measured by electrodes on the chest and neck) to changes in volume over time and allows tracking of volumetric changes such as those occurring during the cardiac cycle. This method has been applied, although not intensively, to assess cardiac sympathetic influences in steady state conditions during sleep.[23,24] The application to transient sympathetic responses is unfortunately limited, because errors can occur in interpretation in the presence of blood pressure increases, which can

induce a lengthening of PEP (instead of the expected shortening) owing to the longer time required to overcome the external pressure.

Microneurographic Recording of Sympathetic Nerve Activity

Microneurography provides direct information on sympathetic vasomotor and sudomotor activity to muscle and skin. MSNA, usually measured at the peroneal nerve, induces vasoconstriction and is modulated by the baroreflex.[25] MSNA also increases in response to hypoxic and hypercapnic chemoreceptor stimulation.[1] Skin sympathetic nerve activity reflects thermoregulatory output related to sudomotor and vasomotor activity and is affected by emotional stimuli but not by the baroreflex.

Although microneurography provides a direct measure of peripheral sympathetic drive, it is invasive and technically demanding for both operator and patient. In addition, the information provided is limited to regional sympathetic neural activity. In view of the heterogeneity of system-specific innervations, MSNA and skin sympathetic nerve activity assessments may not necessarily reflect global sympathetic tone.

Peripheral Arterial Tone and Pulse Transit Time

Peripheral arterial tone (PAT), as measured from the finger, provides an indirect index of sympathetic vasoconstrictory mechanisms directed to the peripheral vascular bed. It is based on measurement of the pulsatile volume changes in the vascular bed at the fingertip, which decreases secondary to sympathetically mediated alpha-adrenergic vasoconstriction. Accordingly, PAT amplitude declined less in patients with sleep apnea at end-apnea and arousal after administration of the alpha-adrenergic blocker phentolamine.[26] PAT does not provide absolute values. Only within-subject changes in pulse wave analysis during a limited time interval can be evaluated, but these may be sufficient to assess PAT attenuation related with respiratory events and microarousals.[27] PAT is noninvasive, can be monitored continuously during sleep, has been proposed as a measure of the autonomic changes occurring with arousal in adults and children[28-30] and, in combination with actigraphy and oxymetry, has been used in the diagnosis of sleep apnea. REM sleep is associated with increased and hugely variable sympathetic tone. High sympathetic tone corresponds on the PAT signal to a sustained attenuation that has been reported to help identify REM sleep.[31] Moreover, episodic vasoconstriction associated with the occurrence of rapid eye movements is superimposed on this attenuation. Differences in amplitude of the PAT signal and its variability during REM sleep versus NREM sleep have been reported, and an automatic REM scoring algorithm has been developed and validated for such scoring purpose.[32]

Pulse transit time (PTT) refers to the time required for a pulse wave to travel between two arterial sites.[33] In practice, in a noninvasive estimate of PTT, the R wave in the ECG generally is used to indicate the starting point of the measure, and the peripheral waveform (assessed by photoplethysmography at the finger) to indicate the end of the measure. PTT is sensitive to moment-to-moment sympathetic neural activity and shortens when BP increases and lengthens when BP falls. Of note, PTT encompasses several physiologic components difficult to control for, and intersubject comparison is not recommended. Only intraindividual relative PTT changes from a baseline condition (over several readings) are instead recommended for clinical consideration. Like PAT, PTT also can be monitored continuously and has been used in the assessment of sympathetic responses to arousals[29,34] and respiratory events, especially in children.[35,36] During REM sleep, variations in sympathetic activity are spontaneously very high, so the PTT baseline is highly variable. Thus the recognition of true micro arousals during REM sleep is less specific than in other sleep stages. The heart and large vessels are located in the thoracic cavity and consequently are affected by variations in thoracic volume and pressure. During inspiration, the volume of the thoracic cavity increases, reducing intrathoracic pressure, which in turn reduces the compression of the heart and large vessels (vena cava and aorta), decreasing BP and slowing PTT. The opposite is true for expiration: As the intrathoracic pressure increases, the heart is compressed, and BP increases and PTT quickens. PTT may serve as a noninvasive marker of respiratory effort, especially for defining certain respiratory events (hypopneas, respiratory effort–related arousals [RERAs], and central events).[37,38]

Systemic Catecholamines

Measurement of plasma catecholamines—epinephrine and norepinephrine—provides an estimate of global sympathetic activity. However, blood norepinephrine reflects only a small percentage (8% to 10%) of neurotransmitter release during sympathetic activation. Moreover, the relatively rapid clearance of catecholamines from the bloodstream may limit the ability to detect transient changes in sympathetic activity. Consequently, only frequent sampling through sleep may detect changes related to the sleep-wake cycle and sleep stages.[39] Measurement of urinary excretion of catecholamines and their metabolites is a simpler approach to provide an estimate of the cumulative catecholamine secretion over time and has been used widely in the clinical and sleep research settings. Urinary catecholamine excretion is strictly dependent on renal function. Accordingly, a correction of excreted catecholamine for indices of renal function (urinary creatinine) is recommended.

SLEEP-RELATED CARDIOVASCULAR AUTONOMIC CHANGES

Day-Night Changes in Neural Circulatory Control

HR and BP physiologically decrease during nighttime as compared with daytime in ambulant subjects, as well as in subjects kept in the supine position for 24 hours.[8] Specifically, the normal 24-hour BP pattern consists of a 10% or greater systolic blood pressure reduction during sleep compared with daytime, a reduction that is commonly referred to as "dipping." Posture and activity strongly influence HR and BP during the day,[40] whereas posture and sleep affect HR and BP at night.[8] However, the nocturnal sleep-related cardiovascular dipping is evident even in subjects who maintain the supine position for 24 hours,[8] underscoring the importance of sleep in inducing decreases in nighttime HR and BP. Studies investigating the autonomic changes associated with the wake-sleep cycle noted that indices of parasympathetic function, such as RR interval and HF components of RR variability, begin to change as early as 2 hours before sleep onset,[23] whereas indices of cardiac and peripheral sympathetic activity such as LF/HF ratio, preejection period, MSNA, and catecholamines

start to decrease only after sleep onset and continue to decrease with the deepening of sleep.[23,25,39] Morning awakening induces a stepwise activation of the sympathoadrenal system, with increased HR, BP, and plasma catecholamines, with further increases occurring with postural change and physical activity.[23,41]

Studies conducting 24 hours of sleep deprivation with the subjects supine showed that the nocturnal fall in HR and cardiovagal indices is still present, whereas the fall in nocturnal BP and PEP prolongation (i.e., decreased sympathetic activity) are blunted.[23,42] It may be, therefore, that HR and parasympathetic mechanisms are largely under circadian influences and might be implicated in mechanisms preparatory to sleep, whereas sympathetic drive to the heart and vessels is mainly linked to the wake-sleep cycle. Increasing evidence suggests that the mean nocturnal BP level is a major predictor of *cardiovascular* morbidity and mortality irrespective of the 24-hour BP levels.[43] Any deterioration in sleep quality or quantity may be associated with an increase in nocturnal BP that could participate in the development or poor control of hypertension.[44]

Physiologic Responses to NREM and REM Sleep

In healthy subjects, autonomic cardiovascular regulation varies considerably with sleep stage, and different autonomic patterns dominate in NREM versus REM sleep. As NREM sleep progresses from stages N1 to N3 (Stages 3 and 4 on Figure 14-2), the RR, respiratory-mediated HF components of RR variability and PEP increase, whereas BP, LF components in systolic BP variability, and MSNA significantly decrease, compared with wakefulness. These changes suggest an increase in cardiovagal drive and a reduction in cardiac and peripheral sympathetic activity[8,25,45] (Figure 14-2). Baroreflex sensitivity appears also to be increased during NREM sleep

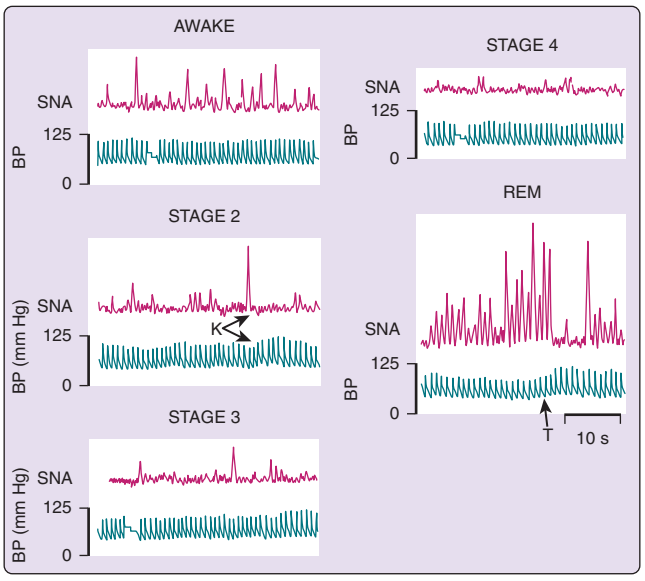


Figure 14-2 Recordings of sympathetic nerve activity (SNA) and mean blood pressure (BP) in a single subject while awake and while in stages 2, 3, and 4 and REM sleep. SNA and BP gradually decrease with the deepening of NREM sleep. Heart rate, BP, and BP variability increase during REM sleep, together with a profound increase in the frequency and amplitude in SNA. K, K-complexes. (Modified from Somers VK, Dyken ME, Mark AL, Abboud FM. Sympathetic-nerve activity during sleep in normal subjects. *N Engl J Med* 1993;328:303–7.)

over that in wakefulness.[46] However, the response is variable. Namely, compared with that in wakefulness, baroreflex gain is heightened in response to BP increments rather than decrements during NREM sleep. This mechanism probably serves to ensure the maintenance of stable low BP and HR during NREM sleep.

By contrast, REM sleep is a state of autonomic instability, dominated by remarkable fluctuations between parasympathetic and sympathetic influences, which produce sudden and abrupt changes in HR and BP.[47] The average HR and BP are higher during REM than in NREM sleep, as is sympathetic neural vasomotor drive.[25] The cardiovascular excitation of REM sleep also is reflected by a significant increase in the low frequency (LF) components (approximately 0.1 Hz) and a shift of the LF/HF ratio toward sympathetic predominance.[8]

RR Interval Variability and Electroencephalographic Coupling

Studies assessing the overnight relationship between RR variability and electroencephalogrphy (EEG) profiles showed that the dynamic of RR interval variability is closely related to the dynamic of EEG, reflecting the depth of sleep. The presence of an ultradian 80- to 120-minute rhythm in the normalized LF, with high levels during rapid eye movement (REM) sleep and low levels during slow wave sleep, has been described.[48] These oscillations were strikingly coupled in a "mirror image" to the overnight oscillations in delta wave activity, which reflect sleep deepening and lightening. Similarly, it was reported that normalized HF components of RR variability were coherent with all EEG spectral bands, with a maximum gain (the ratio of HF amplitude to EEG amplitude was higher) for delta activity and minimum gain (i.e., HF was lower) with beta activity.[49] The two oscillations were coupled with a phase shift of several minutes, with cardiac changes preceding the EEG changes. Although the mechanisms underlying this coupling are not known, it has been hypothesized that a central generator may act to synchronize the oscillatory process in autonomic and sleep modulation, whereby cardiovascular function may anticipate sleep stage changes.[48]

Autonomic Responses Associated with Arousal from Sleep and with Periodic Leg Movements

Arousals

Electrocortical arousal from sleep (i.e., EEG desynchronization with appearance of a low-voltage, high-frequency EEG pattern), either spontaneous or provoked by an exogenous stimuli, or in the context of sleep-disordered breathing, is associated with sympathetic neural surges, leading to transient increases in HR, BP, and MSNA,[50-52] abrupt PTT dips, and PAT attenuations. The typical cardiac response is biphasic, with tachycardia lasting 4 to 5 seconds followed by bradycardia, with HR increasing before cortical arousals. Using time variant analysis, it appears that the surge in sympathoexcitation as represented by LF components of RR variability and BP variability remains substantially elevated above baseline long after the HR, BP, and MSNA return to baseline values.[18] This finding can be particularly relevant in conditions characterized by frequent arousals across the night, conceivably leading to a sustained sympathetic influence on the cardiovascular system. In sleep apnea, an association between repetitive attenuations in peripheral arterial tonometry (PAT) during

sleep and office blood pressure has been reported independently of age, sex, and body mass index.[53] These results suggest that nocturnal sympathetic activity may represent a direct stimulus to chronically elevated blood pressure in humans, even in the daytime.

Auditory stimuli during sleep may result in autonomic and respiratory modifications even in the absence of overt EEG activation (the so-called *autonomic arousal*), or in association with an EEG pattern different from that for conventional arousal, such as K-complexes or bursts of delta waves not followed by EEG desynchronization (the *subcortical arousal*).[51,52] These observations point to a range of partial arousal responses implicating autonomic responses with EEG manifestations different from classical arousals and even without any EEG response. The different EEG patterns and the associated cardiac responses indicate a hierarchical spectrum of increasing strength from the weaker high-amplitude delta burst to a stronger low-voltage alpha rhythm[52] (Figure 14-3).

Periodic Leg Movements during Sleep

Periodic leg movements (PLMs) are described as a repetitive rhythmic extension of the big toe and dorsiflexion of the ankle, with occasional flexion at the knee and hip. PLM can occur during wakefulness as well as during sleep (PLMS). PLMSs occur frequently in several sleep disorders (such as restless legs syndrome, narcolepsy, REM sleep behavior disorder, and sleep apnea) and in patients with congestive heart failure[54] but also are seen in healthy, asymptomatic subjects, especially with advancing age.[55] In the context of sleep apnea, PLMSs may coexist with (and often are difficult to distinguish from) respiratory response–related leg movements, which are part of the arousal response at the end of airway obstruction (in obstructive sleep apnea [OSA]) or at the peak of ventilation (in central sleep apnea). Approximately 30% of PLMSs are associated with cortical arousal, whereas more than 60% are associated with K-complexes or bursts of delta waves.[56] What causes PLMS is still unknown. However, studies of cardiovascular changes associated with PLMSs and their temporal relationship with EEG events are providing new insights into the physiologic mechanisms of PLMS. A stereotypical autonomic response accompanies PLMS, consisting of a rapid rise in HR and arterial BP[56,57] followed by a significant and rapid bradycardia and a return of BP to baseline values (Figure 14-4). Such cardiovascular changes are present whether or not the PLMSs are associated with arousals. The magnitude of the cardiovascular response is greater, however, when PLMSs are associated with cortical arousals. In addition, the amplitude of cardiovascular responses of PLMS is greater during sleep than that associated with spontaneous or simulated PLMSs during wakefulness. These observations suggest that the intensity of cardiovascular responses observed with PLMS is related to the degree of central brain activation (brainstem to cortical activation) that accompanies PLMS and much less to the somatomotor response (i.e., not a classical sensory motor reflex).

Studies assessing the temporal relationship between the leg motor event and autonomic and cortical activation consistently reported that changes in HR and EEG activity precede by several seconds the leg movement.[56,58] Specifically, HR and EEG delta waves rise first, followed by motor activity, and eventually progressive activation of faster EEG frequencies (i.e., in the alpha, beta, and sigma frequencies). A study assessing the dynamic time course of RR variability changes and EEG changes in association with PLMS confirmed the LF components of RR variability to be the first physiologic change to occur, followed by EEG changes in delta frequencies, and thereafter the leg movement with or without faster EEG frequencies.[19] These data corroborate an original hypothesis suggesting the presence of an integrative hierarchy of the arousal response primarily involving the autonomic responses with sympathoexcitation, then progressing towards EEG synchronization (represented by bursts of delta waves) and finally EEG desynchronization (arousal) and eventually awakening.[56] In this view, leg movements are part of the same periodic activation process that is responsible for cardiovascular and EEG changes during sleep.[58]

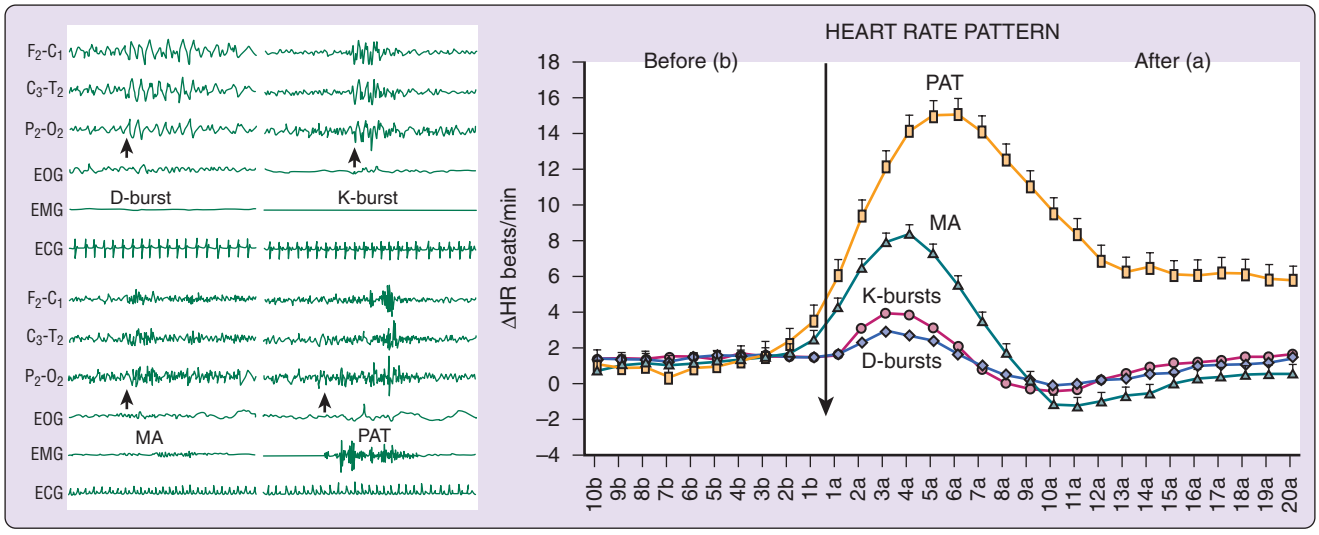


Figure 14-3 Heart rate response (ΔHR) in association with different patterns of EEG activation. K-bursts and D-bursts refer to K-complexes and delta waves, respectively. EEG, Electroencephalogram; EMG, electromyogram; EOG, electrooculogram; MA, microarousal; PAT, peripheral arterial tone. (Modified from Sforza E, Jouny C, Ibanez V. Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response. *Clin Neurophysiol* 2000;111:1611–9.)

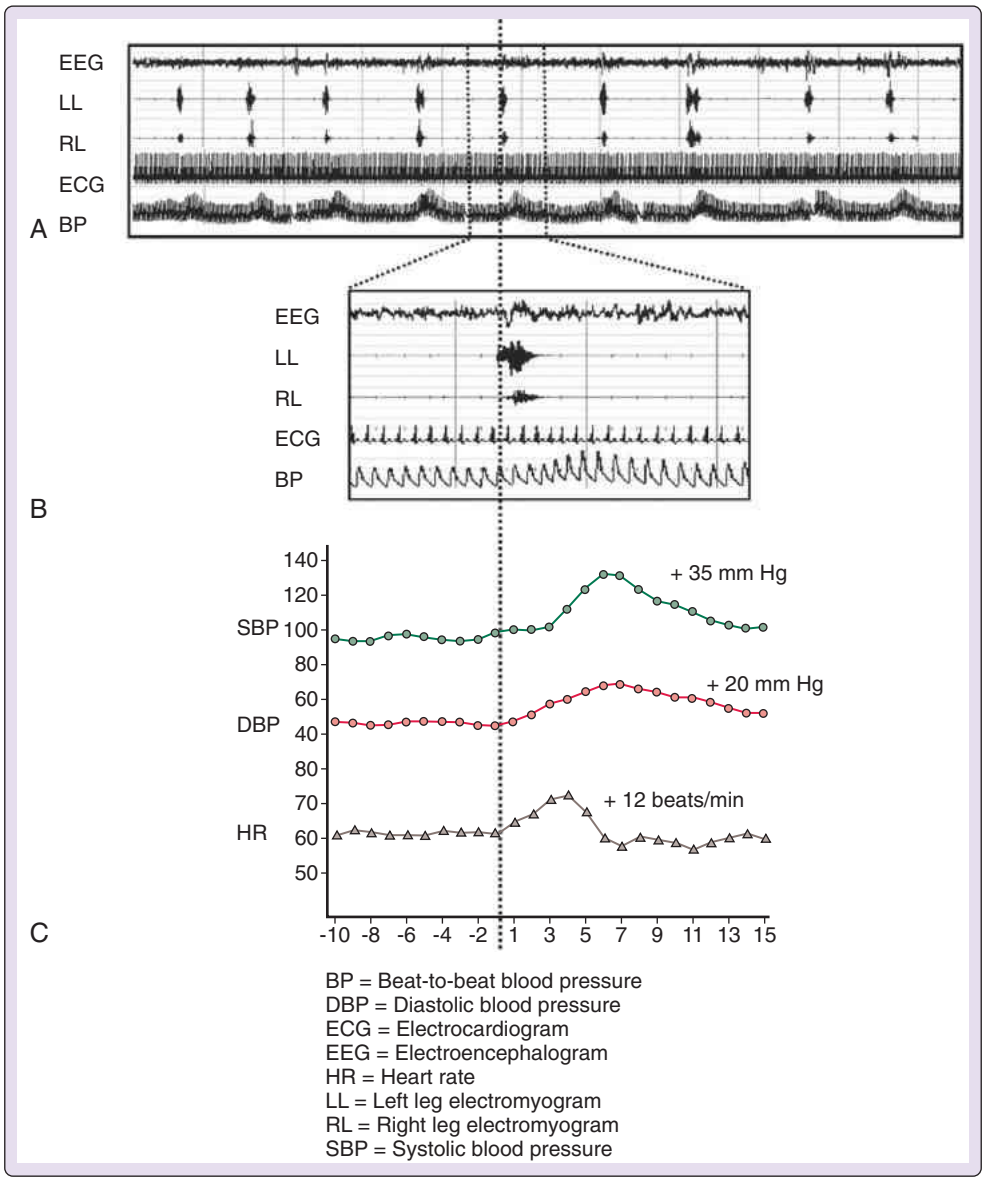


Figure 14-4 ECG, beat-to-beat blood pressure, and polysomnographic recording in a compact window **(A)** and in wider temporal windows (**B** and **C**) in a subject with restless legs syndrome. Significant heart rate and blood pressure increases accompany the periodic leg movements. (From Sforza E, Nicolas A, Lavigne G, et al. EEG and cardiac activation during periodic leg movements in sleep: support for a hierarchy of arousal responses. *Neurology* 1999 10;52:786–91.)

The clinical significance of PLMS has been a subject of debate. Recent findings linked the presence of PLMS to poorer cardiovascular health and outcome. Enhanced central sympathetic outflow or the cardiovascular consequences of repetitive BP surges during sleep could be implicated in this association. Restless legs syndrome is characterized by dysesthesia and leg restlessness occurring predominantly at night during periods of immobility. This syndrome is associated in 80% of the cases with PLMS. A systematic review addressing the association between RLS and hypertension identified 17 mainly cross-sectional studies from 12 countries.[59] Only 10 of the 17 studies supported a positive association between restless legs syndrome and hypertension; this association persisted after adjustment for body mass index, smoking, and sleep problems. These inconsistent findings regarding the association between restless legs syndrome, PLMS, and cardiovascular health may be explained by variations in studied populations, presence of confounding factor, and differences in ascertainment of hypertension and restless legs syndrome. Collectively, these studies indicate that restless legs syndrome might be positively related to hypertension when syndrome-related symptom frequency is high, exceeding 15 days per month, and PLMS index is in the severe range.[60]

IMPACT OF AGING ON NEURAL CIRCULATORY RESPONSE TO NORMAL SLEEP

Aging leads to profound morphologic and functional alterations in the cardiovascular system and its autonomic control.[61] Among these changes, basal central sympathetic drive appears

enhanced (increase in resting plasma catecholamines, MSNA, and LF components of RR variability) but the HR responsiveness to sympathetic stimuli is attenuated, at least in part because of a loss of cardiac receptor sensitivity to catecholamines. The increased central sympathetic drive in older persons is reflected during sleep by a reduction of RR variability and relatively lower parasympathetic influences, which appear to be linked to the loss of slow wave sleep.[62]

The cardiac response to EEG arousals and PLMS also is modified by age. Specifically, the HR increments are attenuated and bradycardia is less profound in older than in younger subjects.[63,64] The attenuated tachycardia can be part of the general age-related attenuation in the cardiac response to sympathetic stimuli, whereas impairment in baroreflex mechanisms, encountered in older persons, could be a factor implicated in the blunted bradycardia.

EFFECTS OF DISORDERED SLEEP AND PRIMARY AUTONOMIC DYSFUNCTION ON DAY-NIGHT AUTONOMIC CHANGES

Effects of Sleep Loss and Sleep Disorders on Nighttime Blood Pressure

As mentioned previously, HR and BP physiologically decrease during nighttime as compared with daytime, a reduction commonly referred to as "dipping." The persistence of high nighttime systolic BP and lack of systolic BP dipping are clinically important and have been linked to precursors of atherosclerosis, including inflammation and endothelial dysfunction.[65] Lack of systolic dipping[66] and, more recently, also lack of HR dipping[67] have been associated with increased cardiovascular mortality, after correction for several confounding variables, including daytime values. Sleep loss and sleep disturbances have been invoked as some of the potential factors underlying these abnormalities.[44] Controlled studies show that during partial sleep deprivation/restriction (allowing 4 hours of sleep), nighttime BP and catecholamine levels remain high, while nighttime nocturnal wakefulness is maintained, and then decrease normally in association with subsequent sleep.[68,69] In the same studies, the morning surge in BP and catecholamines appear to be more pronounced after sleep deprivation than in control conditions, particularly in hypertensive subjects.[68,69] A study in male workers showed that relative to a normal working day allowing 8 hours of sleep, working overtime and sleeping 4 hours induced higher daytime BP on the following day, accompanied by higher LF components of heart rate variability and increased urinary excretion of norepinephrine.[70] Hence it appears that sleep loss (1) is associated with persistence of high sympathetic activity and attenuates physiologic nocturnal BP dipping, as long as nocturnal wakefulness is maintained; (2) may enhance sympathetic activation during morning awakening; and (3) induces sustained sympathetic activation, with increased BP during the following day.

In different cohorts of normotensive and hypertensive subjects without sleep disorders, absence of BP dipping was associated with indices of poor and fragmented sleep, including longer wake-time after sleep onset and higher arousal frequency.[71,72] Increased nighttime BP has been reported in subjects with moderate to severe OSA,[73] the degree of BP alteration being proportional to the severity of sleep apnea.

Insomnia

Insomnia is characterized by subjective dissatisfaction with the quality of sleep and daytime consequences that may or may not be explained by a true reduction in sleep duration. Recent studies have used polysomnography to show that insomnia with objective short sleep duration is associated with a significant risk of hypertension. First, a study using 24-hour beat-to-beat blood pressure recordings concurrently with polysomnography reported that normotensive subjects with chronic insomnia had higher nighttime systolic BP and blunted day-to-night systolic BP dipping compared with aged-matched good sleepers.[74] Vgontzas and colleagues[75] have demonstrated an association of insomnia with prevalent hypertension only in the presence of objectively measured short sleep duration. The prevalence of hypertension increased 3.5-fold when sleep duration was between 5 and 6 hours and 5.1-fold when sleep duration was less than 5 hours per night. Accordingly, chronic insomnia with short sleep duration (less than 6 hours of sleep during polysomnography) was associated with an increased risk for incident hypertension (odds ratio, 3.8) in a general population sample of 786 adults within the Penn State Cohort without hypertension at baseline, assessed over a mean follow-up period of 7.5 years.[76]

Narcolepsy-Cataplexy

In narcolepsy-cataplexy (NC), the sleep-wake cycle is disrupted by the frequent occurrence of REM sleep onset episodes during daytime and by numerous awakenings during nocturnal sleep. The disease is characterized by a marked decrease in the number of hypocretin neurons, which are known to play a role in central autonomic and cardiovascular regulation.[77,78] Only a few cardiovascular studies have investigated human narcolepsy, even though NC is classically associated with obesity, type 2 diabetes, and metabolic syndrome—comorbid conditions leading to an increased cardiovascular risk. Recently, a study compared the 24-hour ambulatory BP monitoring pattern for drug-free patients with narcolepsy against that for control subjects.[79] A "nondipping status" was found in one third of patients with NC, versus in only 4.8% of control subjects. Nondipping of diastolic BP was strongly associated with NC by an odds ratio up to 12-fold, with a significant association with the percentage of REM sleep even after adjustment for confounding factors. Grimaldi and coworkers[80] have nicely demonstrated that systolic BP during nighttime REM sleep was increased in narcolepsy. NC is therefore a unique example of increased nocturnal BP mainly during REM sleep. However, in view of the fact that patients with NC will be treated with psychostimulants for the rest of their lives, a therapy that has direct impact on both the autonomic and cardiovascular systems, preliminary studies demonstrating the effects of NC on the dipping pattern of BP might have important clinical implications. Thus further studies addressing longitudinal associations between NC and hypertension as well as further mechanistic studies are clearly warranted.

Loss of Diurnal Variation in Autonomic Function in Diabetes Mellitus: What Comes First?

Cardiovascular autonomic neuropathy is a serious complication of diabetes mellitus and results from damage to autonomic fibers involved in HR and BP control, in the presence

of impaired glucose metabolism.[81] In patients with insulin-independent diabetes (type 2 diabetes), the 24-hour periodicity of HR and RR variability is lost, with attenuated sympathetic control in the daytime and blunted parasympathetic function during the night.[82] In patients with different degrees of glucose abnormalities without overt diabetes, RR variability and its spectral components appear similar to those in control subjects in the daytime but are significantly altered during sleep, with strikingly higher normalized LF and lower HF, proportional to the degree of insulin resistance.[83] Insulin resistance, characterized by a reduced biologic effect of insulin, and sympathetic overactivity are known to be linked and possibly potentiate each other, with insulin increasing sympathetic activity and neuroadrenergic mechanisms acting to increase plasma glucose availability and to reduce peripheral insulin sensitivity. These findings suggest that a primary alteration in the autonomic nervous system may occur during sleep in these subjects before overt diabetes is evident, and may be linked to the level of insulin resistance. However, one study also observed that selectively altered nighttime autonomic function also was present in nondiabetic offspring of patients with type 2 diabetes, whether or not they had insulin resistance,[84] suggesting that nighttime impaired parasympathetic mechanisms, possibly of genetic origin, may precede metabolic abnormalities.

Type 2 diabetes is a complex disease that derives from the interaction of environmental factors on a background of a genetic susceptibility. Chronic sleep debt, either due to sleep restriction or sleep apnea, has been shown to be one factor that can alter glucose handling[85] and increases the likelihood of developing type 2 diabetes.[86] Little is known about the relationship and interactions between these sleep disturbances and early autonomic dysfunction in subjects with differing severities of glucose abnormalities and their healthy offspring. Patients with type 1 diabetes who demonstrated a nondipping pattern of their nighttime blood pressure had shorter sleep duration than those who exhibited a physiologic nocturnal dip of blood pressure.[87,88] More information on insulin and sleep is found in Chapter 20.

SYMPATHETIC ACTIVATION IN OBSTRUCTIVE SLEEP APNEA

The sympathetic nervous system appears to play a key role in the cardiac pathophysiology of sleep apnea. Even when patients with OSA are awake and breathing normally, and in the absence of any overt cardiovascular disease such as hypertension or heart failure, they exhibit evidence for impaired sympathetic cardiovascular regulation. Specifically, they have high levels of muscle sympathetic nerve activity, increased catecholamines, faster heart rates, and attenuated heart rate variability.[89] Furthermore, even though they are normotensive, they show excessive blood pressure variability.[90] In the setting of apnea, the inhibitory effect of the thoracic afferents is absent, resulting in further potentiation of sympathetic activation. The consequent vasoconstriction results in marked surges in blood pressure, as noted earlier. Sympathetic activity abruptly ceases at onset of breathing owing to the inhibitory effect of the thoracic afferents[91] (Figure 14-5). In a minority of patients with OSA, the diving reflex, described earlier, is activated. These patients may therefore develop marked bradyarrhythmias in association with the obstructive apnea, even though they do not have any intrinsic conduction system abnormality.[4] The bradycardia is secondary to cardiac vagal

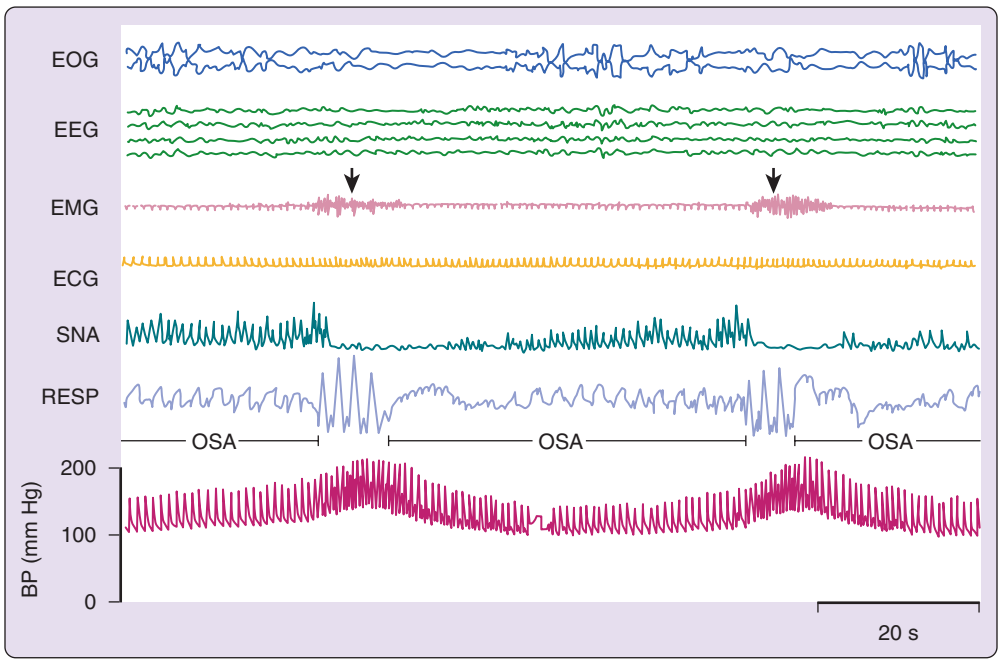


Figure 14-5 Sympathetic nerve activity (SNA) and blood pressure (BP) recordings in association with obstructive sleep apnea (OSA). SNA increases progressively during the apneic episode because of the activation of the peripheral and central chemoreflexes by hypoxemia and hypercapnia. The consequent vasoconstriction results in marked surges in BP, which reaches a peak during the hyperventilation. SNA abruptly ceases at onset of breathing owing to the inhibitory effect of the thoracic afferents. RESP, Respiration. (From Somers VK, Dyken ME, Clary MP, Abboud FM. Sympathetic neural mechanisms in obstructive sleep apnea. *J Clin Invest* 1995;96:1897–904.)

activation owing to the combination of hypoxia and apnea. These acute responses to obstructive apnea may predispose affected persons to longer-term abnormalities in cardiac and vascular structure and function. Several mechanisms have been proposed that could link OSA to cardiovascular diseases through the recurrent nocturnal cycles of hypoxia/reoxygenation. Such changes promote oxidative stress and low-grade inflammation, which are the initiators of a pathophysiologic cascade leading first to sympathetic overactivity. The high vascular sympathetic tone exhibited by patients with OSA results in elevated systemic resistance and hence elevated blood pressure. Impaired arterial vasodilatory capacity may contribute to elevation of blood pressure and lead to vascular disease. Animal models of chronic intermittent hypoxia (CIH) alone or in association with the other stimuli that characterize OSA (i.e., respiratory effort, asphyxia, and arousal from sleep) show elevated blood pressure during the non-CIH portion of the day. These findings suggest that the blood pressure elevation results initially from sympathetic activation. This mechanism requires an intact chemoreflex loop; it also has been demonstrated that in OSA, arterial baroreflex gain is decreased. Although animal models have advanced the current understanding, specific aspects of human physiology may not be adequately represented in such experiments. Accordingly, models of intermittent hypoxia in healthy humans have been developed that induce unstable ventilation and sleep fragmentation similar to those observed in patients with OSA. Healthy humans exposed to 1 or 2 weeks of CIH exhibit an increase in both hypoxic and hypercapnic ventilatory responses, confirming that augmentation of carotid chemoreflex function participates in inducing sustained sympathetic overactivity. After 2 weeks of CIH exposure, MSNA is increased and baroreflex control of sympathetic outflow declines. Consequently, CIH significantly increased daytime ambulatory blood pressure after 2 weeks of exposure (8 mm Hg systolic and 5 mm Hg diastolic)[92] (Figure 14-6).

CLINICAL PEARL

The autonomic nervous system is the mediator of central-cardiovascular interactions occurring during sleep, and its normal function appears to be important in preserving health. Despite the recognized methodologic limitations (the procedure involved is technically demanding, and only cautious interpretation of outcomes of interest, as described in Table 14-1, is possible), broadening sleep polygraphic monitoring to include heart rate and blood pressure recordings may contribute to a better understanding of the physiology and pathology of sleep-related cardiovascular autonomic modulation. This strategy may provide an avenue for innovation in the management of many medical conditions or disorders that are sleep-related (e.g., hypertension, diabetes, metabolic syndrome, periodic limb movements, sleep-disordered breathing).

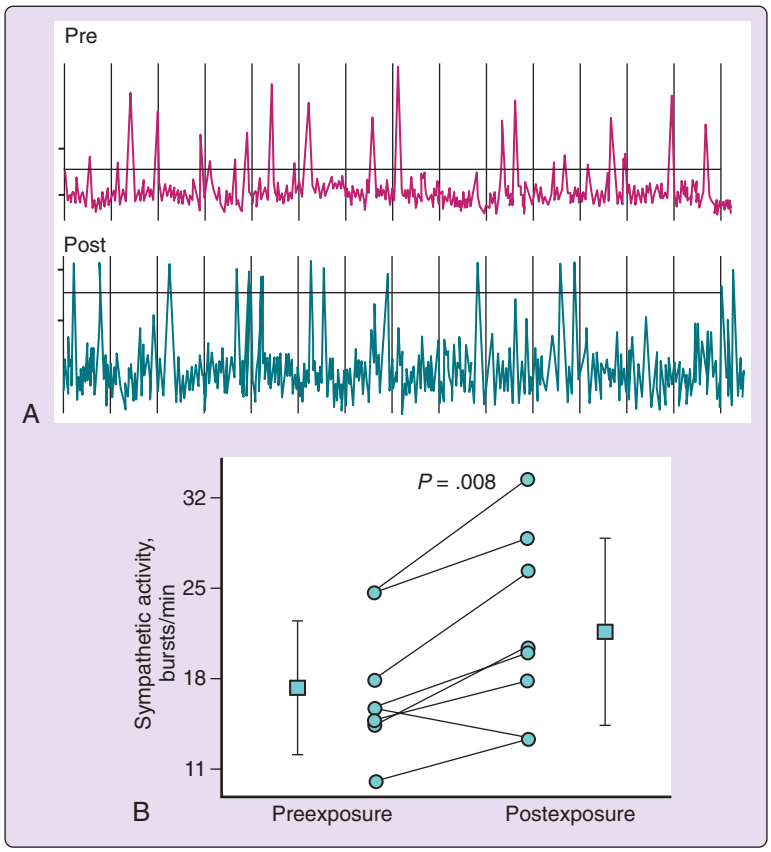


Figure 14-6 Intermittent hypoxia elevates daytime blood pressure and sympathetic activity in healthy humans. **A,** Representative neurograms of muscle sympathetic nerve activity (MSNA) during supine rest while breathing room air before (pre) and after 2 weeks of intermittent hypoxia exposure (post). **B,** MSNA increased across the exposure (17.2 ± 5.1 versus 21.7 ± 7.3 bursts/min; $P < .01$), thus reflecting sympathoactivation.

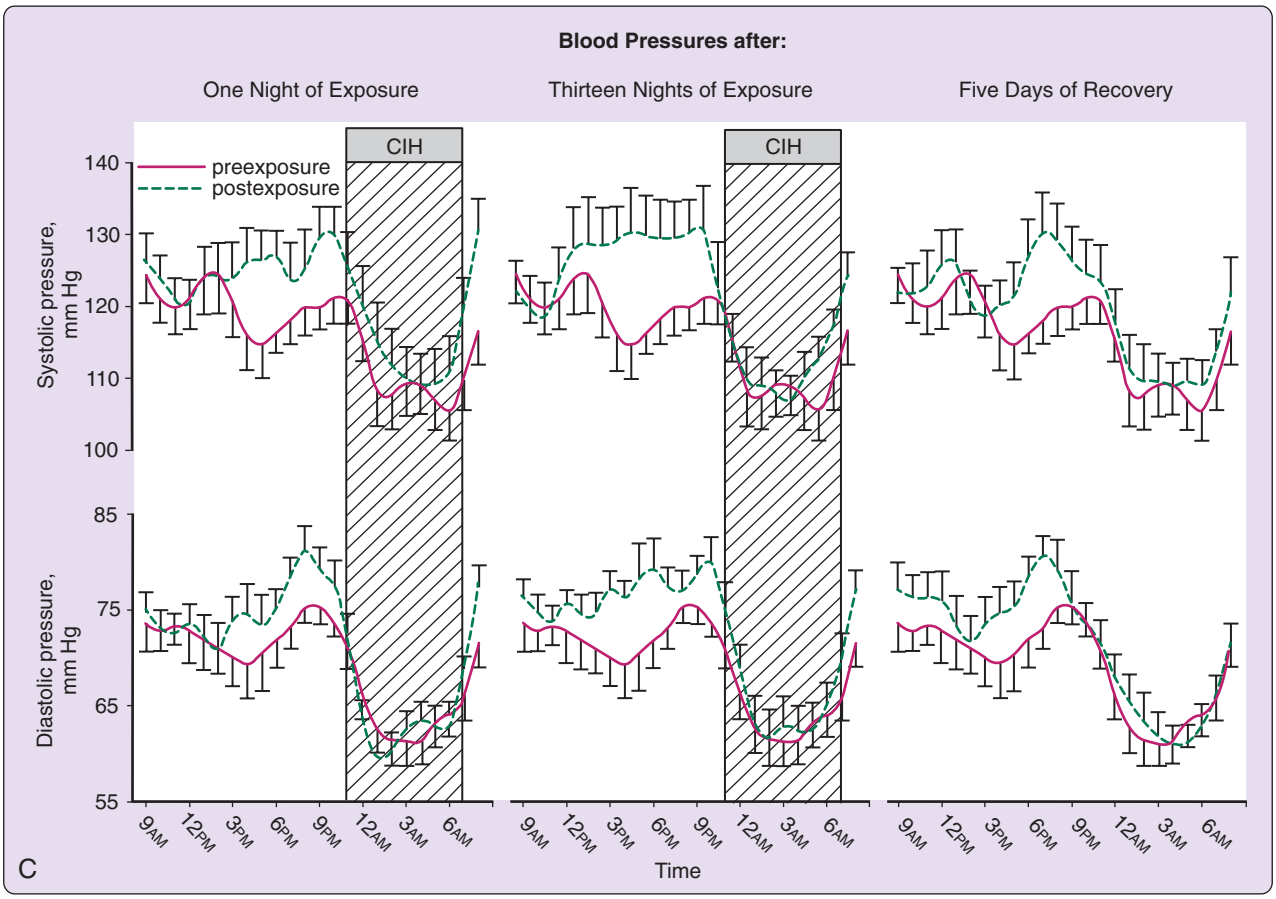


Figure 14-6, cont'd C, Hour-by-hour systolic and diastolic blood pressures during 24 hours of monitoring in healthy humans. Data are presented after 1 night, 13 nights, and recovery from exposure to chronic intermittent hypoxia (CIH) as compared with preexposure values. CIH was associated with significantly increased daytime ambulatory blood pressure after a single night of exposure (3 mm Hg for mean and diastolic pressures), with further increased daytime pressures after 2 weeks of exposure (8 mm Hg systolic and 5 mm Hg diastolic). (From Tamisier R, Pepin JL, Remy J, et al. 14 nights of intermittent hypoxia elevate daytime blood pressure and sympathetic activity in healthy humans. *Eur Respir J* 2011;37:119–28.)

SUMMARY

The autonomic nervous system is intimately linked to central neural state changes. This coupling is especially straightforward for physiologic sleep and sleep disorders. It is clear that although the different stages of physiologic sleep result in structured changes in neural circulatory control, disturbed sleep, such as is seen in patients with OSA, with PLMS, or in sleep deprivation, acts to disrupt the sleep-related physiologic variations in autonomic regulation of heart rate and blood pressure. Current knowledge in this general area is limited by the tools available for comprehensive and direct assessment of the autonomic nervous system in humans. Although microneurography provides a direct measurement of sympathetic neural activity to the peripheral blood vessels, this measurement itself has limitations. The other options available are primarily those that monitor blood and urine levels of catecholamines. Measurements such as heart rate and blood pressure variability, while allowing some insight, provide only indirect information on autonomic cardiovascular control and are of limited usefulness owing to problems with regard to acquisition of data, confounding effects of medications and abnormal breathing patterns, and inconsistencies with regard to interpretation. Rigorous methods in line with standard recommendations[9] are mandatory. Therefore, beyond the body of knowledge regarding neural circulatory control during normal and disordered sleep surveyed in this chapter, the available data are limited, in part because of methodologic shortcomings and also because of the obvious difficulties inherent in nighttime studies of sleep physiology in humans.

Selected Readings

Daly MD, Angell-James JE, Elsner R. Role of carotid-body chemoreceptors and their reflex interactions in bradycardia and cardiac arrest. *Lancet* 1979;**1**(8119):764–7.

Jurysta F, van de Borne P, Migeotte PF, et al. A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men. *Clin Neurophysiol* 2003;**114**(11):2146–55.

Li Y, Vgontzas AN, Fernandez-Mendoza J, et al. Insomnia with physiological hyperarousal is associated with hypertension. *Hypertension* 2015;**65**: 644–50.

Pepin JL, Borel AL, Tamisier R, et al. Hypertension and sleep: overview of a tight relationship. *Sleep Med Rev* 2014;**18**(6):509–19.

Sforza E, Nicolas A, Lavigne G, et al. EEG and cardiac activation during periodic leg movements in sleep: support for a hierarchy of arousal responses. *Neurology* 1999;**52**(4):786–91.

Somers VK, Dyken ME, Mark AL, Abboud FM. Sympathetic-nerve activity during sleep in normal subjects. *N Engl J Med* 1993;**328**(5):303–7.

Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. *Lancet* 1999;**354**(9188):1435–9.

Tamisier R, Pepin JL, Remy J, et al. 14 nights of intermittent hypoxia elevate daytime blood pressure and sympathetic activity in healthy humans. *Eur Respir J* 2011;**37**(1):119–28.

A complete reference list can be found online at ExpertConsult.com.

Respiratory Physiology: Central Neural Control of Respiratory Neurons and Motoneurons During Sleep

Chapter 15

Richard L. Horner

Chapter Highlights

- The *wakefulness stimulus* to breathing, and its withdrawal in sleep, is an enduring principle in respiratory medicine because it is the root mechanism for modeling the effects of sleep on breathing. The neural basis for this wakefulness stimulus is identified.
- Central to understanding breathing in sleep has been delineation of the neurobiology of sleep, its impact on central respiratory neurons and motoneurons, and the important role of tonic excitatory (nonrespiratory) drives in contributing to the overall level of excitability in the respiratory system across sleep-wake states.
- Significant developments have contributed to identifying the neural basis for the suppression of pharyngeal muscle activity in sleep, especially rapid eye movement (REM) sleep. Mechanisms of genioglossus muscle suppression in sleep include withdrawal of excitatory inputs from wakefulness-dependent cell groups and active inhibition, a major component of the latter being mediated through a newly identified pathway.
- Mechanisms of respiratory rhythm generation and factors influencing motor excitability are both essential for the manifestation of effective breathing across sleep-wake states. The neurodepressive effects of commonly administered drugs such as opioids and sedative-hypnotics acting at critical sites in the respiratory network can explain the sometimes severe respiratory depression that can occur during sleep with use of such agents.

RESPIRATORY NEUROBIOLOGY: BASIC OVERVIEW

Medullary Respiratory Neurons and Motoneurons

Bilateral columns of neurons present in the medulla show activity patterns that vary in phase with some component of the respiratory cycle. The dorsal respiratory group (DRG) is located in the dorsomedial medulla, specifically in the ventrolateral nucleus of the solitary tract, and contains predominantly inspiratory neurons[1,2] (Figure 15-1). The DRG and the other subnuclei of the solitary tract also are the primary projection sites for vagal afferents from the lung, and for afferents from the carotid and aortic chemoreceptors and baroreceptors, that exert important reflex influences on breathing. These projections indicate that the nuclei of the solitary tract, including the DRG, are key sites of integration of sensory information from the lung, as well as information regarding the prevailing levels of arterial P_{CO_2}, P_{O_2}, pH, and systemic blood pressure. The ventral respiratory group (VRG) extends from the facial nucleus to the first cervical segment of the spinal cord and contains both inspiratory and expiratory neurons (Figure 15-1).[1,2] The nucleus ambiguus also consists of a rostral to caudal column of neurons expressing respiratory-related activity, with subregions regions containing motoneurons that innervate the muscles of the larynx and pharynx that are not considered part of the VRG per se.[3] In addition to the nucleus ambiguus, from rostral to caudal, the VRG is composed of Bötzinger complex (expiratory) neurons, pre-Bötzinger complex (inspiratory) neurons, rostral retroambigualis (predominantly inspiratory) neurons, and caudal retroambigualis (predominantly expiratory) neurons (Figure 15-1).[1,2]

The VRG and DRG contain both bulbospinal respiratory pre-motoneurons (i.e., neurons that project to spinal motoneurons, which in turn innervate the respective respiratory pump and abdominal muscles of breathing), and propriobulbar neurons (i.e., neurons that project to, and influence the activity of, other medullary respiratory neurons but themselves do not project to motoneurons per se) (Figure 15-1).[1,2] The hypoglossal, trigeminal, and facial motor nuclei also innervate muscles important to pharyngeal motor control and the maintenance of upper airway patency[4] (Figure 15-1). An important point, however, is that the expression of respiratory-related activity is not restricted to neurons of the DRG, VRG, and cranial motoneurons innervating the pharyngeal and laryngeal muscles. For example, neurons expressing respiratory-related activity in the pons, such as the pontine respiratory group (PRG) in Figure 15-1, are thought to play an important role in shaping the activity of medullary respiratory neurons during breathing.[3]

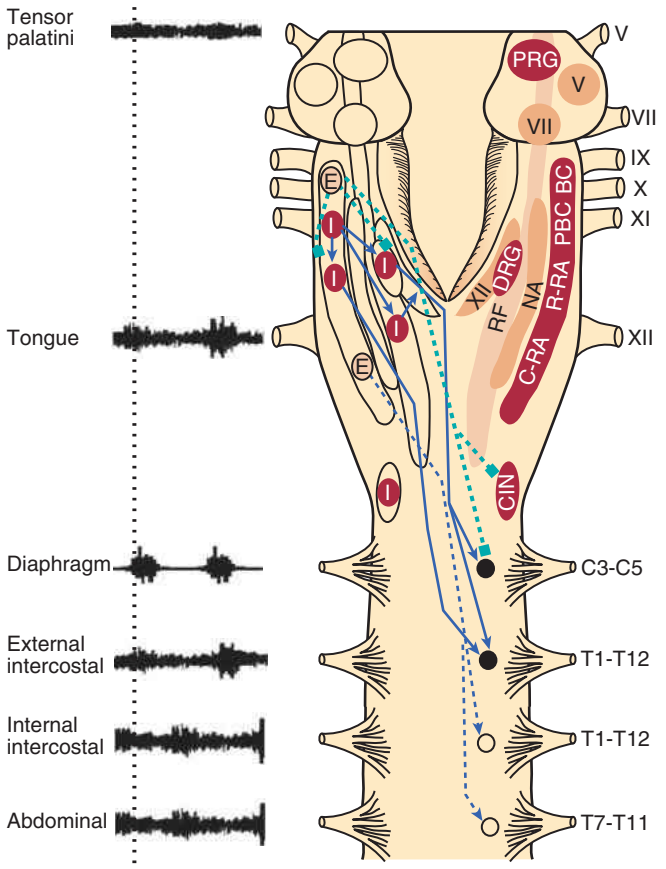


Figure 15-1 Ventral view of the brainstem (with cerebellum removed) showing the main aggregates of respiratory neurons in the dorsal and ventral respiratory groups (DRG and VRG, respectively). The locations of expiratory (E) and inspiratory (I) neurons in the Bötzinger complex (BC), pre-Bötzinger complex (PBC), rostral retroambigualis (R-RA), and caudal retroambigualis (C-RA) are shown. The locations of cervical inspiratory neurons (CIN) and respiratory-related neurons in the lateral reticular formation (RF) projecting to the hypoglossal motor nucleus (XII) also are shown. The projections of inspiratory and expiratory neurons are depicted as *solid* and *dashed lines,* respectively, whereas excitatory and inhibitory synaptic connections are depicted by *arrowhead* and *square symbols,* respectively. The electromyographic activities of various inspiratory-related (e.g., tongue, diaphragm, external intercostal) and expiratory (e.g., internal intercostal, abdominal) muscles are shown. Note that the level of respiratory-related and tonic activities varies for different muscles, with some muscles such as the tensor palatini expressing mainly tonic activity. The onset of muscle activity with respect to the diaphragm is shown by the *dashed line.* The rootlets of cranial nerves V, VII, IX, X, XI, and XII and the cervical (C) and thoracic (T) segments of the spinal cord also are shown, as are the motor nuclei of cranial nerves XII, VII, and V. The locations of the pontine respiratory group (PRG) and the nucleus ambiguus (NA) are shown, although their projections are not included for clarity. See text for further details.

Pre-Bötzinger Complex

Pre-Bötzinger complex neurons have pacemaker-like properties that are thought to be important to the generation of the basic respiratory rhythm, and to the expression of rhythmic neuronal activity elsewhere in the respiratory network[5,6] (Figure 15-1). Respiratory rhythm-generating pre-Bötzinger complex neurons coexpress μ opioid and neurokinin-1 receptors (i.e., the receptors for substance P), which slow and increase respiratory rate, respectively.[6] The development of uncoordinated (ataxic) diaphragm breathing after introduction of lesions of neurokinin-1–expressing pre-Bötzinger complex neurons in animal studies, with this abnormal breathing first appearing in sleep,[7] suggests that pre-Bötzinger complex neurons contribute significantly to normal breathing in vivo. First identified and characterized in rodents, and subsequently in other mammalian species, the pre-Bötzinger complex also has been identified in humans.[8,9] Loss of pre-Bötzinger complex neurons may predispose affected persons to abnormal or ataxic breathing and to central apneas in sleep, such as with aging and in neurodegenerative brainstem diseases.[6,9]

The presence of μ opioid receptors on pre-Bötzinger complex neurons can explain a significant component of the clinically important phenomenon of respiratory rate depression with opioid drugs.[10] The respiratory slowing and central apneas produced by systemically applied opioids are prevented by local application of the μ opioid receptor antagonist naloxone to the pre-Bötzinger complex, showing that this region of medulla is the critical site mediating opioid-induced respiratory rate depression.[10] Moreover, deep non–rapid eye movement (NREM) sleep and general anesthesia are the most vulnerable states for respiratory rate depression produced by opioids at the pre-Bötzinger complex.[10] This observation has significant clinical relevance regarding the potential hazards of administering opioids, for example, in the perioperative setting.

Neuronal Connections

The anatomic connections between the neurons that comprise the essential respiratory network (i.e., respiratory propriobulbar neurons, pre-motoneurons, and motoneurons), and the membrane properties of these cells, are ultimately responsible for the two key components of overall respiratory activity: (1) the generation of respiratory rhythm and (2) the shaping of the central respiratory drive potentials that activate respiratory motoneurons (pattern generation). An analysis of the mechanisms involved in the generation of the basic respiratory rhythm is outside the scope of this chapter; excellent summaries of the concepts underlying pacemaker models (whereby respiratory rhythm is *intrinsic* to some cells, which then drive others in the respiratory network), network models (whereby respiratory rhythm is *dependent* on the inhibitory and excitatory synaptic connections between neurons, and the tonic excitation is derived from both the respiratory chemoreceptors and brainstem reticular neurons), and hybrid models are available in referenced sources.[1,3,6]

With respect to the tonic drive to the respiratory system arising from the respiratory chemoreceptors, this would include both the peripheral and central chemoreceptors, the latter including neurons at the ventral medullary surface such as the retrotrapezoid nucleus, as well as inputs from CO_2-activated sleep state–dependent neurons of the aminergic arousal system (e.g., serotonin and noradrenergic neurons; see the following section).[11] Additional aspects of the organization of the central respiratory network are particularly relevant to understanding the effects of sleep on respiratory neurons and motoneurons, and these concepts are discussed briefly next.

During inspiration the central respiratory drive potential is transmitted to phrenic and intercostal motoneurons via monosynaptic connections from inspiratory pre-motor neurons of the DRG and VRG[1] (Figure 15-1). Bötzinger

complex expiratory neurons have widespread inhibitory connections throughout the brainstem and spinal cord, and these neurons inhibit inspiratory pre-motoneurons and motoneurons during expiration (Figure 15-1). Caudal retroambigualis neurons also increase the excitability of spinal expiratory motoneurons in expiration (Figure 15-1), although this excitation does not necessarily reach the threshold to manifest as expiratory muscle activity.

Of physiologic and clinical relevance, these fundamental aspects of the neural control of spinal respiratory motoneuron activity appear to be different from those for the mechanisms controlling the activity of pharyngeal motoneurons. For example, animal studies show that the source of inspiratory drive to hypoglossal motoneurons is different from the source of drive to phrenic motoneurons, being predominantly from reticular neurons lateral to the hypoglossal motor nucleus (lateral tegmental field) for the former and from bulbospinal VRG and DRG neurons for the latter (Figure 15-1).[1] Of importance, brainstem reticular neurons provide a significant source of tonic drive to the respiratory system, with this drive particularly affected in sleep.[3]

Further differences in the functional control of pharyngeal and diaphragm muscles is shown by the observation that unlike phrenic motoneurons, hypoglossal motoneurons are not actively inhibited in expiration.[1] Accordingly, the activity of the genioglossus muscle in expiration is simply a manifestation of the prevailing tonic inputs. The practical implication of this circuitry is that the overall activation of hypoglossal motoneurons during breathing is composed of an inspiratory drive that adds to a continuous tonic drive that persists in expiration when the inspiratory activation is withdrawn. Moreover, this tonic drive to the pharyngeal muscles, which contributes to baseline airway size and stiffness, is most prominent in wakefulness but withdrawn in sleep, resulting in an upper airspace that is more vulnerable to collapse. Characterization and quantification of this tonic *wakefulness stimulus* have been performed for the pharyngeal muscles in humans.[12] A more detailed analysis of the neural mechanisms controlling the activity of respiratory neurons and motoneurons follows a brief overview of the brain mechanisms modulating the states of wakefulness, NREM sleep, and REM sleep.

SLEEP NEUROBIOLOGY: BASIC OVERVIEW

Although a more detailed discussion of arousal and sleep state regulation is provided in Section 2 of this book, some details are particularly pertinent to the control of respiratory neurons and motoneurons in sleep. Accordingly, a brief overview of the neurobiology of sleep and wakefulness-generating systems is presented next.

Wakefulness

Figure 15-2 shows some of the main neuronal groups contributing to the ascending arousal system from the brainstem that promotes wakefulness. This ascending arousal system includes the cholinergic laterodorsal and pedunculopontine tegmental nuclei that promote cortical activation by way of excitatory thalamocortical projections.[13] The ascending arousal system also incorporates the aminergic arousal system that originates from brainstem neuronal groups principally containing serotonin (dorsal raphe nuclei), norepinephrine (locus coeruleus), histamine (tuberomammillary nucleus), and dopamine (ventral periaqueductal gray). Orexin neurons from the perifornical region of the hypothalamus and cholinergic neurons from the basal forebrain also contribute to this ascending arousal system.[13] Overall, multiple neuronal systems contribute to cortical arousal and wakefulness. These neuronal systems also are positioned to influence respiratory neurons and motoneurons by way of their anatomic projections to the pons, medulla, and spinal cord (Figure 15-2).

NREM Sleep

Sleep is actively generated by neurons in the ventrolateral preoptic area, anterior hypothalamus, and basal forebrain (Figure 15-2).[13] These neurons become active in NREM sleep, an effect influenced by the thermal stimulus that accompanies the circadian rhythm–mediated decline in body temperature at normal bedtime.[14] This circadian-mediated decline in body temperature is mediated by a change in the set-point of hypothalamic temperature-regulating neurons, which initially leads to a relative "warm stimulus" because body temperature is at first higher than the new set-point—that is, before heat loss occurs. This warm stimulus activates NREM sleep-active hypothalamic neurons and so promotes sleep onset. This effect of internal body temperature on sleep is distinct from the influences of ambient environmental temperature on sleep regulation. Activation of ventrolateral preoptic neurons leads to a direct suppression of cortical arousal, this by way of ascending inhibitory cortical projections. Ventrolateral preoptic neurons also promote sleep through descending inhibition of the aforementioned brainstem arousal neurons through release of gamma-aminobutyric acid (GABA) and galanin.[14,15] This effect of GABA explains the sedative-hypnotic effects of barbiturates, benzodiazepines, imidazopyridine compounds, and alcohol, as well as some general anesthetics, all of which enhance GABA-mediated neuronal inhibition through interactions with binding sites on the $GABA_A$ receptor.[16] $GABA_A$ receptors also are strongly implicated in respiratory control and are present throughout the respiratory network,[17] excessive stimulation of which can promote respiratory depression.[18] In summary, sleep onset is triggered by increased GABAergic neuronal activity, and this is accompanied by a massed and coordinated withdrawal of activity of brainstem arousal neurons comprising serotonergic, noradrenergic, histaminergic, and cholinergic neurons. With the widespread projections of these sleep state–dependent neuronal groups, these changes in neuronal activity in sleep also are positioned to influence respiratory neurons and motoneurons (Figure 15-2).[19]

REM Sleep

Decreased serotonergic and noradrenergic activity preceding and during REM sleep withdraws inhibition of the laterodorsal and pedunculopontine tegmental nuclei.[13,15] This effect leads to increased acetylcholine release into the pontine reticular formation and facilitation of transitions into REM sleep.[20,21] Exogenous application of cholinergic agonists or acetylcholinesterase inhibitors (to increase endogenous acetylcholine) into the same region of the pons is used to mimic this process experimentally in animal studies, that is, the "carbachol model" of REM sleep.[20,21] A significant component of the motor suppression of REM sleep is mediated by descending pathways involving activation of medullary

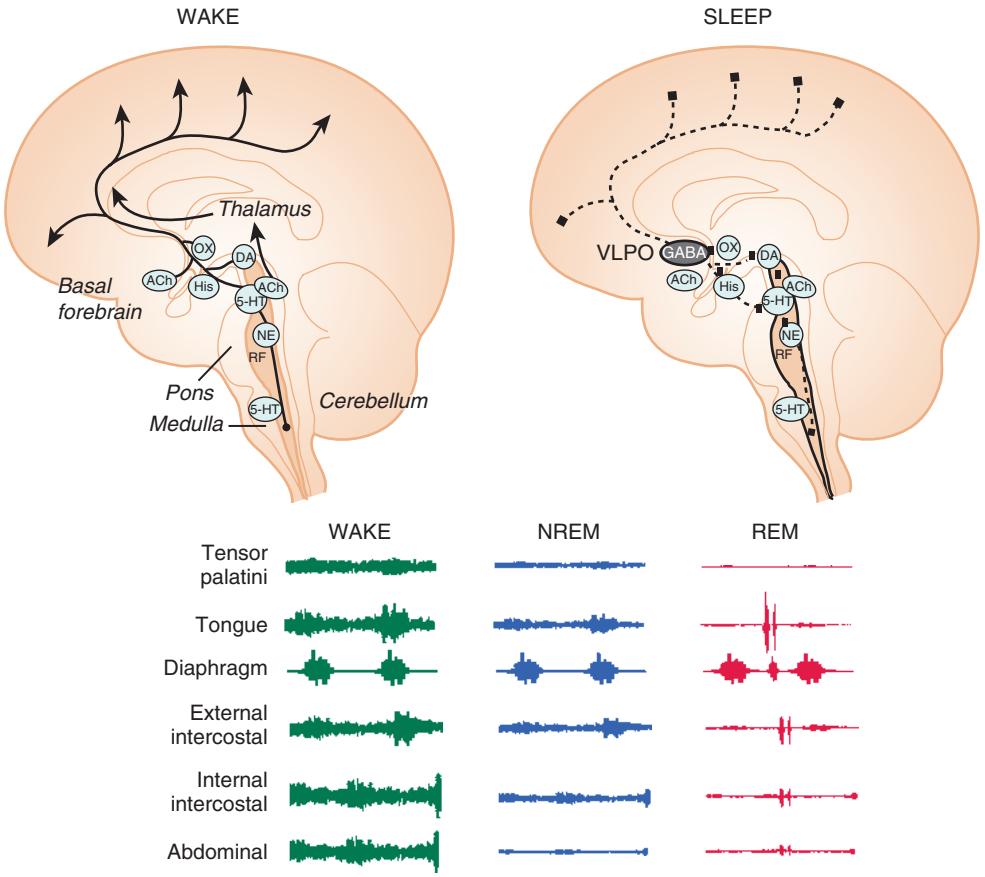


Figure 15-2 Sagittal section of the brain showing the main wake- and sleep-generating neural systems. In wakefulness, acetylcholine (ACh), orexin (OX), histamine (His), dopamine (DA), 5-hydroxytryptamine (5-HT), and norepinephrine (NE) containing neurons contribute to brain arousal (depicted as *black lines* with *arrows*). This ascending arousal system is inhibited in sleep by GABA-containing neurons from ventrolateral preoptic (VLPO) neurons (inhibitory projections shown by *dashed lines* and ■ symbols). By their anatomic projections to the pons, medulla and spinal cord, these wake and sleep-promoting neuronal systems also are positioned to influence respiratory neurons and motoneurons (see Figure 15-1). However, whether the influences of the different arousal-related neurons is excitatory or inhibitory will depend on the receptor subtypes activated (this uncertainty is depicted by the symbol ● in the medulla). Overall, these changes in neuronal activities across sleep-wake states, and their impact on respiratory neurons and motoneurons, mediate the stereotypical changes in the tonic and respiratory components of activity for different respiratory muscles, and their different susceptibilities to motor suppression in sleep. See text and referenced sources[19,20] for further details. GABA, Gamma-aminobutyric acid; RF,reticular formation. (Modified from Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* 437:1257–63, 2005.)

reticular relay neurons[22] that are inhibitory to spinal motoneurons via release of glycine.[23] Despite the strong association and interactions between pontine aminergic and cholinergic neurons in facilitating REM sleep, recent evidence has implicated a glutamatergic-GABAergic mechanism as key to the generation of the REM sleep state per se.[24,25] In addition to the critical contribution of different neural circuits and neuronal interactions to the generation of REM sleep, another key difference between the aminergic-cholinergic and the glutamatergic-GABAergic mechanisms of REM sleep generation is that the motor atonia is produced by different pathways; that is, the latter framework does not require a relay in the medullary reticular region.[25] Rather, in the glutamatergic-GABAergic mechanism of REM sleep induction, the REM sleep-active pontine neurons are thought to lead to suppression of spinal motoneuron activity by way of long glutamatergic projections to the ventral horn of the spinal cord, which then activate local glycinergic interneurons to inhibit motor activity.[25] Such a mechanism is likely to be involved in the strong inhibition of spinal intercostal motoneurons in REM sleep, but whether collaterals from these specific long descending glutamatergic projections also synapse onto glycinergic inhibitory interneurons in the hypoglossal motor pool is not established (see referenced source[19] for further discussion). Recent findings identifying the mechanism of upper airway motor inhibition in REM sleep are discussed later on (see Inhibitory Influences across Sleep-Wake States, under Neuromodulation of Respiratory Motoneurons across Sleep-Wake States). Recent reviews are available for additional details and further discussion.[26,27] In summary, a number of neural systems show changes in activity across sleep-wake states and project to respiratory neurons and motoneurons. Inasmuch as motoneurons are the final common output pathway for the influence of the central nervous system on motor activity, this

chapter initially focuses on the control of respiratory motoneurons across sleep-wake states before addressing the control of the central respiratory neurons that ultimately drive breathing by those motoneurons.

CONTROL OF RESPIRATORY MOTONEURONS

A characteristic and defining feature of mammalian motor activity is that postural muscle tone is highest in wakefulness, decreased in NREM sleep, and minimal in REM sleep, with the hypotonia of REM sleep punctuated by occasional muscle twitches that are associated with vigorous eye movements and "phasic" REM sleep events.[23] Whether *respiratory* muscle activity is affected in the same way as postural muscle activity across sleep-wake states is somewhat complicated by the interaction of the primary influence of sleep state (e.g., producing suppression of muscle tone) and any subsequent respiratory response (e.g., to compensate for any hypoventilation). On balance, however, the overall stereotyped pattern of suppression of postural muscle activity across sleep-wake states also typically occurs in respiratory muscles, with the degree of sleep state–dependent modulation most readily apparent in those muscles that combine respiratory *and* nonrespiratory (e.g., postural and/or behavioral) functions such as the intercostal and pharyngeal muscles.[28] In these respiratory muscles, decreases in activity typically occur immediately at sleep onset,[28] indicating a primary suppressant effect of sleep neural mechanisms on the activity of respiratory motoneurons—that is, before any compensatory increase in activity takes place in response to altered blood gases, mechanical loads, or sleep-disordered breathing. In contrast with those respiratory muscles with both respiratory and nonrespiratory functions, the diaphragm has an almost solely respiratory function and undergoes lesser suppression of activity in NREM sleep and is largely spared the motor inhibition of REM sleep (Figure 15-2).[29] Other chapters in this section provide more detail regarding clinical aspects of the control of breathing and upper airway function during sleep, whereas this chapter describes the fundamental mechanisms underlying these effects of sleep on the respiratory system.

DETERMINANTS OF RESPIRATORY MOTONEURON ACTIVITY

Tonic and Respiratory-Related Inputs to Respiratory Motoneurons

The changes in muscle tone across sleep-wake states ultimately result from the impact of sleep neural mechanisms on the electrical properties and membrane potential of individual motoneurons located in the respective motor pools in the central nervous system. In turn, the excitability of individual motoneurons changes across sleep-wake states because of varying degrees of excitatory and inhibitory inputs to those motoneurons from sleep-wake–related regions in the brain, and from neurons activated during specific behaviors such as purposeful motor acts in wakefulness.[19] At each individual motoneuron, therefore, the relative strengths of and balance between time-varying excitatory and inhibitory inputs ultimately determine net motor output, with neural activity being generated when the membrane potential rises above threshold for the production of action potentials (Figure 15-3). In addition to the excitatory and inhibitory *nonrespiratory* inputs to a motoneuron that alter membrane potential across sleep-wake states, a respiratory motoneuron also receives additional inputs (excitatory and inhibitory) that alter membrane excitability and neural activity in phase with the inspiratory or expiratory phases of the respiratory cycle. In short, a respiratory motoneuron resembles a postural motoneuron in its control and organizational principles except that it receives an additional rhythmic drive related to respiration—that is, the central respiratory drive potential. Figure 15-3 highlights the fact that the electromyographic activity recorded in a given respiratory muscle is critically dependent on the overall sum of the respiratory *and* nonrespiratory (i.e., tonic) inputs to the motoneurons innervating that muscle.

Recognition of the importance of *both* the tonic and respiratory-related inputs to a motoneuron in determining overall motor output is critical to any interpretation of the changes in respiratory muscle activity observed across sleep-wake states. Indeed, periods of hypoventilation, apparent central apnea, and even the sporadic respiratory muscle activations that occur during REM sleep all can result from *independent* effects of sleep neural processes on the tonic and/or respiratory-related inputs to a respiratory motoneuron (Figure 15-3, *A* to *E*). For example, the apparent absence of activity recorded in a respiratory muscle cannot be taken as evidence that the controlling circuitry is inactive; that is, an apparent apnea may not be truly due to a "central" cessation of respiratory drive. Indeed, a simple withdrawal of tonic drive in sleep may be sufficient to take a population of (e.g., otherwise respiratory-related) motoneurons close to, or below, the threshold for the generation of motor activity, such that any excitatory respiratory inputs to the motoneurons are subthreshold for the generation of action potentials and therefore are not revealed as respiratory muscle activity (Figure 15-3, *C*).

In summary, nonrespiratory tonic drives exert important influences on the resting membrane potential of respiratory motoneurons, thereby significantly modulating the excitability of motoneurons in response to the incoming central respiratory drive potential. This significant effect of nonrespiratory tonic drives on the activity of respiratory motoneurons has clear physiologic relevance: When identified experimentally, the tonic drive to respiratory motoneurons typically is reduced from wakefulness to NREM sleep, with consequent important contributions to sleep-related reductions in respiratory muscle activity leading to hypoventilation (Figure 15-3, *A* and *B*).[30,31] Tonic drive to respiratory motoneurons can also be further reduced in REM sleep, although time-varying fluctuations in this tonic drive can produce transient increases or decreases in respiratory muscle activity and contribute to changes in lung ventilation in REM sleep by a mechanism independent of effects on the respiratory-related inputs (Figure 15-3, *E*). Indeed, the presence of endogenous excitatory inputs to respiratory motoneurons in REM sleep (i.e., unrelated to breathing per se and akin to the mechanisms producing phasic muscle twitches in limb muscles), can produce sporadic activation of respiratory muscle and contribute to the expression of rapid and irregular breathing in REM sleep (Figure 15-3, *E*), even in the presence of low CO_2 levels that are otherwise sufficient to produce central apnea in NREM sleep.[30,31]

Electrical Properties of Motoneurons

The electrical properties of the motoneuron membrane also significantly affect the responses of that motoneuron to a

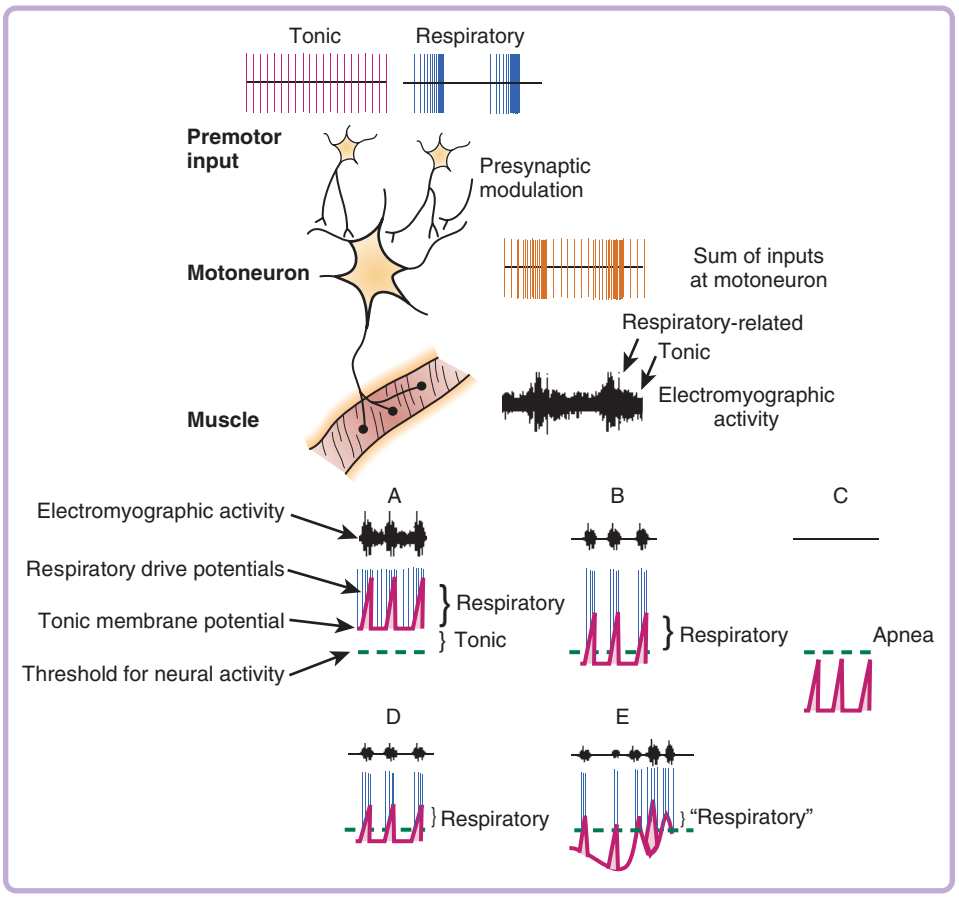


Figure 15-3 Schema depicting how converging tonic (e.g., postural, nonrespiratory) and respiratory inputs to a motoneuron summate to produce the tonic and respiratory components of electromyographic activity. These premotor tonic and respiratory inputs can be excitatory or inhibitory, but here they are shown as excitatory for simplicity. Diagrams **A** to **E** further show how changes in the tonic and respiratory components of respiratory muscle activity can result from *independent* changes in either tonic drives affecting tonic membrane potential (**A, B, C,** and **E**) (such as may occur on transition from wakefulness to NREM and REM sleep) or the magnitude of the respiratory drive potential (**B** versus **D**) (such as may occur in NREM and REM sleep compared with wakefulness). Changes in respiratory drive potential at the motoneuron can result from decreases in the input from respiratory neurons, presynaptic modulation of that input and/or changes in input resistance of the motoneuron membrane per se (see text for further details). In the examples shown in **A** to **E,** respiratory drive is indicated as three depolarizing potentials, each associated with the generation of motoneuron action potentials when the membrane potential exceeds threshold (*dashed line*). Diagram E also shows that time-varying alterations in membrane potential, as occur in REM sleep, for example, can produce respiratory muscle activation unrelated to the prevailing respiratory input (the latter depicted in *gray*). Thus, from peripheral measurements of diaphragm activity or airflow, there appear to be five "breaths," although only three respiratory drive potentials are generated by the central respiratory oscillator.

given synaptic input. For example, reduced motoneuron responses to an incoming respiratory drive potential can be due to the aforementioned effects of reduced tonic drives and consequent membrane hyperpolarization (Figure 15-3), but also to the electrical resistance of the motoneuron membrane itself. The input resistance of a membrane is defined as its voltage response to a given synaptic current, with a decrease in input resistance resulting in less membrane depolarization for a given synaptic drive—that is, a decrease in cell excitability (Figure 15-3). This electrical property of excitable membranes has clear physiologic relevance because a large (~44%) decrease in input resistance of motoneurons occurs in REM sleep compared with NREM sleep and wakefulness.[23] In addition, transient fluctuations in input resistance occur throughout REM sleep episodes, such as the decreased input resistance of somatic motoneurons that occurs in temporal association with the phasic events of REM sleep. Such an effect is likely to contribute to the periods of marked suppression of inspiratory upper airway muscle activity in humans during phasic REM sleep compared with tonic REM sleep.[32]

In summary, a decrease in motoneuron input resistance in REM sleep, especially in association with eye movements, can contribute to decreased motor outflow to the pharyngeal and respiratory pump muscles, leading to periods of increased upper airway resistance and hypoventilation. Moreover, such decreases in respiratory motoneuron activity can occur *despite* the persistence of a continuing, and even heightened, activity of the central respiratory neurons that innervate those motoneurons in REM sleep (Figure 15-4) (see later under Control

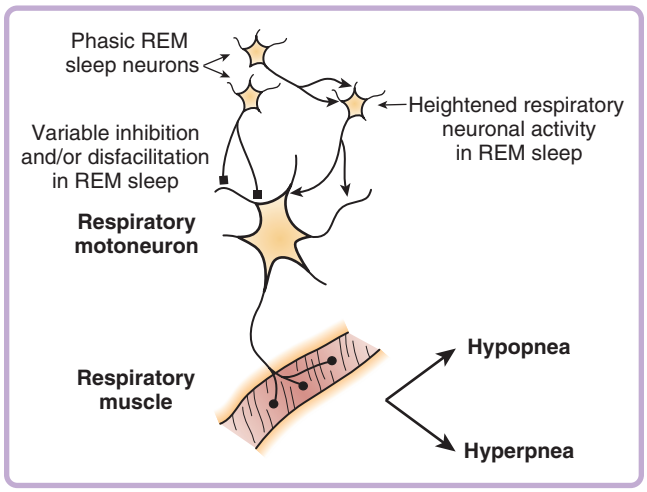


Figure 15-4 Schema depicting how respiratory motoneurons receive competing excitatory (*arrowheads*) and inhibitory (■) drives in REM sleep, the balance of which leads to time-varying increases and decreases in respiratory rate and amplitude, which manifest as hyperpnea and hypopnea, respectively. Additional factors that contribute to this variable lung ventilation in REM sleep are the similar competing excitatory and inhibitory influences at (1) pharyngeal motoneurons in REM sleep that lead to time-varying alterations in upper airway size and resistance, and (2) chest wall and abdominal muscles[23] that modulate resting lung volume and compliance of the chest wall. (Modified from Orem J. Neuronal mechanisms of respiration in REM sleep. *Sleep* 3:251–67, 1980.)

of Respiratory Neurons).[3,31] This observation highlights that a powerful inhibition and/or disfacilitation (i.e., withdrawal of excitation) must be taking place at respiratory motoneurons to explain the periods of reduced motor *output* despite continuing, and even heightened, *inputs* from respiratory neurons in REM sleep (Figure 15-4).[3,33,34]

Presynaptic Modulation

The control of respiratory motoneuron activity by changes in sleep state–dependent neuromodulators and/or inputs from respiratory neurons often emphasizes the postsynaptic effects of released neurotransmitters (see earlier). Such postsynaptic effects do not fully account for the control of motoneuron activity, however, because presynaptic modulation of the prevailing inputs also is important in motor control (Figure 15-3). For example, inhibitory inputs arriving at a nerve terminal before the subsequent arrival of a descending excitatory drive can lead to marked reductions in the release of excitatory neurotransmitters, so leading to the suppression of motoneuron activity. Such presynaptic modulation of neuronal activity is thought to be significant for information processing in neurons innervated by several converging pathways, as is the case for the organization of respiratory motoneurons (Figure 15-3). Accordingly, under specific behaviors, some inputs can be selectively suppressed, whereas others are left unaffected. This presynaptic modulation of specific inputs allows for selective control of motoneuron excitability, an effect that could not be achieved by a generalized postsynaptic modulation that affects the whole cell. This differential control has particular relevance in the control of motoneurons with dual respiratory and nonrespiratory functions, such as hypoglossal motoneurons innervating the genioglossus muscle of the tongue. In hypoglossal motoneurons, the presynaptic inhibition of the incoming central respiratory drive potentials allows for the switching of motor output appropriate for other behaviors such as swallowing, sucking, or speech, without the interference of respiration.[35]

Tonic and Respiratory-Related Activity in Respiratory Muscle

Some respiratory muscles exhibit more respiratory-related activity than others, whereas other muscles are more tonically active and exhibit little respiratory-related activity (Figures 15-1 and 15-2). For example, the genioglossus muscle of the tongue shows both tonic and respiratory-related activity, with the decreased activity of this muscle during sleep strongly linked to the pathogenesis of obstructive sleep apnea.[36] Similarly, the different intercostal muscles show various degrees of respiratory-related and tonic activities related to both the respiratory and postural functions of these muscles, with the expression of this respiratory-related versus tonic activity related to specific anatomic location in the chest wall and ongoing behaviors.[29,37] Suppression of intercostal muscle activity in REM sleep is thought to increase the compliance of the chest wall and to contribute to decreased functional residual capacity, effects that can in turn contribute to hypoventilation, especially in infants because of their already highly compliant chest wall.[29] In contrast with these muscles with respiratory-related activity, the tensor palatini muscle of the soft palate displays mostly tonic activity, which decreases with progression from wakefulness to NREM and REM sleep. The tonic activity in the tensor palatini is thought to enhance stiffness in the segment of the upper airway at the level of the soft palate, a consistent site of airway closure in obstructive sleep apnea.[4] Accordingly, decreases in tonic tensor palatini muscle activity from wakefulness to sleep (Figure 15-2) contribute to increased upper airway resistance and the predisposition to airway occlusion in sleep, with this effect of sleep predominantly affecting breathing by an effect on the tonic (nonrespiratory) inputs to these motoneurons, which receive little or no respiratory input at rest. Ultimately, whether some muscles exhibit respiratory-related activity at rest depends on both the "strength" of the input from respiratory neurons compared with the tonic drives (see Figure 15-5 and later section, Control of Respiratory Neurons)[3] and also the degree of suppression of the respiratory activity by vagal afferents related to lung volume.[4]

NEUROMODULATION OF RESPIRATORY MOTONEURONS ACROSS SLEEP-WAKE STATES

Studies addressing the neurochemical basis for the modulation of respiratory motor activity across natural sleep-wake states, in vivo, have been confined largely to the hypoglossal and trigeminal motor nuclei.[19,20,23,38] This focus on pharyngeal motoneurons is clinically relevant in elucidating the pathogenesis of obstructive sleep apnea, with airway obstructions occurring behind the tongue both at the level of the soft palate and below.[4] In contrast with this focus on pharyngeal motoneurons, similar studies investigating the control of intercostal and phrenic motoneurons in naturally sleeping animals are lacking. Nevertheless, studies of spinal motoneurons have provided important information regarding the control of postural motoneurons across sleep-wake states.[23] Because intercostal motoneurons perform both postural and respiratory functions,

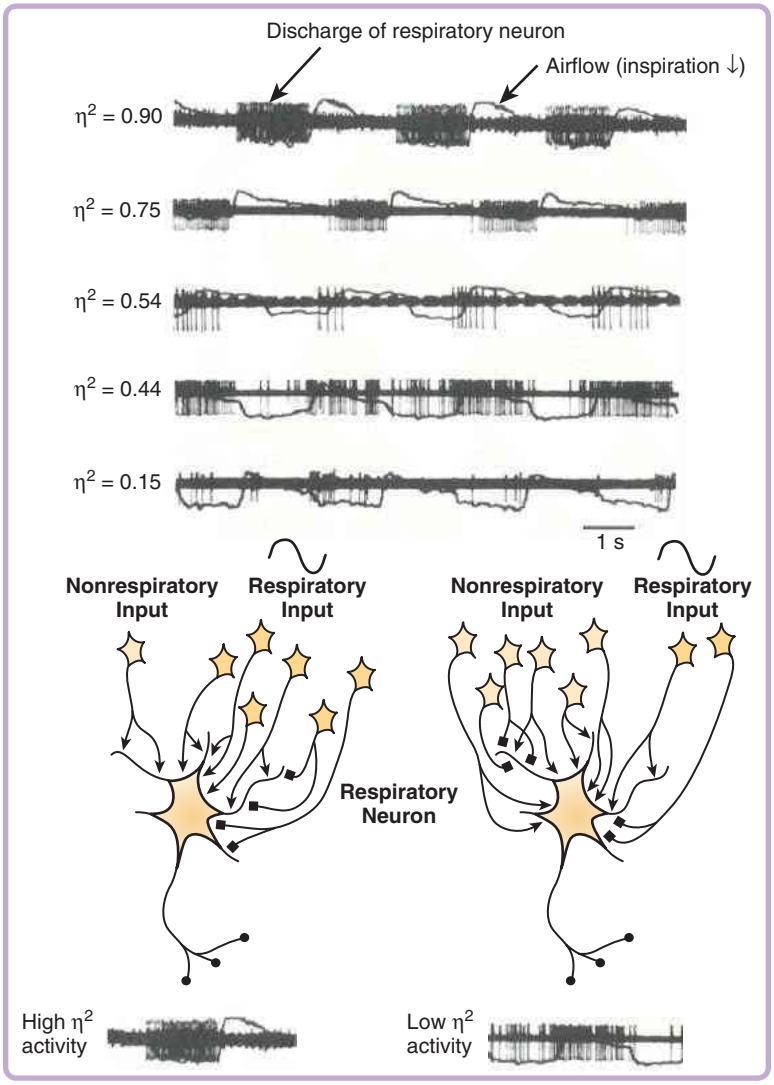


Figure 15-5 Five different medullary respiratory neurons recorded in intact cats in NREM sleep. These neurons vary in the strength of their relationship to breathing, an effect that is quantified by the η^2 statistic with values ranging from 0 (weak relationship) to 1.0 (strong relationship). High η^2 cells are considered to be more strongly influenced by respiratory inputs than nonrespiratory inputs, and vice versa for low η^2 cells. (Modified from Orem J, Kubin L. Respiratory physiology: central neural control. In: Kryger MH, Roth T, Dement WC, editors. *Principles and Practice of Sleep Medicine.* 3rd ed., Philadelphia: WB Saunders; 2000.)

the mechanisms identified at primary postural motoneurons are likely to have close similarities to the mechanisms controlling the nonrespiratory (postural) component of intercostal motor activity. A focus on the control of motoneurons innervating the muscles of the respiratory pump also is relevant within the scope of this chapter because significant hypoventilation can occur in sleep, especially REM sleep, in patients with restrictive lung diseases (e.g., kyphoscoliosis, obesity hypoventilation) and neuromuscular weakness (e.g., postpolio syndrome, muscular dystrophy, amyotrophic lateral sclerosis, partial diaphragm paralysis).[39] Summarized next are the major findings from animal studies addressing the sleep state–dependent modulation of respiratory motor activity. These data derive in large part from studies at the hypoglossal motor nucleus, a model motor pool with dual respiratory and nonrespiratory functions.[19]

Excitatory Influences Across Sleep-Wake States

The concept of a tonic drive activating respiratory muscle in wakefulness but not in sleep (i.e., the wakefulness stimulus for breathing) has been an important and enduring notion in respiratory medicine,[3,29] not least because it is useful in modeling sleep effects on breathing and in elucidating the pathogenesis of sleep-related breathing disorders. Neurons of the aminergic arousal system provide an important source of tonic drive to the respiratory system (Figure 15-2).[3,19,20] Serotonin- and norepinephrine-containing neurons have been of particular attention experimentally because these neurons send excitatory projections to respiratory motoneurons, and because these neurons show their highest activity in wakefulness, reduced activity in NREM sleep, and minimal activity in REM sleep—a pattern that may contribute to reduced

respiratory muscle activity in sleep through withdrawal of excitation.[3,19,20]

Animal studies show that an endogenous noradrenergic drive to the hypoglossal motor nucleus contributes to both the respiratory and tonic components of genioglossus muscle activation in wakefulness, and the residual expression of respiratory-related activity that persists in NREM sleep as the tonic drive is withdrawn.[40] Moreover, this noradrenergic contribution to genioglossus muscle tone was shown to be minimal in REM sleep, thereby explaining, at least in part, the periods of genioglossus muscle hypotonia during REM sleep.[40,41] The identification of an endogenous excitatory noradrenergic drive that contributes to genioglossus muscle activation in wakefulness, but is withdrawn in sleep, is particularly significant because since the first clinical description of obstructive sleep apnea, this was the first identification of a neural drive contributing to the sleep state–dependent activity of a muscle that is central to this disorder. The location of the central noradrenergic neurons that may provide this drive to hypoglossal and other respiratory motoneurons is reviewed elsewhere.[19] As noted previously, with the widespread projections of brainstem aminergic neurons, they also are positioned to provide an endogenous input to other respiratory neurons and motoneurons and thereby influence respiratory pump muscle activity and ventilation across sleep-wake states.[42,43] Recent data also point to a role for endogenous glutamatergic inputs in the tonic excitatory drive that increases pharyngeal muscle activity in wakefulness, the withdrawal of which contributes to reduced activity in sleep.[19,44,45] In contrast with these functionally active tonic inputs, endogenous levels of serotonin at the hypoglossal motor nucleus contribute less to the changes in genioglossus muscle activity in sleeping animals.[19] Whether this minimal influence of endogenous serotonin on genioglossus muscle activity also applies to humans remains to be determined. If so, it may explain (at least in part) the lack of clinically significant effects of selective serotonin reuptake inhibitors on pharyngeal muscle activity and obstructive sleep apnea severity in patients receiving these drugs.[19,46,47]

Local application of serotonergic, noradrenergic, and glutamatergic agonists to the hypoglossal or trigeminal motor nuclei produces robust motor activation in wakefulness and NREM sleep.[19,45] These observations provide "proof of principle" for the notion that it may be possible to develop pharmacologic strategies to increase respiratory muscle activity in sleep—for example, as a potential treatment for obstructive sleep apnea. Of importance, however, a major component of the motor activation observed in response to these agonists in NREM sleep is overcome in REM sleep.[19,45] An important practical implication of this differential modulation of pharyngeal motor responses to otherwise potent excitatory neuromodulators between NREM and REM sleep has been recognized. For example, even if it is possible to effectively target pharyngeal motoneurons with directed pharmacologic manipulations, such as for treatment for obstructive sleep apnea, then different strategies may be required to produce sustained pharyngeal muscle activation throughout both the NREM *and* REM sleep stages, because the neurobiology of motor control is fundamentally different between these two states.[19]

Inhibitory Influences Across Sleep-Wake States

Glycine and GABA are the main inhibitory neurotransmitters in the central nervous system. Glycine and $GABA_A$ receptor stimulation at the hypoglossal motor nucleus in vivo produces the expected depression of genioglossus muscle activity, whereas antagonism of these receptors increases genioglossus activity across all sleep-wake states.[19,26,27] The augmentation of respiratory-related motor activity across *all* sleep-wake states with application of antagonists for these inhibitory neurotransmitters fits best with the notion of a continuous background (i.e., tonic) inhibitory tone that constrains the rhythmic activation via gain modulation.[48] Moreover, any motor-activating effects observed with glycine and GABA receptor blockade at the cranial motor pools are trivial and, of note, are of smallest magnitude in REM sleep compared with wakefulness and NREM sleep. These findings, observed at both the hypoglossal[49,50] and trigeminal[38,51] motor pools, suggest that inhibition by glycine and GABA should not be viewed as a significant mediator of pharyngeal motor inhibition in REM sleep, because the inhibitory tone is present across all sleep-wake states and is weakest of all in REM sleep (see referenced sources[26,27,52] for further details, and see further on for discussion of the strong inhibitory mechanism that operates at the hypoglossal motor pool in REM sleep).

Nevertheless, the tonic inhibitory effects of GABA at respiratory neurons[17] and motoneurons[19] are clinically relevant in view of the widespread use of sedative-hypnotic drugs. For example, benzodiazepine and imidazopyridine drugs commonly are prescribed as sedative-hypnotics (e.g., lorazepam, zolpidem, respectively), and both of these classes of sedatives promote sleep by enhancing GABA-mediated neuronal inhibition through interactions with binding sites on $GABA_A$ receptors.[16] The presence of lorazepam and zolpidem at the hypoglossal motor nucleus also leads to inhibition of genioglossus muscle activity.[19] This inhibitory effect of sedative-hypnotics at respiratory motor nuclei may underlie a component of the respiratory depression observed clinically with excessive $GABA_A$ receptor stimulation, and the predisposition to obstructive sleep apnea in some persons taking sedative-hypnotics and other $GABA_A$ receptor-modulating neurodepressive drugs such as alcohol and certain general anesthetics.[18]

Large inhibitory glycinergic potentials appear to play an important role in the inhibition of spinal motoneuron activity in REM sleep,[23] and this probably explains the inhibition of intercostal respiratory muscle activity in this sleep state.[29] As discussed previously, however, glycine and $GABA_A$ receptor antagonism at the hypoglossal[49,50] and trigeminal[38,51] motoneuron pools fails to reverse the profound *tonic* suppression of genioglossus or masseter muscle activity in REM sleep, although in both cases this antagonism increases the amount and/or magnitude of the sporadic *phasic* motor activations in REM sleep.[19,38] These increases in phasic motor activity during REM sleep with glycine and $GABA_A$ receptor antagonism point to a functional role for sporadic inhibitory neurotransmission in the modulation of hypoglossal and trigeminal motor excitability,[18,25,26] and such inhibitory potentials have been recorded at hypoglossal motoneurons in REM sleep.[53] Based on the findings described here, however, this inhibitory glycinergic and GABAergic mechanism appears not to be as profound as the inhibition demonstrated at spinal motoneurons.[23]

Indeed, the mechanism of hypoglossal motor suppression in REM sleep appears to be different from that for spinal motoneurons. A cholinergic (muscarinic) receptor mechanism

linked to G protein–coupled inwardly rectifying potassium (GIRK) channels mediates the strong inhibition of the tongue musculature in REM sleep.[52] This inhibition is strong enough to counteract the inspiratory excitatory drive to hypoglossal motoneurons that originates from the respiratory network (Figure 15-1), such that respiratory motor activation of the tongue musculature can be abolished during REM sleep even during strong respiratory stimulation with hypercapnia.[54] Moreover, unlike glycine and $GABA_A$ receptor blockade at the hypoglossal motor pool, blockade of this cholinergic-GIRK channel mechanism is capable of reversing the REM sleep–induced hypoglossal motor suppression and of restoring respiratory genioglossus activity throughout REM sleep.[26,52,55]

The degree of suppression observed in a variety of *respiratory pump* muscles in REM sleep appears to be strongly correlated with the muscle spindle density of these different muscles.[37] The diaphragm has few, if any, spindles and little inhibition in REM sleep, whereas different intercostal muscles (especially the external inspiratory intercostals) have significant numbers of muscle spindles and profound suppression of activity in REM sleep, with variation in the degree of suppression in accordance with muscle spindle density.[29,37] Of clinical relevance, *acute* diaphragm paralysis leads to increased reliance on the intercostal and accessory muscles to maintain effective lung ventilation, but this compensation is lost in REM sleep, when the motoneurons innervating these muscles with dual respiratory and postural functions are inhibited.[56] Of interest, however, patients with *chronic* bilateral diaphragm paralysis are able to recruit nondiaphragmatic inspiratory muscle activity during REM sleep, thereby lessening any attendant hypoventilation. This compensation suggests that the central nervous system in these patients is able to functionally reorganize the drives controlling the accessory respiratory muscles such that activity is less suppressed by REM sleep mechanisms in the long term.[57,58]

Mechanisms Operating Across Sleep-Wake States

Current information from experiments in sleeping animals indicates that reduced excitation, largely through withdrawal of endogenous noradrenergic and glutamatergic inputs, is principally responsible for reductions in pharyngeal muscle tone from wakefulness to NREM and REM sleep.[19,20,45] By comparison, an endogenous serotonergic drive plays a lesser role.[19] Increased inhibitory neurotransmission mediated by glycine and GABA also contributes to suppression of pharyngeal motor activity in REM sleep, but the contribution of this mechanism appears to be much less than expected[19,20,45] from studies at spinal motoneurons.[23] Rather, a cholinergic-GIRK channel inhibitory mechanism operates at the hypoglossal motor pool, with the largest inhibitory influence of this mechanism seen in REM sleep and minimal or no effects in waking or NREM sleep. This mechanism is the major cause of inhibition of the tongue musculature in REM sleep.[26,27,52] By contrast, glycine and GABA exert a continuous background (i.e., tonic) inhibitory tone that is present across all sleep-wake states, with constraint of hypoglossal respiratory motor outflow by this tone through gain modulation. Augmentation of this tonic inhibitory GABA tone with commonly administered neurodepressive drugs can lead to further suppression of pharyngeal muscle activity and precipitation of upper airway obstructions in susceptible persons, such as those with anatomically narrow upper airways who already are prone to experience obstructive sleep apnea.

CONTROL OF RESPIRATORY NEURONS

Respiratory Neurons Vary in the Strength of Their Relationship to Breathing

Studies by John Orem and colleagues in sleeping animals led to the fundamental concepts that still best explain the neural basis for the effects of sleep on breathing, including the nature of the so-called wakefulness stimulus, and for the rapid, irregular breathing pattern of REM sleep.[3] Key to this achievement was development of a statistical approach to quantify the consistency and strength of the respiratory-related component of a neuron's activity as related to its overall discharge. The strength of this relationship was quantified by the *eta-squared statistic* (η^2), with η^2 values ranging from 1.0 (strongest relationship) to 0 (i.e., weakest relationship).[3] Of importance, different brainstem respiratory neurons vary in the strength of their relationship to the inspiratory or expiratory phase of the breathing cycle (Figure 15-5).

The interpretation and physiologic meaning of the η^2 value for any given respiratory neuron are best explained in the following quote from Orem, for whom cells with high η^2 values were "quintessentially respiratory . . . , protected from nonrespiratory distortions, perhaps because of rigid sequences of excitatory and inhibitory postsynaptic potentials that preclude activity that is not strictly respiratory."[3] By comparison, the activity of "low η^2-valued cells is the apparent result of mixtures of inputs that have respiratory and nonrespiratory forms."[3] Figure 15-5 further illustrates this concept by showing that the degree of respiratory-related activity of a given respiratory neuron (i.e., its η^2 value) depends on the balance of the respiratory *and* nonrespiratory inputs to that neuron. This is an important concept because respiratory neurons with different η^2 values are differentially affected by sleep-wake state.

Respiratory Neuron Activity in NREM Sleep

The notion that the degree of respiratory-related activity of a particular respiratory neuron depends on the balance of its respiratory and nonrespiratory inputs assumes significant physiologic and clinical relevance with the following experimental observations, made across the sleep-wake cycle:

- Neurons with low η^2 activity—that is, those that are less influenced by the respiratory oscillator but are strongly influenced by nonrespiratory tonic drives—are *most* affected by the transition from wakefulness to NREM sleep, such that their activity can even cease during sleep.
- Neurons with high η^2 activity—that is, those that presumably are strongly coupled to, and controlled by, the respiratory oscillator—are *least* affected by the transition from wakefulness to NREM sleep.

These observations and findings are illustrated in Figure 15-6.[3]

Of note, those respiratory neurons with low η^2 values that become inactive in sleep are not ceasing their activity simply because these neurons lose their respiratory input. That idea is discounted because experimental reexcitation of those low η^2 respiratory neurons that become silenced during NREM sleep restores their rhythmic respiratory activity. This finding shows that the respiratory-related input persists onto those

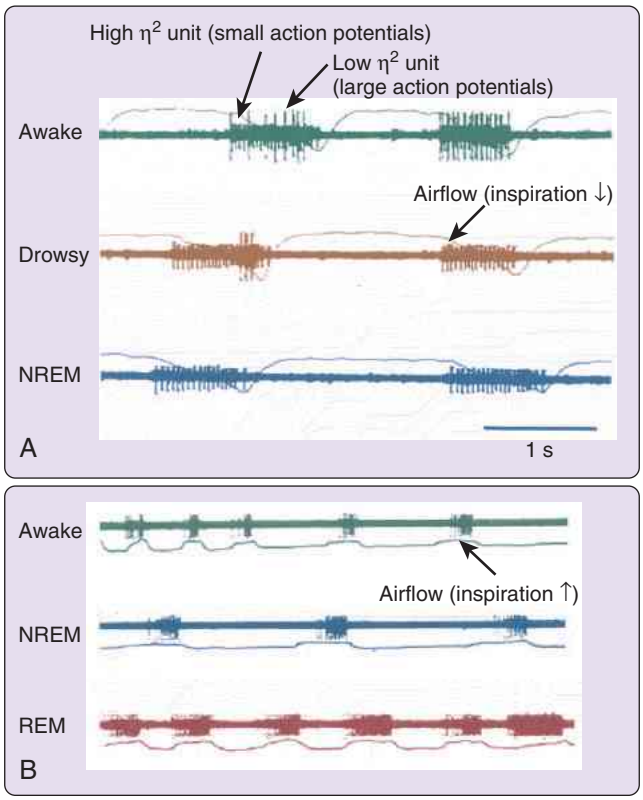


Figure 15-6 A, The activity of high η^2 medullary respiratory neurons is little affected by NREM sleep, whereas the activity of low η^2 cells is significantly suppressed in NREM sleep. This differential effect of NREM sleep on these different classes of respiratory neurons is thought to be due to the particular sensitivity of the tonic nonrespiratory inputs to changes in sleep-wake state, which is the basis of the so-called wakefulness stimulus for breathing. **B,** Electromyographic tracings showing increased and advanced activity of a late inspiratory neuron in REM sleep. (Modified from Orem J, Kubin L. Respiratory physiology: central neural control. In: Kryger MH, Roth T, Dement WC, editors. *Principles and Practice of Sleep Medicine.* 3rd ed., Philadelphia: Saunders; 2000.)

inactive low η^2 respiratory neurons in NREM sleep but that this respiratory signal was subthreshold and therefore did not show itself as motor activation (see Figure 15-3, *C,* for comparison and explanation of this principle).[3,59] The major principle here, well articulated by Orem, is that the magnitude of the "effect of sleep on a respiratory neuron is proportional to the amount of nonrespiratory activity in the activity of that neuron," such that the "wakefulness stimulus to breathing is nonrespiratory in form and affects some respiratory neurons more than others.[3] This principle underscores the key importance of tonic drives in the expression of both tonic *and* respiratory neuronal activities.

Respiratory Neuron Activity in REM Sleep

REM sleep is characterized by (1) overall depression of the ventilatory responses to hypercapnia and hypoxia[29]; (2) periods of profound suppression of motor activity in respiratory muscles (e.g., intercostal and pharyngeal)[19,20] and nonrespiratory (i.e., postural) muscles[23]; and (3) occasional periods of slowing of respiratory rate. Periods of sporadic respiratory slowing in REM sleep are associated with increased release of acetylcholine into the pontine reticular formation.[21] It is not correct, however, to consider REM sleep as a state of *overall* depression of central respiratory neurons because, as for most cells in the central nervous system, the activity of brainstem respiratory neurons typically is *greater* in REM sleep than in NREM sleep. As an example, late inspiratory neurons have increased and advanced activity in REM sleep; that is, cells that discharge in the latter part of inspiration in NREM sleep can be active throughout inspiration in REM sleep (Figure 15-6).[3]

A large degree of variability has been observed in the discharge pattern of respiratory neurons in REM sleep; this variability is associated with tonic and phasic REM sleep events.[3] For example, increased medullary respiratory neuronal activity is associated with increased occurrence of ponto-geniculo-occipital waves, these waves being a defining feature of phasic REM sleep events. This finding suggests that the activity of respiratory neurons in REM sleep is strongly influenced by processes and activities that are peculiar to the neurobiology of the REM sleep state per se, rather than being an intrinsic component of the respiratory network.[3] This notion of significant influences on respiratory network activity by nonrespiratory inputs has similarities to the major influence of tonic drives discussed previously in the context of the wakefulness stimulus to breathing. Together, these concepts serve to highlight that the activity levels of central respiratory neurons and motoneurons is determined by the interaction of their component nonrespiratory *and* respiratory inputs, the former having major influences on overall respiratory activity and being particularly sensitive to changes in sleep-wake state.

As discussed previously for respiratory motoneurons, this effect of REM sleep in activating central respiratory neurons can lead to periods of increased respiratory rate and respiratory muscle activity. Of importance, and as mentioned, these periods of increased respiratory network activity in REM sleep are intimately related to the neural substrate for the REM sleep state per se. As a consequence, they also are largely *unrelated* to processes of respiratory control, including homeostatic feedback regulation and responses to prevailing blood gas tensions.[3,30,31] This increased activity of central respiratory neurons in REM sleep also is likely to be responsible for producing the periods of increased respiratory rate and higher respiratory muscle activity at times when the normally time-varying inhibition of respiratory motoneurons is briefly weakened or withdrawn (Figures 15-3 and 15-4). That REM sleep can lead to periods of heightened diaphragm activity unrelated to prevailing blood gas tensions has particular relevance for the clinical observation that hypocapnic central apneas most commonly occur in NREM sleep but can be absent in REM sleep, when breathing is characteristically erratic.[60,61] Figure 15-4 illustrates how this balance of excitatory and inhibitory influences at respiratory motoneurons can underlie the highly variable respiratory activity in REM sleep, including periods of respiratory depression despite activation of central respiratory neurons.

Neuromodulation of Respiratory Neurons across Sleep-Wake States

Unlike the studies performed at respiratory motor pools,[19,20] no studies have been conducted to identify or otherwise characterize the neurochemicals that may mediate the control of respiratory neurons in vivo as a function of sleep-wake states. Nevertheless, it is a reasonable working hypothesis that the neuronal groups involved in the modulation of respiratory

motoneurons across sleep-wake states also are likely to affect respiratory neurons. Accordingly, influences from brainstem reticular neurons (probably glutamatergic) are positioned to provide a source of tonic drive to respiratory neurons, with alteration of this influence from wakefulness to NREM and REM sleep.[3,19,20,45] Brainstem reticular neurons generally show decreased activity in NREM sleep compared with wakefulness, and increased activity in REM sleep[3,22]—a pattern similar to the changes in respiratory neuron activity discussed earlier. Electrical stimulation of reticular neurons in the midbrain converts the activity of several respiratory motor nerves or muscles from a sleeplike pattern to one more like wakefulness.[3] One key source of the tonic (nonrespiratory) input to medullary respiratory neurons in the awake state (i.e., the wakefulness stimulus) is thought to arise from brainstem reticular neurons.[3] The source or sources of the drives activating central respiratory neurons in REM sleep, however, have not been determined.

Neurons of the aminergic arousal system (serotonergic, histaminergic, and noradrenergic), and other sleep state–dependent neuronal groups, also are positioned to provide a source of tonic (i.e., nonrespiratory) drive to respiratory neurons across sleep-wake states. However, whether these tonic drives would be excitatory or inhibitory to respiratory neurons depends on the receptor subtypes activated, and on the pre- or postsynaptic location of these receptors (see Figures 15-2 and 15-3). This lack of knowledge of the sleep state–dependent neuromodulation of respiratory neurons can be addressed by further research, which also may identify specific pharmacologic approaches that can preserve respiratory neuron activity in sleep and in states of drug-induced brain sedation, so as to minimize respiratory depression. The various brain structures that exert behavioral control of the respiratory system also should be considered as a source of the wakefulness stimulus for breathing,[3] but it is not known if this collection of inputs share the same neurochemicals as for the aforementioned inputs from the brainstem reticular neurons and sleep state–dependent neuronal systems.

CLINICAL PEARL

The withdrawal of the wakefulness stimulus to breathing at the transition from wakefulness to sleep is the principal mechanism underlying the major clinical sleep-related breathing disorders. Current evidence identifies neurons of the aminergic arousal system and reticular neurons as providing the key components of this wakefulness stimulus. Withdrawal of this tonic excitatory drive to the muscles of the upper airway is thought to underlie the normal sleep-related increase in upper airway resistance, and the hypoventilation, flow limitation, and obstructive sleep apnea observed in susceptible persons (e.g., those with already anatomically narrow upper airways). Patients with restrictive lung diseases and neuromuscular weakness rely, to various degrees, on the activation of nondiaphragmatic respiratory muscles to help maintain adequate ventilation in the awake state, but this compensation can be reduced or absent in sleep, leading to severe hypoventilation, as the essential tonic excitatory drive that is present in wakefulness is withdrawn. REM sleep mechanisms also lead to inhibition of respiratory motoneurons, thereby explaining the typically increased severity of abnormal breathing events in REM sleep compared with NREM sleep.

SUMMARY

Sleep is a state of vulnerability for the respiratory system. Central to the pathogenesis of a variety of sleep-related breathing disorders is loss of a wakefulness stimulus that sustains adequate breathing in wakefulness. This loss of the wakefulness stimulus is the root mechanism in understanding sleep effects on breathing. Significant developments have helped identify the neurochemical substrates underlying this wakefulness stimulus. Central to this understanding has been delineation of the neurobiology of sleep, its impact on central respiratory neurons and motoneurons, and the important role of tonic excitatory (nonrespiratory) drives in contributing to overall respiratory system activity. Moreover, in parallel with the realization that sleep onset is not simply the passive withdrawal of wakefulness, breathing during sleep is not simply due to the passive withdrawal of the wakefulness stimulus. NREM sleep and REM sleep are fundamentally different neurobiologic states that exert distinct effects on the control of respiratory neurons and motoneurons. Accordingly, NREM and REM sleep modes pose different problems with breathing during sleep in different people with different pathologic conditions. Understanding these mechanisms is necessary for identifying the physiologic basis for the spectrum of sleep-related breathing disorders and their appropriate clinical management.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported in part by grants from the Canadian Institutes of Health Research, the Ontario Thoracic Society, and the Canada Foundation for Innovation and the Ontario Research and Development Challenge Fund. RLH is supported by a Tier 1 Canada Research Chair in *Sleep and Respiratory Neurobiology.*

Selected Readings

Fuller PM, Saper CB, Lu J. The pontine REM switch: past and present. *J Physiol* 2007;**584**:735–41.

Grace KP, Hughes SW, Horner RL. Identification of the mechanism mediating genioglossus muscle suppression in REM sleep. *Am J Respir Crit Care Med* 2013;**187**:311–19.

Grace KP, Hughes SW, Shahabi S, Horner RL. K^+ channel modulation causes genioglossus inhibition in REM sleep and is a strategy for reactivation. *Respir Physiol Neurobiol* 2013;**188**:277–88.

Horner RL. Neuromodulation of hypoglossal motoneurons during sleep. *Respir Physiol Neurobiol* 2008;**164**:179–96.

Horner RL, Hughes SW, Malhotra A. State-dependent and reflex drives to the upper airway: basic physiology with clinical implications. *J Appl Physiol* 2014;**116**:325–36.

Horner RL, Malhotra A. Control of breathing and upper airways during sleep. In: Broaddus VC, Mason RJ, Ernst JD, editors. *Murray & Nadel's textbook of respiratory medicine*. Philadelphia: Elsevier; 2015. p. 1511–26.

Luppi PH, Gervasoni D, Verret L, et al. Paradoxical (REM) sleep genesis: the switch from an aminergic-cholinergic to a GABAergic-glutamatergic hypothesis. *J Physiol (Paris)* 2006;**100**:271–83.

McGinty D, Szymusiak R. The sleep-wake switch: a neuronal alarm clock. *Nat Med* 2000;**6**:510–11.

Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* 2005;**437**:1257–63.

A complete reference list can be found online at ExpertConsult.com.

Chapter 16

Respiratory Physiology: Understanding the Control of Ventilation

Danny J. Eckert; Jane E. Butler

Chapter Highlights

- In the absence of respiratory disease, the process of breathing typically is afforded little conscious thought. Yet it is clearly fundamental to survival.
- Multiple inputs are capable of regulating the rate and depth of breathing. These are regulated by feedforward and feedback mechanisms that control blood gas levels within relatively narrow limits to maintain homeostasis.
- The physiologic capacity to alter breathing is substantial. When metabolic demand decreases during sleep, even very low levels of ventilation can be tolerated. However, the major changes to the control of breathing that occur during sleep can cause breathing disruption.
- This chapter outlines the key neuroanatomic inputs to breathing, describes the changes that occur in the control of breathing during sleep, including differences between men and women, and highlights how abnormal control of ventilation can contribute to sleep-disordered breathing.

OVERVIEW OF THE CONTROL OF BREATHING

Breathing is controlled by means of highly effective feedforward and feedback mechanisms. Conceptually, the functional organization consists of three key elements: (1) brainstem neurons responsible for respiratory pattern generation (*central control*), (2) respiratory muscles that generate force to move airflow in and out of the lungs (*effectors*), and (3) multiple inputs that relay respiratory sensory information (*sensors*) to brainstem respiratory control centers to allow for adjustments according to the prevailing physiologic conditions (Figure 16-1). A breakdown in or damage to any one of these components can lead to breathing abnormalities. During wakefulness, however, additional inputs to breathing can compensate to maintain breathing and blood gas levels within acceptable levels despite damage to key elements that underlie the control of breathing. Accordingly, breathing problems often only emerge (or worsen) during sleep, when wakefulness compensatory mechanisms are either downregulated or absent. This chapter outlines the key components that underpin the control of breathing and highlights the major changes that occur during sleep. This chapter is a synthesis of many elements described in Chapters 15, 17, 18, and 24, plus Section 14 on Sleep Breathing Disorders, with a perspective aimed at translating the information toward a more comprehensive understanding of respiration and sleep disturbances in humans.

CENTRAL CONTROL OF BREATHING

The precise neuroanatomic locations that contribute to respiratory pattern generation within the brainstem are incompletely understood. The central respiratory control network involves both inspiratory and expiratory neurons. Presented next is a brief summary of some of the key brainstem sites and their interconnections, based primarily on animal models.

Central respiratory control and rhythmicity occur within the pons and medulla. Within the medulla, the dorsal and ventral respiratory groups are particularly important (Figure 16-2). The *dorsal respiratory group* contains the nucleus tractus solitarius (nTS). The nTS is a key cardiorespiratory sensory integration site. Afferent information from phrenic, vagus, and peripheral chemoreceptors (by way of the glossopharyngeal nerve) arrive at the nTS. The nTS has numerous outputs contributing to important control of breathing centers, including the nearby retrotrapezoid nucleus[1,2] (see the following section on chemoreceptors). The ventrolateral region of the nTS is believed to be particularly important for inspiratory activity. Major projections also extend to other key respiratory control centers within the ventral respiratory group. Not yet known, however, is whether direct output to respiratory motoneurons occurs.

The pre-Bötzinger complex forms part of the *ventral respiratory group* (see Figure 16-2). The pre-Bötzinger complex is believed to be the major putative respiratory pacemaker. This stems from findings that show persistence of respiratory rhythmicity within these cells in minimal slice preparations.[3] In support of the importance of this region to respiratory control, the pre-Bötzinger complex has multiple projections to other known respiratory control sites within the brainstem.[4] Adjacent to the pre-Bötzinger is the Bötzinger complex. This area plays an active role during expiration by inhibiting respiratory motor neurons to modulate the overall motor output. The rostral ventral respiratory group also includes inspiratory pre-motoneurons such as those located in the nucleus ambiguus. The nucleus ambiguus provides respiratory motor output to the larynx and pharynx by way of the vagi. The nucleus

retroambiguus also may contribute to respiratory rhythm generation.[5]

Although respiratory rhythm generation neurons are located predominantly within the medulla, the *pontine respiratory group* (previously referred to as the pneumotaxic center) also is importantly involved in central respiratory control[6] (see Figure 16-2). The pontine respiratory group includes the nucleus parabrachialis medialis, containing expiratory active neurons. The parabrachialis lateralis and the Kölliker-Fuse nucleus in the upper pons contain inspiratory neurons. Pontine respiratory group activation can decrease inspiratory activity within the dorsal respiratory group, leading to a decrease in inspiratory time. This "inspiratory-expiratory phase transition" can increase breathing frequency.

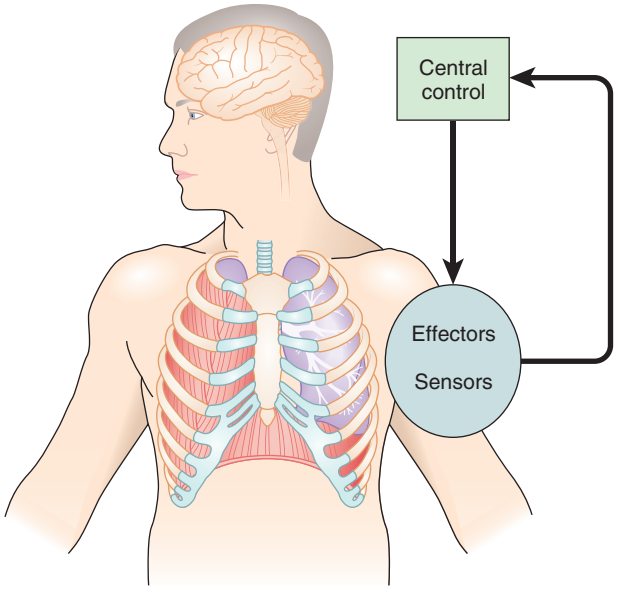


Figure 16-1 Control of Breathing Overview. Breathing is controlled by means of feedforward and feedback mechanisms involving central control, effectors, and sensors. Refer to text for further details.

CHEMICAL CONTROL OF BREATHING

Chemical control is the most important regulator of breathing in healthy persons during quiet breathing. This is true during both wakefulness and sleep. All cells are capable of modifying their activity in response to extreme changes in the chemical environment. Certain cells, however, are highly sensitive to quite minor changes. These chemically sensitive areas can regulate the control of breathing directly or have projections to central control of breathing sites. Accordingly, these groups of cells, known as *chemoreceptors*, are fundamentally important to the control of breathing.

Peripheral versus Central

Chemoreceptors are located peripherally and centrally (Figure 16-3). The main peripheral chemoreceptors lie at the bifurcation of the common carotid arteries. The carotid bodies have long been known to respond to changes in oxygen, carbon dioxide, and hydrogen ion concentration.[7] Detection of these stimuli can lead to rapid alterations in breathing (within 1 or 2 breaths). In addition, recent findings show that the carotid bodies respond to a wide range of other stimuli including potassium, norepinephrine, temperature, glucose, insulin, and immune-related cytokines.[7,8] Repeated exposure to hypoxia can cause plasticity within the carotid bodies.[8] The changes that occur can contribute to pathologic states including an increased propensity for breathing instability during sleep.[7,8] In addition to the carotid bodies, the nearby aortic bodies are also capable of responding to changes in oxygen and other

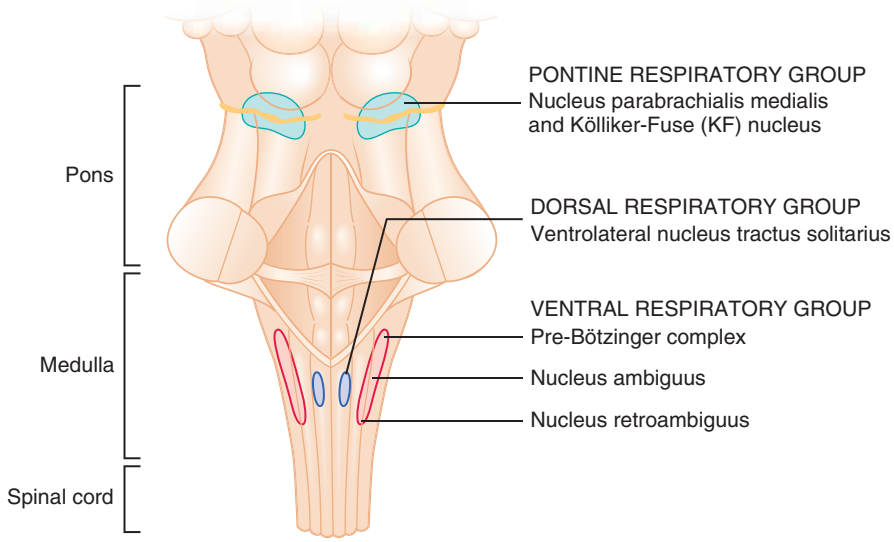


Figure 16-2 Central Control of Breathing. Major regions involved in the central control of breathing lie within the pontine respiratory group, comprising the nucleus parabrachialis medialis and the Kölliker-Fuse nucleus; the ventral respiratory group, consisting of the pre-Bötzinger complex, nucleus ambiguus, and nucleus retroambiguus; and the dorsal respiratory group, comprising the ventrolateral nucleus tractus solitarius. Refer to text for further details. (From Eckert DJ, Roca D, Yim-Yeh S, Malhotra A. Control of breathing. In: Kryger M, editor. *Atlas of clinical sleep medicine*, vol. 2. 2nd ed. Philadelphia: Saunders; 2014, p. 45–52.)

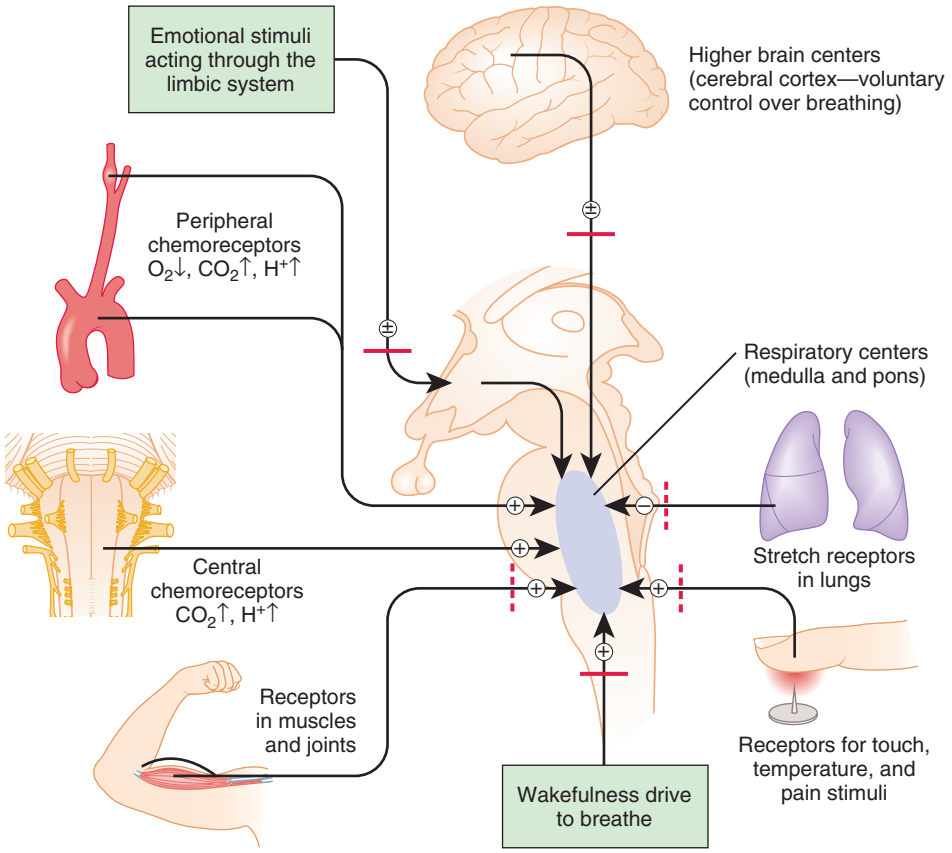


Figure 16-3 Inputs to Breathing. Schematic of the multiple inputs that are capable of regulating breathing. During sleep, many of these inputs are either substantially diminished (*dashed red lines*) or absent (*solid red lines*). Thus the predominant inputs to breathing during sleep are the chemoreceptors, which themselves also are downregulated and affected by state. *Note*: For simplicity, voluntary control of breathing is shown to act by way of the respiratory centers. Whether this is in fact the case or whether voluntary control acts directly on the respiratory motoneurons, however, has not been established. Refer to text for further details. (Modified from Kehlmann GB, Eckert DJ. Central sleep apnea due to a medical condition not Cheyne Stokes. In: Kushida CA, editor. *Encyclopedia of sleep*, vol. 1. 1st ed. San Diego: Elsevier; 2013, p 244–52; and Eckert DJ, Roca D, Yim-Yeh S, Malhotra A. Control of breathing. In: Kryger M, editor. *Atlas of clinical sleep medicine*, vol. 2. 2nd ed. Philadelphia: Saunders; 2014, p. 45–52).

chemical stimuli. Although the peripheral chemoreceptors are important for moment-to-moment modulation of breathing, the most powerful input to breathing during quiet wakefulness is from the central chemoreceptors.

Located on the ventral surface of the medulla, adjacent to the ventral respiratory group, lies the retrotrapezoid nucleus. This region is particularly important for central chemoreception.[9,10] The retrotrapezoid nucleus has major projections to key respiratory control centers including to the nTS within the dorsal respiratory group.[1,2] The central chemoreceptors respond to P_{CO_2} through changes in the pH of the extracellular fluid. CO_2 diffuses across the blood–brain barrier to increase hydrogen ion concentration in the cerebrospinal fluid. Thus, compared with the relatively fast-responding peripheral chemoreceptors, central chemoreceptors can take up to a minute to respond to changes in chemical stimuli. As discussed later, chemoreceptor response delays are critically important in mediating cyclicbreathing instability during sleep.[11-13]

Although the peripheral and central chemoreceptors are anatomically distinct and have different response characteristics, recent findings indicate complex interconnectivity.[7,8,14] Specifically, the activity of the central chemoreceptors is critically dependent on the activity of the peripheral chemoreceptors, and vice versa.[7,8,14]

OTHER INPUTS TO BREATHING

In addition to input from the chemoreceptors, other important inputs and sensors can contribute to the rate and depth at which we breathe (see Figure 16-3). Receptors in the limb muscles and joints can respond to movement to increase minute ventilation. Similarly, when receptors responsible for touch, temperature, and pain are stimulated, breathing increases. An independent stimulus to breathing known as the *wakefulness drive to breathe* also may be recruited.[15] Conversely, overinflation or excess lung stretch can inhibit minute ventilation by means of the Hering-Breuer reflex.[16] Other inputs can either stimulate or inhibit breathing. These inputs include limbic system input in response to emotional stimuli or voluntary cortical control. It remains uncertain, however, if voluntary override of breathing acts indirectly through

changes in central respiratory pattern generation or directly by way of phrenic motoneurons, or by a combination of both.[17] Nonetheless, the physiologic capacity to alter breathing is substantial. As highlighted later, when metabolic demand decreases during sleep, very low levels of ventilation (less than 5 L/minute) can be tolerated. Conversely, during intense exercise, ventilation can increase to greater than 200 L/minute.

STATE-RELATED CHANGES IN THE CONTROL OF BREATHING

Major changes in the control of breathing occur from wakefulness to sleep. The most significant change that occurs from wakefulness to sleep is that a majority of the inputs capable of modifying breathing are either absent or markedly downregulated (see Figure 16-3). Accordingly, chemical control of breathing is the dominant driver of breathing during sleep. In particular, CO_2 is critical in mediating breathing during sleep. Certain disease states adversely affect the chemical control of breathing and can cause sleep-disordered breathing in susceptible persons. This section outlines key state-related changes in the control of breathing that underlie cyclic breathing instability during sleep.

Sleep Onset

Respiratory control is inherently unstable during the transition from wakefulness to sleep.[18] Several factors contribute to respiratory instability at sleep transition. Certain components of respiratory control change rapidly with sleep onset, whereas others require more time. Mismatch in timing combined with downregulation in important respiratory control mechanisms underlies breathing disturbances during the sleep-onset period. Indeed, brief breathing stoppages at sleep onset are very common, even in otherwise healthy persons.

With respect to mechanical factors, the wakefulness drive to breathe and behavioral influences cease with sleep onset.[19] Movement and excitatory input to breathe from other external sensors become minimal or completely absent. Chemosensitivity also decreases[20] (see Figure 16-3). Accordingly, respiratory pump muscle tone is reduced, leading to a reduction in minute ventilation.[21] An abrupt reduction in upper airway muscle tone and protective reflexes also occurs with sleep onset.[21-25] These changes contribute to increased upper airway resistance.[23] The timing and magnitude of these changes vary among individual subjects. Rapid withdrawal of excitatory drive to breathing, in and of itself, can cause respiratory events as a consequence of the delay required to elicit a compensatory response from the chemoreceptors.[26] Patients who experience sleep apnea appear to be more prone than healthy control subjects to major reductions in the wakefulness drive to breathing.[27] As indicated by these findings, sleep onset affects all components of respiratory control and can cause major "state instability."

Stable Sleep

The removal of most excitatory inputs to breathing that occurs with sleep onset is a feature of stable sleep as well. Respiratory load compensation also is reduced during stable sleep compared with wakefulness.[28] Thus minute ventilation decreases during stable sleep, and the control of breathing becomes dominated by chemical input. However, downregulation in chemosensitivity is not isolated to the sleep-onset period. Ventilatory responses to hypoxia are reduced during stage 2 (N2) and slow wave sleep (N3), compared with wakefulness, such that major decreases in oxygen levels are required to stimulate breathing during sleep.[29-31] Accordingly, CO_2 is the main regulator of breathing during sleep. However, ventilatory responses to hypercapnia also are reduced during sleep compared with wakefulness, albeit to a lesser extent than for hypoxia.[32] Consequently, people can tolerate lower levels of minute ventilation and higher levels of CO_2 during sleep than in wakefulness. Typically, depending on the prevailing metabolic conditions, minute ventilation is reduced by 1 to 2 L/minute, and the partial pressure of carbon dioxide in the blood ($Paco_2$) increases by 3 to 8 mm Hg during stable sleep, compared with wakefulness[33] (Figure 16-4, *A*).

In the absence of respiratory disease, breathing is quite regular during stable non–rapid eye movement (NREM) sleep. Rapid eye movement (REM) sleep, by contrast, is characterized by breathing irregularity. Many medullary central control-of-breathing regions exhibit increased activation during REM compared with NREM sleep.[34] In humans, breathing frequency increases, and major variations are seen in breath-to-breath tidal volume. Active eye movements during REM sleep are associated with inhibition of upper airway dilator muscle activity and decreased tidal volume.[28,35] Protective upper airway reflexes also are inhibited.[36] Accordingly, obstructive apnea is common during REM sleep.

Brief Awakenings (Arousal from Sleep)

Brief cortical arousals from sleep lasting less than 15 seconds occur between 10 to 20 times per hour in healthy subjects. Arousal frequency increases with age.[37] Arousals can occur spontaneously or in conjunction with a sleep disorder such as sleep apnea or periodic limb movement disorder. Historically, arousals were believed to be essential for reopening the upper airway during obstructive breathing events.[38] Indeed, arousal can be beneficial in certain circumstances to rapidly resolve blood gas disturbances and to alleviate the increased work of breathing during flow-limited breathing.[39] However, although the initial physiologic changes associated with arousals may be beneficial for respiratory homeostasis, the rapid switch from sleep to wakefulness and the subsequent resumption of sleep can be highly destabilizing for respiratory control.[39] The extent to which arousals destabilize breathing and contribute to central or obstructive breathing events is dependent on two key features: (1) the subject's threshold for arousal—the *arousal threshold*—and (2) the ventilatory response to arousal.

Arousal Threshold

Whether an arousal occurs spontaneously, with a periodic limb movement, or in association with a respiratory disturbance, a person who wakes up easily (i.e., has a *low arousal threshold*) may be susceptible to sleep-state breathing instability. Specifically, a predisposition to sleep-onset breathing instability coupled with a low arousal threshold may lead to repetitive breathing disturbances as the affected person oscillates between wakefulness and sleep.[11] Approximately one third of patients with obstructive sleep apnea arouse to modest levels of respiratory stimuli (negative airway pressure less than 15 cm H_2O).[39,40] This relatively low threshold is likely to contribute to their sleep-disordered breathing.[39] Increasing the arousal threshold in these at-risk patients can stabilize

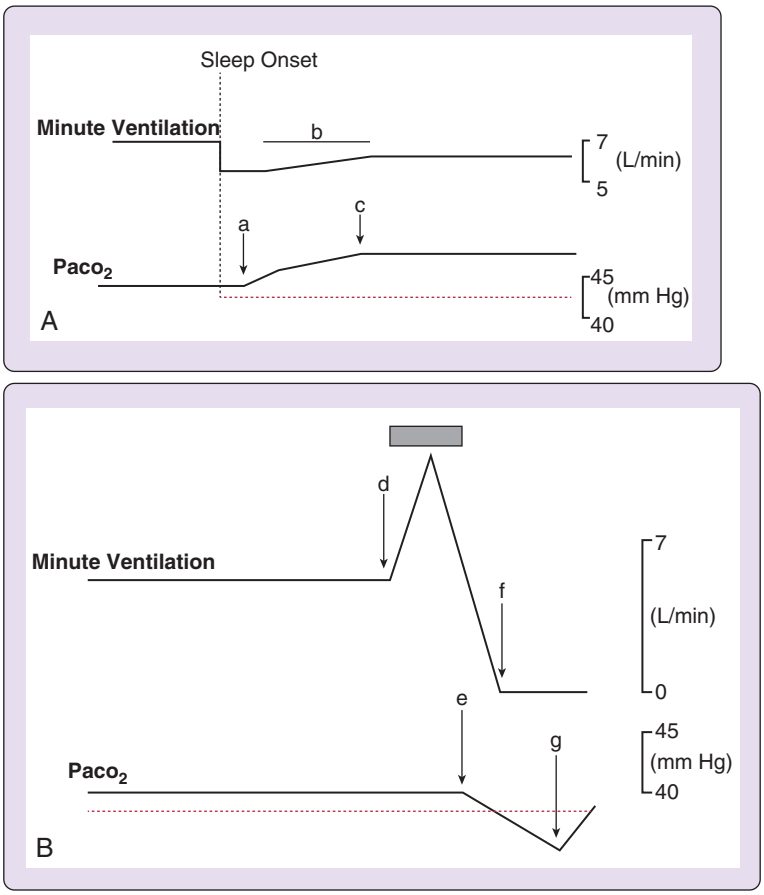


Figure 16-4 Sleep State Changes to the Control of Breathing. **A,** Schema showing typical changes in minute ventilation and $PaCO_2$ from wakefulness to sleep. At sleep onset (*dashed vertical line*), a rapid reduction in minute ventilation (from 7 to 5 L/min) occurs. A delay between the reduction in ventilation and changes in $PaCO_2$ (sleep onset to point a) also is seen. As CO_2 rises, upper airway muscles may be recruited, and minute ventilation may increase somewhat (period b), until a new eucapnic sleeping minute ventilation (5.5 L/min) and $PaCO_2$ (45 mm Hg) are reached (point c). The *horizontal red line* represents the theoretical apnea threshold (in this case, 39 mm Hg). **B,** Schematic representation of a central apnea after an arousal from sleep. At point d, a brief arousal from sleep occurs (arousal duration represented by the *gray box*). Hyperventilation occurs in association with reintroduction of wakefulness stimuli (ventilatory response to arousal). The hyperventilation lowers $PaCO_2$. However, a delay between the change in ventilation and the change in $PaCO_2$ can be seen (point d to point e). As the patient returns to sleep, the reduction in $PaCO_2$, caused by the ventilatory response to arousal, falls below the apnea threshold (which in this example is very close to the eucapnic sleeping $PaCO_2$ level), and apnea occurs (point f). The apnea leads to an increase in $PaCO_2$ until either an arousal occurs, and the cycle is repeated, or the apnea threshold is crossed and breathing resumes. Refer to text for further details. (From Kehlmann GB, Eckert DJ. Central sleep apnea due to a medical condition not Cheyne Stokes. In: Kushida CA, editor. *Encyclopedia of sleep*, vol. 1. 1st ed. San Diego: Elsevier; 2013, p. 244–52).

breathing.[41] Indeed, although the precise mechanisms remain uncertain, the arousal threshold and upper airway muscle activity increase in deeper stages of sleep, and sleep-disordered breathing severity decreases.[42-44] Not known, however, is whether deeper stages of sleep are intrinsically more stable in terms of respiratory control or if breathing stability allows sleep to deepen.

Ventilatory Response to Arousal

In much the same way in which rapid changes in respiratory control occur during sleep onset, arousal from sleep causes a rapid change in the homeostatic control of breathing. As highlighted, during stable sleep, lower levels of minute ventilation and higher levels of CO_2, compared with wakefulness (~3 to 8 mm Hg higher), can be tolerated. With arousal, the wakefulness chemical control of breathing is reinstated, and the increased levels of CO_2 that were tolerated during sleep suddenly become excessive. Upper airway motoneurons are activated, and sleep-related upper airway resistance is rapidly resolved.[45] The wakefulness drive to breath also is reintroduced. Accordingly, arousal from sleep is associated with a rapid increase in breathing. The magnitude of the ventilatory response to arousal is dependent on the integrative effects of the various aforementioned factors and may be further augmented by an independent wakefulness reflex.[46] Indeed, this ventilatory arousal response varies substantially among subjects.[47] As outlined next, on the resumption of sleep, the previous ventilatory response to arousal can drive $Paco_2$ levels

below a critical level known as the apnea threshold[48] (see also Figure 16-4, *B*).

APNEA THRESHOLD

Multiple compensatory mechanisms act to oppose breathing cessation even with quite major reductions in $Paco_2$ during wakefulness (see Figure 16-3). During sleep, however, this is not the case. Specifically, if $Paco_2$ falls below a critical level during sleep, breathing ceases. The apnea threshold ranges between 2 and 6 mm Hg below the stable sleep $Paco_2$ level. Evidently, then, the apnea threshold is similar to the wakefulness $Paco_2$ level[49,50] (see Dempsey[51] for details). The difference between the wakefulness $Paco_2$ level and the apnea threshold often is termed the CO_2 reserve. The reduction in $Paco_2$ required to cause apnea is importantly dependent on the peripheral chemoreceptors.[52] Schematic examples outlining important state-related changes in the control of breathing are displayed in Figure 16-4.

Loop Gain

As outlined in this chapter, many inputs contribute to the control of breathing. Loop gain is one approach to conceptualize and quantify the overall sensitivity of the ventilatory control system (see also Chapters 15, 17, and 18 for respiration in high altitude). Specifically, the gain of the ventilatory control feedback loop can be quantified as the ratio of a ventilatory response to a ventilatory disturbance.[11,26,53] Loop gain has three major components: (1) *plant gain* (the efficiency of breathing to remove CO_2, which is determined by the properties of the lungs, blood, and body tissues), (2) *mixing and circulation delays* (the time required for a change in alveolar CO_2 to mix with the blood in the heart and the arteries before reaching the chemoreceptors), and (3) *controller gain* (the sensitivity of the chemoreceptors). Because CO_2 is the predominant modifier of ventilatory control during sleep, determining the loop gain during sleep provides important insight into the overall sensitivity of the ventilatory control system and allows for comparisons to be made between individual subjects and patient groups. Accordingly, techniques have been developed to quantify the steady state loop gain during sleep.[54,55] If certain elements that contribute to loop gain are abnormal (e.g., plant or controller gain), breathing instability can occur. Circulation delay is an integral component of breathing instability; without it, cyclic breathing would not occur. Of note, however, is that although increasing circulation delay increases the length and duration of breathing instability, increased circulation delay alone does not cause breathing instability.

Sex Differences

Sleep-disordered breathing in adults is more common in men than in women. Respiratory control differences between the sexes may contribute to this difference, at least in part. Progesterone is a respiratory stimulant, and sleep-disordered breathing is more common in women after menopause. Although ventilatory responses to CO_2 and hypoxia vary throughout the menstrual cycle, ventilatory responses during sleep to chemical stimuli do not appear to be systematically different between the sexes.[56,57] Consistent with these earlier observations, overall steady state loop gain is not different between men and women.[58,59] In accordance with increased vulnerability to breathing instability, however, important differences in breathing during sleep onset, the ventilatory response to arousal, and the apnea threshold have been observed between men and women.[18,60-62] Whether or not men have systematically lower arousal thresholds remains unclear.

CLINICAL MANIFESTATIONS

Altered respiratory control can contribute to various forms of sleep-disordered breathing. Many causes of abnormal respiratory control have been recognized. These topics are covered elsewhere in this book (Chapters 15, 17, and 24 and Section 14, Sleep Breathing Disorders) and have been the focus of comprehensive reviews.[11-13,63,64] Briefly stated, an abnormality in one or more of the components that importantly contribute to respiratory control as outlined in this chapter can cause breathing instability during sleep. Damage to central respiratory control centers or drugs that impair its function (e.g., certain brain tumors, Chiari type I malformation, morphine) can directly affect central respiratory control.[11-13] Congenital central hypoventilation syndrome is associated with major loss of chemosensitive neurons within the retrotrapezoid nucleus.[12] Heart failure is associated with heightened peripheral chemosensitivity and increased vulnerability to onset of apnea (i.e., crossing the apnea threshold). Conversely, patients with obesity-hypoventilation syndrome have blunted ventilatory responses to chemical stimuli and experience sustained hypoventilation and major blood gas disturbances during sleep.

As indicated by these findings, high and low loop gain can be problematic and can contribute to both obstructive and central breathing instability during sleep.[63] Indeed, approximately one third of patients who experience obstructive sleep apnea demonstrate abnormally high loop gain, which is likely to be an important contributor to the pathogenesis of their obstructive apnea.[40]

CLINICAL PEARL

Sleep is a particularly vulnerable time for respiratory control instability. Many of the potential compensatory inputs to breathing are markedly diminished or absent during sleep. Accordingly, regardless of the underlying cause, abnormality in one or more of the important contributors to respiratory control can cause sleep-disordered breathing. The sleep-related breathing instability that ensues is dependent on the extent to which the respiratory control system is altered and on which of the components of the respiratory control system are involved.

SUMMARY

An understanding of the control of ventilation provides important insight into the causes of various forms of sleep-disordered breathing. Ventilatory control is regulated by means of highly effective feedforward and feedback mechanisms that control blood gas levels within relatively narrow limits to maintain homeostasis. Many inputs have been recognized to regulate ventilatory control. Although these processes are predominantly under autonomic control,

voluntary modulation of breathing also is possible in various circumstances.

The dorsal, ventral, and pontine respiratory groups are key regions within the medulla and pons responsible for central respiratory control. Central (e.g., retrotrapezoid nucleus) and peripheral (e.g., carotid bodies) chemoreceptors provide essential sensory information to modify breathing. Other sensory systems also can provide input to alter the rate and depth of breathing. Most such systems, however, are either downregulated or absent during sleep. Accordingly, the chemical control of breathing—in particular, by CO_2—is the dominant input to ventilatory control during sleep. Sleep onset is particularly destabilizing to ventilatory control. Arousal from sleep and high loop gain can lead to marked fluctuations in CO_2 and to breathing cession during sleep if the apnea threshold is crossed. Abnormalities in one or more of the components that contribute to ventilatory control can contribute to both central and obstructive breathing events during sleep.

ACKNOWLEDGMENT

DJE and JEB are supported by the National Health and Medical Research Council of Australia.

Selected Readings

Burke PG, Kanbar R, Basting TM, et al. State-dependent control of breathing by the retrotrapezoid nucleus. *J Physiol* 2015.

Carberry JC, Hensen H, Fisher LP, et al. Mechanisms contributing to the response of upper-airway muscles to changes in airway pressure. *J Appl Physiol* 2015;**118**:1221–8.

Dempsey JA, Smith CA. Pathophysiology of human ventilatory control. *Eur Respir J* 2014;**44**:495–512.

Dempsey JA, Smith CA, Blain GM, et al. Role of central/peripheral chemoreceptors and their interdependence in the pathophysiology of sleep apnea. *Adv Exp Med Biol* 2012;**758**:343–9.

Eckert DJ, Jordan AS, Merchia P, Malhotra A. Central sleep apnea: pathophysiology and treatment. *Chest* 2007;**131**:595–607.

Eckert DJ, Roca D, Yim-Yeh S, Malhotra A. Control of breathing. In: Kryger M, editor. *Atlas of clinical sleep medicine*, vol. 2. 2nd ed. Philadelphia: Saunders; 2014. p. 45–52.

Guyenet PG, Stornetta RL, Bayliss DA. Central respiratory chemoreception. *J Comp Neurol* 2010;**518**:3883–906.

Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;**3**:141–63.

Kehlmann GB, Eckert DJ. Central sleep apnea due to a medical condition not Cheyne Stokes. In: Kushida CA, editor. *Encyclopedia of sleep*, vol. 1. 1st ed. San Diego: Elsevier; 2013. p. 244–52.

Khoo MC, Kronauer RE, Strohl KP, Slutsky AS. Factors inducing periodic breathing in humans: a general model. *J Appl Physiol* 1982;**53**:644–59.

Kumar P, Prabhakar NR. Peripheral chemoreceptors: function and plasticity of the carotid body. *Compr Physiol* 2012;**2**:141–219.

A complete reference list can be found online at ExpertConsult.com.

Chapter
17

Physiology of Upper and Lower Airways

Raphael Heinzer; Frédéric Sériès

Chapter Highlights

- Sleep has an impact on ventilation and gas exchanges mediated through an increase in airway resistance and a decrease in lung volume and thoracopulmonary compliance.
- Upper airway stability can be altered during sleep because of its effects on upper airway muscle control and chest mechanics.
- When upper airway anatomy is compromised, these sleep-related effects can trigger obstructive disordered breathing.

This chapter focuses on the physiologic determinants of respiration used to estimate breathing function in normal persons. The ultimate goal is to allow readers who are not familiar with the field of respiratory medicine to benefit from the more subtle concepts that will help them to better understand sleep-disordered breathing. Chapters 15 and 16 describe the basis of the physiology and applied concepts of respiratory function during wake and sleep.

The dual aim of breathing is to provide oxygen to the different body parts and to eliminate carbon dioxide resulting from cell metabolism. This is achieved through continuous gas exchange between inspired and exhaled air and the blood in the pulmonary circulation. After blood coming from the right side of the heart has been loaded with oxygen, it passes to the left side of the heart, which sends it to every part of the body through the arterial system. The different organs then take up oxygen from the arterial blood and remove carbon dioxide. Blood loaded with carbon dioxide travels through the venous system to reach the pulmonary circulation, where carbon dioxide passively diffuses through the alveolocapillary membrane into the airway, whence it is exhaled. Survival depends on the integrity of this physiologic process, and death can occur if respiratory function stops for more than a few minutes. Maintenance of normal arterial blood gases involves several physiologic systems—control of breathing, thoracopulmonary mechanics, circulatory components, and blood transport—that are intimately linked to one another. This chapter considers only the mechanical properties of the chest and the upper and lower airways, which influence ventilation during sleep (Box 17-1).

ANATOMY AND PHYSIOLOGY

The upper airway, which includes the nasal cavities, pharynx, and larynx, serves to moisten and warm the air and conduct it to the trachea and lungs. Upper airway muscles also are involved in phonation and swallowing. A very subtle regulation of vocal cord tension also allows humans to speak and sing during exhalation. It is hypothesized that the evolution of speech, which requires substantial mobility of the pharynx, led to a loss of the rigid support of the upper airway, which makes it more collapsible in humans than in most mammals.

Breathing is possible through either the nose or the mouth, but nasal breathing is the physiologic breathing route. The lower airway includes the trachea and the lungs (bronchi and alveoli). Thin blood vessels, the capillaries lining the alveoli, allow gas exchange between inspired air and blood. The rib cage provides protection for the lungs and also allows them to change volume from a minimum of approximately 1.5 L to a maximum of 6 to 8 L, depending on the height and sex of the person.[1]

The ribs articulate with the transverse processes of the thoracic vertebrae and have flexible anterior cartilaginous connections with the sternum. The lungs are covered by thin visceral pleura. The inner aspects of each hemithorax are lined with parietal pleura. The virtual space between the visceral and the parietal pleura contains a few milliliters of lubricating fluid, which allows these layers to slide against each other easily during ventilation. Owing to its proximity to the pleural tissue, the esophageal pressure varies in parallel with the changes in pleural pressure and often is used to quantify respiratory efforts.

Respiratory Muscles

The diaphragm is the main muscle of respiration. It is a dome-shaped muscle that separates the thoracic and abdominal cavities. The diaphragm is innervated by the phrenic nerves. During inspiration, the neural outflow coming from the central respiratory centers leads to diaphragm contraction; the shortening of those muscle fibers flattens the diaphragm, with consequent loss of its dome shape, thereby increasing intrathoracic volume. Intercostal muscles also can increase the intrathoracic volume by elevating ribs and increasing the anteroposterior diameter of the thorax (Figure 17-1). Accessory breathing muscles such as the scalene or sternocleidomastoid are not active during normal breathing, but they can be recruited during an effort or in the presence of thoracopulmonary disorders.

Box 17-1 SOME DEFINITIONS USED IN RESPIRATORY MECHANICS

Chest or Lung Compliance

Change in volume per change in pressure: $\Delta V/\Delta P$

Minute Ventilation

Tidal volume times respiratory rate: $Vt \times RR$

Laminar Flow

Change in pressure per resistance: $\Delta P/R$

Turbulent Flow

Pressure drop along the airway is proportional to flow and its square values: $\Delta P \propto aV + bV^2$ (V is air flow, and a and b are constants)

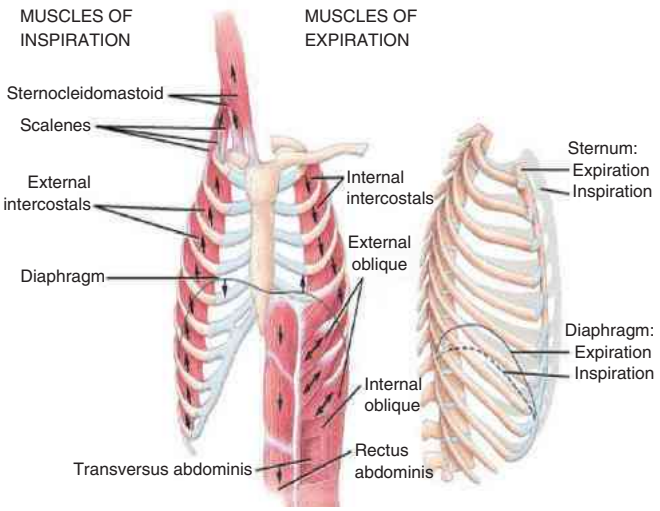


Figure 17-1 Drawing of inspiratory and expiratory muscles from abdomen to neck. The main inspiratory muscles include the diaphragm and external intercostal muscles. Accessory inspiratory muscles include the scalene and sternocleidomastoid muscles. Expiration usually is a passive process. However, internal intercostals and abdominal muscles are recruited during forced expiration. (Reproduced from Netter FH. *Atlas of human anatomy.* Philadelphia: Saunders; 2006.)

Elastic Forces and Lung Volumes

An isolated lung (not surrounded by the thoracic cage) will tend to contract until it eventually collapses owing to the large amount of elastic fibers inside the lung tissue. The lung is thus submitted to a constant recoil force. By contrast, the isolated thoracic cage tends to expand to a volume approximately 1 L more than its natural, in vivo resting position. In a relaxed subject with an open airway and no airflow, the inward elastic recoil of the lungs will be balanced by the outward resting force coming from the thoracic cage. *Lung compliance* or *distensibility* is defined as the change in lung volume per unit change in transmural pressure gradient:

$$\text{compliance} = \Delta V/\Delta P$$

where V is volume and P is pressure.

The lung volume, in the natural resting end-expiratory position, is its *functional residual capacity* (FRC). *Total lung capacity* (TLC) is reached when the thoracic cage and lungs are fully expanded (maximal inspiratory effort). *Residual volume* (RV) represents the volume remaining in the lungs at the end of a forced expiration. *Vital capacity* (VC) is the maximum amount of air that can be expelled after the lungs have been fully inflated. *Tidal volume* (V_T) is the volume of air inspired or expired during each quiet breathing cycle (Figure 17-2). The typical V_T value is 500 mL, but it can dramatically increase during exercise. Only approximately two thirds of inspired air participates in oxygen and carbon dioxide exchange, because the volume corresponding to upper airway, trachea, and bronchi does not contribute to gas exchange; this area is the *dead space* (V_{DS}) of the respiratory tract.[1]

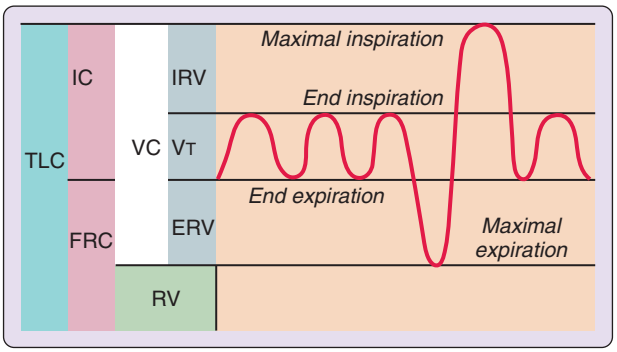


Figure 17-2 Schematic illustration of the static lung volumes determined by a spirometer in which airflow velocity does not play a role. Lung capacity is estimated by the sum of two or more lung volume subdivisions. ERV, Expiratory reserve volume; FRC, functional residual capacity; IC, inspiratory capacity; IRV, inspiratory reserve volume; RV, residual volume; TLC, total lung capacity; VC, vital capacity; V_T, tidal volume. (Reproduced with permission from American Association for Respiratory Care. AARC clinical practice guideline: static lung volumes: 2001 revision & update. *Respir Care* 2001;46:531–9.)

Breathing Cycle and Minute Ventilation

Air always flows from an area of higher pressure to one of lower pressure, to achieve equilibrium. The pressure inside the pleural space is generated by the forces developed during inspiration and expiration and is proportional to the amount of respiratory effort. The pleural pressure represents the driving pressure. During inspiration, the diaphragm and intercostal muscles contract and the pressure inside the thorax decreases below the atmospheric pressure (negative transpulmonary pressure gradient). This gradient is responsible for air movement from the nose (atmosphere) to the tracheobronchial tree down to the alveoli. During expiration, the inspiratory muscles relax, making resting expiration a passive phenomenon. However, during active expiration (volitional or during exercise), the contraction of abdominal and external intercostal muscles enhances the changes in intrathoracic pressure. This causes an abrupt increase in pleural pressure to a less-negative value, with a corresponding rise in alveolar pressure by the same amount. These changes generate a positive pressure gradient from the alveoli to the mouth, which is responsible for exhalation. Lung and chest volume decrease as air flows out, causing lung recoil pressure to fall until a new equilibrium is reached at FRC.

Respiratory rate, or breathing frequency, represents the number of breaths per minute. Average respiratory rate in a healthy adult subject at rest is approximatelyt 12 (range, 10 to

18) breaths/minute. Minute ventilation ($\dot{V}$) can be calculated using the following equation:

$$\dot{V} = V_T \times RR$$

where RR is the respiratory rate. During quiet breathing, a typical value is 6 L/minute, but the volume can rise to 180 L/minute during exercise.

Resistance

Different profiles of airflow may be observed inside the airways, depending on airway anatomy (in accordance with the specific division of the tracheobronchial tree) and mechanical properties (caliber, shape, collapsibility) of the airway structures and on the amount of driving pressure. With a constant laminar flow regimen, the resistance is directly proportional to the pressure gradient along the tube:

$$\text{flow} = \Delta P / R$$

where ΔP is the pressure difference and R is the resistance. Airflow is described as *turbulent* when the pressure drop along the airway is proportional to flow and its square values:

$$\Delta P \propto aV + bV^2$$

where ΔP is the pressure difference and V is the airflow. Airflow along airways is complex and usually consists of a mixture of laminar and turbulent flow. In normal lungs, respiratory resistance depends mainly on airway diameter. The velocity of airflow and airway diameter decrease in successive airway generations, from a maximum in the trachea to almost zero in the smallest bronchioles.

A third flow regimen is represented by flow limitation, whereby flow plateaus once the driving pressure has reached a given level. In this regimen, the flow value depends on the difference between intraluminal and extraluminal pressures, as well as on the compliance of the specific airway. Flow limitation can occur during expiration when the pressure generated by expiratory forces increases intraluminal pressure and induces an external compression of the airway walls at the same time. This pattern of flow is, however, more prone to be seen during inspiration at the level of the upper airway. Upper airway resistance depends on nasal and pharyngeal anatomy, position of the vocal cords, and lung volume (see later).

EFFECTS OF OBESITY AND BODY POSTURE ON LUNG VOLUMES

In an awake normal and healthy subject, a reduction in FRC and TLC is observed in the supine position in comparison with the upright position, both in adults[2] and in children.[3] This reduction is thought to be due to an increase in intrathoracic blood volume or to the gravitational effect of abdominal contents pushing the relaxed diaphragm into a more rostral position.[4] The change in diaphragm position reduces its ability to contract, as suggested by a decreased maximal inspiratory pressure in the supine posture relative to that in the upright and sitting positions.[5] Moreover, this restrictive defect in lung volume increases the work of breathing and deteriorates gas exchange by decreasing the ventilation-perfusion ratio in the dependent parts of the lungs. Decreased lung volume also can increase upper airway resistance by reducing the caudal traction of the mediastinum and trachea on the pharyngeal walls, making them more collapsible during inspiration (as discussed further later on).[6-9]

In obese subjects, a restrictive defect in lung volume also is observed in the sitting position. A further small decrease of 70 to 80 mL from approximately 2.4 L (for an average-sized man) in FRC and TLC occurs when obese subjects lie supine.[2] In view of the effects of abdominal volume on lung function in sitting obese subjects, a greater reduction in lung volume with adoption of the supine position compared with that in lean persons might be expected. However, a lesser decline in FRC and TLC in obese subjects in the supine position has been documented.[2,4,10] One possible explanation is that in sitting obese subjects, the diaphragm is already shifted in a more rostral position and cannot move much farther in the supine position. Two experimental studies also suggest a possible protective or adaptive mechanism against large changes in end-expiratory lung volume during wakefulness[11] and sleep.[12]

Maximal minute ventilation, expiratory reserve volume, FVC, and, to a lesser extent, forced expiratory volume in 1 second (FEV_1) also are affected by obesity.[13] The estimated reduction of FVC is 17.4 mL/kg weight gain for men and 10.6 mL/kg weight gain for women.[14] Men show more impairment of FVC with weight gain than women, possibly because of differential patterns of fat deposition: Waist circumference is negatively associated with FVC and FEV_1. On average, a 1-cm increase in waist circumference was associated with a 13-mL reduction in FVC.[15] All of these effects observed with change from the upright to the supine position and in obese persons may contribute to the exacerbation of respiratory disturbances in the presence of sleep-disordered breathing, as described in later chapters on that topic (Section 14).

EFFECTS OF SLEEP ON LUNG VOLUME

A modest but significant decrease in FRC occurs during sleep in most healthy subjects. FRC decreases by approximately 200 mL in stage 2 non–rapid eye movement (NREM) sleep and by 300 mL during slow wave sleep and rapid eye movement (REM) sleep when measured with a helium dilution technique, in comparison with normal FRC obtained with the subject awake (approximately 2.4 L for an average-sized man).[16] When plethysmography is used to measure differences in lung volume, a 440- to 500-mL decrease in lung volume has been reported in NREM sleep (stages 2 to 4), with a similar decrease in REM sleep.[17] Possible mechanisms of the decrease in FRC during sleep are rostral displacement of the diaphragm secondary to diaphragmatic hypotonia, alteration of the respiratory timing from the central generator of breathing, decrease in lung compliance, decrease in thoracic compliance, and central pooling of blood (see Chapters 15 and 16).

A reduction in tidal volume by approximately 6% to 15% has been reported during NREM sleep (stages N2 and N3), with a further decrease during REM sleep (approximately 25% lower than during wakefulness).[18,19] Minute ventilation is significantly lower during all NREM sleep stages compared with wakefulness and decreases further during REM sleep,

especially during phasic REM (approximately 84% of the level during wakefulness)[18-21] (Figure 17-3). The decrease presumably is due to a faster and shallower breathing pattern in all sleep stages with a lower tidal volume, especially during REM sleep. This explanation is, however, controversial, because another study showed no significant change in V_T between wakefulness and any sleep stage and suggested that the decrease in minute ventilation (8% in NREM and 4% in REM sleep) is due to a decrease in respiratory rate.[22] Nevertheless, most studies agree that during NREM sleep, the rib cage's contribution to V_T increases, in association with an approximately 34% increase in the activity of intercostal muscles.[21-23] There is thus an apparent contradiction between the increase in electromyogram (EMG) activity of thoracic muscles and a decrease in minute ventilation. A possible explanation is that even though muscle activity increases, the actual negative thoracic pressure decreases because of a decrease in the efficiency of muscle contraction during NREM sleep.[24] During REM sleep, the relative contribution of the rib cage and abdomen is not significantly different from that during wakefulness.[21] Age and sex do not seem to significantly alter sleep-related changes in lung volume.

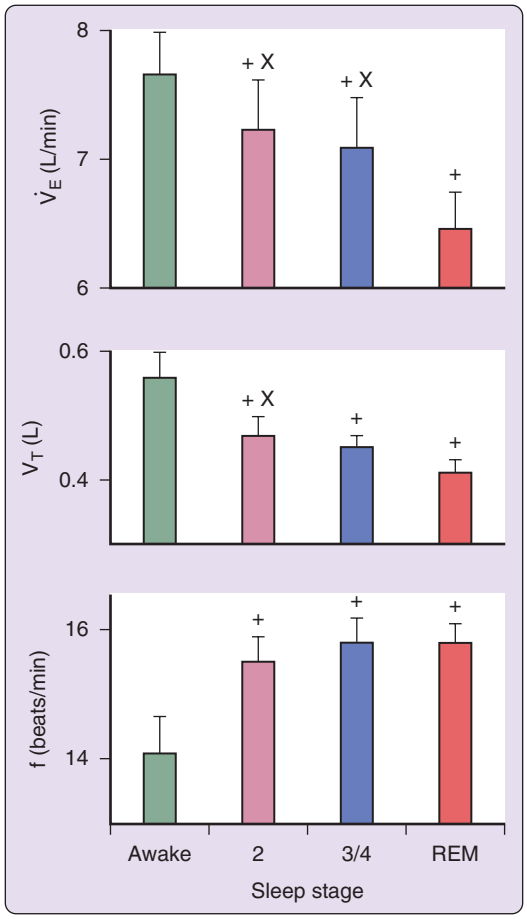


Figure 17-3 Effects of sleep on ventilation and lung volumes. Minute ventilation ($\dot{V}_E$), tidal volume (V_T), and breathing frequency (f) during wakefulness and different sleep stages are illustrated. $\dot{V}_E$ is reduced during NREM sleep, with a further reduction in REM sleep. +, $P < .05$ versus awake; X: $P < .05$ versus REM sleep. (Reproduced with permission from Douglas NJ, White DP, Pickett CK, et al. Respiration during sleep in normal man. *Thorax* 1982;37:840–4.)

EFFECTS OF SLEEP ON BREATHING PATTERN AND BLOOD GASES

During NREM sleep, the decrease in minute ventilation induces a drop in Pao_2 of 3 to 9 mm Hg and an increase in $Paco_2$ and $Paco_2$ levels ranging from 2 to 4 mm Hg.[25,26] During stable NREM sleep, the breathing pattern usually is regular. However, periodic breathing with waxing and waning ventilation commonly is observed at sleep onset (unstable NREM sleep).[27,28] Complete cessation of breathing for more than 10 seconds with respiratory effort (obstructive sleep apnea) or without respiratory effort (central sleep apnea) can even occur at this time in healthy persons. In these circumstances, the transient periodic breathing seems to be due to an unstable ventilatory feedback loop (loop gain). A low arousal threshold during this stage also can induce instability in the sleep-wake cycle and contribute to unstable breathing. Because of the higher CO_2 set point during sleep, arousals are associated with a sudden increase in ventilation, which will then decrease CO_2 level. If the CO_2 level is below the apnea threshold (below which the central respiratory drive is abolished) when sleep resumes, an apneic interval can occur and breathing will resume only when the CO_2 level again reaches the sleep set point. The magnitude and the breathing fluctuation depend on several factors such as chemoreceptor sensitivity (controller gain), lung-to-chemoreceptor circulation delay, and the efficiency of the respiratory system in inducing changes in CO_2 level (plant gain) (see also Chapter 16).[29] The relative effects of each loop gain component can be evaluated using a validated model.[30]

During REM sleep, ventilation is notably variable in both amplitude and frequency. This heterogeneity seems to be directly related to the intensity of phasic activity, as indicated by bursts of eye movements. Specifically, phasic REM activity, characterized by a high density of rapid eye movements and muscle twitches, seems to have an inhibitory influence on ventilation.[19] Overall alveolar ventilation tends to fall by approximately 20% compared with wakefulness, mainly because of a fall in tidal volume.[21]

Upper Airway

Among the mechanical determinants of ventilation just summarized, the upper airway plays a unique role because its mechanical properties are dramatically affected by sleep.

The airway can be divided into intrathoracic and extrathoracic components. These include the upper part of the trachea, the larynx, and the different pharyngeal (nasopharynx, velopharynx, orophrarynx, hypopharynx) segments. The upper airway corresponds to the pharyngeal and laryngeal structures. The airway should remain open throughout the respiratory cycle. The intrathoracic, tracheal, and laryngeal airway structures are supported by cartilaginous structures that prevent them from collapsing during tidal breathing in normal persons. Pharyngeal airways do not have such rigid support and are prone to close in conditions of imbalance between the forces that tend to dilate or close them.

From a mechanical standpoint, the upper airway behaves as a Starling resistor, in which the pharyngeal airway represents the collapsible segment and is situated between two noncollapsible structures (larynx and nasopharynx). The flow

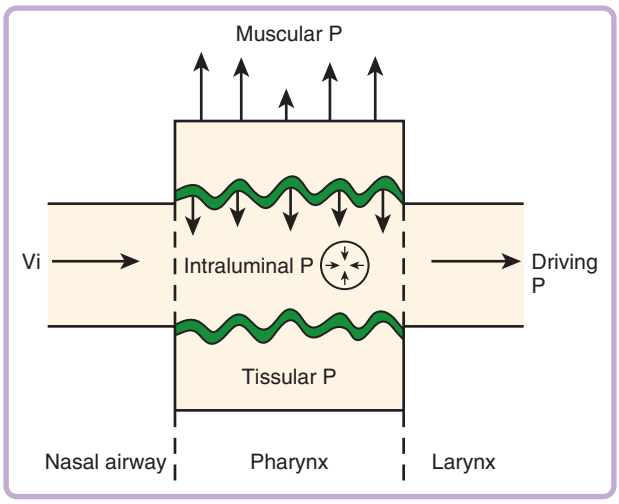


Figure 17-4 Schematic representation of the upper airway (UA) and of the forces applied to the pharyngeal airway. The muscular pressure represents the dilating force coming from the tonic and phasic activity of UA dilator muscles. The intraluminal pressure and the tissue pressure both tend to occlude the UA. P, Pressure; Vi, inspiratory flow volume.

pattern depends on the forces applied inside and outside the collapsible segment. The transmural pressure gradient is the net pressure difference between all of these opposite forces. The collapsing forces are represented by the negative inspiratory transmural pressure gradient and the pressure applied by upper airway tissue (Figure 17-4). The contraction of upper airway stabilizing muscles (upper airway dilators) is the main dilating force, the other being represented by tracheal traction (Figure 17-4). Therefore the amount and timing of the neuromuscular activation process of upper airway stabilizing muscles and the mechanical properties of upper airway tissues play a pivotal role in determining upper airway stability.

According to the Starling resistance model,[31] inspiratory flow increases with rising inspiratory efforts (driving pressure) up to a maximal value and then plateaus independently of respiratory efforts (Figure 17-5, *A*). These features of the flow-pressure relationship characterize a flow limitation regimen. The steepness of the initial rise in flow depends on the resistance upstream and downstream of the collapsing site. The pressure at which flow begins to plateau depends on upper airway mechanical properties. The critical pressure (Pcrit) represents the pressure at which the dilating forces cannot overcome the collapsing ones, leading to upper airway closure. The changes in maximal inspiratory flow with modifying upstream pressure can be used to determine Pcrit and resistance upstream to the collapsing site. A linear positive relationship between these variables can be shown (Figure 17-5, *B*). The slope of the relationship corresponds to the reciprocal of upstream resistance, and the pressure at which flow is zero represents Pcrit. In a given subject, an increase in the propensity for the upper airway to occlude translates the flow-pressure relationship to the right, without changes in slope (i.e., slope a to slope b), making the Pcrit value more positive. In the situation of a decrease in upstream resistance, the steepness of the slope will rise (greater changes in flow occur with changing upstream pressure), but the Pcrit value will remain unchanged (i.e., slope a to slope c on Figure 17-5, *B*).

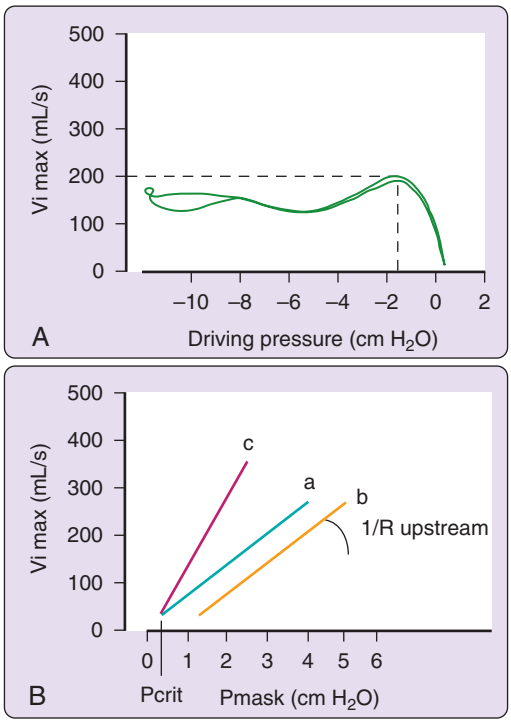


Figure 17-5 **A,** The *upper panel* is a representative example of the relationship between the respiratory flow over the driving pressure during a flow-limited breath. The instantaneous flow value reaches a maximum value and then plateaus despite the continuous decrease in driving pressure. **B,** Typical relationship between flow and upstream pressure during a series of flow-limited breaths. An increase in the instability of the upper airway will be accompanied by a right shift of the flow-pressure curve (slope b). A decrease in upstream resistance will increase the slope of the flow-pressure curve (slope c). Pcrit, Critical pressure; Pmask, mask pressure; 1/R upstream, reciprocal of upstream resistance; Vimax, maximal inspiratory flow.

Collapsing Forces

The negative intrathoracic pressure generated by diaphragmatic contraction is transmitted to the whole airway, from the alveoli to the nose, to create inspiratory flow. At the pharyngeal level, the difference between intraluminal and peritissue pressures (transmural pressure) represents a suction force that tends to dynamically close the upper airway. According to the Bernoulli principle, the pressure along the walls of a tube drops with the increase in its velocity, making the intraluminal pressure decrease (become more negative) with increasing inspiratory flow. Changes in flow from a laminar to a turbulent pattern increase air velocity near airway walls, which will further reduce intraluminal pressure.

The weight of upper airway tissue significantly influences upper airway stability. In animals, upper airway critical pressure increases proportionally to the weight applied to the hyoid arch.[32] This correlation could account for the fact that positive pressure needs to be applied to open the upper airway during anesthesia with paralysis in patients with sleep apnea,[33] who are known to have large amounts of muscular and adipose tissue surrounding the upper airway.[34] See Chapter 148 for more information on anaesthesia effects in sleep apnea patients. On the other hand, negative pressure applied around the neck significantly unloads the upper airway,[35] and resection of upper airway tissue improves Pcrit in patients with sleep apnea.[36]

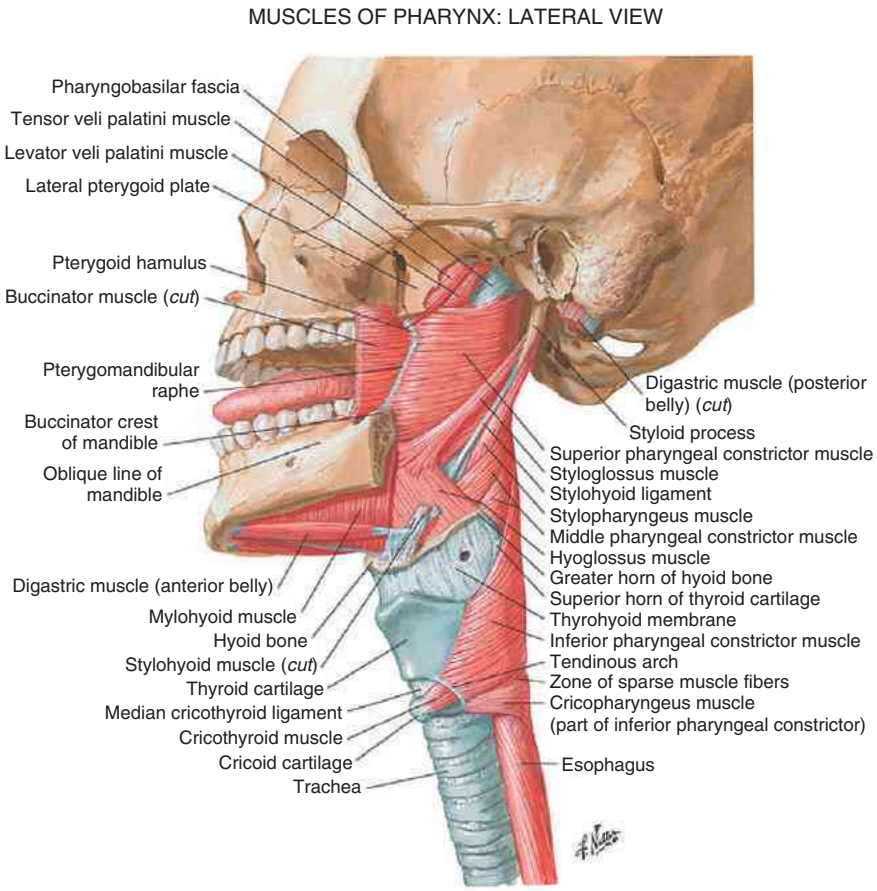


Figure 17-6 Drawing of upper airway muscles. (Reproduced from Netter FH. *Atlas of human anatomy.* Philadelphia: Saunders; 2006.)

Dilating Forces

Contraction of inspiratory muscles—diaphragm, intercostals, and accessory muscles—leads to lung inflation. The downward movement of the diaphragm produces a longitudinal traction of the bronchi and of the trachea. This traction is transmitted to the upper airway, where it contributes to unloading of that region.[37] From a dynamic perspective, tracheal traction improves upper airway stability by unfolding upper airway soft tissue and by decreasing extraluminal airway pressure.[38,39]

Numerous upper airway stabilizing muscles (such as the genioglossus, levator palatini, tensor palatini, geniohyoid, musculus uvulae, and palatopharyngeus) contribute to the maintenance of upper airway patency (Figure 17-6). Activation of masseter and pterygoid muscles also may contribute to stabilizing the upper airway by their influence on the position of the mouth and the mandible.[40] The activation profile of the upper airway muscles is characterized by their tonic activity and the respiratory-related and afferent reflex–mediated phasic activities.[41] This last factor is an important determinant of activity of the upper airway muscles, the negative pressure developed inside the upper airway having a positive feedback on muscle activity through activation of tensoreceptor and mechanoreceptor pathways.[42]

Tonic activity contributes to the maintenance of the upper airway aperture, its obligatory fall during sleep leading to a reduction in upper airway volume.[43,44] Inspiratory phasic activity has an automatic component that is linked with the central respiratory activity through projections of premotor inspiratory neurons to the hypoglossal motor nucleus (as detailed in Chapter 15).[45] Neuromodulators—serotonin, norepinephrine, glutamate, thyrotropin-releasing hormone, and substance P—play a key and complex role in the activity of upper airway muscles.[46–49] In lean animals, resting tonic and phasic activities of the genioglossus muscles mainly depend on endogenous norepinephrine, rather than serotonin drive on hypoglossal motor nucleus,[50,51] but these neuromodulators have similar stimulating effects.[50]

However, the influence of the serotonin drive on upper airway stabilizing muscle activity may be enhanced if upper airway patency is compromised, as demonstrated by the detrimental effects of serotonin antagonists (ritanserine) on upper airway caliber and stability, and on the occurrence on breathing abnormalities in animal models of obstructive sleep apnea.[52,53] Such changes in the balance of the norepinephrine-serotonin drive could result from facilitating hypoglossal nerve activities induced by intermittent hypoxia,[54] or from the relative vulnerability of norepinephrine and serotonin neurons to intermittent severe hypoxia.[55,56] Stimulation of peripheral chemoreceptors by intermittent hypoxia can lead to a prolonged rise in minute ventilation (long-term facilitation)[57,58] and a decrease in upper airway resistance (see also Chapter 16).[59,60] These ventilatory and upper airway facilitation effects are thought to be mediated by the serotonin-driven changes in activity of the phrenic and hypoglossal nerves[54,61] through

plasticity. In humans, posthypoxia upper airway facilitation is observed during sleep in conditions of flow-limited breathing (as in snorers and persons with sleep apnea)[60,62] but is not observed during wakefulness[63-65] unless periodic desaturation is associated with hypercapnia.[66]

Apart from the influence of the extent of phasic activation of upper airway muscles, the dynamic profile of this phasic activity plays a key role in the maintenance of upper airway patency. Phasic activation of upper airway muscles precedes and reaches its peak value earlier than that of respiratory muscles.[67,68] Phasic activity and the preactivation delay increase with increasing central respiratory activity[67,69] and with decreasing upper airway pressure.[70] This activation pattern decreases upper airway resistance and prevents upper airway inspiratory collapse. The occurrence of upper airway obstruction in normal awake subjects when this preactivation of upper airway stabilizing muscles is lost (as with diaphragmatic pacing, phrenic nerve stimulation, or iron lung ventilation)[71] further supports the importance of the upper airway muscle preactivation pattern in maintaining upper airway patency. The link that exists between ventilatory and upper airway stability (see later on) could result from the common activation process of respiratory and upper airway stabilizing muscles originating from the central pattern generator that would be responsible for the fine tuning in the amplitude and activation pattern of these different muscle groups.

Another phasic component comes from the reflex activation of upper airway muscles linked with the decrease in upper airway pressure during inspiration.[70] Upper airway mechanoreceptor afferents contribute to modulation of the different components of upper airway muscle activity, as suggested by the effects of local anesthesia on tonic and phasic activities[72] and on genioglossus reflex–mediated negative pressure response.[73,74] Accordingly, modulating any of these components of the upper airway muscle activation profile can have an influence on upper airway patency[75,76] and stability.[77-80]

EFFECTS OF SLEEP ON UPPER AIRWAY MUSCLE ACTIVITY

The loss of wakefulness stimulus contributes to the sleep-induced decrease in upper airway muscle activity.[81] Tonic and phasic upper airway activities are significantly altered during sleep.[82-84] The impact of sleep on the activation profile of upper airway muscles differs among the various muscles. The tensor palatini has a tonic activity, but the genioglossus, palatoglossus, and levator palatini demonstrate phasic activities. These activity levels are higher during wakefulness, but only tensor palatini activity consistently falls at sleep onset.[85] The decrease in tensor palatini activity correlates with the sleep-induced rise in upper airway resistance, and a compensatory rise occurs in genioglossus activity.[85]

The tensor palatini and genioglossus muscles strongly differ in their response to negative airway pressure during both wakefulness and sleep,[86] with no correlation being found between tensor palatini activity and driving pressure. Even if tensor palatini and genioglossus activities are governed by different efferent motor fibers (trigeminal motor nucleus versus hypoglossal motor nucleus), both activities depend on central neuromodulator drive.[87,88] The preferential decrease in upper airway muscle tonic activity observed during sleep[89] may relate to decrease in central excitatory drive to upper airway motor nuclei stemming from the loss of the awake corticomotor-stimulating drive and from a decrease in the stimulating effects of neuromodulators.[50,51,90,91] Sleep also may compromise upper airway stability by altering the pattern of preactivation of upper airway muscles.[92] The loss of such preactivation is associated with the rise in upper airway resistance and upper airway closure. The reappearance of alpha activity on the electroencephalogram (EEG) restores the normal preactivation pattern with a parallel drop in upper airway resistance and ventilatory resumption.

The neuromuscular activation processes of upper airway and respiratory muscles are closely linked. Tidal inspiration has a facilitating effect—increase in amplitude and reduction in latency of motor response—on diaphragm bulbospinal activity that is enhanced during sleep. This can be attributed to the loss of a wakefulness-related tonic depolarization of phrenic motor neurons, with secondary unmasking of the role of the bulbospinal command on the corticomotor excitability of the diaphragm. It is not known how sleep interacts with the facilitating effect of inspiration on upper airway muscle excitability.[93] On the other hand, some evidence indicates that breathing instability during sleep may promote upper airway closure.

Obstructive breathing disorders are mainly observed during stages N1 and N2 of NREM and REM sleep, when ventilation is physiologically unstable, and rarely during slow wave sleep, when breathing amplitude and frequency are particularly regular.[94] Breathing remains unstable (periodic) after resumption of upper airway obstruction with tracheostomy in patients with obstructive sleep apnea.[95] In normal sleeping subjects, the induction of periodic breathing can lead to partial upper airway obstruction.[96] Ventilatory stimulation with CO_2 decreases the occurrence of obstructed breaths in patients afflicted with sleep apnea.[97] In patients with a moderate increase in upper airway collapsibility, the frequency of obstructive sleep-disordered breathing correlates with the degree of breathing instability.[98]

FACTORS INFLUENCING STABILIZING AND COLLAPSING FORCES

For a given amount of upper airway neuromuscular outflow, the net mechanical effect of the neuromuscular activation process depends on the mechanical effectiveness of the contraction of upper airway stabilizing muscles.[99] Such function depends on factors such as the shape and dimensions of the upper airway. In fact, the amount of phasic activity required to maintain a given upper airway cross-sectional area increases when the upper airway axis converts from a transverse to an anteroposterior orientation.[100,101] Lung volumes influence upper airway dimension, as demonstrated by the decrease in pharyngeal cross-sectional area and the increase in upper airway resistance and collapsibility when lung volume decreases from TLC to residual volume.[102-104] Upper airway dimension also varies throughout the respiratory cycle, being maximal at the beginning of expiration and minimal at end expiration.[105] Vascular tone also interacts with upper airway collapsibility through its effect on upper airway dimension; the decrease in vascular tone or increase in vascular content decreases upper airway caliber but not upper airway collapsibility.[106] In these physiologic situations,

various factors as described can interact with upper airway patency to favor obstruction of the upper airway if upper airway stability is already compromised (i.e., with a highly compliant upper airway).

The mechanical conditions that prevail during muscle contraction also determine the force the involved muscles can develop. The suctioning effect of negative intraluminal pressure can result in a lengthening of upper airway muscles during inspiration (eccentric contraction)[107] that interferes with their ability to dilate the upper airway and leads to upper airway muscle fatigue and structural damage.[108-110] The characteristics of the soft tissues surrounding the upper airway muscles also influence the ability of these muscles to improve upper airway patency, the increase in tissue stiffness impeding the transmission of the dilating force to the upper airway structure.[111]

CONCLUSIONS

Numerous factors are involved in the regulation of normal breathing, including a predominant role of different muscles such as respiratory and upper airway muscles as well as the mechanical conditions that determine the effectiveness of their contraction. Sleep can interfere with several determinants of normal ventilation such as ventilatory control, skeletal muscle activity, and lung volumes. Therefore, because of the influence of thoracopulmonary mechanics on upper airway patency and the close link between respiratory and upper airway muscles, sleep also has a strong impact on upper airway aperture and mechanical properties. Careful delineation of sleep-related changes in respiratory physiology is key to improving our knowledge of sleep-disordered breathing, because these principles are involved in all nocturnal breathing disturbances: hypoventilation, periodic breathing, central apnea, and upper airway closure.

CLINICAL PEARL

Numerous factors contribute to ventilation and mechanical properties of the thoracopulmonary system. Because sleep interacts with several of these factors, it has an impact on ventilation and gas exchanges through its effect on airway resistance, thoracopulmonary compliance, and lung volumes. As a consequence of its effect on upper airway muscle control and chest mechanics, sleep has a strong influence on upper airway stability. Accordingly, persons with compromised upper airway anatomy are at increased risk for development of obstructive sleep-induced disordered breathing, especially during the transition between wakefulness and sleep.

SUMMARY

The respiratory system can be divided into two compartments, the upper and lower airways. The mechanics of both compartments are strongly influenced by sleep. Lung volume, rib cage muscle activity, and minute ventilation tend to decrease during sleep, as does the activity of upper airway stabilizing muscles. The upper airway also plays a critical role in determining ventilation and breathing pattern during sleep. Its patency is influenced not only by pharyngeal and orofacial muscle activity but also by thoracopulmonary mechanics. Sleep therefore has a strong impact on upper airway aperture and mechanical properties. Even though obesity and susceptible pharyngeal anatomy are important contributors to the development of sleep-induced disordered breathing, sleep plays a key role in generating upper airway instability and therefore in determining the underlying pathophysiology.

Selected Readings

Deacon NL, Catcheside PG. The role of high loop gain induced by intermittent hypoxia in the pathophysiology of obstructive sleep apnoea. *Sleep Med Rev* 2015;**22**:3–14.

Dempsey JA, Veasey SC, Morgan BJ, O'Donnell CP. Pathophysiology of sleep apnea. *Physiol Rev* 2010;**90**:47–112.

Gederi E, Nemati S, Edwards B, et al. Model-based estimation of loop gain using spontaneous breathing: a validation study. *Respir Physiol Neurobiol* 2014;**201**:84–92.

Heinzer RC, Stanchina ML, Malhotra A, et al. Lung volume and continuous positive airway pressure requirements in obstructive sleep apnea. *Am J Respir Crit Care Med* 2005;**172**(1):114–17.

Horner RL, Hughes SW, Malhotra A. State-dependent and reflex drives to the upper airway: basic physiology with clinical implications. *J Appl Physiol* 2014;**116**:325–36.

Series F, Cormier Y, Desmeules M. Influence of passive changes of lung volume on upper airways. *J Appl Physiol* 1990;**68**(5):2159–64.

Stanchina M, Robinson K, Corrao W, et al. Clinical use of loop gain measures to determine continuous positive airway pressure efficacy in patients with complex sleep apnea. A pilot study. *Ann Am Thorac Soc* 2015;**12**(9):1351–7.

Stradling JR, Chadwick GA, Frew AJ. Changes in ventilation and its components in normal subjects during sleep. *Thorax* 1985;**40**(5):364–70.

Trinder J, Whitworth F, Kay A, Wilkin P. Respiratory instability during sleep onset. *J Appl Physiol* 1992;**73**(6):2462–9.

White DP, Younes MK. Obstructive sleep apnea. *Compr Physiol* 2012;**2**:2541–94.

A complete reference list can be found online at ExpertConsult.com.

Chapter
18

Respiratory Physiology: Sleep at High Altitudes

Philip N. Ainslie; Keith R. Burgess

Chapter Highlights

- Ventilatory acclimatization to altitude involves cellular and neurochemical reorganization in the peripheral chemoreceptors and central nervous system.
- Sleep at high altitude is disturbed by various factors including a change of sleep environment, snoring, and insomnia; periodic breathing during sleep, however, probably causes the most disturbances and occurs in a majority of people above 3500 m.
- The extent of periodic breathing during sleep at altitude intensifies with duration and severity of exposure and is explained in part by elevations in loop gain. A dimensionless value is a measure of the propensity for a system governed by feedback loops to develop unstable behavior.[1]
- Because periodic breathing may elevate rather than reduce mean arterial oxygen saturation (Sao_2) during sleep, this may represent an adaptive rather than a maladaptive response to altitude.
- Although new mechanical and pharmacologic management techniques are emerging, an oral acetazolamide regimen remains so far the most effective and practical means to reduce periodic breathing in altitude.

OVERVIEW

High altitude can disturb sleep in many ways (see also Chapter 122). Sojourners to high altitude often report restless and sleepless nights. Others describe a feeling of suffocation on awakening from sleep. Additional reported factors include physical discomfort from cold or unsatisfactory bedding and the noise of other people's snoring. At altitudes above 2500 m, a pattern of periodic breathing (as described next) often is seen, and the higher the altitude, the more common it is. Above 5000 m, it is almost universal and becomes the most common cause of sleep disturbance.[2-5]

Periodic breathing is manifested as a pattern of two to four breaths, separated by a brief interval with no breathing from the next burst of two to four breaths, which closely resembles the breathing pattern seen in the premature infant.[6] This periodic breathing during sleep was first described by Mosso[7] in 1886 (Figure 18-1), with further observations by other investigators a few decades later.[8] Of note, periodic breathing at high altitude is different from the typical waxing and waning of tidal volume observed in the periodic breathing of heart failure, or the somewhat chaotic or irregular pattern of apneas associated with opiate use (see Chapter 24 for more information on opioid effects on sleep and breathing).[9] Periodic breathing is more common in male subjects than in female subjects at 5400 m.[10]

During non–rapid eye movement (NREM) sleep, hyperventilation begins immediately on hypoxic exposure and intensifies over time.[11,12] After approximately 10 minutes of hypoxia in the sleeping human, tidal volume begins to oscillate in a waxing and waning pattern. These oscillations keep increasing in magnitude as hypoxia is maintained, and the partial pressure of arterial carbon dioxide ($Paco_2$) falls further to the level of the apneic threshold. Typically an augmented inspiration occurs, and the subject begins overt periodic breathing cycles approximately 15 to 25 seconds in duration, characterized by two to four large (e.g., three to four times higher than normal) tidal-volume breaths followed by an apneic interval of 5 to 15 seconds (Figure 18-2), as well as large swings in blood pressure that drive oscillations in cerebral blood flow (CBF) (Figure 18-3, middle and right panels). During these periodic cycles, arterial hemoglobin oxygen saturation (Sao_2) also oscillates, and often, depending on the altitude, values lie (dangerously) on the steep part of the oxygen dissociation curve. The bursts of breathing (i.e., hyperpneas) are sometimes associated with arousal from sleep and sometimes full wakefulness, which, at least at moderate altitude, may lead to fatigue during the day and cognitive impairment,[13] similar to that from other causes of sleep disruption.

ACCLIMATIZATION

As summarised perfectly by Houston and Riley in 1947, "Acclimatization to high altitude consists of a series of integrated adaptations which tend to restore the tissue oxygen pressure towards normal sea level values in spite of lowered oxygen pressure of the atmosphere."[14] This acclimatization process has two major components: ventilatory adaptation to hypoxia and the renal excretion of bicarbonate, allowing further ventilatory adaptation. Detailed reviews on this topic are available.[15,16]

In brief, acute exposure to high altitude results in the following sequence of physiologic events: (1) Initial changes consist of a decrease in the alveolar partial pressure of oxygen

(Po_2) and, correspondingly, in the partial pressure of arterial oxygen (Pao_2). (2) This decrease in oxygen tension results in stimulation of the peripheral chemoreceptors (predominantly at the carotid sinus), with a resultant increase in ventilation. (3) This initial increase in ventilation (the hypoxic ventilatory response) decreases Pco_2 and increases Po_2 in the alveolar gas according to the alveolar ventilation and gas equations. (4) The decrease in arterial Pco_2 (and consequent increase in arterial pH—defining conditions of respiratory alkalosis) acts to inhibit the peripheral chemoreceptor. In addition, because CO_2 is freely diffusible across the blood-brain barrier, a decrease in cerebrospinal fluid CO_2 occurs, thereby raising cerebrospinal fluid and brain extracellular fluid pH, causing inhibition at the central chemoreceptors. (5) Finally, both of these effects act to return ventilation back toward sea level values. Over a period of hours to days at high altitude, however, the body compensates for the respiratory alkalosis by increasing bicarbonate excretion in the kidney and increasing bicarbonate removal from the extracellular fluid by the choroid plexus. (6) Thus the inhibition at the central and peripheral

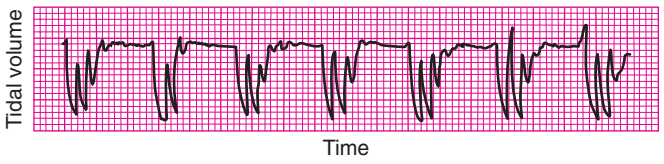


Figure 18-1 Periodic breathing during sleep in the Regina Margherita Hut (at 4559 m in the Italian Alps), as described by Mosso in 1898.[7]

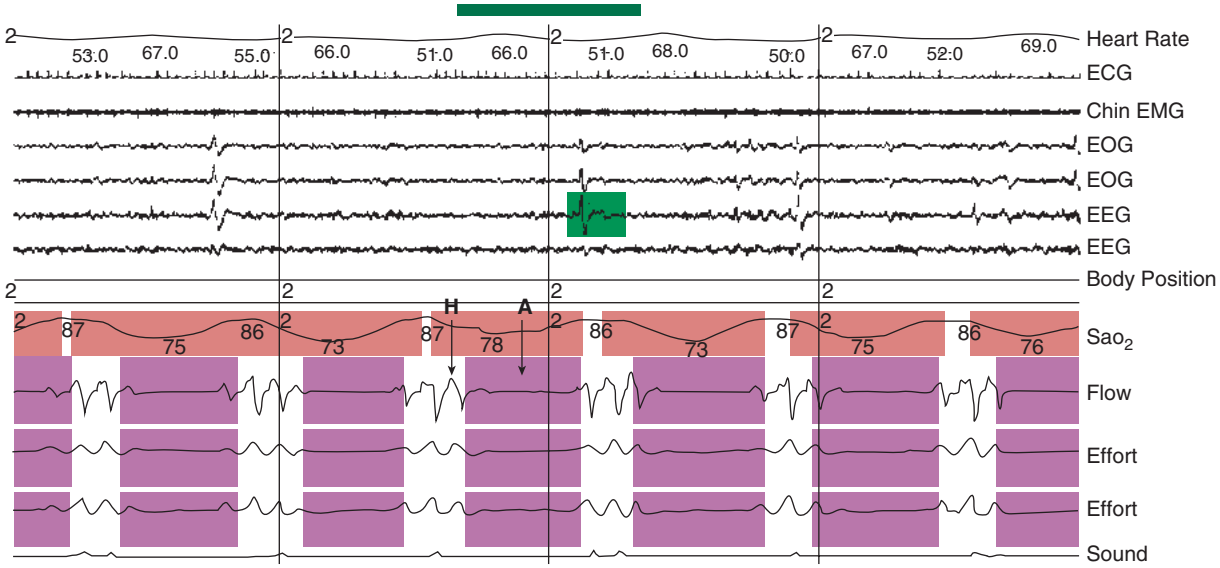


Figure 18-2 A 2-minute epoch from a polysomnogram recorded from one subject during sleep at 5050 m showing central sleep apnea (CSA). *Arrows*: H indicates the period of hyperpnea, and A, the period of apnea. Arterial oxygen saturation (Sao_2) reading shows periods of desaturation. Nasal airflow was measured using a pressure-transduced nasal cannula. Respiratory effort was measured by means of piezoelectric bands. ECG, Electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram.

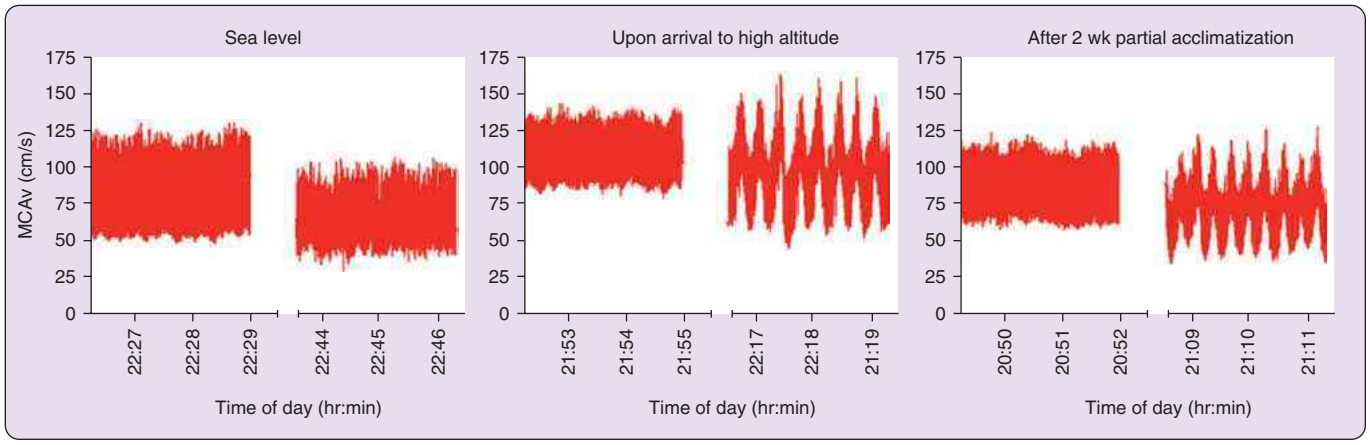


Figure 18-3 A typical profile of the observed changes in cerebral blood flow, as indexed by middle cerebral artery blood velocity (MCAv), before sleep onset (*left-hand trace*) and during stage 2 sleep (*right-hand trace*) at sea level, then on arrival, and after 2 weeks at high altitude, as recorded in one participant. Note the elevation in MCAv on arrival, compared with that after 2 weeks of acclimatization. (Modified from Burgess KR, Lucas SJ, Shepherd KL, et al. Worsening of central sleep apnea at high altitude—a role for cerebrovascular function. *J Appl Physiol* 2013;114:1021–8.)

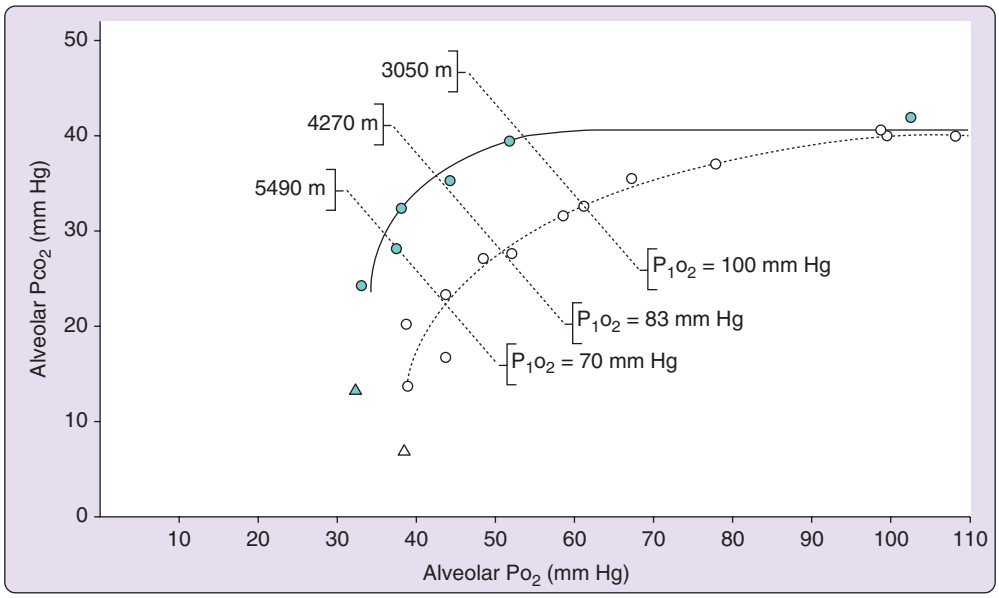


Figure 18-4 Effect of altitude acclimatization on alveolar gas composition. *Blue symbols* represent unacclimatized data. *Open* (*white*) *symbols* represent acclimatized data. Iso-altitude lines were determined using the ideal alveolar gas equation and an assumed respiratory exchange ratio of 0.85. (Data from Rahn H, Otis AB. Man's respiratory response during and after acclimatization to high altitude. *Am J Physiol* 1949;157:445–559; West JB, Hackett PH, Maret KH, et al. Pulmonary gas exchange on the summit of Mount Everest. *J Appl Physiol* 1983;55:678–87; Malconian MK, Rock PB, Reeves JT, et al. Operation Everest II: gas tensions in expired air and arterial blood at extreme altitude. *Aviat Space Environ Med* 1993;64:37–42; and Wagner PD, Sutton JR, Reeves JT, et al. Operation Everest II: pulmonary gas exchange during a simulated ascent of Mt. Everest. *J Appl Physiol* 1987;63:2348–59.)

chemoreceptors is removed and the ventilation once again increases. A direct influence of hypoxia on the central nervous system also may act to drive these progressive elevations in ventilation (as recently reviewed[16]). As shown in Figure 18-4, acclimatization at high altitude is reflected in reductions in $Paco_2$ and an increase in Pao_2. Although these changes tend to mitigate the deleterious effects of the hypoxic environment, it should be noted that restoration of Pao_2 back to sea level values can never occur.

SLEEP ARCHITECTURE

Sleep architecture has an effect on breathing at high altitude. The effects of altitude on sleep architecture were first reported in 1970 by Joern,[17] who studied only two men in Antarctica at an altitude of approximately 3500 m. Since then, more than 20 studies have been published, which investigated sleep architecture at high altitude. The "sample size" typically has been very small, with only a few subjects each, but two somewhat larger studies included 15 and 19 subjects.[18,19] (A review of the smaller studies is provided in the 19-subject report.[19]) Most of the studies found that duration of light sleep (stage 1 NREM) increased and the duration of slow wave sleep decreased with increasing altitude. The effects on REM sleep were variable. The study with 19 subjects reported no significant change in the percentage of REM sleep with increasing altitude.[19] No differences were found in the percentages of time in slow wave sleep and REM sleep between subjects with periodic breathing and those without. Sleep arousal indices were higher in subjects with periodic breathing than in those without, probably explaining the associated poorer subjective sleep quality. Another observation was that after a cortical arousal, episodes of central sleep apnea (CSA), designated "postarousal centrals," may occur in any sleep stage. These events appear to be more common at high altitude than at sea level, even in subjects in whom sustained CSA did not develop, presumably owing to relatively greater frequency of sleep arousal–related transient hyperventilation. Because sustained CSA occurs only in the lighter sleep stages (stage 1 and stage 2 NREM), the reduced duration of slow wave sleep and increased amount of stage 1 NREM sleep at high altitudes facilitate the onset of CSA.

MECHANISMS CAUSING PERIODIC BREATHING

The mechanisms causing periodic breathing are discussed in a general context in earlier chapters (see Chapters 15 to 17), but some altitude-specific comments are warranted. The principal reason for the occurrence of apnea and periodic breathing during sleep in hypoxic environments is believed to be elevations in controller or feedback gain, as evidenced by the steep increase in the CO_2 response slope above and below eupnea and the greatly narrowed CO_2 reserve.[20-22] These aspects that determine the CO_2 reserve below eupnea are based on mathematical modeling concepts of *plant gain* and *controller gain*.[23-25] Another important factor that has been postulated to influence ventilatory stability and thus periodic breathing is the poststimulus short-term potentiation, or what was initially called the "ventilatory afterdischarge."[26,27] Although these concepts have been described in detail with sleep apnea,[22] a brief review of likely changes experienced at high altitude and the implications for periodic breathing is presented here.

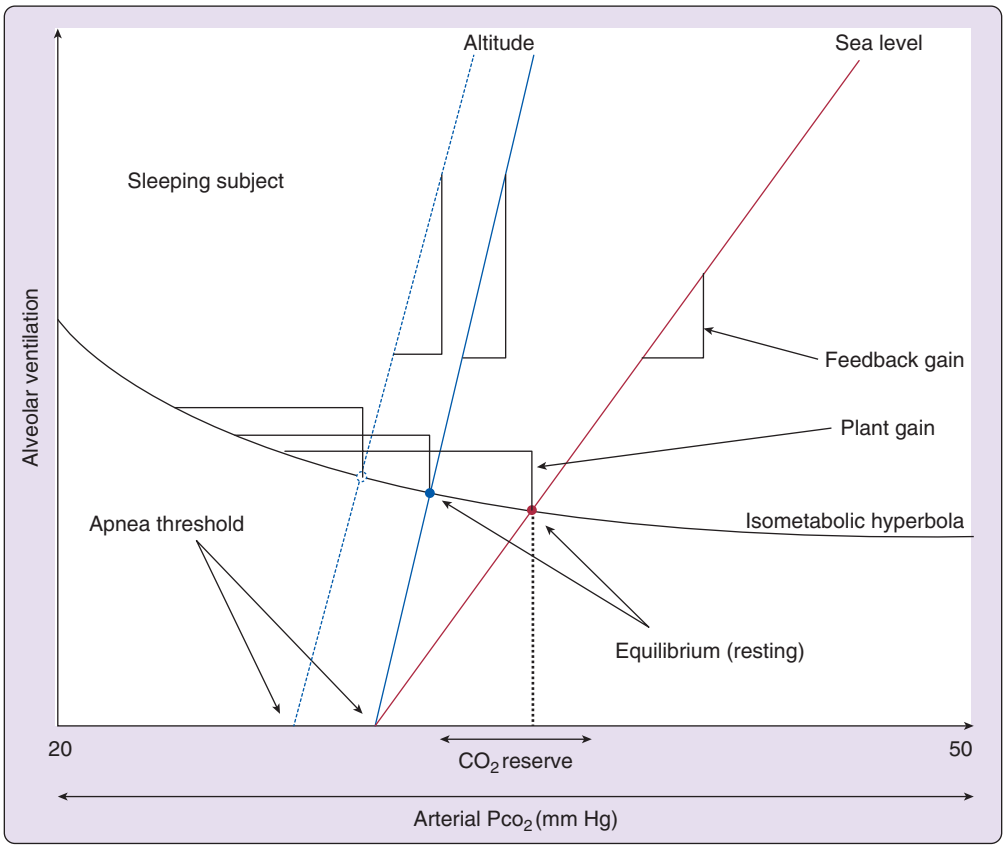


Figure 18-5 Illustration of the relationship between alveolar ventilation and alveolar P_{CO_2} at a fixed CO_2 production (e.g., 250 mL/min). Ascent to altitude increases the chemoreflex slope (*solid blue line*) but does not necessarily change the apnea threshold; the increase in slope moves the equilibrium to an increased ventilation and lower P_{CO_2}, thereby decreasing plant gain. This effect—chronic hyperventilation induced by the high altitude and subsequent reductions in plant gain—indicates that a greater transient increase in alveolar ventilation (V_A) and corresponding reduction in Pa_{CO_2} is required to reach the apneic threshold than would be the case under conditions of normocapnia. Therefore this reduction in plant gain acts to stabilize breathing. Should the apnea threshold also be decreased with acclimatization at altitude (*dotted blue line*), then ventilation increases and P_{CO_2} decreases, and plant gain is further decreased. For a given background Pa_{CO_2}, alterations in the slope of the change in V_E per change in Pa_{CO_2} relationship below eupnea would alter the CO_2 reserve (i.e., the amount of reduction in Pa_{CO_2} required to cause apnea). Changing the slope of the ventilatory response to CO_2 above eupnea would alter susceptibility for transient ventiltory overshoots. Although at altitude the chronic hyperventilation-induced hypocapnia may be "protective" against apnea and breathing instability through reductions in plant gain, the other chemoreceptor (e.g., controller gain) and nonchemoreceptor (e.g., increased pulmonary pressures, behavioral drives, awake-to-sleep transitions, locomotion feedback/forward stimuli) factors may contribute, potentially negating this response. (Modified from Ainslie PN, Lucas SJ, Burgess KR. Breathing and sleep at high altitude. *Respir Physiol Neurobiol* 2013;188:233–56.)

Plant Gain at Altitude

The term *plant gain* is defined as the effectiveness of ventilation in changing the blood gases.[28] This plant gain is determined by the intersection of the chemoreflex response and the isometabolic hyperbola that defines resting conditions of minute ventilation ($\dot{V}_E$) and $Paco_2$ (Figure 18-5). At high $Paco_2$, the equilibrium point is located on a relatively flat portion of the metabolic hyperbola, so the plant gain (i.e., $\Delta Paco_2/\Delta\dot{V}_E$) is high. With the reduced $Paco_2$ with ventilatory accclimatization to hypoxia, the equilibrium point is located on a steeper portion of the isometabolic hyperbola, so the plant gain is low, thereby protecting against instability. What this implies is that hyperventilation per se at high altitude (because of the reduction in $Paco_2$) seemingly "protects" against apnea and ventilatory instability. Therefore, although chronic hyperventilation-induced hypocapnia should theoretically be protective against apnea and breathing instability, other chemoreceptor (e.g., controller gain) and nonchemoreceptor (e.g., increased pulmonary pressures, behavioral drives, awake-to-sleep transitions, locomotion feedback/forward stimuli) factors must contribute, potentially negating this response.

Controller Gain at Altitude

The other mechanism by which the magnitude of the CO_2 reserve decreases to below eupnea is through alterations in the slope or sensitivity (see the blue lines in Figure 18-5) of the reductions in $\dot{V}_E$ below eupnea in response to transient hypocapnia. An increased CO_2 response slope below eupnea during NREM sleep occurs at high altitude.[29] In other words, an elevated chemosensitivity causes a more vigorous response to the rise in $Paco_2$ during the apneic interval (synonymous with a higher controller gain), which is sufficient to overcome the

relatively reduced baseline $Paco_2$ (indicating a lower plant gain), thereby further destabilizing ventilation.[22,30]

Short-term Potentiation and Periodic Breathing

As described previously, another factor that has been postulated to influence ventilatory stability and thus periodic breathing is poststimulus short-term potentiation, or "afterdischarge."[26,27] This short-term potentiation reflects the maintenance of ventilation after cessation of a stimulus despite hypocapnic inhibition. Of note, in a number of human studies it has been reported that short-term potentiation is reduced during hypoxic exposure in the awake state[31] and also in NREM sleep,[32] so that periodic breathing is more likely to occur under these conditions. For the development of periodic breathing during sleep at altitude, however, the combination of transient hypocapnic inhibition and sustained hypoxia presumably overrides or abolishes any meaningful influence from short-term potentiation or after discharge.[28]

Other Factors Influencing Periodic Breathing

Periodic breathing in hypoxia occurs in breath "clusters," with tidal volume increasing from zero to three to four times control levels, almost instantaneously following each apneic interval. This pattern has been suggested to reflect the presence of a transient arousal state at apnea termination that would further augment the responsiveness of the respiratory control system and produce the sudden ventilatory overshoot.[33] Another possibility that also could influence periodic breathing is a direct influence of brain hypoxia.[30] Other evidence, also based on findings in animals,[34,35] supports the notion that breathing instability may also involve pulmonary J receptors. These receptors are stimulated by pulmonary congestion/lung edema at high altitude and evoke reflex inhibition of ventilation which prolongs the apnea. Moreover, acute pulmonary hypertension (as reflected in elevations in left atrial pressure by 5.7 mm Hg in the well-controlled animal model) during sleep results in a narrowed CO_2 reserve and thus predisposes affected subjects to apnea/unstable breathing.[36] It seems likely that the periodic breathing-induced oscillations in CBF also act to destabilize breathing by provoking large swings in brain tissue pH and hence central chemoreflex stimulation and inhibition. Although clear evidence for these complex pathways at high altitude is still lacking, it is known that at sea level, periodic breathing during sleep is more pronounced in patients with pulmonary hypertension than in those without.[37]

Periodic Breathing and Hypoxic Ventilatory Response

If elevations in controller gain are the principal precipitating mechanism for periodic breathing, then, other things being equal, persons with the highest hypoxic and hypercapnia ventilatory responses should have more severe periodic breathing. Although many studies cite the classic Lahiri study[42] to provide evidence of the correlation of hypoxic ventilatory response (HVR) and periodic breathing, this relationship was largely created by the inclusion of a Sherpa group with a blunted HVR. Upon examination, no obvious relationship was found between HVR and periodic breathing within the so-called lowlander population. This absence of a relationship was further confirmed, albeit in a subgroup ($n = 5$), at 6300 and 8050 m.[4] These findings are consistent with those of Masuyama and associates: In their study, two of nine mountaineers did not develop CSA at altitude despite normal values for HVR.[39] More recently, absence of a relationship between HVR and periodic breathing at 5050 m has been reported.[5] By contrast, at 4400 m in a small sample size ($n = 4$), it was shown that the respiratory stimulant almitrine doubled the HVR and elevated periodic breathing compared with acetazolamide or placebo.[40] Nevertheless, the known blunted HVR and diminished periodic breathing in Sherpas[38] lend support to a role for HVR in periodic breathing. A number of potential explanations exist for these discrepant and variable findings, including evidence that the hypoxic and CO_2 response are not always similar above and below eupnea[22]; differences in awake versus sleep respiratory control; variable acid-base status; and methodologic differences (e.g., chemoreflex testing and hence inclusion of CBF using steady state or rebreathing methods, natural versus simulated altitude, or other means). At least on the basis of rebreathing measures in humans at high altitude,[41-43] it is not clear if actual wakefulness chemoreflex gain differs above and below resting equilibrium. Collectively, these findings highlight the multifactorial complexity of characterizing and studying periodic breathing at high altitude.

Periodic Breathing Changes with Both Magnitude and Duration of Hypoxic Stimulus

It was originally considered that the amount of periodic breathing in sleep is greatly reduced over time in normobaric hypoxia.[11,12,44] At least at high altitude, however, recent evidence derived using full polysomnography shows the opposite: that periodic breathing intensifies over time (12 to 15 days) at a given altitude, in all subjects[5] (Figure 18-6). As highlighted in Figure 18-7, it is clear that periodic breathing increases proportionally with altitude, and as illustrated in Figure 18-8, a small but progressive decrease occurs in the average duration of the apnea-hypopnea events at altitude. Because the development of CSA is almost exclusive to NREM sleep (especially during stage 1 and 2 light sleep), collectively, this information allows determination of the theoretical ceiling of CSA at high altitude. Naturally this limit would vary depending on individual differences in cycle duration and percentage of time in REM sleep. Nevertheless, on the basis of this information, it is possible to ascertain when apnea-hypopnea cycling has reached the maximal theoretical value. At this point, which may occur with acclimatization, these calculations are important, because they indicate that the development of CSA becomes independent of key factors affecting its severity (e.g., controller gain, apneic threshold, and cerebrovascular influences; see further on).

A Role of Cerebral Blood Flow in Breathing Stability at Altitude

The supportive evidence for a putative role of CBF and reactivity on breathing stability is now clear. First, pharmacologic blunting of CBF and its reactivity to CO_2 leads to elevations in controller gain, reduced CO_2 reserve, and subsequent increased susceptibility to onset of apnea and breathing instability during sleep.[45] These changes also are evident during wakefulness.[46] Moreover, acute elevations in CBF velocity and reactivity to $Paco_2$, induced by intravenous acetazolamide, have been demonstrated to be related to improvements in breathing stability at high altitude during wakefulness[47] and

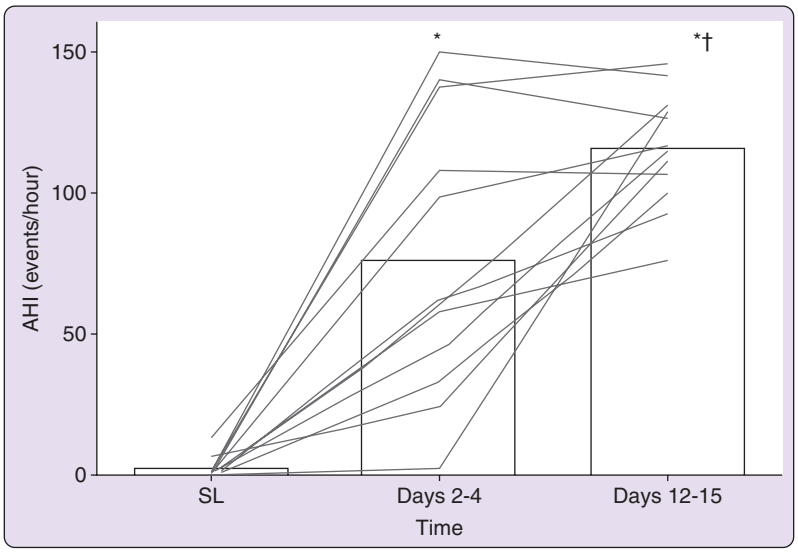


Figure 18-6 Increase in apnea-hypopnea index with time at 5050 m. *Lines* represent individual subjects. Asterisks and dagger indicate level of statistical significance, such that * = 0.05, † = 1.01. SL, Sea level. (Modified from Dempsey JA, Forster HV. Mediation of ventilatory adaptations. *Physiol Rev* 1982;62:262–346.)

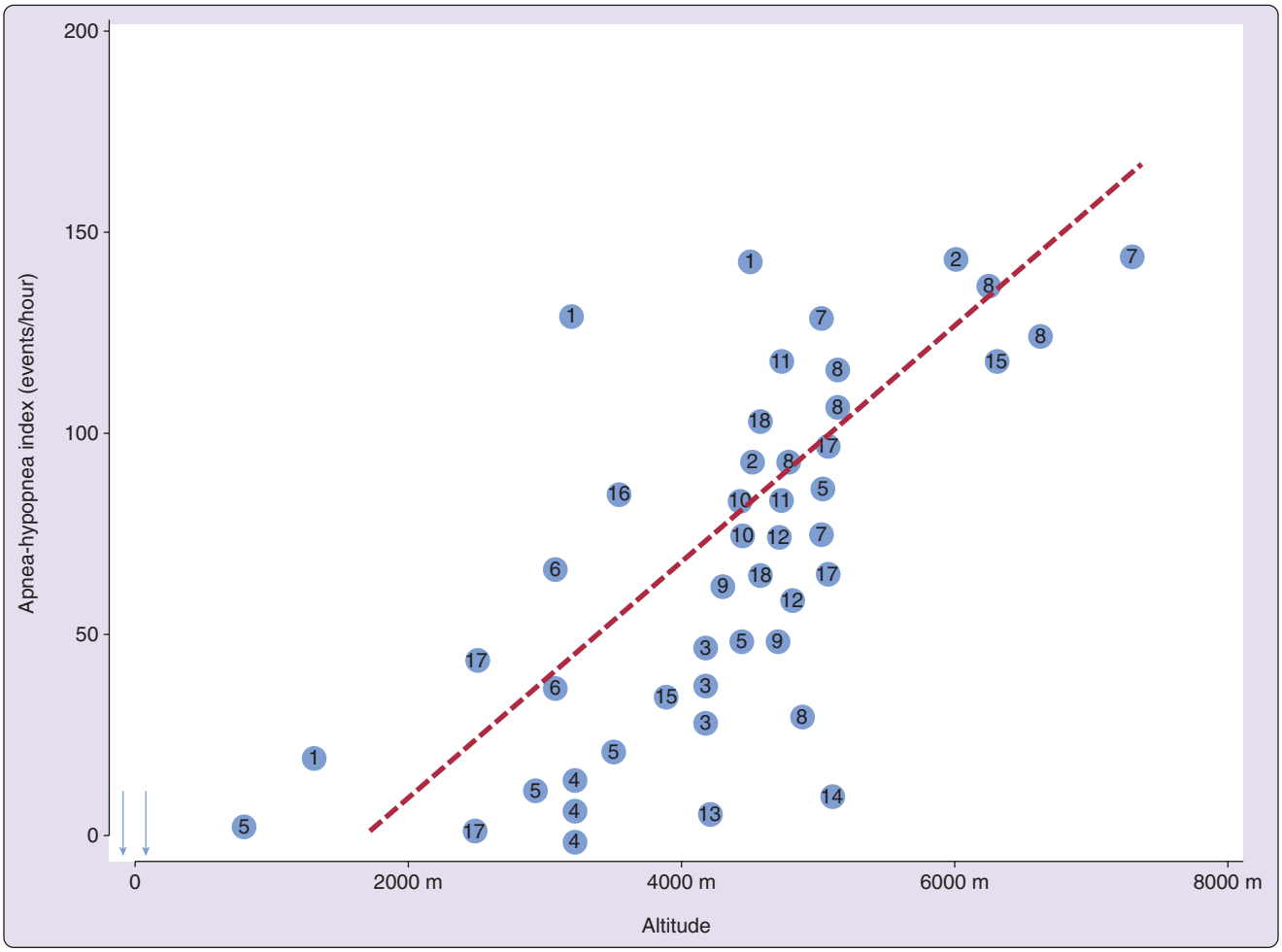


Figure 18-7 Relationship between altitude and apnea-hypopnea index. Each *number* refers to a different study, and the *red line* denotes the regression between studies. (Modified from Dempsey JA, Forster HV. Mediation of ventilatory adaptations. *Physiol Rev* 1982;62:262–346.)

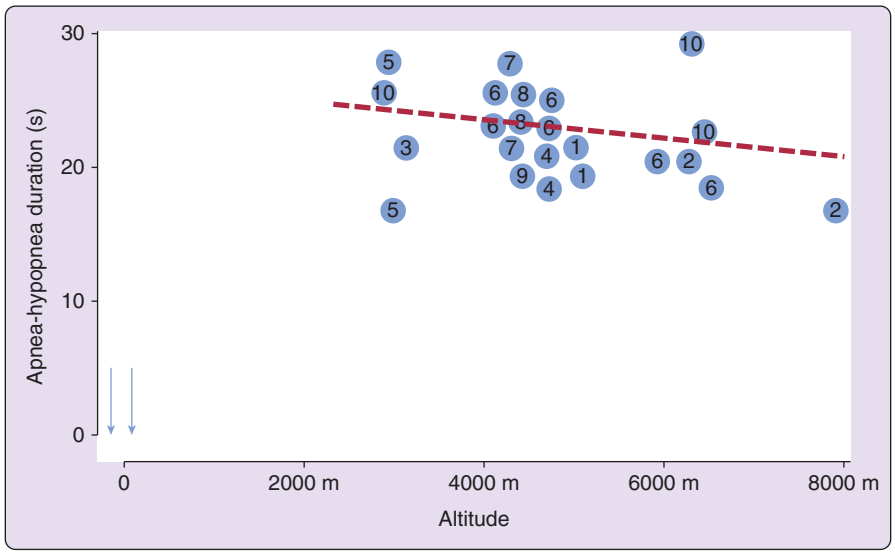


Figure 18-8 Relationship between altitude and apnea-hypopnea duration. Each *number* refers to a different study, and the *red line* denotes the regression between studies. (Modified from Dempsey JA, Forster HV. Mediation of ventilatory adaptations. *Physiol Rev* 1982;62:262–346.)

sleep.[48] In support, modeling studies have shown that theoretically, doubling of cerebrovascular reactivity to CO_2 leads to a marked dampening of respiratory oscillations in conditions of sleep at high altitude.[49] Conversely, when cerebrovascular reactivity to CO_2 was halved after a sigh transformed a stable breathing pattern into a periodic breathing pattern, restoration of reactivity restored stability.[49] Thus CBF and its related CO_2 reactivity, through its influence on central chemosensitivity, provide an important mechanism in the pathophysiology of CSA. As mentioned earlier, CBF is elevated on initial arrival to high altitude. We speculate that this elevated blood flow provides a protective effect on CSA during initial exposure to high altitude via effective buffering of changes in central P_{CO_2} and hence reductions in controller gain (i.e., chemosensitivity). Moreover, after partial acclimatization, CBF and its reactivity decline, resulting in a further increase in hypercapnic ventilatory response (HCVR) and universally severe CSA at altitude[5] (see arrows in Figure 18-8)—changes ultimately mediated by elevations in controller gain[50] and reduced CO_2 reserve.

The stimulation of the central chemoreceptors is from the CO_2 that is actually produced in the surrounding brain. Arterial blood flow to the brain washes the CO_2 away from the chemoreceptors. This means that CBF during sleep may be very important in ventilatory control. In the absence of the wakefulness drive to breathe, marked oscillations in CBF occur as a consequence of the periodic breathing, similar in nature to that reported in patients who experience sleep apnea at sea level.[51,52] Previous studies[5,53] demonstrated a relation between the decline in CBF from awake to NREM sleep, albeit being only a modest predictor of CSA. Of interest, the relation was stronger after 2 weeks at high altitude, when absolute perfusion was lower (both awake and during sleep), further supporting the idea that reduced $[H^+]$ washout within the brain enhances chemoreceptor activation (see earlier). Moreover, in view of the link between breathing pattern and CBF,[54,55] these oscillations in CBF are likely to be important in the pathophysiology of periodic breathing. Indeed, regardless of the causation of the first apneic episode, that is, whether alterations in basal CBF[56] or in cerebral or arterial P_{CO_2}[57-59] (or a combination of these and other factors) start the apnea cycle, the large swings in CBF that ensue seem likely to exacerbate the under- and overshooting of the ventilatory drive that characterizes the CSA disorder.[30]

The potential role for alterations in CBF during sleep at high altitude in the control of breathing has been further supported by the results of artificially increasing and reducing CBF during sleep by the use of medications. In a group of 12 normal volunteers at 5050 m, the administration of oral indomethacin 100 mg, which reduced CBF by approximately 23%, increased the severity of CSA by 16%. The HCVR also increased by 66%, suggesting that the reduction in CBF may have caused the increase in HCVR, which in turn increased the severity of CSA. Conversely, the administration of intravenous acetazolamide, which increased CBF by 28% without changing acid-base balance in the short term, reduced the severity of CSA by approximately 47%.[60] These results are consistent with modeling studies and suggest that CBF through its influence on the central chemoreceptors, not only in support of ventilation at rest but also in stabilized breathing in situations such as sleeping at altitude.

Role of Arousal from Sleep

Arousal from sleep is a very common feature of CSA, typically occurring during the hyperpneic phase of the cyclic breathing. Traditional thinking was that the hyperpnea of CSA was the cause of such arousals, and that the arousal was crucial in perpetuating the instability of breathing that typifies CSA.[61] In earlier studies at similar altitudes, the arousal index has tracked fairly closely with the increase in apnea-hypopnea index (AHI) with increasing altitude (as summarized in a contemporaneousreview). Although data indicate that benzodiazepines can reduce CSA by reducing arousals,[62] more recent findings indicate that this role has been overemphasized.[5] The cyclic changes in CBF may

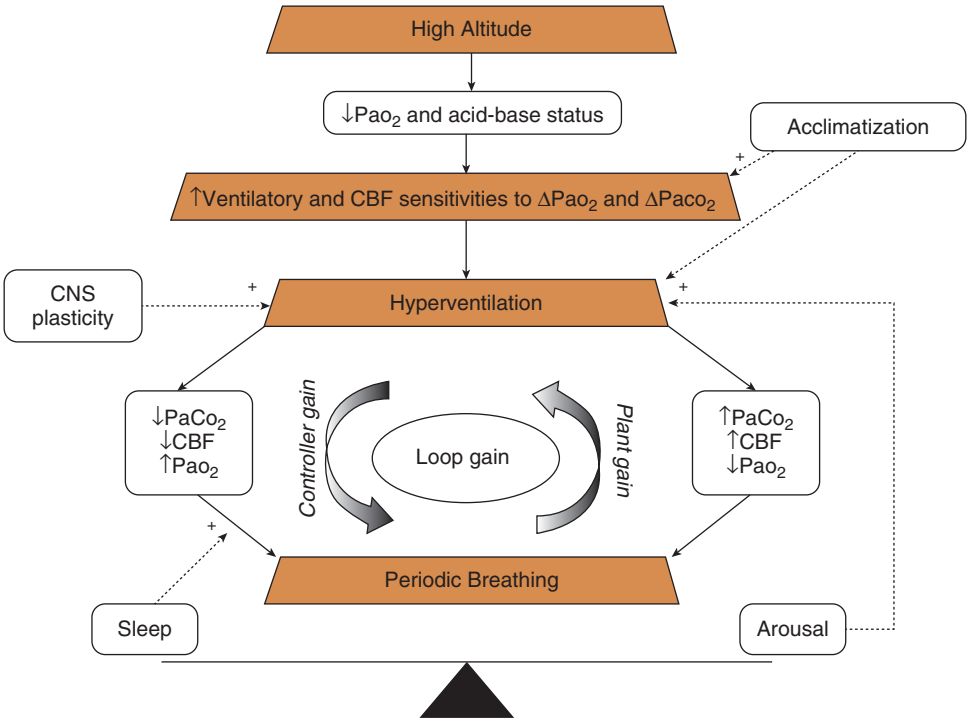


Figure 18-9 Schematic diagram showing various mechanisms by which high-altitude exposure leads to the development of periodic breathing during sleep. Initial effects of high-altitude exposure include a reduction in the partial pressure of arterial oxygen (Pa_{O_2}) and acid-base adjustments. These changes lead to alterations in chemoreflex control and cerebrovascular responses to changes in arterial blood gases. Overall, these complex cellular and neurochemical changes in chemoreflexes, acid-base status, and the central nervous system (CNS) lead to hyperventilation. Acclimatization, at least in lowlanders, magnifies these changes. Elevations in loop gain outweigh the improvements in plant gain caused by the chronic hypocapnia, leading to periodic breathing. Sleep and arousals lead to greater breathing instability. Apnea, which is associated with an increase in Pa_{CO_2} and decrease in Pa_{O_2} (and/or arousal), restimulates the PCRs and consequently ventilation. These changes in blood gases also lead to marked alterations in cerebral blood flow (CBF) (see Figure 18-7), which in turn may result in a sudden elevation (with reduced CBF) or reduction (with increased CBF) in brainstem pH. (Modified from Ainslie PN, Duffin J. Integration of cerebrovascular CO_2 reactivity and chemoreflex control of breathing: mechanisms of regulation, measurement, and interpretation. *Am J Physiol Regul Integr Comp Physiol* 2009;296:R1473–95.)

therefore be more important than repeated arousal from sleep in perpetuating CSA at high altitude. Because arousals without a clear temporal relationship to periodic breathing cycles during sleep also have been reported,[23,63] it seems likely that periodic breathing does not always cause arousal, and is not the only cause of arousals and fragmented sleep at altitude. Figure 18-9 summarizes the various mechanisms that have been outlined by which high-altitude exposure leads to the development of periodic breathing during sleep.

MANAGEMENT OF PERIODIC BREATHING AT ALTITUDE

Because of the strong correlation between absolute altitude and severity of CSA (see Figure 18-7), the obvious treatment would be to reverse that process and descend. If this strategy is not desired or practical, a number of options are available for management of periodic breathing. These can be broadly divided into three different categories: medical gases, pharmacologic interventions, and devices. The evidence of effectiveness of each of these treatment approaches is summarized next.

Medical Gases

Lahiri and associates have shown elegantly the curative effects of supplemental oxygen therapy in a subject with sustained CSA at 5300 m[38] (Figure 18-10). On rapid restoration of normoxic Sa_{O_2} by increasing the fraction of inspired oxygen (Fi_{O_2}), periodic breathing continues with prolonged apneic periods until hyperventilation is gradually reduced and Pa_{CO_2} returns to normal. Stabilizing effects of small incremental increases in fraction of inspired carbon dioxide (Fi_{CO_2}) also have been reported.[11] The mechanism probably involves blunting the degree of fall in Pa_{CO_2} during the hyperpnea phase of the CSA. West has proposed the addition of modest amounts of supplemental oxygen (i.e., using a device; see section on device below) into the sleeping quarters of high-altitude residents as a means of improving sleep quality and daytime performance.[64]

Pharmacologic Interventions

A number of studies have used pharmacologic manipulation at high altitude to decrease the occurrence and severity of periodic breathing, with agents such as acetazolamide, dexamethasone, various hypnotics, and theophylline. (Hypnotics are discussed later with insomnia.)

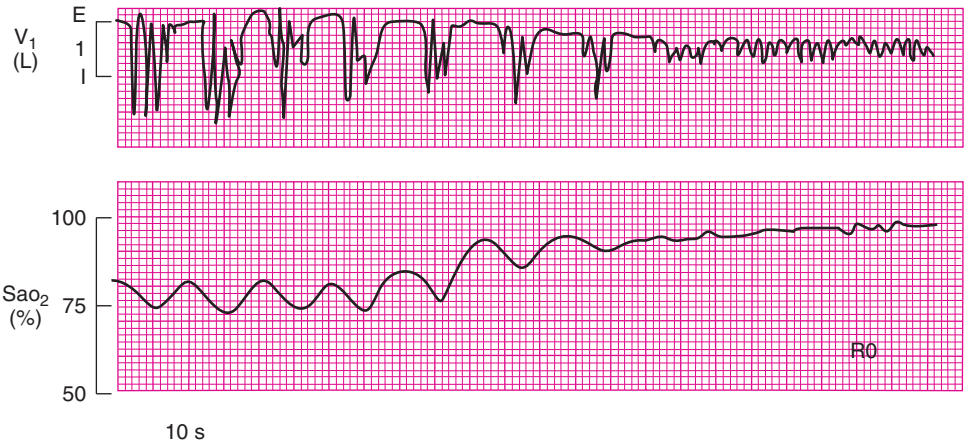


Figure 18-10 Polygraphic tracing. The effect of oxygen on periodic breathing and arterial oxygen saturation during sleep at 5400 m. As oxygen arterial saturation increases, periodic breathing is replaced by shallow and continuous breathing.[38] E, Expiration; I, inspiration. (Data from Lahiri S, Maret K, Sherpa MG. Dependence of high altitude sleep apnea on ventilatory sensitivity to hypoxia. *Respir Physiol* 1983;52:281–301.)

Oral acetazolamide has been shown by a number of investigators to effectively suppress CSA by 50% to 80% at high altitude.[40,65-67] The efficacy of carbonic anhydrase (CA) inhibitors is the result of ventilatory stimulation and better arterial oxygenation driven by the metabolic acidosis and the slight CO_2 retention from vascular CA inhibition and any partial red cell CA inhibition. Depending on the dose and route (intravenous or oral) of administration, CA inhibitors improve ventilatory control instability by increasing the tonic output of the central chemoreceptors and lowering their apneic threshold, thus rendering them less responsive to periodic reductions in $Paco_2$. In addition, inhibition of CA in the peripheral chemoreceptors reduces both the magnitude of hypoxic and hypercapnic sensitivity and the rate at which these signals arrive at the respiratory controller.[68-72] Consistent with this notion, peripheral chemoreceptor stimulants such as almitrine aggravate periodic breathing at high altitude.[40] Other benefits attributed to CA inhibitors in altitude adaptation include mild diuresis, reduced cerebrospinal fluid formation and increased CBF. Although these factors have received little attention at high altitude, studies have shown that acute intravenous acetazolamide induce elevations in CBF velocity and reactivity to $Paco_2$ are related to improvements in breathing stability at high altitude during wakefulness[47] and sleep.[48]

It has recently been reported that, at least in subjects susceptible to high-altitude pulmonary edema (HAPE), dexamethasone (at 8 mg/day in two divided doses) taken before ascent prevents severe hypoxemia and sleep disturbances, while dexamethasone taken 24 hours after arrival at 4559 m increases oxygenation and deep sleep.[73] Whether dexamethasone affects sleep on ascent to high altitude in otherwise healthy persons is unknown, although this seems unlikely. Dexamethasone has been repeatedly studied with regard to its possible beneficial effects on acute mountain sickness (AMS) and HAPE, but rarely with regard to sleep architecture per se. A passing reference suggests that it has no separate effect from acclimatization.[74]

It also has been shown, in a placebo-controlled trial, that low-dose (300 mg/day), slow-release theophylline reduces symptoms of AMS in association with alleviation of events of periodic breathing and oxygen desaturation at 4559 m.[75] In another randomized, double-blind, placebo-controlled study, the effects of theophylline (250 mg/day in two divided doses) on periodic breathing were compared with those of acetazolamide (also at 250 mg/day in two doses) after fast ascent to high altitude (3454 m) (n = 30).[76] Polysomnographic measurements were performed during two consecutive nights, and AMS, pulse rate, oxyhemoglobin saturation, and arterial blood gases were assessed three times a day. Both theophylline and acetazolamide normalized sleep-disordered breathing (median AHI, 2.5/hour versus 4.2/hour; range, 0 to 19, respectively) and reduced oxyhemoglobin desaturations during sleep (median desaturation index, 41.5/hour for placebo versus 6.5/hour for acetazolamide versus 8.5/hour for theophylline; range, 3 to 32). In contrast with theophylline, acetazolamide significantly improved basal oxyhemoglobin saturation during sleep (86% versus 81%). It was concluded that both oral slow-release theophylline and acetazolamide are effective to normalize high-altitude sleep-disordered breathing.[76]

Devices

Recently a number of devices have shown potential to treat periodic breathing at altitude, including bilevel positive airway pressure and the simple addition of dead space using a modified facemask.

A very different treatment, bilevel ventilation, recently was shown in a pilot study to halve the severity of CSA in seven volunteers at 3800 m at the White Mountain Research Center in California.[77] Unfortunately arterial blood gases, ventilatory responses, or CBF measurements were not collected, so the underlying mechanisms for those effects are uncertain. One could speculate that the ventilation further reduced $Paco_2$ and raised Pao_2; however, a further fall in $Paco_2$ would be expected to exaggerate periodic breathing during sleep. Noninvasive positive-pressure ventilation, such as continuous positive airway pressure (CPAP), raises functional residual capacity, which would increase oxygen stores and hence lower loop gain. Indeed, Edwards and colleagues have shown a reduction in loop gain in premature lambs by the application of CPAP, with resolution of CSA.[78] Increasing oxygen stores seems the more likely mechanism.

The simple addition of a 500-mL dead space also has been shown to improve sleep in some subjects at 3500 m.[79] This

study was conducted in 12 unacclimatized persons using full polysomnography. In random order, half of the night was spent with a 500-mL increase in dead space through a custom-designed full face mask and the other half without it. Although the dead space had no effect on individuals with AHI less than 30 events/hour, it did lead to marked reductions in AHI (from 70 to down to 30 events/hour) and oxygen desaturation index (from 73 to 43). Thus a 500-mL increase in dead space through a fitted mask may improve nocturnal breathing in those with severe altitude-induced sleep-disordered breathing.[79] Similar to the aforementioned studies above that have used elevations in $Fico_2$ to improve periodic breathing, the mechanism via elevations in dead space is likely through the stabilizing influence of elevations in $Paco_2$ on the CO_2 reserve.

OTHER SLEEP-RELATED CONDITIONS AFFECTED BY HIGH ALTITUDE

A number of sleep-related conditions are recognized to be influenced by ascent to high altitude, including nasal obstruction, obstructive sleep apnea (OSA), and insomnia.

Nasal Obstruction and High Altitude

Nasal obstruction is known to predispose affected persons to snoring and OSA and poor sleep quality at sea level.[80] Nasal obstruction is common in travelers to high altitude because of the usually dusty environment and the high prevalence of viral infections, at least in newcomers such as trekkers. Snoring is therefore very common at high altitude in sojourners.

OSA in nonnative subjects is recognized to resolve with increasing altitude[81] and the passage of time at high altitude when CSA develops, which has been attributed to increased central respiratory drive. Nasal obstruction and mouth breathing may persist, but snoring seems to lessen with time and increasing altitude regardless—perhaps because of increased central respiratory drive, which tends to stiffen the upper airway, thereby reducing obstruction and making the soft palate and other tissues less susceptible to vibration.[82]

Obstructive Sleep Apnea and High Altitude

Although research into CSA has been substantial, less is known about OSA at high altitude or the effects of altitude on subjects with OSA. It initially was recommended in popular high-altitude medicine publications that persons with OSA should avoid ascending to high altitude because of the likelihood that the condition will worsen.[83] On the basis of pathophysiologic considerations and uncontrolled observations in a small number of patients, a stay at altitude was thought to aggravate sleep-related breathing disturbances in patients with OSA; however, the data reported were elevations in the number of episodes of central apnea and reductions in obstructive apnea, not a worsening of OSA.[84,85] In a recent randomized controlled trial, it was shown that altitude exposure (up to 2590 m) in untreated patients with OSA aggravates hypoxemia, increases sleep-related breathing disturbances due to frequent central apneas/hypopneas, impairs driving simulator performance, and induces cardiovascular stress.[86] Again, the increase in respiratory events was due to an increase in central events, not to a worsening of OSA. It was found that a combination of acetazolamide (750 mg/day) and auto-CPAP therapy, compared with auto-CPAP alone, resulted in improvement in nocturnal oxygen saturation and AHI.[87]

The influence of OSA and optimum treatment approaches at higher elevations are largely unknown. An interesting finding is that mild OSA (approximately 5 events/hour) is abolished at higher altitudes (5050 m) and replaced by CSA.[2] The same effect also has been shown in patients with known OSA, many on CPAP therapy, exposed to a simulated high-altitude environment of 2750 m using normobaric hypoxia.[82]

Treatments for OSA and Snoring at High Altitude

In patients unable to use CPAP, or if electrical power is not available, an optimally fitted mandibular advancement device may be an alternative treatment option (to be confirmed by evidence) that can be combined with acetazolamide during altitude sojourns.[88] Evidently, however, acetazolamide alone also is beneficial and better than no treatment at all, because it improves oxygen saturation, curtails breathing disturbances, and potentially obviates the excessive blood pressure elevation in patients with OSA traveling to altitude.[89]

Insomnia at High Altitude

Part of the disruption to sleep at high altitude is the insomnia (both sleep onset and sleep maintenance) due to repeated arousals and awakenings from the hyperpnic phase of the periodic breathing. Several investigators have studied the effects of hypnotic medications in placebo controlled trials in the field. There is a theoretical risk that sedating medications might suppress ventilatory responsiveness and hence lead to worsened arterial oxygen saturation during sleep, which might also impair sleep quality and exacerbate AMS; however, evidence for this effect is lacking.

The effectiveness of various hypnotics has received attention. Both Dubowitz[90] and Nickol and colleagues[62] have used temazepam at 5400 m and reported a subjective improvement in sleep quality, but with variable effects on saturation and CSA severity. In Dubowitz's study, a group of 11 subjects showed no change in mean arterial saturation but appeared to show a reduction in "desaturation events," probably indicating a reduction in CSA severity linked to arousal from sleep, although no measurements of sleep state were recorded.[90] Nickol and coworkers,[62] on the other hand, demonstrated a modest but significant reduction in CSA index, from 16 to 9 events/hour, in a group of 33 healthy volunteers. Other reported benefits included a small reduction in mean saturation from 78% to 76% and improvement in AMS scores. New nonbenzodiazepine sedative-hypnotics also have been studied at high altitude.[13] Sleep quality was improved, but no direct data were provided about effects on CSA, although no change in oxygen desaturation index was seen.

Headache at High Altitude

Headache is a cardinal feature of AMS; hence it is very commonly experienced by sojourners to high altitude. Anecdotal evidence also suggests that high-altitude exposure increases the frequency of migraine attacks. Administration of paracetamol, with or without codeine, is the usual treatment, but acetazolamide and/or dexamethasone may be required if moderate to severe AMS is present. Opioid medication should be avoided because the possible (in the

most vulnerable subject) depressant effect on the ventilatory responses to both hypoxia and hypercapnia would predispose to lower saturation during sleep (although tending to suppress CSA).

Sleep in High-Altitude Natives

Native populations of the Tibetan and Andean plateaus both are descended from early colonizers. Tibetans arrived approximately 25,000 years ago, whereas the Andean populations arrived much later, around 11,000 years ago. Both populations have therefore been exposed to the opportunity for natural selection for traits to offset the unavoidable environmental stress of severe lifelong exposure to high altitude. The physiologic and genetic consequences of this environmental stress have been elegantly reviewed.[91,92]

CLINICAL PEARL

At high altitudes, say, above 3500 m, the most common cause of disturbed sleep in sojourners is periodic breathing due to the associated hypoxia. Surprisingly, the severity increases over time for at least 1 month at the same altitude, during the ongoing acclimatization. Established treatments, apart from descent to a lower altitude, include regular oral acetazolamide, which reduces CSA severity (as well as improving Pa_{O_2} and thereby decreasing the symptoms of AMS), and hypnotic medications, which reduce sleep disturbance from arousals.

SUMMARY

Sleep at high altitude is disturbed by various factors including a change of sleep environment, snoring, and insomnia; however, periodic breathing during sleep probably causes the most disturbances and occurs in almost everyone above 5000 m. Ventilatory acclimatization to altitude involves cellular and neurochemical reorganization in the peripheral chemoreceptors and central nervous system. The extent of periodic breathing during sleep at altitude intensifies with duration and severity of exposure; this increase is explained in part by elevations in loop gain. Although new mechanical and pharmacologic management techniques are emerging, oral acetazolamide remains the most effective and practical means to reduce periodic breathing. Use of benzodiazepine and other hypnotic agents appears to be a safe way to improve sleep quality at very high altitudes. Dexamethasone is a proven treatment for AMS (and associated sleep disturbance) but probably has no other effect on sleep quality.

Selected Readings

Ainslie PN, Lucas SJ, Burgess KR. Breathing and sleep at high altitude. *Respir Physiol Neurobiol* 2013;**188**(3):233–56.

Andrews G, Ainslie PN, Shepherd K, et al. The effect of partial acclimatization to high altitude on loop gain and central sleep apnea severity. *Respirology* 2012;**17**(5):835–40.

Burgess KR, Johnson PL, Edwards N. Central and obstructive sleep apnoea during ascent to high altitude. *Respirology* 2004;**9**(2):222–9.

Burgess KR, Lucas SJ, Shepherd K, et al. Influence of cerebral blood flow on central sleep apnea at high altitude. *Sleep* 2014;**37**(10):1679–87.

Dempsey JA. Crossing the apnoeic threshold: causes and consequences. *Exp Physiol* 2005;**90**(1):13–24.

Furian M, Latshang TD, Aeschbacher SS, et al. Cerebral oxygenation in highlanders with and without high-altitude pulmonary hypertension. *Exp Physiol* 2015;**100**(8):905–14.

Hackett PH, Roach RC, Harrison GL, et al. Respiratory stimulants and sleep periodic breathing at high altitude. Almitrine versus acetazolamide. *Am Rev Respir Dis* 1987;**135**(4):896–8.

Lahiri S, Maret K, Sherpa MG. Dependence of high altitude sleep apnea on ventilatory sensitivity to hypoxia. *Respir Physiol* 1983;**52**(3):281–301.

Nickol AH, Leverment J, Richards P, et al. Temazepam at high altitude reduces periodic breathing without impairing next-day performance: a randomized cross-over double-blind study. *J Sleep Res* 2006;**15**(4):445–54.

Rexhaj E, Rimoldi SF, Pratali L, et al. Sleep disordered breathing and vascular function in patients with chronic mountain sickness and healthy high-altitude dwellers. *Chest* 2015 Nov 5. doi: 10.1378/chest.15-1450.

Swenson ER, Leatham KL, Roach RC, et al. Renal carbonic anhydrase inhibition reduces high altitude sleep periodic breathing. *Respir Physiol* 1991;**86**(3):333–43.

White DP, Gleeson K, Pickett CK, et al. Altitude acclimatization: influence on periodic breathing and chemoresponsiveness during sleep. *J Appl Physiol* 1987;**63**(1):401–12.

Xie AL, Skatrud JB, Barczi SR, et al. Influence of cerebral blood flow on breathing stability. *J Appl Physiol* 2009;**106**(3):850–6.

A complete reference list can be found online at ExpertConsult.com.

Chapter 19

Sleep and Host Defense

Mark R. Opp; James M. Krueger

Chapter Highlights

- That sleep is altered during sickness has been known for millennia. Yet systematic and controlled studies aimed at elucidating the extent to which sleep is altered in response to immune challenge have only been conducted during the past 30 years.
- Substances historically viewed as components of the innate immune system are now known to be involved in the regulation or modulation of physiologic sleep-wake behavior, in the absence of immune challenge. Changes in sleep during immune challenge are actively driven and result from amplification of these physiologic mechanisms.
- Although the precise changes in sleep-wake behavior depend on the pathogen, route of infection, timing of infection, host species, and other factors, altered sleep during immune challenge is generally characterized by periods of increased non–rapid eye movement (NREM) sleep, increased delta power during NREM sleep, and suppressed rapid eye movement (REM) sleep. Infection-induced alterations in sleep are often accompanied by fever or hypothermia.
- Altered sleep has been studied in humans during pathologies and infections with pathogens, including human immunodeficiency virus/acquired immunodeficiency syndrome, rhinovirus (common cold), streptococci, trypanosomes, prions, and sepsis. Laboratory animal models include sepsis, influenza, and other viruses (gammaherpesvirus, vesicular stomatitis virus, rabies, feline immunodeficiency virus), several bacterial species, trypanosomes, and several prion diseases.
- Mechanisms that link sleep to innate immunity involve a biochemical brain network composed of cytokines, chemokines, growth factors, transcription factors, neurotransmitters, enzymes, and their receptors. Each of these substances and receptors is present in neurons, although interactions with glia are critical for host defense responses to immune challenge. Redundancy, feedforward, and feedback loops are characteristic of this biochemical network. These attributes provide stability and flexibility to the organismal response to immune challenge.

Most individuals have experienced the lethargy, malaise, and desire to sleep that may occur at the onset of infection. Further, most have been admonished to "get plenty of rest, or you will get sick." Conventional wisdom and personal experience suggest a connection between sleep and host defense systems; our sleep is perceptively different when sick and insufficient sleep may predispose to getting sick. These beliefs are not new. Indeed, Hippocrates, Aristotle, and many of our predecessors acknowledged such a relationship. But only within the past 30 years have modern science and medicine systematically investigated relationships between sleep and host defense systems. This chapter is organized around four main themes related to sleep and host defense: (1) the acute phase response and host defense, (2) infection-induced alterations in sleep, (3) effects of sleep loss on immune function, and (4) mechanisms linking sleep and immunity. Finally, in the Clinical Pearl section, we briefly present sleep as a recuperative process during sickness.

THE ACUTE PHASE RESPONSE AND HOST DEFENSE

Rapidly after infection, trauma, or during some malignant conditions, a complex response involving many cell types and peripheral organs is evoked that is collectively referred to as the *acute phase response* (APR). Markers of the APR include changes in serum concentrations of acute phase proteins. Measurement of acute phase proteins, such as C-reactive protein, is useful in clinical practice because they indicate inflammation. In addition to changes in serum concentrations of acute phase proteins, the APR includes physiologic changes, such as fever and increased vascular permeability, and other metabolic and pathologic changes. A major theme of this chapter is that altered sleep as a host defense also is part of the APR to inflammatory challenge. Altered sleep during inflammatory challenge is actively driven by multiple mediators and systems, many of which are shared with other facets of the APR.

Recent advances in our knowledge of central nervous system (CNS) innate immunity provide a framework for understanding many of the shared mechanisms underlying the APR in general as well as the specific alterations in sleep that occur during immune challenge. The APR is a critical innate immune response[1] that follows any inflammatory challenge, such as an infection or traumatic injury. Inflammatory challenges that are localized, for example, a minor cut or splinter, may activate a low-level APR that manifests as redness at the site of injury and may not be perceived by the subject. But with increased injury severity or response to an infectious challenge, the full systemic APR develops. The APR to infection by invading pathogens develops within a matter of hours, and the subject feels sick. In the case of infections, the function of the APR is to alert the host to the invasion and mobilize systemic protective responses, isolate and destroy invading pathogens, and remove tissue debris. The systemic inflammatory response activates the brain, liver, and bone marrow to react in a stereotypic manner. The APR includes physiologic and behavioral responses (e.g., fever, excess sleep, anorexia) as well as biochemical responses (e.g., C-reactive protein, serum amyloid A, mannose binding protein). Increased secretion of a broad array of endocrine hormones, including the stress hormones, also occurs. This complex of responses leads to host protective behaviors (e.g., social withdrawal),[2] physiologic responses (e.g., fever, which can increase efficiency of the immune response and inhibit growth of some microorganisms),[3,4] and immune responses (e.g., mobilization of leukocytes and natural killer [NK] cells).[1] Hormonal changes (e.g., prolactin regulation of antimicrobial nitric oxide levels)[5] and biochemical changes (e.g., potentiation of microbial phagocytosis)[6] also contribute to host defense. Although physical barriers (skin, mucosa) are the first line of defense, the APR is the first responder of host defense and is the trigger for acquired immunity, mediated by specific antibodies and cytotoxic T lymphocytes.[7]

A major class of proteins, cytokines, initiates the APR. Cytokines are generally associated with immune cells, but they are made by most cell types. More than 100 of these intercellular signaling molecules have been identified, and the complexity of their interactions rivals that of the CNS. Cytokines induce their own production and the production of other cytokines, and they form biochemical cascades characterized by much redundancy. Cytokines are classified into two major groups: type I cytokines that promote inflammation (proinflammatory) and type II cytokines that suppress it (antiinflammatory).[8] Three proinflammatory cytokines appear to be primary triggers of the APR. These early responder cytokines are interleukin-1β (IL-1β), tumor necrosis factor-α (TNF-α), and IL-6, each of which is implicated in the regulation and modulation of sleep. The class II cytokines include interferon-α (IFN-α), IFN-β, IL-4, and IL-10. These cytokines damp the APR and may also modulate sleep responses; for example, IL-4 and IL-10 inhibit spontaneous non–rapid eye movement (NREM) sleep (Figure 19-1). Cytokines can act in an autocrine, juxtacrine, paracrine, or endocrine manner to activate numerous APRs through such effectors as nitric oxide, adenosine, and prostaglandins.

A major advance in our understanding of the APR was the recognition that all known microorganisms have one or more biologically stable and chemically unique structural components.[9] These unique structural components are termed

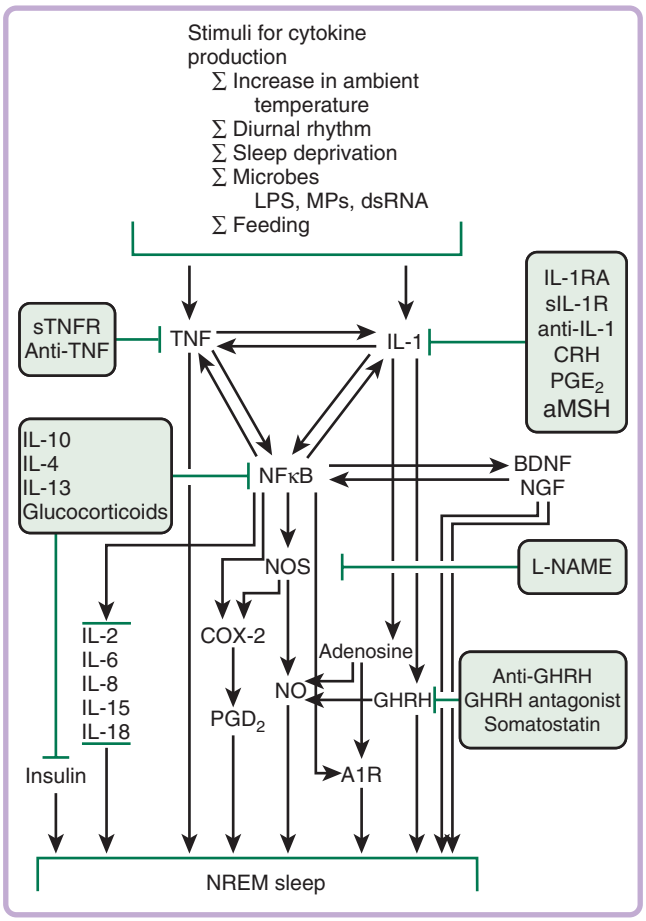


Figure 19-1 Interleukin (IL)-1β and tumor necrosis factor (TNF)-α are part of a brain biochemical network that regulates physiologic sleep and links multiple facets of innate immunity to sleep regulation. Much is known about mechanisms by which IL-1 and TNF directly or indirectly regulate and modulate non–rapid eye movement (NREM) sleep. Less is known about mechanisms of action for the REM sleep–suppressing effects of immune challenge. Current knowledge of the biochemical network that translates information about environmental perturbation into host responses that actively drive changes in sleep-wake behavior is much more complicated than depicted, and sites of action are not indicated (but see[53]). This biochemical cascade included cytokines, chemokines (not included), growth factors, transcription factors, neurotransmitters, and enzymes and their receptors. Because the network is redundant and parallel, inhibition of any single component does not result in complete sleep loss, nor does it block altered sleep in response to immune challenge. Such redundant pathways provide stability to the sleep regulatory system and alternative mechanisms by which sleep-promoting or sleep inhibitory stimuli may affect sleep. Substances in boxes inhibit NREM sleep and inhibit either the production of or the actions of substances in downstream pathways. The receptor and intracellular signaling systems for all these substances are found in neurons. Also not depicted in this schema are interactions of components of this biochemical network with glial cells. Gliotransmission is implicated in the modulation of physiologic sleep, and is likely to play a critical role in brain responses to immune challenge that result in altered sleep-wake behavior (see[100]). The symbol → indicates stimulation or upregulation; ⊥ indicates inhibition or downregulation. A1R, Adenosine A1 receptor; anti-IL-1, anti-IL-1 antibody; anti-TNF, anti-TNF antibody; anti-GHRH, anti–growth hormone–releasing hormone antibody; BDNF, brain-derived neurotrophic factor; COX-2, cyclooxygenase-2; CRH, corticotropin-releasing hormone; dsRNA, double-stranded RNA; GHRH, growth hormone–releasing hormone; IL-1RA, IL-1 receptor antagonist; L-NAME, an arginine analogue; LPS, lipopolysaccharide; MPs, muramyl peptides; αMSH, α-melanocyte-stimulating hormone; NFκB, nuclear factor kappa B; NGF, nerve growth factor; NO, nitric oxide; NOS, nitric oxide synthase; PGD_2, prostaglandin D_2; PGE_2, prostaglandin E_2; sIL-1R, soluble IL-1 receptor; sTNFR, soluble TNF receptor.

pathogen-associated molecular patterns, or PAMPs. The innate immune system recognizes PAMPs using specialized receptors (pathogen recognition receptors, or PRRs) that are either membrane bound or cytoplasmic.[7] These PRRs include Toll-like receptors (TLRs) and nucleotide-binding domain and leucine-rich repeat domain receptors (NLRs; more commonly designated as nucleotide-binding oligomerization domain, or Nod, proteins).[10] The PRR binding of microbial PAMPs induces cytokines, and these cytokines in turn upregulate PRRs and cytokines in neighboring tissues, resulting in amplification of the initial response. Thus in infectious illness, pathogens induce cytokines, then cytokines activate the APR and thereby facilitate host defense through dozens of protective mediators and activated immune cell types.[9] Altered sleep (increased NREM sleep and suppressed rapid eye movement [REM] sleep) is one outcome of this cytokine cascade during infectious illness.

INFECTION-INDUCED ALTERATIONS IN SLEEP

The impact of infection on sleep has been determined for viral, bacterial, and fungal pathogens; prion-related diseases; and protozoan parasites. Most studies to date have used virus and bacteria as the infectious agent, and as such this chapter focuses primarily on altered sleep in response to these pathogens.

Viral Infections and Altered Sleep

Viral diseases that cause CNS lesions or systemic inflammation alter sleep.[11] In von Economo's seminal paper,[12] he related the postmortem location of brain lesions of patients suffering encephalitis lethargica to specific changes in sleep patterns. This work led to the concept that sleep was an active process, not simply the withdrawal of sensory stimuli, and to the idea that there was some degree of localization of neural networks regulating sleep. Although von Economo's encephalitis was commonly thought to have been caused by the 1918 influenza virus pandemic ("Spanish flu"), recent analyses reveal that the disease preceded the 1918 pandemic and was probably an autoimmune complication of streptococcal infections affecting the basal ganglia.[13,14] Despite the importance of von Economo's work, many years passed before the direct effects of viral infections on sleep were experimentally determined.

During the early stages of infection with human immunodeficiency virus (HIV), and before patients are symptomatic for acquired immunodeficiency virus (AIDS), sleep is altered such that excess stage 4 NREM sleep occurs during the latter half of the night.[15] Other CNS viral diseases, such as rabies[16] or viral encephalitis in rodents after vesicular stomatitis virus infection,[17] also are associated with altered sleep. In these CNS infections, it is difficult to know whether sleep is altered by direct actions on sleep regulatory mechanisms or whether altered sleep results from virus-induced brain lesions. One model that has been frequently used to determine effects of viral infections on sleep is influenza. Influenza virus localizes to the respiratory tract during the early stage of disease and does not cause brain lesions. In addition, influenza infections pose tremendous public health burdens owing to the millions of lives lost each year and the threat of pandemics. Smith and colleagues[18] report that low doses of influenza in humans increase sleep and cognitive dysfunction; these symptoms appear after low viral doses that fail to induce the better known characteristics of the APR, such as a fever. However, in that study indexes of behavior, not polysomnography, were used. Drake and colleagues[19] demonstrated in healthy human volunteers that infection with rhinovirus 23 increases total sleep time and impairs cognitive performance. (Rhinoviruses are the predominant cause of the "common cold.") In rabbits, intravenous injections of influenza virus are also associated with large increases in NREM sleep and suppressed REM sleep, even though the virus does not replicate in this species.[11]

Studies in mice infected with influenza virus demonstrate profound changes in sleep through the course of disease progression.[20-22] Changes in sleep of mice during influenza infection share some features of sleep responses to bacterial infections (described later). As a preclinical model, influenza infection of mice is clinically relevant because mouse-adapted strains of this virus can be introduced into the respiratory tract and can fully replicate in the lungs, causing a severe APR. Mice challenged intranasally with influenza virus display profound increases in NREM sleep and inhibition of REM sleep, which last 3 or more days.[20] Macrophages appear to be the critical immune cell type driving increased NREM sleep, whereas NK cells, neutrophils, and T lymphocytes do not play a significant role.[23] There are strain differences in responses of mice to this challenge,[24] indicating a genetic component affecting the sleep response to influenza virus. Genetic regulation of the inflammatory response to influenza in mice and humans has been reviewed elsewhere.[25]

One generic viral PAMP that increases NREM sleep and initiates other facets of the APR is virus-associated double-stranded RNA (dsRNA). All viruses examined to date produce virus-associated dsRNA, which is generally derived from the annealing of viral replication products rather than from the virus itself.[26] Virus-associated dsRNA, recognized by the PRR TLR3, induces numerous cytokines, including IL-1, IL-6, TNF, and IFN. Virus-associated dsRNA can be extracted from lungs of infected mice[27] and is capable of inducing an APR in naïve rabbits that is similar to that of live virus. Similarly, rabbits given short double-stranded (but not single-stranded) oligomers that correspond to a portion of influenza gene segment 3 also exhibit large increases in NREM sleep.[28] Synthetic dsRNA (polyriboinosinic polyribocytidylic acid, or poly I:C), when inoculated into the lungs of mice primed with IFN-α, induces an APR that is virtually identical to that after influenza virus.[26] These observations suggest that virus-associated dsRNA is sufficient to initiate the APRs seen in influenza-infected mice.

Poly I:C administered into rabbits also induces an influenza-like APR, but the corresponding single strands of poly I or poly C are inert. In rabbits, poly I:C can substitute for virus in the induction of a hyporesponsive state to viral challenge.[28] Rabbits challenged with viable virus or poly I:C have increased plasma antiviral activity that occurs concomitantly with the changes in sleep. The antiviral activity is attributed to IFN-α and other cytokines. Injection of IFN-α into rabbits also induces sleep responses similar to those induced by virus, poly I:C, or the double-stranded viral oligomers.[29] High doses of IFN-α increase NREM sleep in other species as well,[30] and low doses that simulate concentrations of IFN-α comparable to those observed during an infection inhibit both NREM and REM sleep in humans.[31]

Interferons play a major role in viral symptoms. Knockout (KO) mice have been widely used to better understand the

role of specific cytokines or hormones in host defense. Mice genetically deficient for the receptor that binds both IFN-α and IFN-β (the type I receptor) respond to poly I:C with altered sleep and a hypothermic response that is similar to that seen in infected wild-type mice. However, in influenza-infected IFN receptor KO mice, the APR occurs earlier,[32] suggesting that type I IFNs may modulate the APR, presumably by regulating proinflammatory cytokine production. Influenza-infected IFN receptor KO mice are less ill later in the infection and recover sooner.[32] Sleep modulatory cytokines, in addition to IFNs, likely mediate the sleep responses to influenza virus. For example, although the duration of altered NREM and REM sleep is the same in both strains after viral challenge, mice deficient in the 55-kd and 75-kd TNF receptors manifest reduced electroencephalogram (EEG) delta power, a measure of sleep intensity, whereas in wild-type control mice delta power increases.[33] IL-1 signaling in brain requires a brain-specific receptor accessory protein.[34] Mice lacking the IL-1 receptor brain-specific accessory protein have higher morbidity and mortality after influenza inoculation and sleep less during the infection than wild-type mice.

Another mediator that plays a role in sleep and host defense is nitric oxide, which is synthesized by multiple nitric oxide synthetases (NOSs). Mice deficient in either neuronal NOS or inducible NOS have attenuated NREM sleep responses to influenza challenge compared with infected wild-type controls.[35]

Mice and rats with natural mutations of the growth hormone–releasing hormone (GHRH) receptor express a dwarf phenotype and altered spontaneous NREM sleep.[36] The GHRH receptor is a candidate gene for regulating NREM sleep increases in response to influenza virus.[37] Dwarf mice with nonfunctional GHRH receptors (called lit/lit mice) fail to respond to influenza virus with increased NREM sleep or EEG delta power.[38] Instead, infected lit/lit mice manifest a pathologic state with EEG slow waves, enhanced muscle tone, and increased mortality.[38] Such results indicate that single genes can substantially modify sleep responses to infectious challenge. Importantly, results from lit/lit mice also demonstrate that the sleep responses forming part of the APR correlate with survival.

Influenza virus is a frequently used model for APR studies, in part because it was assumed that the virus does not invade the brain or lead to the complications associated with the use of neurovirulent viruses. Recent studies, however, demonstrate that the strain of influenza most commonly employed in preclinical studies rapidly invades the olfactory bulb of the mouse brain following intranasal inoculation.[39] The virus activates microglia in the outer layer of the olfactory bulb and upregulates IL-1 and TNF at times that correspond to the postinfection time period when the systemic APR begins. These studies suggest that cytokines made in the olfactory bulb could affect the CNS components of the APR to influenza virus, including sleep responses.

Bacterial Challenge

Altered sleep is also observed after bacterial infection. Indeed, results obtained after inoculating rabbits with the gram-positive bacteria *Staphylococcus aureus* were the first to suggest that NREM sleep responses were part of the APR.[40] In those experiments, rabbits were given *S. aureus* intravenously to induce septicemia; within a few hours of the inoculation NREM sleep was twice the amount as during comparable periods after control inoculation. Associated with the increase in NREM sleep were increases in amplitude of EEG slow waves. EEG slow wave (0.5 to 4.0 Hz) amplitudes are thought to indicate the intensity of NREM sleep. This initial phase of increased duration and intensity of NREM sleep lasted about 20 hours; it was followed by a more prolonged phase of decreased NREM sleep and decreased EEG slow wave amplitudes.[40] During both phases of the NREM sleep changes, REM sleep was inhibited and animals were febrile. Other changes characteristic of the APR, for example, fibrinogenemia and neutrophilia, occurred concurrently with the changes in sleep.[40] In subsequent studies in which gram-negative bacteria and other routes of administration were used, a similar general pattern of biphasic NREM sleep responses and REM sleep inhibition was observed.[41] However, the timing of sleep responses depends on the bacterial species and the route of administration. For example, after intravenous administration of *Escherichia coli,* NREM sleep responses are rapid in onset, but increased NREM sleep lasts only 4 to 6 hours. The subsequent phase of reduced NREM sleep and reduced amplitude of EEG slow waves is sustained for relatively long periods. In contrast, if the gram-negative bacterium *Pasteurella multocida* (a natural respiratory pathogen in rabbits) is given intranasally, a different time course of sleep responses is observed. In this case, the increased NREM sleep responses occur after a longer latency, and the magnitude of the increases in NREM sleep is less than the effects of this pathogen given by other routes of administration.

The intestinal lumen of mammals contains large amounts of many different bacteria species. Bacteria translocate into the intestinal lymphatics under normal conditions. Of importance to this discussion, intestinal permeability is altered after sleep deprivation, resulting in increased release of bacterial products into the lymphatics. Local lymph node macrophages phagocytose and digest these bacterial products,[42] releasing PAMPs that can trigger sleep responses. This mechanism operates at a low basal rate under normal conditions and is amplified during systemic inflammation. The phagocytosis by macrophages of bacterial products is also likely to be involved in sleep responses induced by sleep deprivation and excess food intake. A role for gut bacteria in sleep modulation is also evidenced by observations that reducing bacterial populations in the intestine is associated with reduced sleep.[43]

The first bacterial PAMP demonstrated to alter sleep was a specific muramyl peptide derived from bacterial cell wall peptidoglycans isolated from the brain and urine of sleep-deprived subjects.[44] Peptidoglycan components are recognized by certain NLRs and appear to play a major role in the pathogenesis of inflammatory mucosal diseases. The sleep-promoting activity of muramyl peptides is dependent on their chemical structure.[45] Many muramyl peptides are also immune adjuvants and pyrogenic, although the structural requirements for these biologic activities are distinct from those required for sleep-promoting activity.[45]

Another bacterial product that is involved in sleep responses to gram-negative bacteria is the lipopolysaccharide (LPS) component of cell wall endotoxin. LPS is the dominant PAMP associated with endotoxin, and it binds to TLR4. LPS

has been intensively studied in animal models[46] and humans volunteers[47] with respect to effects on sleep. LPS and its toxic moiety, lipid A, alter sleep in animals and humans.[45,48] The toxic moiety, lipid A, alters sleep, and modification of the lipid A structure alters LPS activity and reduces its sleep-altering properties. Healthy human volunteers injected with LPS manifest sleep changes, fever, cytokine expression, and hormonal changes[47] somewhat similar to those seen in animals. However, the impact of LPS on the human EEG differs from that observed in rabbits or rats, and in humans it requires a higher LPS dose to increase NREM sleep than it does to suppress REM sleep.

Most experimental studies of bacterial infections and sleep have used inoculation of a single pathogen species as the infectious challenge. The gut microbiome, however, is polymicrobial, and many infections result from invasion by multiple pathogen species. Such is the case in sepsis, during which polymicrobial infections are routinely the cause. Clinical studies demonstrate EEG anomalies in patients in the intensive care unit who become septic.[49] The etiology of sepsis is complex, and sepsis may result from many different kinds of insult. As a consequence, several preclinical models have been developed to study sepsis. Although each model used has strengths and limitations, the model currently considered to be the gold standard is cecal ligation and puncture (CLP).[50] CLP produces a polymicrobial infection that is considered clinically relevant because of its time course, because it reproduces the dynamic changes in cardiac function observed in human patients, and because there is a progressive release of inflammatory mediators. The severity of the ensuing infection is readily titrated in this model. Sleep is altered during the acute phase of CLP sepsis, which occurs 1 to 4 days after sepsis induction.[51] In this period, the NREM and REM sleep phases of rats increase during the dark period but are reduced during the light period. These changes in sleep coincide with increased cytokine messenger RNA (mRNA) and protein in the brain.[52] Of interest, effects of sepsis on body temperature and activity rhythms persist long after the animal has recovered and is no longer at risk for dying.[52] These observations suggest that sepsis alters brain function and are in agreement with observations that patients surviving sepsis often suffer severe and debilitating cognitive impairment.

Other microbes, for example, protozoan parasites such as *Trypanosoma brucei brucei,* express their own PAMPs, which bind to specific TLRs and induce sleep responses.[21] Trypanosomiasis in rabbits is associated with periods of increased NREM sleep that occur about every 7 days. Trypanosomes undergo antigenic variations in the host; the proliferating new antigenic variants stimulate new host immune responses, and such periods are accompanied by increased NREM sleep.[21] As with bacteria and viruses, protozoans induce cytokine production by the host.

In summary, infectious challenge is associated with profound changes in sleep. As mentioned in the overview of the APR, PRRs such as the TLR and NLR receptor families detect the various PAMPs capable of altering sleep. Detection of PAMPS by the innate immune system explains, in part, why diverse microbial pathogens activate stereotypic host defense responses such as fever, anorexia, and altered sleep. Microbe-induced alterations in sleep, like the other components of the APR, are adaptive.[53]

EFFECTS OF SLEEP LOSS ON IMMUNE FUNCTION

Sleep is altered during immune challenge, yet whether sleep loss alters immune function has been more difficult to demonstrate. There are multiple systems associated with immunity, each with myriad mediators and modulators. There are positive and negative feedback control mechanisms that interact in complex ways. This complexity of the immune system makes it difficult to determine what measurements one should use to assess immune function. From a functional perspective, the most important question is whether sleep loss renders the animal more vulnerable to infection, tumor formation, or systemic inflammatory diseases. (We already know that sleep loss renders one more vulnerable to accidental injury.) Although few studies have been conducted within the context of sleep, some suggest relationships between sleep and functional immune outcomes. For example, among 12 mammalian species sampled, those with longer daily sleep times have the greatest number of white blood cells and are least susceptible to parasites.[54] Susceptibility to infection has been used as an end point in some studies of human subjects. Shift workers are considered chronically sleep deprived, and a large population study reveals increased incidence of infections in those who experienced the most shift changes.[55] However, sleep time was not quantified in these individuals, and many other variables (including circadian rhythm disruption and stress) confound the interpretation of these results. Self-reports of sleep duration and sleep efficiency before controlled challenge with a "cold" virus suggest that individuals sleeping less than 7 hours per night, or with sleep efficiency of less than 92%, are more likely to develop colds.[56] Although this study was carefully conducted, self-reports may not adequately capture information about sleep duration, and mechanisms underlying these associations cannot be determined.

Vaccinations are effective only when the antigenic challenge (the vaccination) induces a sufficient antibody response (acquired immunity) such that on subsequent exposure to the same or similar pathogen, there is already an immune memory. Some individuals do not respond to vaccination with an antibody response sufficient to confer protection, and factors contributing to nonresponders are not well understood. Several studies of human volunteers have examined the effects of sleep loss on subsequent antibody responses to vaccines. The first of these studies used a protocol in which subjects were restricted to 4 hours of sleep opportunity per night for 4 nights and then given a flu shot.[57] Sleep restriction continued for 2 nights after the vaccination. Subjects then were allowed 12 hours of sleep opportunity each night for the next 7 days. Control subjects were allowed ad libitum sleep but otherwise followed the same protocol. Sleep-restricted subjects 10 days after vaccination produced less than half the antibody titers of control subjects who were allowed 8 hours of sleep opportunity per night. In a different study, subjects were given a hepatitis A vaccination and then deprived of sleep for one night. Antibody titers in sleep-deprived subjects 4 weeks after the vaccination were reduced by about 50% relative to those of control (non–sleep-deprived) subjects.[58] A similar study by the same group demonstrated that effects of a single night of total sleep loss reduced antigen-specific helper T cells and antigen-specific antibody for at least 1 year.[59] Collectively, these aforementioned controlled laboratory studies of healthy volunteers

suggest that sleep loss impairs acquired immunity and that there are functional outcomes. There are few community-based studies relating sleep duration to antibody production after vaccination, but at least one[60] demonstrates a direct relationship between sleep duration (defined by actigraphy) and antibody titers in response to hepatitis B vaccination. This study[60] also demonstrated that short sleep duration was associated with decreased clinical protection from hepatitis B at the conclusion of the three-vaccination series. Studies of clinical populations focused on similar questions are difficult to control and interpret. At least one clinical study failed to demonstrate differences in the antibody response to influenza vaccine in moderate to severe obstructive sleep apnea patients compared with controls.[61]

In contrast to functional studies in which effects of sleep or sleep loss on resistance to parasites or antibody production is the primary outcome, the most widely used approach to determine effects of sleep loss on immune function is to select one or more parameters of interest, for example, NK cell activity or plasma cytokine concentrations, and determine whether sleep loss alters these outcomes. Often such results leave the reader uninformed as to whether the outcome is adverse or beneficial for the host. Sleep deprivation or sleep disruption can be associated with stress, and many factors influence the impact of stress on sleep.[62] The most commonly measured "stress hormone" is cortisol in humans and corticosterone in rodents. Cortisol or corticosterone is a critical negative feedback regulator of cytokine production in brain. As such, nonspecific increases in this hormone could affect the outcome measure of interest. In addition to stressor-induced elevation of cortisol or corticosterone, sleep deprivation protocols often increase locomotor activity, alter feeding patterns, and disrupt normal variation in other hormones and body temperature. Each of these variables is known to affect immune function. Despite these limitations, data derived from human volunteers or laboratory animals suggest that sleep loss does indeed influence the immune system.

Results from studies of laboratory animals subjected to short-term sleep deprivation are consistent with most human studies. Toth and colleagues challenged rabbits with *E. coli* before or after 4 hours of sleep deprivation. They concluded that sleep deprivation failed to exacerbate *E. coli*–induced clinical illness, although the combination of sleep deprivation and bacterial infection altered some facets of sleep responses compared with either manipulation alone.[63] Furthermore, mice immunized against influenza virus and then rechallenged with influenza just before sleep deprivation failed to clear the virus from their lungs.[64] However, in a similar study[65] sleep loss failed to alter preexisting mucosal and humoral immunity in either young or senescent mice. The variation in the effects of sleep loss on outcomes in mice subjected to influenza virus likely results from differences in the sleep deprivation protocols, end points analyzed, and influenza models employed. Little research has focused on sleep deprivation and clinical responses to bacteria, but mortality is greater in mice in which sleep is disrupted after they are made septic by CLP.[66] Collectively, these studies suggest that acute sleep loss impairs or alters host defense.

The effects of long-term sleep loss on host defense in laboratory rodents are more striking. If rats obtain only about 20% of their normal sleep when deprived by the disk-over-water method,[67] they die after a period of 2 to 3 weeks.[68] Yoked control rats, which manage to maintain about 80% of their normal sleep during the protocol period, survive. The experimental rats, but not the yoked controls, develop septicemia.[68] Bacteria cultured from the blood are primarily facultative anaerobes indigenous to the host and environment. These results demonstrate that, using this method, innate host defenses in the rat are compromised by long-term sleep loss. These results suggest that prolonged sleep loss likely amplifies the normally occurring process of gut permeability to bacteria and bacterial products. However, these interactions may be somehow unique to rats because other species do not die when subjected to sleep deprivation by the disk-over-water method.

Sleep disruption may induce low-grade inflammation or may render the animal more susceptible to inflammatory challenge. We recently demonstrated that disrupting daytime sleep of mice for prolonged periods (9 days) exacerbates febrile responses to LPS.[69] The exacerbated febrile response to LPS under the conditions of this study may be due to sleep disruption per se because no other parameters measured (corticosterone, food or water intake, body weight) differed substantially from either home cage control animals or animals housed on the sleep-disruption device but allowed ad libitum sleep.

An independent literature clearly demonstrates that sleep loss is associated with changes in parameters normally associated with inflammation and the immune response. Cytokines such as IFN, IL-1, and TNF are well known for their roles as immunomodulators, and their production is altered by sleep deprivation or sleep disruption.[70] For example, sleep deprivation enhances TNF production in streptococcus-stimulated white blood cells. Other stressors, unlike sleep deprivation, fail to prime for systemic production of TNF, whereas sleep loss increases the ability of LPS-stimulated monocytes to produce TNF. The ability of cultures of whole blood to produce IL-1 and IFN in response to LPS is maximal at the time of sleep onset. In humans or animals, sleep deprivation leads to enhanced nocturnal plasma levels of IL-1–like activity.

Several reports show that in healthy volunteers plasma levels of cytokines are related to the sleep-wake cycle. Such relationships were first described by demonstrating that plasma IL-1–like activity was related to the onset of slow wave sleep.[71] Plasma concentrations of TNF vary in phase with EEG slow wave amplitudes.[72] There is also a temporal relationship between sleep of healthy human volunteers and IL-1 activity.[73] Several clinical conditions associated with sleepiness, such as sleep apnea, chronic fatigue syndrome, chronic insomnia, preeclampsia, postdialysis fatigue, psychoses, rheumatoid arthritis, and AIDS, are associated with enhanced plasma levels of TNF and other cytokines.[70] Only those sleep apnea patients showing elevated TNF activity experience fatigue.[74]

Other facets of the immune response are also linked to sleep. About 40 years ago, altered antigen uptake after sleep deprivation was reported.[75] Studies carried out in the 1970s also showed a decrease in lymphocyte DNA synthesis after 48 hours of sleep deprivation and a decrease in phagocytosis after 72 hours of sleep deprivation.[76,77] Sleep deprivation also induces changes in mitogen responses. Circulating immune complexes fall during sleep and rise again just before getting out of bed, and in mice sleep deprivation reduces immunoglobulin G (IgG) catabolism, resulting in elevated IgG levels. In contrast, one study failed to show an effect of sleep deprivation on spleen cell counts, lymphocyte proliferation, or

plaque-forming cell responses to antigens in rats.[78] In a comprehensive study of human volunteers, 64 hours of sleep deprivation reduced CD4, CD16, CD56, and CD57 lymphocytes after 1 night of sleep loss, although the number of CD56 and CD57 lymphocytes increased after 2 nights of sleep loss.[79] Another group also showed that 1 night of sustained wakefulness reduced counts of all lymphocyte subsets measured.[80]

Sleep and sleep loss are associated with changes in NK cell activity. NK cell activity is reduced in patients with insomnia[81] and decreases after partial night sleep restriction.[82,83] In contrast, increased NK cell activity increases after 64 hours of total sleep deprivation.[79] Circulating NK cell activity, as well as NK cell activity in a variety of tissue compartments, may be sensitive to sleep, although the exact nature of relationships between NK cell activity and sleep likely depend on the specific experimental conditions used to elucidate them.

In summary, determination of sleep deprivation effects on immune function may be confounded by stress and other coincident physiologic responses in animals. Concurrent physiologic changes (other than stress) also complicate sleep deprivation studies in humans. Sleep deprivation protocols are not standardized in animal or human studies, making comparison of results difficult. Despite the problems with available data, collectively the extensive literature on sleep deprivation and immune changes suggests that short-term deprivation potentiates immune function, whereas long-term deprivation leads to functional immune suppression.

MECHANISMS LINKING SLEEP AND IMMUNITY

Substantial evidence now suggests that IL-1 and TNF are involved in physiologic sleep regulation.[70,84] Furthermore, IL-1 and TNF mRNA and protein change during pathologies characterized by altered sleep. Sleep deprivation is associated with enhanced sleepiness, sleep rebound, sensitivity to kindling and pain stimuli, cognitive and memory impairments, performance impairments, depression, and fatigue. Exogenous administration of IL-1 or TNF induces all of these symptoms associated with sleep loss.[46,70] Further, chronic pathologies associated with sleep loss such as metabolic syndrome, chronic inflammation, and cardiovascular disease are also characterized by changes in IL-1 and TNF activity,[46,70] and in some cases these pathologies are attenuated if these cytokines are inhibited.[85-87] Clinically available inhibitors of either IL-1 (e.g., the IL-1 receptor antagonist, anakinra) or TNF (e.g., the TNF-α soluble receptor, etanercept) alleviate fatigue and excess sleepiness in humans with pathologies such as sleep apnea or rheumatoid arthritis.[85,86,88] The IL-1 receptor antagonist and TNF soluble receptor are normal gene products found in blood and brain, and their concentrations are altered by sleep.[46]

In addition to being immunocyte products, the production of which is amplified by viral and bacterial products, IL-1 and TNF are also found in normal brain.[46,70] IL-1and TNF mRNA have diurnal rhythms in the brain, with the highest values being associated with periods of maximum sleep. TNF protein also has a sleep-associated diurnal rhythm in several brain areas, and IL-1 in cerebrospinal fluid varies with the sleep-wake cycle.[89] Cortical expression of TNF is enhanced by afferent nerve activity,[90] which may be part of the process that is responsible for local use-dependent sleep.[46]

Administration of either IL-1 or TNF promotes NREM sleep.[46,48,70] The increase in NREM sleep after either IL-1 or TNF administration is physiological in the sense that sleep remains episodic and readily reversible if animals are disturbed. Further, IL-1 or TNF enhances NREM sleep intensity, as measured by the amplitude of EEG delta waves. The effects of IL-1 on sleep depend on dose and the time of day it is given.[91] IL-1 and TNF inhibit the binding of the BMAL/CLOCK complex in the suprachiasmatic nucleus[92]; this action may be responsible for the differential effects of these cytokines at different times of the day. Finally, knockout strains of mice that lack either the type I IL-1 receptor,[93] the 55-kd TNF receptor,[94] or both of these receptors[95] sleep less than control strains.

NREM sleep increases after sleep deprivation, excessive food intake, or acute mild increases in ambient temperature. The somnogenic actions of each of these manipulations are associated with enhanced production of either IL-1 or TNF. After sleep deprivation, circulating IL-1 increases, brain levels of IL-1 mRNA increase, and the NREM sleep rebound that would normally occur after sleep deprivation is greatly attenuated if either IL-1 or TNF is blocked using antibodies or soluble receptors.

The actions of excessive feeding on NREM sleep and liver and brain production of IL-1 represent physiologic changes, yet they likely involve the actions of bacterial cell wall products. Gut permeability to bacteria and bacterial products is influenced by dietary factors,[96] and the gram-negative bacteria cell wall product, endotoxin, is a normal constituent of portal blood.[97] Endotoxin stimulates IL-1 production in liver and elsewhere. Other bacterial cell wall products, for example, muramyl peptides, also have the capacity to stimulate IL-1 and TNF production[45] and to cross the intestinal wall into lymph. NREM sleep responses induced by muramyl dipeptide are attenuated if animals are pretreated with either blockers of IL-1 or TNF.[45,98] As mentioned previously, prolonged sleep deprivation of rats by the disk-over-water method results in bacteremia. It thus seems likely that the interaction of those bacteria with liver macrophages results in the amplification of the physiologic processes that are also associated with excessive food intake.

IL-1 and TNF act within a biochemical network (see Figure 19-1). For example, IL-1 and TNF stimulate nuclear factor kappa B (NFκB) production. NFκB is a DNA-binding protein involved in transcription. Other sleep-altering cytokines, such as acidic fibroblast growth factor, epidermal growth factor, and nerve growth factor, also stimulate NFκB production. NFκB promotes IL-1 and TNF production and thus forms a positive feedback loop. Sleep deprivation is associated with the activation of NFκB in the cerebral cortex, basal forebrain cholinergic neurons, and lateral hypothalamus. Activation of NFκB also promotes IL-2, IL-6, IL-8, IL-15, and IL-18 production, each of which promotes sleep in rats.[46,48,70]

GHRH is likely involved in IL-1 promotion of NREM sleep. There is an independent literature implicating GHRH in sleep regulation.[31,70] Administration of GHRH promotes NREM sleep, whereas antagonizing GHRH or its receptor inhibits spontaneous NREM sleep and blocks the increase in NREM sleep induced by IL-1. Finally, as mentioned previously, the GHRH receptor seems necessary for an effective response to viral challenge.[38]

The mechanisms by which sleep regulatory substances (SRSs) are regulated and induce sleep are beginning to be understood. TNF and IL-1 neuronal expression is enhanced

in response to afferent nerve activity. For instance, excessive stimulation of rat facial whiskers for 2 hours enhances IL-1 and TNF immunoreactivity in the cortical layers of the somatosensory cortical columns that receive the enhanced afferent input.[90]

What is it about neuronal activity or wakefulness that causes the enhanced SRS activity? Neuronal activity manifests as presynaptic and postsynaptic events that act in both the short and long term. Neuronal activity in presynaptic neurons results in the release of transmitters and adenosine triphosphate (ATP).[99] In turn, some of that ATP is converted to adenosine, and some ATP acts on purine P2X7 receptors on glia to release TNF and IL-1.[46,100] ATP also acts to release cytokines in immunocytes.[101] The extracellular adenosine derived from ATP interacts with neurons through the adenosine A_1 receptor (A1AR). The TNF released in response to ATP activates NFκB in postsynaptic and presynaptic neurons.[46] NFκB enhances the A1AR, thereby rendering the cell more sensitive to adenosine. NFκB also enhances production of a subunit of the AMPA receptor gluR1 mRNA. The time courses of enhanced mRNA for receptors or ligands are much slower than the direct actions of adenosine or TNF; the subsequent production of protein offers a way for the brain to keep track of prior neuronal network activity and translate that activity into a greater sleep propensity. The various time courses of action of the neurotransmitters (milliseconds), the conversions of ATP to adenosine and its actions (seconds), and the actions of ATP-induced release of cytokines and their subsequent effects on gene transcription and translation (minutes to hours) provide a mechanism for activity-dependent oscillations of neuronal assembly sleep.[102]

There is a growing literature demonstrating direct effects of IL-1 and TNF on neural substrates implicated in the regulation of sleep. Some of these mechanisms include interactions with classic neurotransmitters such as glutamate, serotonin, acetylcholine, gamma-aminobutyric acid, histamine, and dopamine.[103] For example, IL-1 increases serotonergic activity in brain regions implicated in sleep regulation,[104] and an intact serotonergic system is required for the full effects of IL-1 on sleep to manifest.[105,106] IL-1 inhibits discharge rates of serotonergic[107,108] and cholinergic[109] neurons in brainstem. Within the hypothalamus, IL-1 increases c-Fos[110] and inhibits wake-active neurons.[111] TNF promotes sleep if microinjected into the anterior hypothalamus, whereas injection of a soluble TNF receptor into this area reduces sleep.[112] TNF also alters sleep if injected into the locus coeruleus,[113] effects likely related to interactions with α_2-adrenergic receptive mechanisms and norepinephrine release.[114] Interestingly, TNF or IL-1, if applied locally onto the surface of the cerebral cortex unilaterally, enhances EEG delta activity on the side to which it is applied but not the contralateral side.[115,116] Conversely, application of the TNF soluble receptor unilaterally onto the cortex of sleep-deprived rats attenuates sleep loss–induced EEG delta activity on the side injected, but not on the opposite side. Further, unilateral application of a TNF siRNA (inhibits TNF) reduces spontaneous cortical TNF expression and EEG slow wave activity on the ipsilateral side. These latter studies suggest that TNF acts locally within the cortex (in addition to its somnogenic actions in the hypothalamus) to enhance EEG synchronization and possibly sleep intensity. In fact, application of TNF directly to the cortex enhances the probability of individual cortical columns entering into a sleep-like state.[90]

CLINICAL PEARL

Although physicians routinely prescribe bed rest to aid in recuperation from infections and other maladies, as yet there is little direct evidence that sleep aids in recuperation. Such studies are difficult to perform because the recovery from an infection, for instance, is influenced by the baseline severity of the infection (i.e., differences in exposure or innate resistance that determine the replication level and clearance of the invading microbe) as well as by what the patient does during the infection. Physicians will continue to prescribe bed rest, and often this is just what the patient desires. It seems likely that such advice is beneficial because enhanced sleep is part of the adaptive APR. The only evidence of which we are aware that is relevant to this issue is consistent with the concept that sleep aids in recuperation; after infectious challenge, animals that have robust NREM sleep responses have a higher probability of survival than animals that fail to exhibit NREM sleep responses.[117] Although strictly correlative, these data suggest that sleep does indeed facilitate recovery. Perhaps our grandmothers' folk wisdom pertaining to the preventive and curative attributes of sleep and sickness is correct, although much additional research is needed before we know whether this admonishment has a biologic basis.

SUMMARY

Sleepiness, like fever, is commonly experienced at the onset of an infection or other cause of systemic inflammation. Changes in sleep in response to microbes appear to be one facet of the acute phase response. Typically, soon after infectious challenge, time spent in NREM sleep increases and REM sleep is suppressed. The exact time course of sleep responses depends on the infectious agent, the route of administration, and the time of day the infectious challenge is given.

There is a common perception that sleep loss renders one vulnerable to infection. Some studies demonstrate that sleep loss impairs acquired immunity, and many studies have shown that sleep deprivation alters selected aspects of the innate immune response. A few studies have combined sleep deprivation with infectious challenge. After mild sleep deprivation, several immune system parameters (e.g., NK cell activity) change, and resistance to a viral challenge is decreased in individuals who spontaneously sleep less. Studies have not yet been done to determine the effects of sleep deprivation on recovery from an infection.

The molecular mechanisms responsible for the changes in sleep associated with infection appear to be an amplification of a physiologic sleep regulatory biochemical cascade. Sleep regulatory mechanisms and the immune system share regulatory molecules. The best characterized are IL-1 and TNF, which are involved in physiologic NREM sleep regulation. IL-1 and TNF are key players in the development of the acute phase response induced by infectious agents. During the initial response to infectious challenge these proinflammatory cytokines are upregulated, leading to the acute phase sleep response. This chain of events includes well-known immune response modifiers such as prostaglandins, nitric oxide, and adenosine. Each of these substances, and their receptors, is a normal constituent of the brain, and each is involved in physiologic sleep regulation.

ACKNOWLEDGMENTS

During the writing of this chapter, the authors were supported, in part, by grants from the National Institutes of Health: NS-25378 and HD 36520 (JMK) and AG041827 (MRO).

Selected Readings

Besedovsky L, Lange T, Born J. Sleep and immune function. *Pflugers Arch* 2012;**463**:121–37.

Carroll JE, Irwin MR, Merkin SS, Seeman TE. Sleep and multisystem biological risk: a population-based study. *PLoS ONE* 2015;**10**:e0118467.

Frank MG. Astroglial regulation of sleep homeostasis. *Curr Opin Neurobiol* 2013;**23**:812–18.

Haus EL, Smolensky MH. Shift work and cancer risk: potential mechanistic roles of circadian disruption, light at night, and sleep deprivation. *Sleep Med Rev* 2013;**17**:273–84.

Imeri L, Opp MR. How (and why) the immune system makes us sleep. *Nat Rev Neurosci* 2009;**10**:199–210.

Ingiosi AM, Opp MR, Krueger JM. Sleep and immune function: glial contributions and consequences of aging. *Curr Opin Neurobiol* 2013;**23**:806–11.

Lange T, Dimitrov S, Born J. Effects of sleep and circadian rhythm on the human immune system. *Ann N Y Acad Sci* 2010;**1193**:48–59.

Mahlios J, De la Herran-Arita AK, Mignot E. The autoimmune basis of narcolepsy. *Curr Opin Neurobiol* 2013;**23**:767–73.

McCusker RH, Kelley KW. Immune-neural connections: how the immune system's response to infectious agents influences behavior. *J Exp Biol* 2013;**216**:84–98.

Mehra R, Redline S. Sleep apnea: a proinflammatory disorder that coaggregates with obesity. *J Allergy Clin Immunol* 2008;**121**:1096–102.

Motivala SJ. Sleep and inflammation: psychoneuroimmunology in the context of cardiovascular disease. *Ann Behav Med* 2011;**42**:141–52.

Opp MR, Krueger JM. Sleep and immunity: a growing field with clinical impact. *Brain Behav Immun* 2015. pii: S0889-1591(15)00082-3. doi:10.1016/j.bbi.2015.03.011; [Epub ahead of print].

Porkka-Heiskanen T. Sleep homeostasis. *Curr Opin Neurobiol* 2013;**23**:799–805.

Preston BT, Capellini I, McNamara P, et al. Parasite resistance and the adaptive significance of sleep. *BMC Evol Biol* 2009;**9**:7.

Zielinski MR, Krueger JM. Sleep and innate immunity. *Front Biosci (Schol Ed)* 2011;**3**:632–42.

A complete reference list can be found online at ExpertConsult.com.

Chapter
20

Endocrine Physiology in Relation to Sleep and Sleep Disturbances

Eve Van Cauter; Esra Tasali

Chapter Highlights

- Sleep and circadian rhythmicity have distinct modulatory effects on endocrine and metabolic function and affect activity of the hypothalamic-pituitary axes, carbohydrate metabolism, appetite regulation, and the hormone control of blood pressure and body-fluid balance.
- Sleep curtailment has become an endemic behavior in modern society. Current evidence from both epidemiologic and laboratory studies suggests that insufficient sleep due to either sleep curtailment or sleep disorders has deleterious effects on hormones, glucose metabolism, and body weight regulation.
- Reciprocally, the most common endocrine disorders, including obesity, diabetes, and polycystic ovary syndrome, are associated with a higher prevalence of and risk for sleep disorders, particularly obstructive sleep apnea.
- This chapter reviews the effects of sleep and sleep disturbances on the endocrine system, the impact of reduced sleep duration and quality on hormonal and metabolic function, age-related alterations in sleep and endocrine function, and the adverse metabolic consequences of sleep disturbances in obesity, type 2 diabetes, and polycystic ovary syndrome.

This chapter is divided in three main sections. We start with a review of the interactions between sleep and endocrine release in the hypothalamic-pituitary axes and the roles of sleep in carbohydrate metabolism, appetite regulation, and hormone control of body-fluid balance in healthy adults. Table 20-1 provides basic information about the hormones that will be discussed in this chapter. We then summarize the growing body of evidence linking decrements of sleep duration or quality that occur with sleep restriction, in sleep disorders, or as a result of normal aging with disturbances of endocrine and metabolic function. Lastly, we review recent evidence linking disorders of sleep-wake regulation with metabolic and endocrine diseases, including obesity, type 2 diabetes, and polycystic ovary syndrome (PCOS). For a review of sleep abnormalities in other endocrine diseases, the reader is referred to other sections in this book.

MODULATION OF ENDOCRINE FUNCTION BY SLEEP-WAKE HOMEOSTASIS AND CIRCADIAN RHYTHMICITY

In healthy adults, reproducible changes of essentially hormonal and metabolic variables occur during sleep and around wake-sleep and sleep-wake transitions. These daily events reflect the interaction of central circadian rhythmicity and sleep-wake homeostasis. Pathways by which circadian rhythmicity and sleep-wake homeostasis affect peripheral endocrine function and metabolism include the modulation of the activity of the hypothalamic releasing and inhibiting factors, the autonomous nervous system control of endocrine organs, and the 24-hour periodicity of circulating glucocorticoids. Findings from genome-wide association and epidemiologic studies also support a role of circulating melatonin levels on specific endocrine targets, including the pancreatic beta cells.[1-4]

Circadian oscillations can be generated in many peripheral organs, including tissues that release endocrine signals such as adipocytes, liver, adrenal glands, and pancreatic beta cells.[5,6] These "local" oscillators appear to be under the control of the central pacemaker in the suprachiasmatic nuclei either directly through neural or endocrine signals, or indirectly through its control of behavioral rhythms such as the sleep-wake cycle and feeding schedule.

To differentiate between effects of circadian rhythmicity and those subserving sleep-wake homeostasis, researchers have used experimental strategies that take advantage of the fact that rhythms primarily under the control of the central circadian pacemaker take several days to adjust to a large sudden shift of sleep-wake and light-dark cycles (such as occur in jet lag and shift work). Such strategies allow for the effects of circadian modulation to be observed in the absence of sleep and for the effects of sleep to be observed at an abnormal circadian time. Figure 20-1 illustrates mean profiles of hormonal plasma concentrations, glucose levels, and insulin-secretion rates (ISRs) observed in healthy subjects who were studied before and during an abrupt 12-hour delay of the sleep-wake and dark-light cycles, for normal-deprived-recovery sleep periods. To eliminate the effects of feeding, fasting, and postural changes, the subjects remained recumbent throughout the study, and the normal meal schedule was replaced by intravenous glucose infusion at a constant rate.[7]

As shown in Figure 20-1, this drastic manipulation of sleep had only modest effects on the wave shape of the cortisol profile, in sharp contrast with the immediate shift of the growth hormone (GH) and prolactin (PRL) rhythms that followed the shift of the sleep-wake cycle. The temporal

Table 20-1 Origin and Main Action of Hormones

Hormone	Main Secreting Organ	Main Action in Adults
Growth hormone (GH)	Pituitary gland	Anabolic hormone that regulates body composition
Prolactin (PRL)	Pituitary gland	Stimulates lactation in women; pleiotropic actions
Adrenocorticotropic hormone (ACTH)	Pituitary gland	Stimulates release of cortisol from adrenal cortex
Cortisol	Adrenal cortex	Stress hormone, antiinsulin effects
Thyroid-stimulating hormone (TSH)	Pituitary gland	Stimulates the release of thyroid hormones from the thyroid gland
Luteinizing hormone (LH)	Pituitary gland	Stimulates ovarian and testicular function
Follicle-stimulating hormone (FSH)	Pituitary gland	Stimulates ovarian and testicular function
Testosterone	Gonads	Stimulates spermatogenesis
Estradiol	Ovaries	Stimulates follicular growth
Insulin	Pancreas	Regulates blood glucose levels
Melatonin	Pineal gland	Hormone of the dark that transmits information about the light-dark cycle
Leptin	Adipose tissue	Satiety hormone regulating energy balance
Ghrelin	Stomach	Hunger hormone regulating energy balance

organization of thyroid-stimulating hormone (TSH) secretion appears to be influenced equally by circadian and sleep-dependent processes. Indeed, the evening elevation of TSH levels occurs well before sleep onset and has been shown to reflect circadian phase. During sleep, as further described later, an inhibitory process prevents TSH concentrations from rising further. Consequently, in the absence of sleep, the nocturnal TSH elevation is markedly amplified. Both sleep and time of day clearly modulated glucose levels and ISRs. Nocturnal elevations of glucose and ISRs occur even when the subjects are sleep deprived, and recovery sleep at an abnormal circadian time is also associated with elevated glucose level and ISR but at a lower amplitude. This pattern of changes in glucose levels and ISRs reflects changes in glucose use because, when glucose is infused exogenously, as described earlier and illustrated in the study in Figure 20-1 (central section of glucose secretion pattern, pink bar period), endogenous glucose production is largely inhibited.

The Growth Hormone Axis

Pituitary release of GH is stimulated by hypothalamic growth hormone–releasing hormone (GHRH) and inhibited by somatostatin. In addition, the acylated form of ghrelin, a peptide produced predominantly by the stomach, binds to the growth hormone secretagogue receptor and is a potent endogenous stimulus of GH secretion.[8] There is a combined and probably synergic role of GHRH stimulation, elevated nocturnal ghrelin levels, and decreased somatostatinergic tone in the control of GH secretion during sleep. Although sleep clearly involves major stimulatory effects on GH secretion, the hormones of the somatotropic axis, including GHRH, ghrelin, and GH itself, in turn appear to be involved in sleep regulation.[9]

In healthy adult subjects, the 24-hour profile of plasma GH levels consists of stable low levels abruptly interrupted by bursts of secretion. The most reproducible GH pulse occurs shortly after sleep onset.[10] In men, the sleep-onset GH pulse is generally the largest, and often the only, secretory pulse observed over the 24-hour span. In women, daytime GH pulses are more frequent, and the sleep-associated pulse, although still present in most individual profiles, does not account for the majority of the 24-hour secretory output. Sleep onset elicits a pulse in GH secretion whether sleep is advanced, delayed, or interrupted and reinitiated. The mean GH secretion profile shown in Figure 20-1 illustrates the maintenance of the relationship between sleep onset and GH release in subjects who underwent a 12-hour delay shift of the sleep-wake cycle. There is a consistent relationship between the appearance of delta waves in the electroencephalogram (EEG) and elevated GH concentrations, and maximal GH release occurs within minutes of the onset of slow wave sleep (SWS).[10,11] In healthy young men, there is a quantitative correlation between the amount of GH secreted during the sleep-onset pulse and the duration of the slow wave episode.[12] Pharmacologic stimulation of SWS increases in GH secretion.[13,14] Sedative hypnotics that are ligands of the gamma-aminobutyric acid A receptor, such as benzodiazepines and imidazopiridines, do not increase nocturnal GH release, consistent with their lack of stimulation of slow wave activity.[15]

The robust relationship between early sleep and GH release is consistent with a synchronization between anabolic processes in the body and a state when behavioral rest occurs and cerebral glucose use is at its lowest point.[16] There is good evidence to indicate that stimulation of nocturnal GH release and stimulation of SWS reflect, to a large extent, synchronous activity of at least two populations of hypothalamic GHRH neurons.[16] Sleep-onset GH secretion appears to be primarily regulated by GHRH stimulation occurring during a period of decreased somatostatin inhibition of somatotropic activity. Indeed, in humans, GH secretion during early sleep may be nearly totally suppressed by administration of a GHRH antagonist.[17] The late evening and nocturnal hours coincide with the trough of a diurnal variation in hypothalamic somatostatin tone[18] that is likely to facilitate nocturnal GH release. It is also possible that ghrelin plays a role in causing increased GH secretion during sleep because the postdinner rebound of

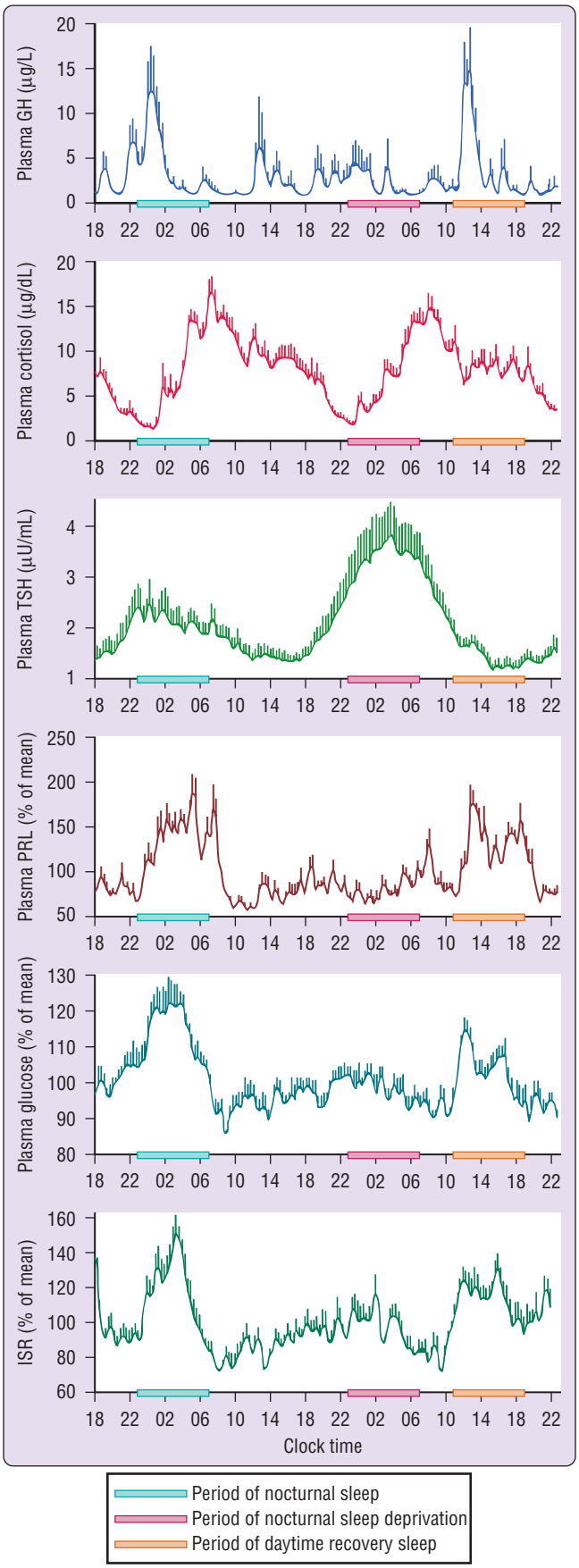


Figure 20-1 *From top to bottom:* Mean 24-hour profiles of plasma growth hormone (GH), cortisol, thyrotropin (TSH), prolactin (PRL), glucose, and insulin secretion rates (ISR) in a group of eight healthy young men (20 to 27 years old) studied during a 53-hour period including 8 hours of nocturnal sleep (*blue horizontal bar*), 28 hours of sleep deprivation (*red bar*), and 8 hours of daytime sleep (*orange bar*). The *vertical bars* on the tracings represent the standard error of the mean (SEM) at each time point. The *blue bars* represent the sleep periods. The *red bars* represent the period of nocturnal sleep deprivation. The *orange bars* represent the period of daytime sleep. Caloric intake was exclusively under the form of a constant glucose infusion. Shifted sleep was associated with an immediate shift of GH and PRL release. In contrast, the secretory profiles of cortisol and TSH remained synchronized to circadian time. Both sleep-dependent and circadian inputs can be recognized in the profiles of glucose and ISR. (Modified from Van Cauter E, Spiegel K. Circadian and sleep control of endocrine secretions. In: Turek FW, Zee PC, editors. *Neurobiology of sleep and circadian rhythms.* New York: Marcel Dekker; 1999; and Van Cauter E, Blackman JD, Roland D, et al. Modulation of glucose regulation and insulin secretion by sleep and circadian rhythmicity. *J Clin Invest* 1991;88:934–42.)

ghrelin levels results in maximal concentrations during the early part of the night.[19-21]

The upper panel of Figure 20-1 shows that the secretion of GH is increased during sleep independently of the circadian time when sleep occurs and that sleep deprivation results (the pink bar period on the figure) in greatly diminished release of this hormone. However, a slight increase may be observed during nocturnal sleep deprivation, suggesting the existence of a weak circadian component that could reflect, as discussed earlier, lower somatostatin inhibition. Following a night of total sleep deprivation, GH release is increased during the daytime such that the total 24-hour secretion is not significantly affected.[22] Again, the mechanisms underlying this compensatory daytime secretion are unknown, but they could involve decreased somatostatinergic tone or elevated ghrelin levels, as have been observed in experimental studies of partial or total sleep deprivation.[23,24]

Marked rises in GH secretion before the onset of sleep have been reported by several investigators.[25-27] Presleep GH pulses may reflect the presence of a sleep debt because they occur consistently after recurrent experimental sleep restriction.[28] The short-term negative feedback inhibition exerted by GH on its own secretion may also explain observations of an absent or reduced GH pulse during the first slow wave period, when a secretory pulse occurred before sleep onset. Awakenings interrupting sleep have an inhibitory effect on GH release.[29,30] Thus sleep fragmentation generally decreases nocturnal GH secretion.

The Corticotropic Axis

Activity of the corticotropic axis—a neuroendocrine system associated with the stress response and behavioral activation—may be measured peripherally through plasma levels of the pituitary adrenocorticotropic hormone (ACTH) and of cortisol, the adrenal hormone directly controlled by ACTH stimulation. The plasma levels of these hormones decline from an early morning to maximal level throughout the daytime and are near the lower limit of most assays in the late evening and early part of the sleep period. Although the rhythm of ACTH reflects a circadian variation in corticotropin-releasing hormone (CRH) activity, itself under control by the central circadian pacemaker, a peripheral clock in the adrenals enhances the rhythm of glucocorticoid release, one of the

largest and most robust rhythms in humans.[31,32] Sleep is normally initiated when corticotropic activity is quiescent. Reactivation of ACTH and cortisol secretion occurs abruptly a few hours before the usual waking time.

The mean cortisol secretion profile shown in Figure 20-1 illustrates the remarkable persistence of this diurnal variation even when sleep is manipulated. Nonetheless, modulatory effects of sleep or wake have been clearly demonstrated. Indeed, a number of studies have indicated that sleep onset is reliably associated with a short-term inhibition of cortisol secretion,[7,33] although this effect may not be detectable when sleep is initiated at the time of the daily highest corticotropic activity, that is, in the morning.[34] Under normal conditions, because cortisol secretion is already quiescent in the late evening, this inhibitory effect of sleep, which is temporally associated with the occurrence of slow wave sleep,[35-37] results in a prolongation of the quiescent period. Therefore under conditions of sleep deprivation (pink bar period, Figure 20-1), the nadir of cortisol secretion is less pronounced and occurs earlier than under normal conditions of nocturnal sleep. Conversely, awakening at the end of the sleep period is consistently followed by a pulse of cortisol secretion.[7,30,38]

During sleep deprivation, the rapid effects of sleep onset and sleep offset on corticotropic activity are obviously absent, and, as may be seen in the profiles shown in Figure 20-1 (left side of cortisol secretion pattern), the nadir of cortisol level is slightly higher than during nocturnal sleep (because of the absence of the inhibitory effects of the first hours of sleep), and the morning maximal peak is slightly lower (because of the absence of the stimulating effects of morning awakening). Overall, the amplitude of the rhythm is reduced by about 15% during sleep deprivation compared with normal conditions. In addition to the immediate modulatory effects of sleep-wake transitions on cortisol levels, nocturnal sleep deprivation, even partial, results in an elevation of cortisol levels on the following evening (not shown in Figure 20-1).[39] Sleep loss thus appears to delay the normal return to evening quiescence of the corticotropic axis.

The Thyroid Axis

Daytime levels of plasma TSH are low and relatively stable until the initiation of a rapid elevation in the early evening resulting in maximal concentrations around the beginning of the sleep period.[37,40] The later part of sleep is associated with a progressive decline in TSH levels, and daytime values resume shortly after morning awakening. The first 24 hours of the study illustrated in Figure 20-1 are typical of the diurnal TSH rhythm. Because the nocturnal rise of TSH occurs well before the time of sleep onset, it probably reflects a circadian effect. A marked effect of sleep on TSH secretion may be seen during sleep deprivation (as shown in Figure 20-1), when nocturnal TSH secretion is increased by as much as 200% over the levels observed during nocturnal sleep. Thus sleep exerts an inhibitory influence on TSH secretions, and sleep deprivation relieves this inhibition.[37,41]

Interestingly, when sleep occurs during daytime hours, TSH secretion is not suppressed significantly below normal daytime levels, indicating once again the interaction between the effects of circadian time and sleep effects. When the depth of sleep at the habitual time is increased by prior sleep deprivation, the nocturnal TSH rise is more markedly inhibited, suggesting that SWS is probably the primary determinant of the sleep-associated fall.[37] Awakenings interrupting nocturnal sleep appear to relieve the inhibition of TSH and are consistently associated with a short-term TSH elevation.

Circadian and sleep-related variations in thyroid hormones have been difficult to demonstrate, probably because these hormones have long half-lives and are bound to serum proteins. Thus their peripheral concentrations are affected by diurnal variations in hemodilution caused by postural changes. However, under conditions of sleep deprivation, the increased amplitude of the TSH rhythm may result in a detectable increase in plasma triiodothyronine (T_3) levels, paralleling the nocturnal TSH rise.[42] If sleep deprivation is continued for a second night, the nocturnal rise of TSH is markedly diminished compared with that occurring during the first night.[42,43] It is likely that following the first night of sleep deprivation, the elevated thyroid hormone levels, which persist during the daytime period because of the prolonged half-life of these hormones, limit the subsequent TSH rise at the beginning of the next nighttime period. Data from a study of 64 hours of sleep deprivation suggest that prolonged sleep loss may be associated with an upregulation of the thyroid axis, with lower levels of TSH and higher levels of thyroid hormones.[44] Findings of elevations in free thyroxine (T_4) index and of peripheral levels of free T_3 and free T_4 in subjects submitted to experimental sleep restriction or total sleep deprivation are consistent with this hypothesis.[45-48]

Prolactin Secretion

Under normal conditions, PRL levels undergo a major nocturnal elevation starting shortly after sleep onset and culminating around midsleep. Decreased dopaminergic inhibition of PRL during sleep is likely to be the primary mechanism underlying this nocturnal PRL elevation. In adults of both sexes, the nocturnal maximum is about twofold higher than mean daytime levels.[42] Morning awakenings and awakenings interrupting sleep are both consistently associated with a rapid inhibition of PRL secretion.[42]

Studies of the PRL profile during daytime naps or after shifts of the sleep period have consistently demonstrated that sleep onset, irrespective of the time of day, has a stimulatory effect on PRL release. This is well illustrated by the profiles shown in Figure 20-1, in which elevated PRL levels occur both during nocturnal sleep and during daytime recovery sleep, whereas the nocturnal period of sleep deprivation was not associated with an increase in PRL concentrations. However, the sleep-related rise of PRL may still be present, although with a reduced amplitude, when sleep does not occur at the normal nocturnal time. Maximal stimulation is observed only when sleep and circadian effects are superimposed, suggesting that circadian rhythm is not the main driver of PRL release.[49-51] A close temporal association between increased PRL secretion and slow wave activity is apparent.[52] However, in contrast to the correlation between slow wave activity and amount of GH release that has been evidenced in men, no such "dose-response" relationship has been demonstrated for PRL in either men or women.

Benzodiazepine and imidazopiridine hypnotics taken at bedtime may cause an increase in the nocturnal PRL rise, resulting in concentrations near the pathologic range for hyperprolactinemia for part of the night.[53,54] A potential mechanism is a greater suppression of dopaminergic activity under zolpidem versus placebo because dopamine has been

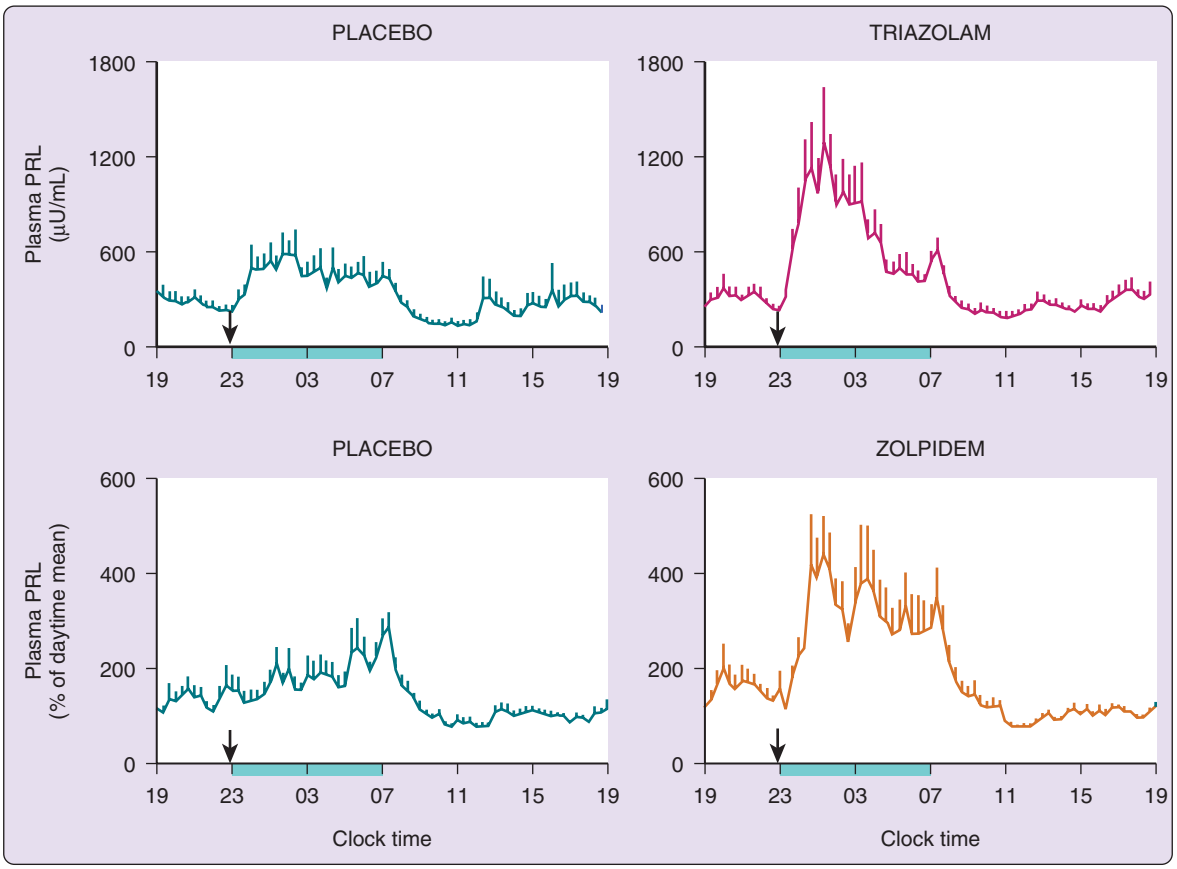


Figure 20-2 Effects of commonly used hypnotics on the 24-hour profile of plasma prolactin (PRL) in healthy young subjects. Data are mean plus standard error of the mean. Samples were collected at 15- to 20-minute intervals. Sleep was polygraphically recorded. *Top,* Effects of bedtime administration of triazolam (0.5 mg). *Bottom,* Effects of bedtime administration of zolpidem (10 mg). Both benzodiazepine and nonbenzodiazepine hypnotics cause transient hyperprolactinemia during the early part of sleep. Time in bed is denoted by the *black bars*. *Arrows* denote time of drug administration. (Data from Copinschi G, Van Onderbergen A, L'Hermite-Balériaux M, et al. Effects of the short-acting benzodiazepine triazolam taken at bedtime on circadian and sleep-related hormonal profiles in normal men. *Sleep* 1990;13:232–44; Copinschi G, Akseki E, Moreno-Reyes R, et al. Effects of bedtime administration of zolpidem on circadian and sleep-related hormonal profiles in normal women. *Sleep* 1995;18:417–24; and Van Cauter E, Spiegel K. Circadian and sleep control of endocrine secretions. In: Turek FW, Zee PC, editors. *Neurobiology of sleep and circadian rhythms.* New York: Marcel Dekker; 1999.)

identified as a PRL-inhibiting factor. This is illustrated for triazolam and zolpidem in Figure 20-2. Neither triazolam nor zolpidem has any effect on the 24-hour profiles of cortisol, melatonin, or GH. Chronic treatment of insomnia with the melatonin receptor agonist ramelteon increases PRL release in women, but not in men.[55]

There is evidence from animal studies that PRL is involved in the humoral regulation of REM sleep.[56] The primary effect is a stimulation of REM sleep, which appears to be dependent on time of day. PRL-deficient mice have decreased REM sleep.[57]

The Gonadal Axis

The relationship between sleep and the 24-hour patterns of gonadotropin release and gonadal steroid levels varies according to age and sex (for review, see Copinschi and Challet[42]). Before puberty, luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are secreted in a pulsatile pattern, and an augmentation of pulsatile activity is associated with sleep onset in a majority of both girls and boys. The increased amplitude of gonadotropin release during sleep is one of the hallmarks of puberty.

During the transition from puberty to adulthood, the amplitude of daytime LH pulses increases, and in adult men, the day-night variation of plasma LH levels is dampened or even undetectable. During the sleep period, LH pulses appear to be temporally related to the rapid eye movement (REM) sleep and non–rapid eye movement (NREM) sleep cycle.[58] Despite the low amplitude of the nocturnal increase in gonadotropin release, a marked diurnal rhythm in circulating testosterone levels is present, with minimal levels in the late evening, a robust rise following sleep onset, and maximal levels in the early morning.[59,60] Thus the robust circadian rhythm of plasma testosterone may be partially controlled by factors other than LH. The nocturnal rise of testosterone appears temporally linked to the duration of the first NREM period[61] because plasma levels continue to rise until the first REM episode occurs. A robust rise of testosterone may also be observed during daytime sleep, suggesting that sleep, irrespective of time of day, stimulates gonadal hormone release.[62]

Experimental sleep fragmentation in young men resulted in attenuation of the nocturnal rise of testosterone, particularly in subjects who did not achieve REM sleep.[63] Androgen concentrations in young adults decline significantly during periods of total sleep deprivation and recover promptly after the sleep of the subjects is restored.[62,64] In contrast, pharmacologic suppression of testosterone in healthy men appears to have no effect on the total amount and overall architecture of nighttime sleep.[65] In older men, the amplitude of LH pulses is decreased, but the frequency is increased and no significant diurnal pattern can be detected.[66-68] With aging, the circadian variation of testosterone persists, although it is markedly dampened.[68] The sleep-related rise is still apparent in older men, but its amplitude is lower and the relationship to REM latency is no longer apparent.[69] It is likely that decreases in slow wave activity as occurs in aging as well as in sleep disorders (e.g., obstructive sleep apnea [OSA]) plays a role in the dampening of the sleep-related testosterone rise. In otherwise healthy older men, morning testosterone levels are partly predicted by the amount of nighttime sleep, whether measured at home or in the laboratory,[70] suggesting that habitual sleep duration should be taken into account in the diagnosis of androgen deficiency.

In women, the 24-hour variation in plasma LH is markedly modulated by the menstrual cycle.[71,72] In the early follicular phase, LH pulses are large and infrequent, and a marked slowing of the frequency of secretory pulses occurs during sleep, suggestive of inhibitory effect of sleep on pulsatile GnRH release. Awakenings interrupting sleep are usually associated with a pulse of LH concentration.[73] In the midfollicular phase, pulse amplitude is decreased, pulse frequency is increased, and the frequency modulation of LH pulsatility by sleep is less apparent. Pulse amplitude increases again by the late follicular phase. In the early luteal phase, at the opposite, the pulse amplitude is markedly increased, the pulse frequency is decreased, and nocturnal slowing of pulsatility is again evident. In the mid and late luteal phase, pulse amplitude and frequency are decreased and there is no modulation by sleep. In postmenopausal women, gonadotropin levels are elevated, but they show no consistent circadian pattern.[74] A recent well-documented study has demonstrated a causal relationship between the elevation of gonadotropin levels, hot flashes, and decreases in objective and subjective sleep quality.[75] A number of studies[76-78] have indicated that estrogen replacement therapy has modest beneficial effects on subjective and objective sleep quality, particularly in the presence of environmental disturbance[79] or sleep-disordered breathing.[76,77,80]

Glucose Regulation

The consolidation of human sleep in a single 7- to 9-hour period implies that an extended period of fast must be maintained overnight. Despite the prolonged fasting condition, glucose levels remain relatively stable across the night. In contrast, if subjects are awake and fasting in a recumbent position, glucose levels fall by an average of 0.5 to 1.0 mmol/L (±10 to 20 mg/dl) over a 12-hour period.[81] Thus a number of mechanisms that operate during nocturnal sleep must intervene to maintain stable glucose levels during the overnight fast.

The lower panels of Figure 20-1 show profiles of blood glucose and insulin ISRs obtained in normal subjects who were studied under conditions of constant glucose infusion, a condition that results in a marked inhibition of endogenous glucose production. Thus changes in plasma glucose levels illustrated in the figure mainly reflect changes in glucose use. A marked decrease in glucose tolerance (reflected in higher plasma glucose levels under this condition of continuous challenge by the exogenous glucose infusion) is apparent during nighttime as well as daytime sleep. A smaller elevation of glucose and insulin also occurs during nocturnal sleep deprivation, indicating an effect of circadian-dependent mechanisms. Recovery sleep was associated with a robust increase in glucose and insulin, owing to the release of GH linked to sleep onset.

During nocturnal sleep, the overall increase in plasma glucose ranged from 20% to 30%, despite the maintenance of rigorously constant rates of caloric intake, that is, constant glucose infusion. Maximal levels are reached around the middle of the sleep period. During the later part of the night, glucose tolerance begins to improve, and glucose levels progressively decrease toward morning values. The mechanisms underlying these robust variations in set-point of glucose regulation across nocturnal sleep are different in early sleep and late sleep.

It is estimated that about two thirds of the fall in overall body glucose use during early sleep is due to a decrease in brain glucose metabolism[82] related to the predominance of slow wave stages, which are associated with a 30% to 40% reduction in cerebral glucose metabolism compared with the waking state (see Chapter 18).[83] The remainder of the fall would then reflect decreased peripheral use, including diminished muscle tone and rapid hyperglycemic effects of the sleep-onset GH pulse. Furthermore, the nocturnal elevation of melatonin levels could contribute to the nocturnal decrease in glucose tolerance because of an inhibitory effect of melatonin on insulin release from beta cells.[2,84] During the later part of the sleep period, glucose levels and insulin secretion decrease to return to presleep values, and this decrease appears to be partially due to the increase in wake and REM stages.[85] Indeed, glucose use during the REM and wake stages is higher than during NREM stages.[83] In addition, several other factors may also contribute to the decline of glucose levels during late sleep. These include an increase in insulin sensitivity due to a delayed effect of low cortisol levels during the evening and early part of the night.[86]

Sleep and Appetite Regulation

Sleep plays an important role in energy balance. In rodents, food shortage or starvation results in decreased sleep[87] and, conversely, total sleep deprivation leads to marked hyperphagia.[88] The identification of hypothalamic excitatory neuropeptides, referred to as hypocretins or orexins, that have potent wake-promoting effects and stimulate food intake has provided a molecular basis for the interactions between the regulation of feeding and sleeping.[89,90] Orexin-containing neurons in the lateral hypothalamus project directly to the locus coeruleus and other brainstem and hypothalamic arousal areas, where they interact with the leptin-responsive neuronal network involved in balancing food intake and energy expenditure. Orexin-containing neurons are active during waking and quiescent during sleep. Orexin activity is inhibited by leptin, a satiety hormone, and stimulated by ghrelin, an appetite-promoting hormone. Multiple peptides derived from the gut and adipose tissues participate in the control of hunger and satiety. A relationship with sleep has

been demonstrated in some studies for two of them, leptin and ghrelin.[91,92]

Leptin, a hormone released by the adipocytes, provides information about energy status to regulatory centers in the hypothalamus.[93] Circulating leptin concentrations in humans show a rapid decline or increase in response to acute caloric shortage or surplus, respectively. These changes in leptin levels have been associated with reciprocal changes in hunger. The 24-hour leptin profile shows a marked nocturnal rise.[94] The upper panel of Figure 20-3 shows a typical 24-hour profile of plasma leptin levels in a normal man. The nocturnal elevation of leptin has been thought to suppress the hunger during the overnight fast. Although daytime food intake plays a major role in the nocturnal rise of leptin, a study using continuous enteral nutrition to eliminate the impact of meal intake showed the persistence of a sleep-related leptin elevation, although the amplitude was lower than during normal feeding conditions.[95] Prolonged total sleep deprivation results in a decrease in the amplitude of the leptin diurnal variation.[96]

Ghrelin, a peptide produced predominantly by the stomach, is also involved in regulating energy balance[8] and stimulates appetite.[97] Daytime profiles of plasma ghrelin levels are primarily regulated by the schedule of food intake: levels drop sharply after each meal intake and rebound in parallel with increased hunger until the initiation of the following meal.[98] The 24-hour profile of ghrelin levels shows a marked nocturnal rise, which is only modestly dampened when subjects are sleep deprived.[19] The nocturnal ghrelin rise partly represents the rebound of ghrelin following the dinner meal. Despite the persistence of the fasting condition, ghrelin levels do not continue to increase across the entire sleep period and instead decrease during the later part of the night, consistent with an inhibitory effect of sleep on ghrelin release.[21] The lower panel of Figure 20-3 illustrates a representative 24-hour profile of ghrelin from a normal subject who ingested three carbohydrate-rich meals.

Water and Electrolyte Balance during Sleep

Water and salt homeostasis is under the combined control of vasopressin, a hormone released by the posterior pituitary, the renin-angiotensin-aldosterone system, and the atrial natriuretic peptide. Urine flow and electrolyte excretion are higher during the day than during the night, and this variation partly reflects circadian modulation. In addition to this 24-hour rhythm, urine flow and osmolarity oscillate with the REM-NREM cycle. REM sleep is associated with decreasing urine flow and increasing osmolarity.

Vasopressin release is pulsatile but without apparent relationship to sleep stages.[99] Levels of atrial natriuretic peptide are relatively stable and do not show fluctuations related to the sleep-wake or REM-NREM cycle.[100] Whether the levels of plasma atrial natriuretic peptide exhibit a circadian variation is still a matter of controversy.[100]

Figure 20-4 illustrates the 24-hour rhythm of plasma renin activity (PRA) in a subject studied during a normal sleep-wake cycle and following a shift of the sleep period.[101] The initiation of sleep, irrespective of time of day, is associated with a robust increase in PRA. In contrast, the nocturnal elevation

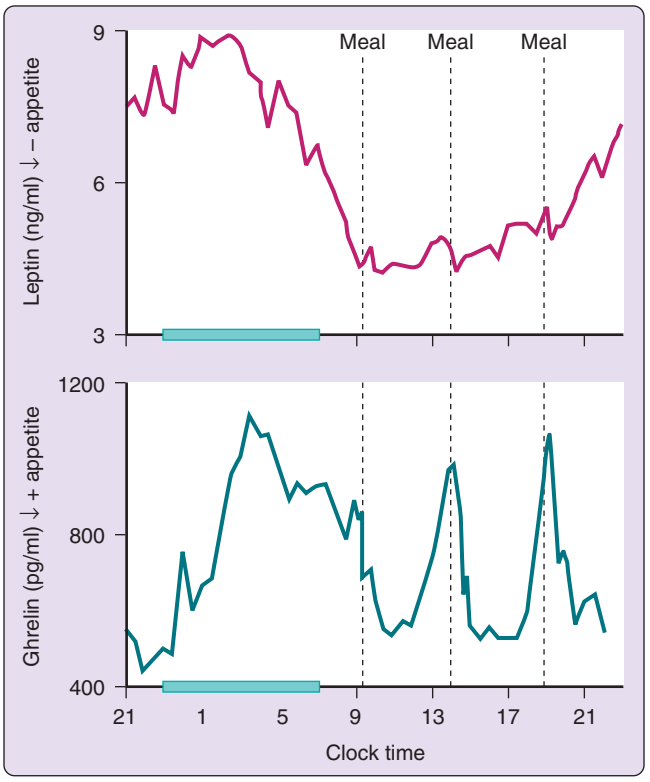


Figure 20-3 Typical 24-hour profiles of plasma leptin (*top*), an appetite-suppressing hormone, and ghrelin (*bottom*), a hunger-promoting hormone from a healthy lean young man. Time in bed is denoted by the *turquoise bars*. The *vertical lines* denote the time of presentation of identical high-carbohydrate meals.

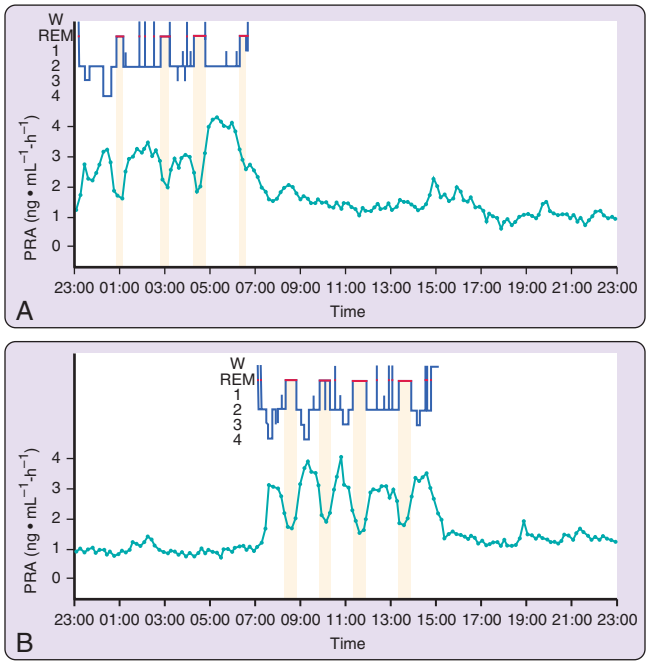


Figure 20-4 The 24-hour profiles of plasma renin activity sampled at 10-minute intervals in a healthy subject. **A,** Nocturnal sleep from 23:00 to 07:00. **B,** Daytime sleep from 07:00 to 15:00 after a night of total sleep deprivation. The temporal distribution of stages wake (W); REM; 1, 2, 3, and 4 are shown above the hormonal values. The oscillations of plasma renin activity are synchronized to the REM-NREM cycle during sleep. (From Brandenberger G, Follenius M, Goichot B, et al. Twenty-four hour profiles of plasma renin activity in relation to the sleep-wake cycle. *J Hypertens* 1994;12:277–83.)

of PRA does not occur when the subject is sleep deprived (lower panel of Figure 20-4). A well-documented study[102] has delineated the mechanisms responsible for the elevation of PRA during sleep. The initial event is a reduction in sympathetic tone, followed by a decrease in mean arterial blood pressure and an increase in slow wave activity. The rise in PRA becomes evident a few minutes after the increase in slow wave activity. During REM sleep, sympathetic activity increases, whereas renin and slow wave activity decrease and blood pressure becomes highly variable. This pattern of changes in PRA during sleep drives the nocturnal profile of aldosterone levels.[103,104] Acute total sleep deprivation eliminates the nocturnal PRA rise, dampens the nighttime elevation of plasma aldosterone, and increases natriuresis.[105] A close relationship between the beginning of REM episodes and decreased activity has been consistently observed for both PRA and aldosterone.[99,101,106-108] This relationship was confirmed in studies with selective REM-sleep deprivation in healthy subjects.[109] Increases in PRA parallel increases in slow wave EEG activity.[104] In conditions of abnormal sleep architecture (e.g., narcolepsy, sleeping sickness), the temporal pattern of plasma renin activity faithfully reflects the disturbances of the REM-NREM cycle.[99]

RECURRENT SLEEP RESTRICTION: IMPACT ON ENDOCRINE AND METABOLIC FUNCTION

Voluntary sleep curtailment has become a very common behavior in modern society. Data from the 2008 "Sleep in America" poll indicate that although working adults report a sleep need of an average of 7 hour and 18 minutes to function at best, 44% of them sleep less than 7 hours and 16% sleep less than 6 hours on a typical weeknight.[110] Sleep times in European countries appear to follow a similar trend.[111] For a substantial portion of the adult population, the cumulative sleep loss per workweek may correspond to as much as 1 full night of sleep deprivation. Several laboratory studies involving extension of the bedtime period for prolonged periods of time have provided evidence that the "recommended 8-hour night" does not meet the sleep need of healthy young adults, who may carry a substantial sleep debt even in the absence of obvious efforts at sleep curtailment.[112-114]

The following subsections review, respectively, the laboratory evidence supporting an adverse impact of recurrent partial sleep restriction on pituitary and pituitary-dependent hormones, glucose metabolism, the neuroendocrine control of appetite, food intake, and energy expenditure. The epidemiologic evidence for an adverse impact of short sleep on the risk for diabetes and obesity is then summarized.

Laboratory Studies of Experimental Sleep Restriction

Figure 20-5 summarizes the hormonal and metabolic findings of the first "sleep debt study,"[45] which examined the impact of 6 days of sleep restriction to 4 hours per night compared with 6 days of sleep extension to 12 hours per night in a group of healthy young men.[28,45,46] The findings suggested that sleep restriction has adverse effects on multiple endocrine axes as well as on glucose metabolism. Multiple observational studies and randomized controlled trials have since been conducted to further examine the hormonal and metabolic consequences of insufficient sleep. It is not possible to provide an exhaustive description of all studies in the present chapter. For excellent recent reviews, the reader is referred to references.[91,92,115-117]

Pituitary and Pituitary-Dependent Hormones

As may be seen in the upper panel of Figure 20-5, sleep restriction results in an alteration in nocturnal GH release such that a GH pulse occurs consistently before sleep onset. There was a negative correlation between presleep GH secretion and sleep-onset GH release. This profile of GH release is quite different from that observed during acute total sleep deprivation (back to top panels of Figure 20-1), where minimal GH secretion occurs during nocturnal wakefulness and GH secretion rebounds during daytime recovery sleep.

When compared with the fully rested condition, the state of sleep debt was associated with alterations of the 24-hour profile of cortisol, including a shorter quiescent period and elevated levels in the evening (Figure 20-5, second panel shaded areas). This alteration was similar to that observed after 1 night of acute total or partial sleep deprivation[39] and may reflect decreased efficacy of the negative feedback regulation of the hypothalamic-pituitary-adrenal axis.[45] Several studies that have assessed the profile of plasma or saliva cortisol levels across the daytime period in individuals submitted to 2 to 7 days of sleep restriction by 4 to 5 hours per night have similarly observed an elevation of cortisol concentrations,[118-121] but there have been well-documented negative studies as well.[122,123] The severity and duration of sleep restriction may play a role in the discrepancies. Recovery sleep after one workweek of mild sleep restriction was shown to reduce daytime cortisol levels relative to baseline.[124]

Restriction and extension of sleep duration were also associated with clear changes in thyrotropic function. The nocturnal elevation of plasma TSH was dampened and thyroid hormone levels were higher in the sleep debt state.[45] Previous studies have demonstrated that total sleep deprivation is initially associated with a marked increase in TSH secretion (see Figure 20-1), which becomes smaller when sleep deprivation continues, presumably because of negative feedback effects from slowly rising levels of thyroid hormones. Similar mechanisms are likely to underlie the alterations in thyrotropic function after recurrent partial sleep restriction. Findings of elevations in free T_4 index and of peripheral levels of free T_3 and free T_4 in subjects submitted to experimental sleep restriction or total sleep deprivation are consistent with this hypothesis.[45-48] In middle-aged overweight adults exposed to moderate sleep restriction over a 14-day period, TSH and free T_4 levels were lower after 14 days of sleep restriction compared with normal sleep.[125]

Evidence implicating an adverse impact of insufficient sleep on the gonadal axis has been obtained for testosterone levels in men. One week of partial sleep restriction (5 hours in bed) in healthy young men has been shown to result in a 10% to 15% decrease in afternoon and evening testosterone levels (Figure 20-6), concurrent with increased levels of subjective sleepiness.[123] A similar trend was observed in a study of 5 nights of sleep restriction to 4 hours in bed.[120] In a study examining morning testosterone levels after 1 night of total sleep deprivation or following sleep restricted to the first 4.5 hours of the night, testosterone levels were also reduced by about 20%.[126] Taken together, these findings suggest that obtaining an estimation of habitual sleep duration as well as sleep duration during the night before testosterone testing

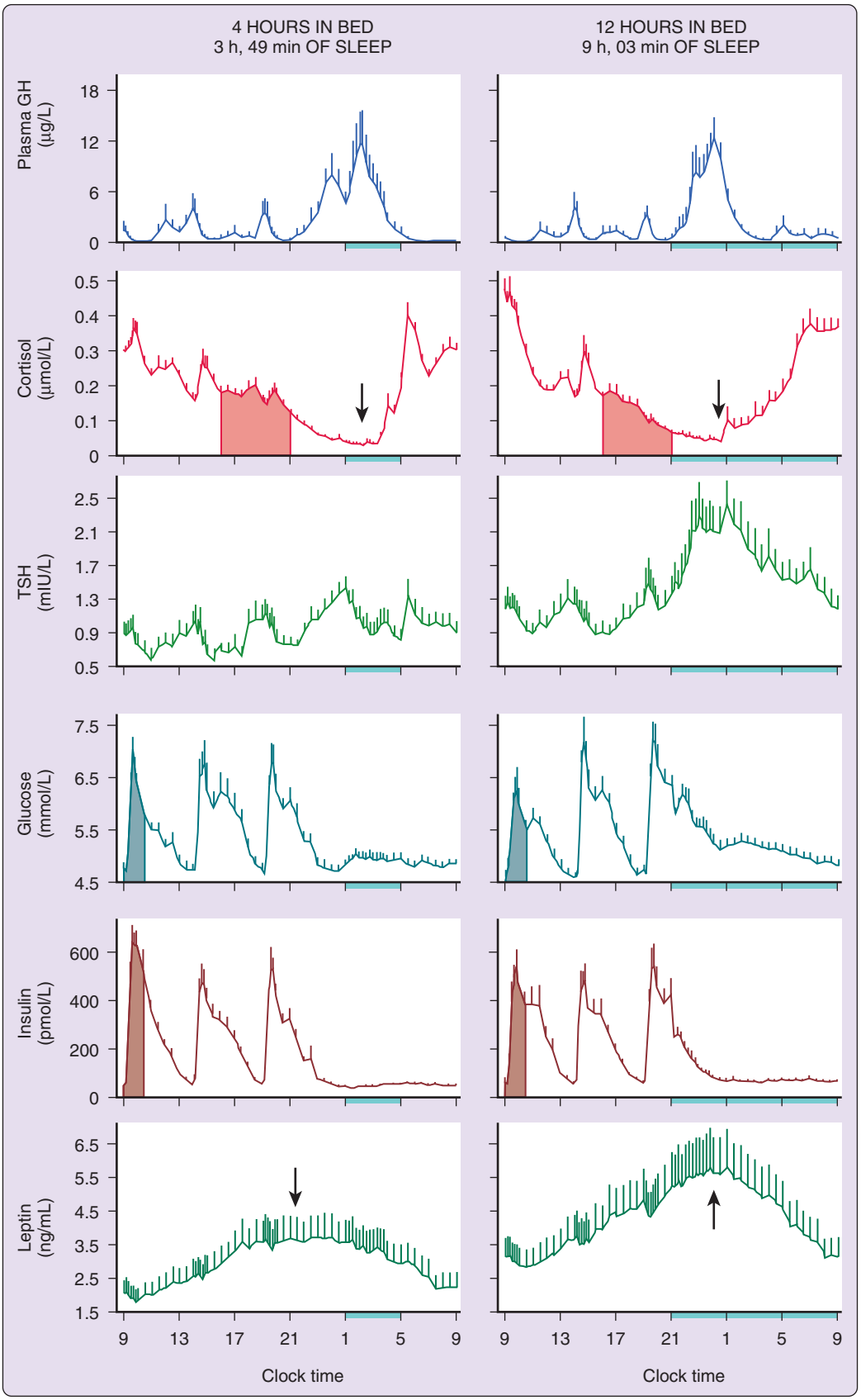


Figure 20-5 The 24-hour profiles of plasma growth hormone (GH), plasma cortisol, plasma thyrotropin (TSH), plasma glucose, serum insulin, and plasma leptin levels in 11 healthy young men who were studied after 1 week of bedtime restriction to 4 hours per night (*left panels*) and 1 week of bedtime extension to 12 hours per night (*right panels*). The *turquoise bars* represent the bedtime period. On the cortisol profiles, the *blue areas* show the increase in evening cortisol levels, and the *arrows* indicate the timing of the nadir. On the glucose and insulin profiles, the *blue area* shows the response to the morning meal. On the leptin profiles, the *arrows* indicate the timing of the nocturnal acrophase. (From Spiegel K, Leproult R, Van Cauter E. Impact of a sleep debt on metabolic and endocrine function. *Lancet* 1999;354:1435–9; Spiegel K, Leproult R, Colecchia E, et al. Adaptation of the 24-hour growth hormone profile to a state of sleep debt. *Am J Physiol* 2000;279:R874–83; and Spiegel K, Leproult R, L'Hermite-Balériaux M, et al. Leptin levels are dependent on sleep duration: relationships with sympathovagal balance, carbohydrate regulation, cortisol, and thyrotropin. *J Clin Endocrinol Metab* 2004;89:5762–71.)

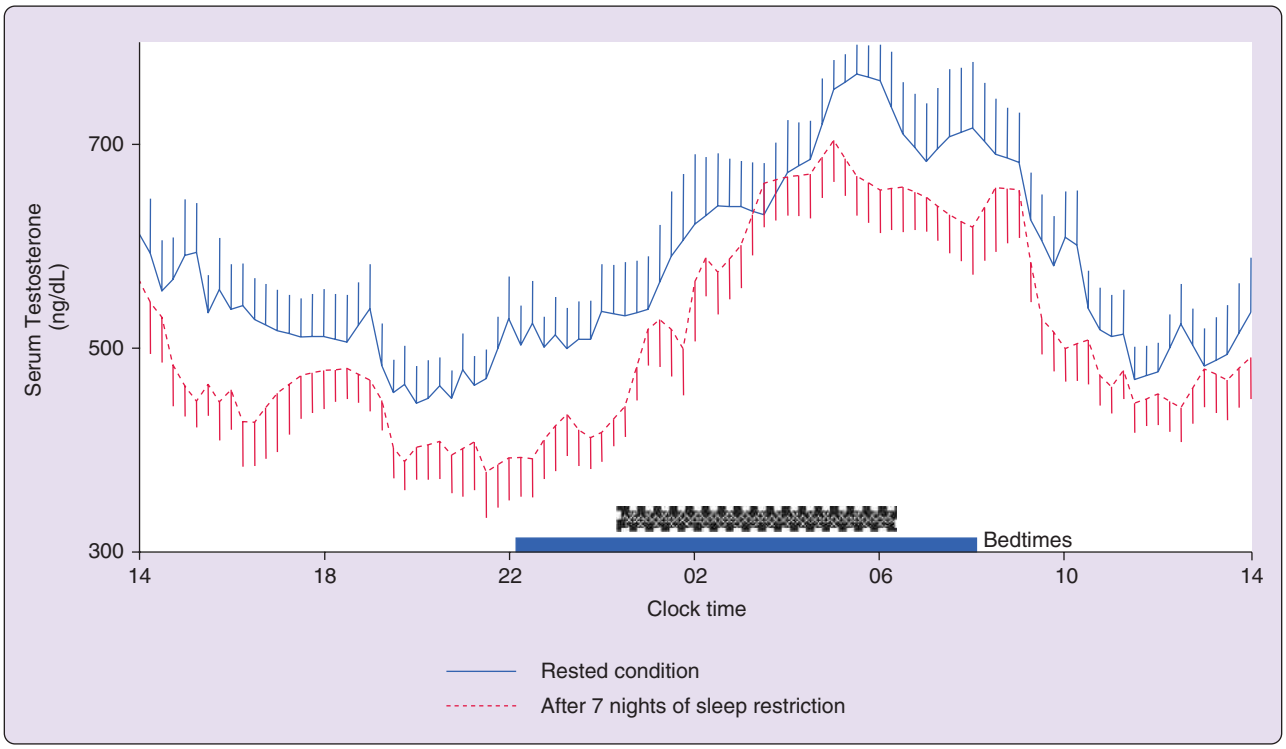


Figure 20-6 Twenty-four hour testosterone profiles observed in 10 healthy young men who each spent 11 days in the research laboratory. *Solid lines* depict the mean (+SEM) 24-hour profile observed in the rested condition (after 2 days of 10-hour bedtimes 2200 to 0800), and *dashed lines* depict the mean (–SEM) profile after 1 week of sleep restriction to 5 hours per night (bedtimes 0030 to 0530). (Data From Leproult R, Van Cauter E. Effect of 1 week of sleep restriction on testosterone levels in young healthy men. *JAMA* 2011;305(21):2173–4).

may be important in the diagnosis of androgen deficiency. Because prescriptions for exogenous testosterone replacement for complaints of low energy, fatigue, and reduced libido in adult men have increased dramatically in the past few years, the possibility that partial sleep restriction, a condition that can produce these symptoms, may be involved in producing or exacerbating the condition should be considered.

Glucose Metabolism

In the original "sleep debt" study,[45] 5 days of bedtime curtailment resulted in a higher glucose response to breakfast despite similar insulin secretion (see Figure 20-5, lower panels). The difference in peak postbreakfast glucose levels between the sleep debt and fully rested conditions (i.e., ±15 mg/dL) is consistent with a state of impaired glucose tolerance. Intravenous glucose tolerance testing confirmed this deterioration in glucose tolerance.[45] Reduced glucose tolerance was found to be the combined consequence of a decrease in glucose effectiveness, a measure of non–insulin-dependent glucose use, and of a reduction in the acute insulin response to glucose despite decreased insulin sensitivity. The product of insulin sensitivity and acute insulin response to glucose, that is, the disposition index, a validated marker of diabetes risk,[127] was decreased by nearly 40% in the state of sleep debt, reaching levels typical of populations at an elevated risk for diabetes.[128,129] Of note, the impact of recurrent sleep restriction was only seen on the responses to meals, and intravenous glucose fasting levels were unchanged.

These findings were confirmed in a number of subsequent randomized control trials that involved recurrent sleep restriction in the laboratory and included assessments of glucose tolerance and insulin levels or sensitivity during a glucose challenge.[119,122,130-132] In a randomized crossover study[132] comparing 4 days of 4.5 hours in bed versus 8.5 hours in bed, biopsies of subcutaneous abdominal fat were obtained from each participant at the end of each sleep condition. Adipocytes were exposed in vitro to incremental insulin concentrations to examine the ability of insulin to increase the phosphorylation of Akt, a crucial step in the insulin-signaling pathway. The insulin concentration needed to achieve the half-maximal phosphorylation of Akt response was nearly threefold higher when subjects had restricted sleep compared with normal sleep, indicating that sleep is an important modulator of energy metabolism in this peripheral tissue. A 2010 study involving 23 young men submitted to either 5 nights of sleep restriction to 4 hours per night or 8-hour bedtimes observed an increase in the ratio of insulin to glucose under fasting conditions.[133] This study further suggested that this metabolic alteration was partly corrected after 2 nights of recovery sleep.

Two randomized crossover design studies have examined the impact of repeated sleep restriction versus normal sleep in subjects submitted to caloric restriction who lost weight over the course of both short and long sleep interventions.[134,135] The findings were consistent in that daytime glucose levels were unchanged in either study. The study that had the most severe caloric restriction and the longest period of sleep loss found lower insulin levels in the short sleep condition, suggestive of an improvement rather than a deterioration of systemic insulin sensitivity.[134] Findings from intravenous glucose tolerance testing in the same subjects were, however, in the opposite

direction, with a 26% decrease in insulin sensitivity in the short versus normal sleep condition. Differences in the counterregulatory responses of cortisol, GH, and epinephrine during the ivGTT were proposed to explain the inconsistency.[134] A putative role of sleep restriction to lower incretin responses and thus postprandial insulin release is another possibility. Clearly, the interaction of insufficient sleep and dietary restriction is worthy of additional research because millions of individuals are attempting to follow a weight loss diet without consideration of the potential impact of their habitual sleep duration. In a randomized crossover laboratory study of 14 days with extended or restricted sleep and moderate caloric restriction, insufficient sleep resulted in a decrease in the proportion of weight lost as fat and an increase in the loss of fat-free mass.[136] Consistent findings were reported in a study including 123 overweight and obese adults who underwent a weight loss intervention involving caloric restriction, in which a significant relationship between self-reported sleep duration and loss of body fat was detected, after adjusting for age, sex, baseline body mass index (BMI), length of the intervention, and change in energy intake.[137] Recently, a change in sleep duration from 6 hours or less to between 7 and 8 hours was found to be associated with less visceral fat accumulation over 6 years.[138]

Individuals with a family history of type 2 diabetes have a greater than twofold increased risk for developing diabetes themselves. A 2011 study[139] showed that among adults with a parental history of diabetes, those who have habitual sleep duration of 6 hours or less have increased insulin resistance, making them more susceptible to develop diabetes.

Neuroendocrine Control of Appetite

In the sleep debt study mean levels of the satiety hormone leptin were reduced by 20% to 30% under sleep restriction compared with extension (see Figure 20-5, lowest panels), and the amplitude of the circadian rhythm was decreased.[46] This effect size of sleep restriction is comparable to that occurring after 3 days of dietary restriction by approximately 900 kcal/day under normal sleep conditions.[140] Further, there was a clear dose-response relationship between sleep duration and characteristics of the leptin profile.[46] Importantly, these differences in leptin profiles occur despite identical amounts of caloric intake, similar sedentary conditions, and stable weight. Four independent studies examining the leptin profiles after sleep restriction in lean young adults (mostly male) under conditions of controlled caloric intake also found a reduction of leptin levels or amplitude after sleep restriction.[24,96,141,142] In the most recent study, sleep loss was associated with circadian misalignment.[142] Findings regarding the impact of sleep restriction on leptin profiles or on isolated leptin levels in research participants who were exposed to ad libitum food intake or in whom body weight changed across the study period have been inconsistent, as summarized in several recent reviews.[91,92,117,143] Findings of epidemiologic cross-sectional studies examining the relationship between sleep duration and leptin levels have been similarly inconsistent.[92]

In a randomized crossover study of 2 nights of 4 hours in bed versus 2 nights of 10 hours in bed, in which the only source of caloric intake was a constant glucose infusion, daytime profiles of the hunger hormone ghrelin were measured, and the subjects completed validated scales for hunger and appetite for various food categories.[24] Daytime ghrelin levels were increased by 28% and the ghrelin-to-leptin ratio increased by more than 70%. Hunger showed a 23% increase, and appetite for nutrients with high carbohydrate content was increased by more than 30% when sleep was restricted. There was an excellent correlation between the change in the ghrelin-to-leptin ratio and the increase in self-reported hunger. Subsequent studies examining ghrelin levels in response to partial sleep restriction had variable findings, with no change detected in several studies. Differences in the demographics of the participants, length and severity of sleep restriction, nutritional status, sampling frequency, and assay methodology make it difficult to clearly summarize the current literature.[91,92,143]

Two epidemiologic studies reported reduced leptin levels, after controlling for BMI or adiposity, in habitual short sleepers.[144,145] Higher ghrelin levels have also been also associated with short sleep.[144] A subsequent study involving only postmenopausal women did not confirm the link between sleep duration, leptin, and ghrelin levels,[146] but very few participants had short sleep durations. Lastly, an ambulatory study of 80 obese adults found no cross-sectional association between fasting leptin levels and measures of adiposity.[147]

A reasonable conclusion regarding the roles of leptin and ghrelin as mediators of appetite dysregulation under conditions of insufficient sleep is that both pathways have been shown to be operative under certain experimental conditions but not uniformly. Sleep loss is likely to alter multiple other pathways involved in the control of energy intake.

Hunger, Satiety, and Food Intake

As summarized by Morselli and colleagues,[91] findings of laboratory studies that have examined hunger, satiety, or food intake under ad lib conditions have been more consistent than those focusing on alterations of the neuroendocrine control of appetite. A randomized crossover study of overweight middle-aged adults who were submitted to 2 weeks of 1.5 hours of sleep extension or restriction was the first to clearly demonstrate an increase in food intake from snacks during sleep restriction.[148] A subsequent study of 5 days with 4-hour bedtimes compared with 5 days of 9-hour bedtimes found that participants consumed on average nearly 300 kcal more when sleep restricted, mostly from fat.[149] As in previous studies of recurrent partial sleep restriction, short bedtimes resulted mainly in a loss of stage 2 and REM sleep. Linear mixed model analysis revealed a positive association between stage 2 duration and resting metabolic rate. Greater loss of stage 2 or REM sleep was associated with more hunger, more appetite for sweet as well as salty foods, and more energy consumed.[150] One recent study has addressed the possibility that extending sleep in short sleepers may decrease appetite.[151] In this home-based 2-week intervention, young adults obtained 1 hour and 36 minutes more sleep per day on average and reduced their overall rating of appetite by 14%, whereas the desire for sweet and salty foods was decreased by 62%.[151]

In the past few years, four studies used functional magnetic resonance imaging to examine brain function in subjects after normal sleep, 1 night of total sleep deprivation, or repeated partial sleep deprivation.[152-155] These studies have been consistent in showing that sleep loss increases neuronal activity in brain areas involved in the reward system in response to presentation of food stimuli or decreases neuronal activity in cortical regions involved in food choices.

Energy Expenditure

A logical explanation to the increased hunger and food intake associated with sleep restriction is that they occur in response to the caloric needs of extended wakefulness. Several studies used the doubly labeled water method to assess changes in energy expenditure during sleep restriction. Surprisingly, all three studies failed to detect an increase in energy expenditure.[148,149,156] However, when the subjects were confined to a calorimetry room to monitor minute-to-minute energy expenditure during normal sleep and total sleep deprivation, the caloric cost of wakefulness under recumbent conditions compared with sleep averaged only 17 kcal/hour.[157] A recent study involving 5 days of partial sleep restriction, similar to a workweek, under controlled laboratory conditions observed that the approximate 5% increase in daily energy expenditure was overcompensated by energy intake, particularly at night.[158] Another calorimetry room study comparing 3 nights of 4 hours in bed versus 3 nights of 8 hours in bed found that energy expenditure per 24-hour period was increased by 92 kcal in the 4-hour bedtime condition, thus 23 kcal/hour of extended wakefulness.[159] Taken together, these whole-room indirect calorimetry studies suggest that the stimulation of hunger and food intake far exceeds the caloric needs of extended wakefulness. Additionally, there is evidence that individuals who have insufficient sleep have lower levels of physical activity.[160,161]

Epidemiologic Studies Linking Habitual Short Sleep and the Risk for Obesity and Diabetes

Over the past 10 years, a large number of studies have examined associations between sleep duration and the prevalence and incidence of obesity and type 2 diabetes. Nearly all these studies explored existing data sets that included self-reported sleep duration, and none of them determined whether short sleep was the result of bedtime curtailment or was due to the presence of a sleep disorder or other comorbidities. Further, self-reported sleep duration is strongly dependent on demographics (sex, age, race or ethnicity), socioeconomic factors (income, occupation, education), and mental health status.[162] Nonetheless, by mid-2012, more than 60 epidemiologic studies, most with a cross-sectional design, had examined the relationship between sleep duration and obesity, BMI, or weight gain in adults, and most had found significant associations. In longitudinal studies in adults, the findings have been more mixed, and systematic reviews found either 8 out of 13 positive studies[163] or 8 out of 10 positive studies.[92] Findings from prospective studies in children have been more consistent in indicating that insufficient sleep increases the risk for weight gain or obesity.[163]

To date, 14 prospective studies in adults including a total of 583,263 participants have examined the relative risk (RR) of developing type 2 diabetes associated with short sleep duration,[164-177] and 8 of them reported significantly elevated RR for short sleep (≤5 hours; RR range: 1.51 to 2.94) relative to normal sleep (7 to 8 hours). A meta-analysis published in 2010 that included 10 of the 14 currently available studies concluded that short sleep increases the risk for type 2 diabetes by 28%.[178] Of note, the risk was significantly higher in men (RR: 2.07) than in women (RR: 1.07), and long sleep (≥9 hours) was also found to be associated with a higher risk for incident diabetes (RR: 1.48). All studies relied on self-report of sleep duration, and it is highly likely that different factors mediate the association of diabetes with short versus long sleep.

This body of epidemiologic evidence supports the hypothesis that sleep curtailment may be a nontraditional lifestyle factor contributing to the epidemics of obesity and type 2 diabetes.

REDUCED SLEEP QUALITY AND SLEEP DISORDERS: IMPACT ON ENDOCRINE AND METABOLIC FUNCTION

Experimental Reduction of Sleep Quality

Early studies have been consistent in showing that experimentally induced full awakenings interrupting nocturnal sleep consistently trigger pulses of cortisol secretion.[30,179,180] Furthermore, in an analysis of cortisol profiles during daytime sleep, it was observed that 92% of spontaneous awakenings interrupting sleep were associated with a cortisol pulse.[180]

The initiation of SWS is associated with a decrease in cerebral glucose use, stimulation of GH secretion, inhibition of cortisol release, decreased sympathetic nervous activity, and increased vagal tone. All these correlates of SWS affect total body glucose regulation, suggesting that low amounts of SWS may be associated with reduced glucose tolerance. Our group tested this hypothesis by selectively suppressing SWS (using acoustic stimuli) in healthy young adults and examining the impact on the response to intravenous glucose injection.[181] The amount of SWS was reduced by nearly 90%, similar to what occurs over the course of four decades of aging. Such low levels of SWS are also typical of moderate to severe OSA. Importantly, this intervention did not reduce total sleep duration. Slow wave activity was markedly reduced in each experimental night compared with baseline (left panels of Figure 20-7). After 3 nights of SWS suppression, insulin sensitivity was decreased by about 25% (right panels of Figure 20-7), reaching the level reported in older adults and in populations at high risk for diabetes.[182] This decrease in insulin sensitivity was not compensated for by an increase in insulin release because acute insulin response to glucose remained virtually unchanged. Consequently, diabetes risk, as assessed by the disposition index, was lower, and glucose tolerance was reduced, reaching the range typical of impaired glucose tolerance. These laboratory findings demonstrate that reduced sleep quality, without change in sleep duration, may adversely affect glucose regulation. In this study where SWS was suppressed while carefully avoiding full awakenings, cortisol levels were not affected at any time of the day or night.[181] An increase in daytime sympathic-vagal balance, as assessed by spectral analysis of heart rate variability, was identified as one of the possible mechanisms mediating the adverse impact of SWS suppression on glucose metabolism. In another study, nonselective sleep fragmentation for 2 nights by acoustic stimuli was associated with a decrease in insulin sensitivity and non–insulin-dependent glucose disposal.[183] Notably, nonselective sleep fragmentation resulted in marked reductions in slow wave sleep, whereas other sleep stages were minimally affected.[183] The importance of SWS for the maintenance of glucose homeostasis has also been confirmed by a more recent experimental study in healthy adults.[184]

A randomized crossover study with 1 night of either fragmented or nonfragmented sleep found decreased subjective

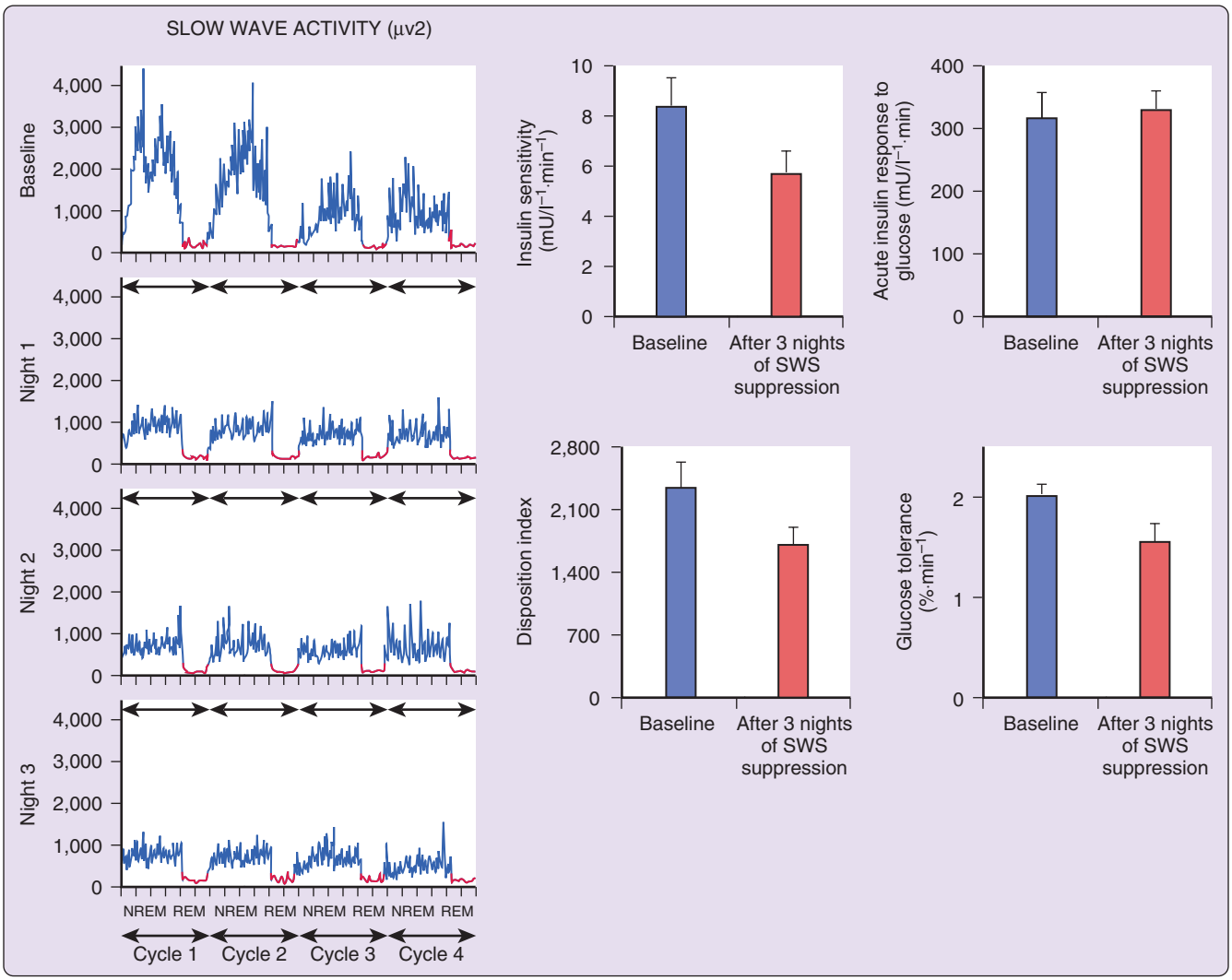


Figure 20-7 *Left panel,* Mean (±SEM) profiles of slow wave activity (μV^2) for the first four NREM-REM sleep cycles (NREM1, NREM2, NREM3, and NREM4) during baseline and in each experimental night of slow wave sleep (SWS) suppression (night 1, night 2, night 3). Slow wave activity was markedly and similarly reduced in each experimental night compared with baseline, and the largest reductions were achieved during the first two NREM cycles. *Right panel*, Mean (±SEM) insulin sensitivity, acute insulin response to glucose, disposition index, and glucose tolerance (Kg) at baseline and after 3 nights of SWS suppression. (From Tasali E, Leproult R, Ehrmann D, Van Cauter E. Slow wave sleep and the risk of type 2 diabetes in humans. *Proc Natl Acad Sci U S A* 2008;105(3):1044–9).

fullness with reduced REM sleep and preservation of SWS.[185] In another study, 2 nights of fragmented sleep with reductions in both REM sleep and SWS were associated with unchanged total energy expenditure, increased carbohydrate oxidation, and decreased fat oxidation, which may predispose to weight gain.[186]

Prospective Epidemiologic Studies Linking Poor Sleep Quality to Diabetes Risk

Multiple epidemiologic studies have provided evidence for an association between self-reported poor sleep quality and the prevalence or incidence of diabetes, after controlling for age, BMI, and various other confounders. Of note, in six of seven prospective studies that examined self-reported problems (e.g., difficulty initiating or maintaining sleep, use of sleeping pills, or insomnia complaint), poor sleep quality was associated with an increased risk for diabetes.[166,170,173,187-190] Meta-analysis of these studies found that self-reported difficulty in initiating sleep was associated with an increased risk for diabetes (RR: 1.57; ~18,000 participants), and self-reported difficulty in maintaining sleep also predicted the development of diabetes (RR: 1.84; ~24,000 participants).[178]

Insomnia

There have been remarkably few studies of hormonal and metabolic variables in subjects with physician-diagnosed insomnia. A well-documented study[191] in patients with insomnia revealed that those with decreased total sleep time have higher cortisol levels across the night (Figure 20-8). A few other studies have also shown that insomnia is associated with increased levels of cortisol and norepinephrine.[192-194] It is unclear whether this relative hypercortisolism is the result of sleep fragmentation and the associated sleep loss or, alternatively, whether hyperactivity of the corticotropic axis is causing

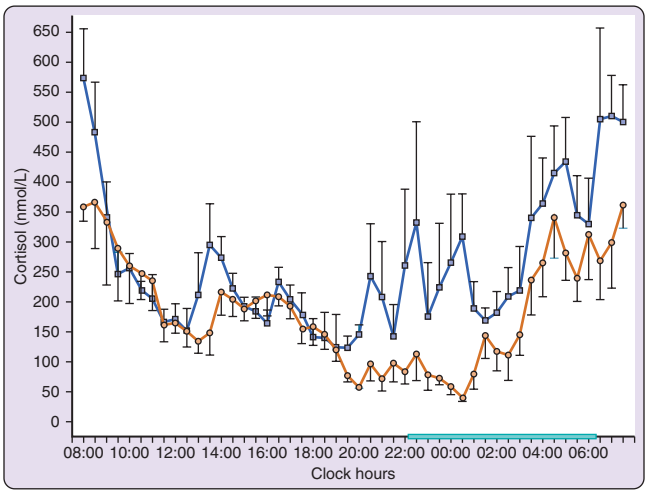


Figure 20-8 Mean 24-hour profiles of plasma cortisol in young insomniacs with low total sleep time (*blue squares*) compared with young insomniacs with high total sleep time (*orange circles*). The *turquoise bar* indicates the sleep-recording period. The error bars indicate standard error of the mean (SEM). (From Vgontzas A, Bixler EO, Lin HM, et al. Chronic insomnia is associated with neurohumoral activation of the hypothalamo-pituitary-adrenal axis: clinical implications. *J Clin Endocrinol Metab* 2001;86:3787–94).

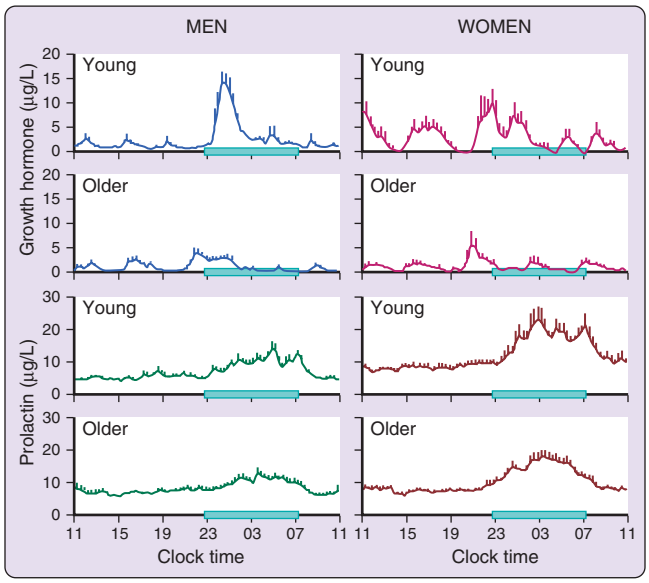


Figure 20-9 *Upper panels,* Mean 24-hour profiles of plasma growth hormone in healthy young (18 to 33 years) and older (51 to 72 years) men (*left*) and women (*right*). Young women were studied in the follicular phase of the menstrual cycle. Older women were postmenopausal and not on hormone replacement therapy. The *turquoise bars* represent the sleep periods. *Lower panels,* Mean 24-hour profiles of plasma prolactin in the same subjects. (From Van Cauter E, Plat L, Copinschi G. Interrelations between sleep and the somatotropic axis. *Sleep* 1998;21:533–66; Latta F, Leproult R, Tasali E, et al. Sex differences in nocturnal growth hormone and prolactin secretion in healthy older adults: relationship with sleep EEG variables. *Sleep* 2005;28:1519–24; and Caufriez A, Leproult R, L'Hermite-Balériaux M, et al. A potential role for endogenous progesterone in modulation of growth hormone, prolactin and thyrotropin secretion during normal menstrual cycle. *Clin Endocrinol* 2009;71(4): 535–42).

hyperarousal and insomnia. Recent views on chronic insomnia propose that it is a disorder of hyperarousal during both the night and the daytime, with associated hyperactivity of the hypothalamic-pituitary-adrenal axis.[195,196] A population-based study involving a total of 1741 men and women found that insomnia with short sleep duration was associated with increased odds of diabetes.[197] A small study involving 14 patients with insomnia found decreased nocturnal ghrelin levels, providing evidence for a possible dysregulation of energy balance in this patient population.[198]

Obstructive Sleep Apnea

There is substantial evidence linking OSA to abnormalities of glucose metabolism, including insulin resistance, glucose intolerance, and increased risk for type 2 diabetes. For a summary of the present state of knowledge, the reader is referred to Section 14 of the present volume as well as to recent reviews.[199,200]

OSA is also associated with disturbances in the control of weight and neuroendocrine regulation of appetite. Indeed, patients with OSA appear more predisposed to weight gain than similarly obese subjects without OSA.[201] Ghrelin levels were found to be increased in patients with OSA compared with controls in most,[202-204] but not all,[205] studies. Elevated leptin levels in OSA, after controlling for BMI, were reported in earlier studies,[201,206] whereas more recent studies found no difference between apneic patients and BMI-matched controls.[205] Hyperleptinemia in OSA is thought to reflect leptin resistance.[201]

Although most studies have shown reduced ghrelin levels after continuous positive airway pressure (CPAP) treatment of OSA,[204,207,208] one study found no difference in ghrelin levels after CPAP.[209] Leptin levels were also consistently found to be decreased after CPAP treatment.[204,209] However, the findings on the effect of CPAP on body weight or visceral adiposity are mixed. Weight loss was reported in one study after 6 months of CPAP,[210] whereas other studies found weight gain after CPAP use.[208,211] In a randomized controlled multicenter trial, the greatest weight gain was found in those most compliant with CPAP.[212] CPAP therapy added to a weight reduction program has not resulted in greater weight loss.[213,214] If weight loss is important, loss of visceral fat is by far more relevant from a metabolic point of view. Again, the studies have yielded conflicting results.[215-219]

Age-Related Sleep Alterations: Implications for Endocrine Function

Normal aging is associated with pronounced age-related alterations in sleep quality, which consist primarily of a marked reduction of SWS (stages 3 and 4), a reduction in REM stages, and an increase in the number and duration of awakenings interrupting sleep (see Chapter 3). There is increasing evidence that these alterations in sleep quality may result in neuroendocrine disturbances, suggesting that some of the hormonal hallmarks of aging may partly reflect the deterioration of sleep quality.[220]

Growth Hormone Axis

There are mutual interactions between somatotropic activity and sleep that are evident both in young and older age. Sex and age differences are illustrated in the upper panels of Figure 20-9. In normal young men, there is a dose-response relationship between SWS or slow wave activity and GH secretion, and the sleep-onset GH pulse is often the largest pulse observed over the 24-hour span. In normal young

women, daytime GH pulses are more frequent, and the sleep-onset pulse, although generally present, is smaller.[221,222] In healthy older adults, in both gender groups, a significant amount of GH secretion occurs in the late evening, before habitual bedtime, at a time when GH secretion is usually quiescent in young adults.[223] Such presleep GH pulses may appear in young subjects when studied in a state of sleep debt.[28] In older men, but not women, the quantitative relationship between SWS/delta activity and sleep-onset GH release persists. In contrast, in older women, presleep GH release inhibits both the amount of GH secreted during sleep and sleep consolidation, as evidenced by negative correlations between presleep GH secretion and sleep maintenance.[223]

The impact of aging on the amount of SWS and on GH release in healthy men occurs with a similar chronology characterized by major decrements from early adulthood to midlife (Figure 20-10).[224] Reduced amounts of SWS were found to be a significant predictor of reduced GH secretion in middle life and late life, independently of age. The observation that in older adults, levels of insulin-like growth factor I, the hormone secreted by the liver in response to stimulation by GH, are correlated with the amounts of SWS[225] is consistent with this finding. The relative GH deficiency of elderly adults is associated with increased fat tissue and visceral obesity, reduced muscle mass and strength, and reduced exercise capacity. The persistence of a consistent relationship between SWS and GH secretion in older men suggests that drugs that reliably stimulate SWS in older adults may represent a novel strategy for GH replacement therapy.

Prolactin Secretion

In both men and women, most of the daily release of PRL occurs during sleep, irrespective of age. The lower panels of Figure 20-9 illustrate typical profiles in healthy nonobese young and older men and women. The sex difference is apparent both during daytime and nighttime in young adulthood, but in older age only nighttime levels are affected. A nearly 50% dampening of the nocturnal PRL elevation is evident in elderly men and women.[226] This age-related endocrine alteration may partly reflect the increased number of awakenings (which inhibit PRL release) and decreased amounts of REM stages (which stimulate PRL release).[223]

Besides its role in the control of lactation and parental behavior, PRL has multiple actions, including on metabolism and immunoregulation. Age-related alterations in sleep architecture and their impact on nocturnal PRL release could thus impact healthy aging.

Pituitary-Adrenal Axis

There are highly consistent, sex-specific alterations in the diurnal pattern of basal cortisol secretion across the lifetime.[227] Figure 20-9 shows 24-hour profiles typical of young and older men and women. In young adulthood, overall cortisol levels are lower in women than in men because the female response to the early morning circadian signal is slower and of lesser magnitude and the return to quiescence is more rapid. In men the nocturnal quiescent period is shorter, and the early morning elevation is higher and more prolonged. During aging there seems to be a progressive decline in the endogenous inhibition of nocturnal cortisol secretion in both men and women, as reflected by a delay of the onset of the quiescent period and higher nocturnal cortisol levels.

In contrast to the rapid decline of SWS and GH secretion from young adulthood to midlife, the impact of age on REM sleep, sleep fragmentation, and evening cortisol levels does not become apparent until later in life. As illustrated in Figure 20-10, REM sleep, wake after sleep onset, and evening cortisol levels follow the same chronology of aging, that is, no alteration until midlife and then a steady rise from midlife to old age.[224] There is a significant negative relationship between the loss of REM sleep in old age and the inability to achieve or maintain the quiescence of the corticotropic axis. Both animal and human studies have indicated that deleterious effects of HPA hyperactivity are more pronounced at the time of the trough of the rhythm than at the time of the peak. Therefore, modest elevations in evening cortisol levels could facilitate the development of central and peripheral disturbances associated with glucocorticoid excess, such as memory deficits and insulin resistance, and further promote sleep fragmentation.

Pituitary-Gonadal Axis

A progressive decline in testosterone levels occurs with aging in normal men. Starting at 30 to 40 years of age, testosterone concentrations decrease by 1% to 2% per year. In elderly men, the diurnal variation of testosterone is still detectable, but the nocturnal rise is markedly dampened.[59] A recent study indicated that the considerable interindividual variability of testosterone levels in healthy elderly men might be partly related to differences in sleep quality.[70] Indeed, both total and free (i.e., biologically active) morning testosterone levels were significantly correlated with total sleep time achieved during a night of laboratory polysomnography. A difference in total sleep time between 4.5 and 7.5 hours translated into a clinically meaningful difference in total testosterone levels because concentrations around 200 to 300 ng/dL are considered to be borderline-low for older men, and concentrations around 500 to 700 ng/dL represent mid-normal values typical of healthy young adults. A similar robust correlation was found with the usual amount of nighttime sleep monitored by actigraphy at home.[70] Thus it is important to enquire about poor or insufficient sleep in the interpretation and management of low testosterone levels in older men.

Sleep Disturbances in Metabolic and Endocrine Disorders

Obesity

Obesity is a major risk factor for OSA.[228] Complaints of daytime sleepiness may be present in obese subjects even in the absence of OSA.[229-232] In obese subjects without OSA, there may be disturbances in sleep architecture, including lighter and more fragmented sleep compared with nonobese controls.[230] Severely obese patients without OSA may have significantly shorter sleep latencies than lean age-matched controls.[229] Excessive daytime sleepiness has been found in 35% of obese subjects (BMI: 40 ± 6 kg/m^2) without OSA compared with 2.7% in age-matched nonobese controls.[232] It has been proposed that excessive daytime sleepiness and fatigue (i.e., tiredness without increased sleep propensity) in obese individuals without OSA could be due to a disruption of sleep homeostasis caused by elevated levels of somnogenic proinflammatory cytokines released by visceral fat (interleukin-6 and tumor necrosis factor-α).[233] In a cohort of 1300 middle-aged men and women who had 1 night of laboratory polysomnography, 47% of obese subjects reported

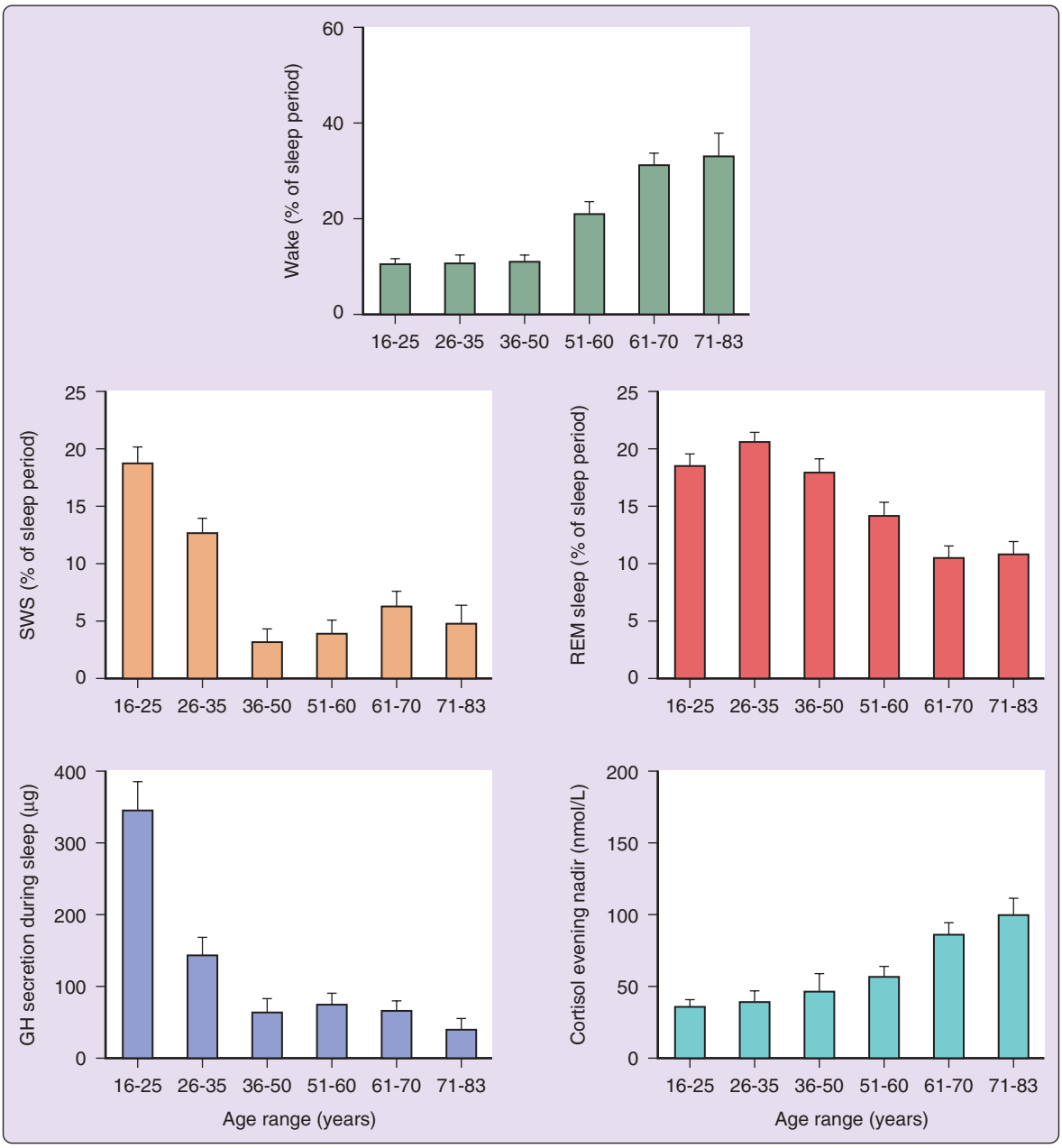


Figure 20-10 Mean (SEM) amounts of wake after sleep onset (*top panel*), slow wave sleep (stages III and IV, *middle left panel*), REM sleep (*middle right panel*), growth hormone (GH) secretion during sleep (*lower left panel*), and nadir of plasma cortisol concentrations (*lower right panel*) by age group in 149 healthy nonobese men. Sleep stages are expressed as a percentage of the sleep period, defined as the time interval between sleep onset and final morning awakening. (From Van Cauter E, Leproult R, Plat L. Age-related changes in slow wave sleep and REM sleep and relationship with growth hormone and cortisol levels in healthy men. *JAMA* 2000;284:861–8.)

subjective sleep disturbances (insomnia, sleep difficulty, excessive daytime sleepiness) compared with 26% of nonobese individuals. Thus the association between short sleep and high BMI evidenced in multiple epidemiologic studies may partly reflect the high prevalence of sleep disturbances and emotional stress.[234]

Type 2 Diabetes

Two clinic-based studies have examined the relationship between sleep duration and quality and glycemic control in type 2 diabetes. The first study administered the Pittsburgh Sleep Quality questionnaire to 161 African American diabetic patients.[235] Higher perceived sleep debt or lower sleep quality were associated with poorer glycemic control after controlling for age, sex, BMI, insulin use, and the presence of complications.[235] Importantly, the magnitude of these effects of sleep duration or quality was comparable to that of commonly used oral antidiabetic medications. The second study used actigraphy in 47 diabetic patients and 23 nondiabetic controls under free-living conditions. After adjusting for age, gender, and

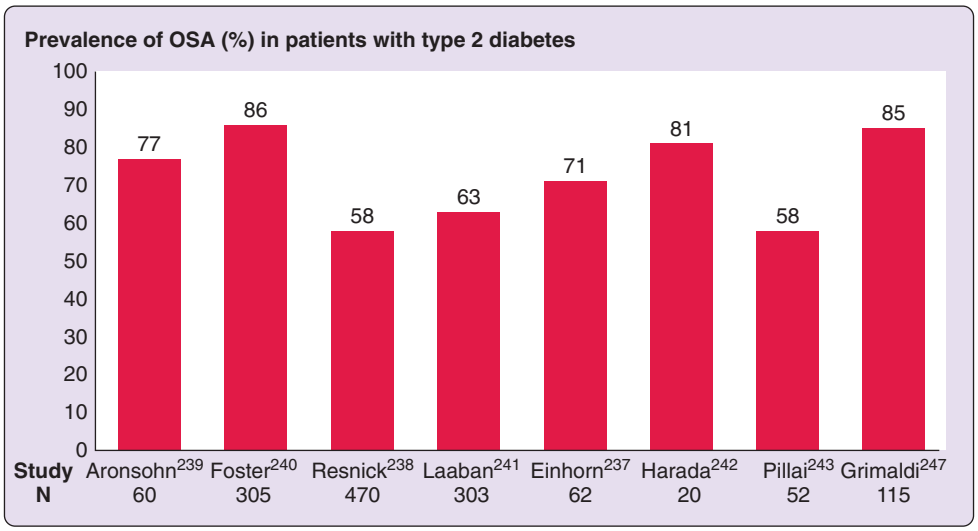


Figure 20-11 Prevalence of obstructive sleep apnea (OSA) as assessed by polysomnography among type 2 diabetes patients from eight independent studies (listed with the references). (Modified from Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of disease. *Ann N Y Acad Sci* 2014;131:151–73.)

schooling, measures of sleep fragmentation were significantly higher in the patients with diabetes, and glycemic control correlated inversely with sleep efficiency.[236]

In patients with type 2 diabetes, the prevalence of OSA as assessed by polysomnography was found to be high, ranging from 58% to 86% (Figure 20-11).[237-244] A retrospective analysis of a total of 16,066 diabetic patients from 27 primary care practices found that only 18% were diagnosed with OSA, suggesting that OSA may remain untreated in most diabetic patients.[245] There is also evidence to suggest that the presence and severity of untreated OSA may be associated with poor glucose control in type 2 diabetic patients.[199,239,243,246,247] In the CARDIA Sleep Study, participants with type 2 diabetes who had more fragmented sleep or insomnia had higher fasting glucose and insulin levels.[246] A recent study indicated that obstructive events in REM sleep rather than NREM sleep may have more adverse metabolic effects in diabetic patients.[247] A number of interventional studies have examined whether CPAP treatment of OSA has beneficial effects on glycemic control in type 2 diabetic patients. Although uncontrolled studies were generally positive,[248-252] one randomized controlled trial found no beneficial effects of CPAP on glycemic control in patients with type 2 diabetes.[253] Notably, this negative study reported an average nightly therapeutic CPAP use of only 3.6 hours.[253]

Polycystic Ovary Syndrome

PCOS, the most common endocrine disorder of premenopausal women, is characterized by hyperandrogenism, obesity, insulin resistance, and an elevated risk for type 2 diabetes. Insulin resistance is often referred to as a "hallmark" of PCOS. OSA is present in at least 50% of PCOS women.[254-259] In one study about two thirds of PCOS women were found to have poor sleep quality, and 45% had chronic daytime sleepiness.[258] In a study involving 52 women with PCOS and 21 women without PCOS of similar age and BMI, OSA was found to be an important determinant of insulin resistance, glucose intolerance, and type 2 diabetes in PCOS.[257] Both the prevalence of impaired glucose tolerance and the degree of insulin resistance increased in direct proportion to the severity of OSA.[257] Eight weeks of CPAP treatment of OSA in obese women with PCOS led to improvement in insulin sensitivity and decreased sympathetic output as assessed by 24-hour profiles of plasma catecholamines.[260] The magnitude of these beneficial effects was modulated by the hours of CPAP use and the degree of obesity.[260] Although the current evidence points to the importance of systematic identification and treatment of OSA in the management of PCOS patients, most clinicians who treat PCOS today are not yet aware of the high risk for OSA in this patient population.[261]

CLINICAL PEARL

Sleep exerts marked modulatory effects on most components of the endocrine system and has an important impact on glucose regulation. There is rapidly accumulating evidence from both laboratory and epidemiologic studies indicating that sleep loss and poor sleep quality are associated with hormonal disturbances and an increased risk for obesity and diabetes. Sleep disorders may also exacerbate the severity of an existing condition. Findings suggest that part of the constellation of hormonal and metabolic alterations that characterize normal aging may reflect the deterioration of sleep quality. Strategies to improve sleep quality may have beneficial effects on endocrine and metabolic function.

SUMMARY

Sleep exerts important modulatory effects on the endocrine system. Sleep timing, duration, and quality may also affect the circadian system and its control of hormone release and action. Pathways mediating the impact of sleep on endocrine function and metabolism include the activity of the hypothalamic releasing and inhibiting factors on pituitary hormone release and the autonomous nervous system control of endocrine organs. Modulatory effects of sleep are not limited to the

hormones of the hypothalamic-pituitary axes; these effects are also observed for the hormones controlling carbohydrate metabolism, appetite regulation, and water and electrolyte balance. Sleep loss is associated with disturbances of hormone secretion and metabolism, which may have clinical relevance, particularly as voluntary partial sleep curtailment has become a highly prevalent behavior in modern society. Reduced sleep quality also adversely affects endocrine release and metabolism. Major metabolic diseases such as obesity, type 2 diabetes, and polycystic ovary syndrome are all associated with sleep disturbances, which may promote the development or exacerbate the severity of the condition. Strategies to reverse decrements in sleep duration or quality may have beneficial effects on endocrine and metabolic function.

Selected Readings

Aurora RN, Punjabi NM. Obstructive sleep apnoea and type 2 diabetes mellitus: a bidirectional association. *Lancet Respir Med* 2013;**1**(4):329–38.

Broussard JL, et al. Impaired insulin signaling in human adipocytes after experimental sleep restriction: a randomized, crossover study. *Ann Intern Med* 2012;**157**(8):549–57.

Markwald RR, et al. Impact of insufficient sleep on total daily energy expenditure, food intake, and weight gain. *Proc Natl Acad Sci U S A* 2013; **110**(14):5695–700.

Morris CJ, Yang JN, Garcia JI, et al. Endogenous circadian system and circadian misalignment impact glucose tolerance via separate mechanisms in humans. *Proc Natl Acad Sci U S A* 2015;**112**:E2225–34.

Morselli LL, Guyon A, Spiegel K. Sleep and metabolic function. *Pflügers Arch* 2012;**463**(1):139–60.

Nedeltcheva AV, et al. Insufficient sleep undermines dietary efforts to reduce adiposity. *Ann Intern Med* 2010;**153**(7):435–41.

Nedeltcheva AV, Scheer FA. Metabolic effects of sleep disruption, links to obesity and diabetes. *Curr Opin Endocrinol Diabetes Obes* 2014;**21**(4): 293–8.

Pamidi S, Tasali E. Obstructive sleep apnea and type 2 diabetes: is there a link? *Front Neurol* 2012;**3**:126.

Quan SF, et al. Impact of treatment with continuous positive airway pressure (CPAP) on weight in obstructive sleep apnea. *J Clin Sleep Med* 2013; **9**(10):989–93.

Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. *Ann N Y Acad Sci* 2014;**1311**:151–73.

Schmid SM, Hallschmid M, Schultes B. The metabolic burden of sleep loss. *Lancet Diabetes Endocrinol* 2015;**3**(1):52–62.

Spiegel K, Tasali E, Leproult R, et al. short sleep, poor sleep: impact on glucose metabolism and obesity risk. *Nat Rev Endocrinol* 2009;**5**(5): 253–61.

Vgontzas AN, Basta M, Fernandez-Mendoza J. Subjective short sleep duration: what does it mean? *Sleep Med Rev* 2014;**18**(4):291–2.

A complete reference list can be found online at ExpertConsult.com.

Chapter
21

Thermoregulation in Sleep and Hibernation

Kurt Kräuchi; Tom Deboer

Chapter Highlights

- Despite the long-term awareness that thermoregulation and sleep are intimately coupled, there is still a lack of knowledge about the crucial mechanisms. Based on the fact that sleeping pills exhibit undesirable side effects, there is an increasing need for nonpharmacologic therapies such as thermal interventions.
- This chapter describes normal sleep in humans in relation to circadian regulation of core body temperature. Based on this correlation, human and animal experimental intervention studies changing ambient temperature or sleep pressure are described to gain information about more causative mechanisms.
- In addition to sleep, the torpid state in animals is described. Because this state is entered through normal sleep, it may be a valuable model to further investigate the relationship between thermoregulation and sleep.

The objective of this chapter is to cover the physiology of the relationship between the thermoregulatory system and the sleep regulatory system. Both are driven, independently, by two interacting physiologic principles, homeostasis and circadian regulation. The chapter is divided into three main sections: (1) a brief introduction into the circadian regulation of core body temperature (CBT); (2) the interaction of sleep and thermoregulatory mechanisms; and (3) hibernation, a special condition displayed by a limited number of mammalian species.

Animals increase survival by residing in a safe sleeping site and have used sleep to maximize energy savings by reducing body and brain energy consumption and to conduct a variety of recuperative processes.[1,2] Knowledge about thermophysiology and its relation to sleep leads to the hope that temperature-related interventions can alleviate sleep disturbances and be helpful to cure certain aspects of sleep and alertness problems in the general population.

A vast amount of knowledge is found in the literature on the variability in rest and sleep states and on thermophysiology across the animal kingdom.[1,2] To limit the scope of this chapter, only findings from humans, rats, ground squirrels, and hamsters are reviewed.

CIRCADIAN REGULATION OF CORE BODY TEMPERATURE

More than 50 year ago, Aschoff[3] showed that the human body consists of two thermophysiologic compartments: the heat-producing, homeothermic core; and the heat-loss-regulating, poikilothermic shell. The size of the latter is largely dependent on environmental temperature. In a warm environment, the shell is small; in a cool environment, it is large and thus acts as a buffer to protect the core from dangerous cooling. All peripheral tissues, such as fat, the skin, and in particular the skeletal muscles of the legs and arms, can contribute substantially to the size of the shell, provided that peripheral blood flow is low. Therefore rates of blood flow through muscles and skin are the main determinants of shell size variability and hence of peripheral insulation. The distal skin regions, in particular fingers and toes, are the main thermoeffectors to lose core body heat because they possess the physical and physiologic properties to best serve the function of heat loss. They have ideal surface shapes (round, small radius) for good heat transfer to the environment; the surface-to-volume coefficient increases from proximal to distal skin sites. The distal skin temperatures therefore provide a good measure of the shell size.

CBT comprises the temperature of the brain and the abdominal cavity, including inner organs (e.g., liver, heart, kidney).[3] In most placental mammals CBT is regulated around 37° C, whereas the brain is the main target for homeothermy, allowing control of all behavioral and physiologic processes over a broad environmental temperature range. A detailed description of the thermoregulatory system can be found elsewhere.[4]

CBT is regulated between thermoeffector thresholds, which are subject to circadian oscillations.[5] Circadian rhythms in mammals are generated by the self-sustaining central pacemaker localized in the suprachiasmatic nuclei (SCN) of the hypothalamus and are usually entrained to the 24-hour solar day mainly by the synchronizer light.[6] A rostral projection from the SCN to the preoptic anterior hypothalamus (POAH) conveys the circadian signal to the thermoregulatory system.[6] The regulation of CBT results from the concerted action of the homeostatic and circadian processes. In humans, the daily decline of CBT in the evening results from a regulated decline in the thermoregulatory thresholds of heat production and heat loss; the inverse happens in the morning. When heat production surpasses heat loss, body heat content increases and

vice versa. Depending on environmental temperature, about 70 to 90% of body heat content is located in the body core. Therefore changes in CBT reflect to a great extent changes in body heat content. Heat production and heat loss are modified by activities such as muscular exertion and fluid and food intake that are not randomly distributed over the circadian cycle. These behaviors induce so-called masking effects and differentially modify the endogenous rhythm of CBT.[7]

To disentangle circadian effects and influences from masking effects of an overt diurnal activity pattern, the constant routine (CR) protocol was developed in humans.[8] With this protocol it was shown that the time course of heat production precedes heat loss and that CBT varies as an intermediate resultant.[8] Heat production and heat loss are separated not only in time but also in space in the body.[3] Under resting conditions, about 70% of heat production depends on the metabolic activity of inner organs, whereas body heat loss is initiated by heat redistribution from the core to the shell through blood flow to the distal skin regions.[3] Thermoregulatory distal skin blood flow is regulated by the autonomic nervous system by constriction or dilation of arteriovenous anastomoses. These are shunts between arterioles and venules, exclusively found in nonhairy distal skin regions such as the toes and fingers.[3] When they are open, warm blood flows rapidly and directly from arterioles to the dermal venous plexus, enabling an efficient heat exchange from the core to the distal skin. Sympathetic nerve activity seems to be crucial for peripheral vasoconstriction, but the exact neural process by which this regulation is achieved is still a matter of debate.[9] The endogenous time course of distal skin temperatures (hands and feet), measured during a CR, exhibits an inverse circadian rhythm in comparison with CBT, with a phase advance of about 100 minites[8] (i.e., in the evening, distal skin temperatures rise before CBT declines).[8] The amplitude of these distal skin temperature rhythms is about three times larger than that of CBT.[8,10] In contrast, temperatures of proximal skin regions (e.g., thigh, infraclavicular region, stomach, forehead) change in parallel with CBT, and the amplitudes are of similar magnitude.[8,10] This inverse relation between distal and proximal skin temperature rhythms reflects the differences in thermophysiologic regulatory mechanisms, as described earlier.[3] The distal minus proximal skin temperature gradient (DPG) therefore provides a selective measure of distal skin blood flow and hence body heat loss through the extremities.[3]

Nocturnal secretion of the pineal hormone melatonin, which is under control of the SCN, plays a crucial role in the endogenous downregulation of CBT in the evening.[11] Administration of melatonin in the afternoon, when endogenous melatonin levels are low, provokes exactly the same thermophysiologic effects, as observed naturally in the evening.[11] Whether melatonin induces distal vasodilation in humans by acting directly on blood vessel receptors, indirectly through modulation of sympathetic nerve activity, or both, remains to be determined.[11] In addition, both subjective ratings of sleepiness and the level of activity in the electroencephalogram (EEG) theta and alpha rhythms as an objective outcome of the sleep-wake state are increased.[11] Moreover, it is noteworthy that rise in melatonin secretion in the evening belongs to a well-orchestrated circadian physiologic regulation controlled by the SCN, which in turn downregulates CBT, increases sleepiness, and promotes sleep.

RELATIONSHIP BETWEEN THE SLEEP REGULATORY AND THE THERMOREGULATORY SYSTEM

The most evident explanation for whether and why the sleep regulatory and thermoregulatory systems are interrelated is a teleologic one: sleep is for energy conservation.[2,12,13] All species sleep or rest when their energy expenditure is low. Rest or quiet wakefulness is a prerequisite for sleep in all species.[1,12,13] These observations represent the starting point of all energetic explanations of the function of sleep. Human sleep evolved from ancestral sleep, and it is quite possible that earlier forms of sleep were linked to energy conservation in ancestors with a smaller body size.[14]

There are two mechanisms that explain how sleep can conserve energy. One is that sleep reduces energy expenditure indirectly by reducing activity. This mechanism would also work when animals only exhibit quiet wakefulness. Alternatively, sleep induces an additional decline in energy expenditure below that accomplished by quiet wakefulness by a change in physiology. Human sleep is only accompanied by a modest additional decline in energy expenditure.[13,15,16] However, energy conservation through sleep may be particularly important in small animals and infants.[2,13,17] Their high surface-to-body mass ratio is ideal to dissipate heat and renders energy conservation achieved by sleep highly adaptive.[13] When body size increases and sensory-motor systems mature, in the course of infant development a parallel decrease in sleep time occurs.[2,17]

COVARIATION OF SLEEP AND THERMOPHYSIOLOGIC VARIABLES

In the following subsections, two lines of evidence are presented to clarify the relationship between sleep and thermoregulatory systems: at baseline thermal comfort condition and after various conditions such as circadian, temperature, and sleep pressure changes. Furthermore, recent research has provided new insights into the relationship between thermoregulation and sleep on the basis of neuroanatomic studies showing significant interaction of the two systems.

Baseline Conditions

To compare the sleep and thermoregulatory systems, it is crucial to separate circadian from masking components of an overt diurnal pattern. This is much more easily accomplished in humans than in animals. Despite this advantage, the most neglected factor in human research is the so-called laying down effect. A change from standing to supine body position induces redistribution of blood, together with heat, from the core to the periphery, thereby increasing skin temperatures, decreasing CBT, and increasing sleepiness.[18] This effect lasts about 1 to 2 hours[18] and significantly confounds the endogenous time course of CBT in a classic human sleep recording protocol where laying down occurs about 20 to 30 minutes before lights off. The temporal relationship between thermophysiologic variables, heart rate, subjective ratings of sleepiness, and salivary melatonin secretion under CR conditions before habitual bedtime and for the following sleep episode is summarized in Figure 21-1. The only thing that changed during this protocol was that the low-intensity lights were switched off with the implicit permission to fall asleep. Before

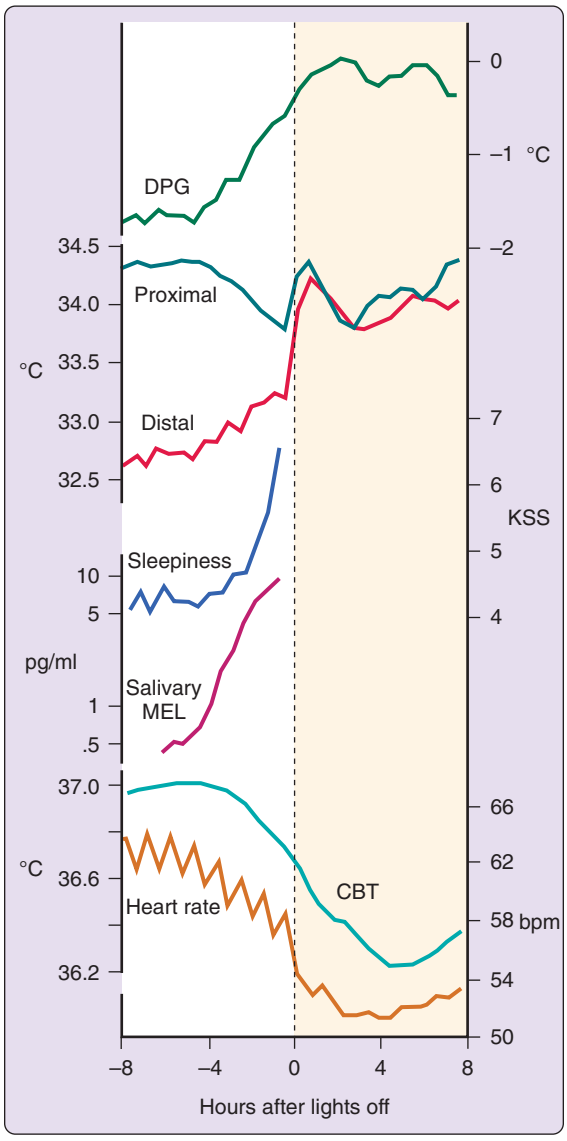


Figure 21-1 Time course of heart rate and core body temperature (CBT) (see lower traces) and changes in salivary melatonin concentration, sleepiness ratings, distal and proximal skin temperatures, and the distal-proximal skin temperature gradient (DPG) in a baseline 7.5-hour constant routine followed by a 7.5-hour sleep period, *yellow area.* Continuously measured data are plotted in 30-minute bins. Mean values of $N = 18$ male subjects. Subjective ratings of sleepiness: KSS, Karolinska sleepiness scale; MEL: melatonin; heart rate: bpm. *Note:* Distal and proximal skin temperatures exhibit inverse time course before lights off but are nearly indistinguishable approximately 1 hour later. Heart rate reflects the study protocol rhythm of one hourly food and water intake before lights off and declined sharply thereafter. Mean sleep onset latency: 12 ± 4 minutes. (Modified from Kräuchi K, Cajochen C, Werth E, Wirz-Justice A. Functional link between distal vasodilation and sleep-onset latency? *Am J Physiol Regul Integr Comp Physiol* 2000;278:R741–8.)

lights off, the previously described endogenous pattern of CBT downregulation is already visible. In the evening, heart rate (an indirect measure of intrasubject variation of heat production) declined first, followed by heat loss and finally by a decrease in CBT. Subjective ratings of sleepiness increased in parallel with DPG and salivary melatonin levels. The proximal skin temperature exhibited a similar pattern as CBT. Immediately after lights off and before sleep stage 2, the distal and proximal skin temperature increased and heart rate declined.[19] In addition, an increase in sweating is often observed, depending on CBT.[20]

The typical increase in distal skin temperature, as shown in Figure 21-1,[21] is caused by redistribution of heat from the core to the shell. Similar findings at sleep onset have been described in the lower leg.[22] However, CBT exhibited only a slight but significant increase in the rate of change after lights off,[3,21] leading to approximately 0.3°C lower CBT values during sleep compared with quiet wakefulness.[23] In contrast to the fast changes in skin temperature, the decline in CBT is slow, which can be explained by the reduced cardiac output during sleep initiation impeding a faster heat loss during the sleep episode under thermoneutral conditions.[19] The magnitude of the decrease in CBT is negatively correlated with environmental temperature.[24] A distal minus proximal skin temperature gradient (DPG) of 0°C indicates that during sleep the thermoregulatory shell has disappeared, resembling a state similar to that of the human body in the awake state in a warm environment (e.g., 35°C).[3] Heat redistribution from the core to the shell is completed within about 1 hour after lights off. Such a completely relaxed one-compartment body, when core and shell are fused, is prone to a fast cooling when sleep occurs in a cool environment. Under normal conditions, CBT is protected because humans and animals try to occupy a sleep berth in a comfortable thermal environment.[13] When humans initiate sleep outside the natural temporal niche by taking an afternoon nap, similar thermophysiologic changes occur right after lights off and before the initiation of sleep stage 2.[25] There are subjects, mostly women, who exhibit a proneness to cold hands and feet and therefore to having a large shell.[26,27] These subjects show alteration in some of the macrostructure variables of sleep such as a significant prolonged sleep onset latency to stage 2 (SOL2).[26,27] In fact, it has been shown that subjects with sleep onset insomnia respond well to a mild heating with reduced thermoregulatory heat loss from their fingers.[28] Within this context recent studies have shown that wrist skin temperature best predicts thermal sensation, especially in women, and therefore seems useful as a physiologic parameter to thermoregulatory behavior[29] such as using thermophysiologic remedies (e.g., bed socks).[22]

At the end of sleep, the transition to waking is accompanied by an inverse thermophysiologic pattern.[25,30] This period is named *sleep inertia;* after awakening it takes a certain time interval to recover all physiologic and cognitive functions.[25,30] During that time, a similar but inverse time course in distal vasoconstriction is observed.[25,30] It is noteworthy and of clinical relevance that similar thermophysiologic effects as seen during sleep initiation can be observed after administration of benzodiazepines[31] and with certain relaxation techniques like yoga, autosuggestion of warmth, autogenic training, and meditation without falling asleep.[3,31-33] These techniques induce a withdrawal in muscular and cutaneous sympathetic nerve activity, which leads to increased distal skin temperature, and to a reduction in heart rate, energy expenditure, and CBT.[31,32] Inverse effects were induced after caffeine administration with elevated CBT, distal vasoconstriction, and disturbed daytime recovery sleep and prolonged sleep onset latency after night sleep deprivation.[34] Therefore distal vasodilation followed by a drop in CBT appears to be a thermophysiologic event, which is primarily related to relaxation occurring before sleep onset,[35] and the opposite is true for vasoconstriction.

Studies carried out in humans to describe changes in thermophysiologic variables show that changes in CBT and proximal and distal skin temperature related to the non–rapid eye movement (NREM)–rapid eye movement (REM) sleep cycle are very small.[36,37] Heart rate is clearly increased shortly before and during REM sleep relative to NREM sleep, which is, however, reflected only in a minor increase in energy expenditure during REM sleep.[16] Extensive studies on thermophysiologic alterations regarding the NREM-REM sleep cycle concluded that changes in brain heat production are practically not relevant for changes in brain temperature.[38] To our knowledge, only one human study recorded brain temperature together with sleep EEG data, but no significant systematic changes regarding the NREM-REM sleep cycle were found.[39]

One of the advantages of animal research is the parallel recording of body and brain temperature. In many small mammals (rabbit, rat, Djungarian hamster), NREM sleep is associated with a decrease in brain temperature, whereas REM sleep and waking are associated with an increase[40,41] (Figure 21-2).

In an elegant study performed in the rat, it was shown that heat is redistributed across the body when vigilance states change.[42] At the initiation of NREM sleep, the brain and intraperitoneal temperature decreased, whereas the tail skin temperature increased. The opposite occurred at the transition from NREM sleep to wake. At transitions from NREM to REM sleep, brain temperature rose slightly, whereas intraperitoneal and tail temperature did not change. These data are in accordance with those obtained in humans. Heat is redistributed from the core to the shell at the onset of sleep. Humans thermoregulate by vasodilation and vasoconstriction of blood vessels within the skin of extremities; in rats similar changes are observed in the tail. The main difference lies in the timing of the redistribution of heat relative to the onset of sleep and waking. In humans changes are visible several hours before sleep onset; in the rat the same changes occur at the immediate onset of sleep. This difference is probably related to the smaller body size and to the shorter and repetitive ultradian sleep-wake pattern in the rat, which renders a time lag of several hours to be nonfunctional.

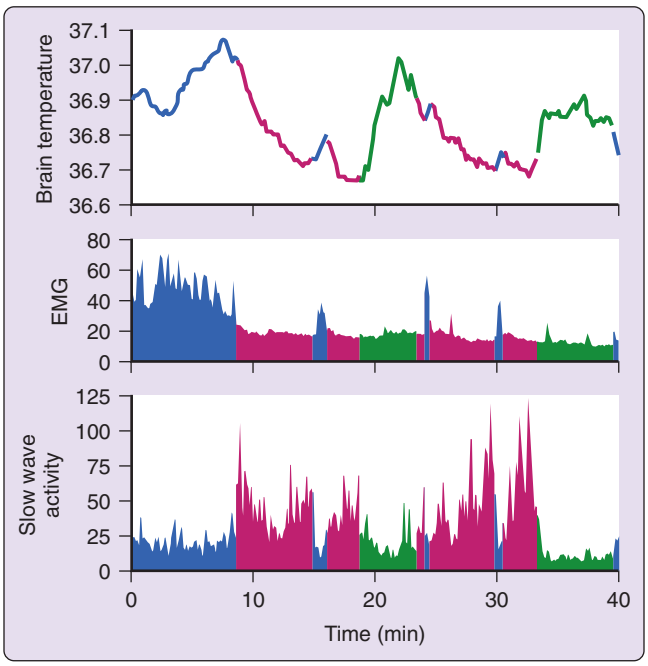


Figure 21-2 A 40-minute record of brain temperature measured at the parietal cortex, integrated electromyogram (EMG) activity from the neck muscles, and electroencephalogram (EEG) slow wave activity (SWA; mean EEG power density between 0.75 and 4.0 Hz) of a Djungarian hamster (*Phodopus sungorus*). *Blue*, Waking; *red*, NREM sleep; *green*, REM sleep. Values are plotted for 8-second epochs. Note the decrease of brain temperature at the entrance into NREM sleep and the increase during REM sleep and waking.

Applying a CR in rodents is not possible. However, on the basis of the relationship between brain temperature and vigilance states, it was possible to subtract the influence of vigilance state changes on brain temperature, rendering a mathematical CR.[43] This study concluded that, in the rat, about 90% of the variance in brain temperature is caused by changes in vigilance. A recent study confirmed that vigilance state–related changes in brain temperature are independent of the functioning of the circadian clock because they remained intact after removing the SCN.[44]

Taken together, there are robust thermoregulatory effects induced by lying down and the relaxing sleep behavior; however, the NREM-REM sleep cycle seems to have minor thermoregulatory function in humans. The thermoregulatory mechanisms, which are active during the wake-sleep transition, redistribute heat from the core to the shell and induce a decline in heart rate, energy expenditure,[15,16] and CBT.[37] Relaxing behavior before sleep in humans and animals belongs inseparably to sleep, and therefore these data do not contradict the energy conservation hypothesis of sleep.[14] The accompanying thermoregulatory effects in humans may be a remnant of their evolutionary past.[14]

Changed Circadian Conditions

It has been observed that subjects living under normal conditions choose their bedtime (lights off) at the maximal rate of decrease in their CBT rhythm.[45] However, when subjects are living on self-selected sleep-wake schedules in a time-free environment, bedtime is phase-delayed close to the CBT minimum, which is an indication that the sleep-wake cycle and the circadian rhythm of CBT are separate but usually entrained (synchronized) oscillatory systems.[46] Unfortunately, neither direct nor indirect measurements of heat loss and heat production were carried out in parallel in these studies. Therefore it is possible that CBT is not the crucial variable for sleep induction, but rather is one of its determinants (i.e., heat loss). Because heat loss seems to be closely linked to sleep initiation, it may be speculated that the circadian rhythm of heat loss is phase-delayed under free-run conditions. The duration of sleep episodes was maximum when initiated at the time when CBT reached its maximum, and at the opposite, minimal sleep lengths occurred when sleep was initiated during the rising phase of the CBT rhythm.[17]

Under most experimental conditions, REM sleep propensity exhibits a strong circadian pattern with a peak about 1 to 2 hours after CBT has reached its circadian minimum.[37,47] There is also a reproducible and robust circadian rhythm in SOL2, which is closely related to the circadian CBT rhythm and thermoregulatory effects described previously.[48] In forced desynchrony studies (i.e., living on a scheduled 28-hour day including a 9.3-/18.7-hour sleep-wake cycle), it was shown

that SOL2 is longest at the circadian phase where CBT reaches its maximum, that is, 1.5 hours before CBT starts to decline and melatonin secretion rises.[47] At this circadian phase, named the *wake-maintenance zone*,[49] inner heat conduction is lowest as indicated by the largest difference between CBT and distal skin temperature and the largest negative DPG values. Thereafter, SOL2 declines rapidly and is minimal around the time when CBT reaches its circadian trough, when inner heat conduction is largest (distal skin temperature is highest and the difference between CBT and distal skin temperature is lowest). However, it remains to be determined whether thermal interventions, like lower leg warming, at the wake-maintenance zone are successful to reduce SOL2, as was shown for melatonin administration.[50]

Taken together, self-selected sleep timing, SOL2, REM sleep latency, REM sleep propensity, and sleep duration are closely associated with CBT. Even though these variables are not fully in phase with CBT, it is possible that one of the determinants of CBT (e.g., heat production, heat loss) is directly interrelated. It still remains to be established whether these rhythms are independently governed by the SCN or causally linked directly to measured thermophysiologic outcomes. These correlative findings lead to the question of how is sleep affected by thermoregulatory challenges.

INTERVENTION STUDIES IN HUMANS

Effects of thermal interventions (heating or cooling) on sleep are not easy to investigate. Thermal interventions, applied either passively or actively by physical exercise, induce significant changes in skin temperatures and CBT.[15,17,37,51,52] The intensity of a thermal intervention is crucial, as are the skin region selected and the time of application. During sleep only passive thermal loads can be applied. It has been shown that sleep reduces the thresholds and gains of the autonomic temperature defense mechanisms and expands the inter-threshold zone (the temperature range for activation of metabolic heat production or evaporative heat loss).[17,52,53] These threshold changes are modest in slow wave sleep (SWS) but much stronger in REM sleep.[17,52] As a consequence, CBT and skin temperatures are more sensitive to changes in environmental temperature during sleep. Maximal total sleep time (TST) is found in the thermoneutral zone (the range of ambient temperature at which temperature regulation is achieved solely by vasomotor responses), whereas REM sleep is more vulnerable to thermal interventions than SWS.[17,52] Too intense thermal interventions induce arousals and awakenings, which in turn can induce thermoregulatory effects, such as elevating CBT.[52] When a thermal load is applied repeatedly, the thermoregulatory system can adapt and the effects on sleep are changed; for example, the arousing effects are reduced. Aborigines in the Central Australian desert and nomadic Lapps in Arctic Finland were experiencing comparable degrees of cold exposure during the night, and both showed lower thermoregulatory thresholds for shivering before modern technology arrived.[54,55] As a consequence, in these subjects CBT was more reduced during sleep, and undisturbed sleep occurred at a lower environmental temperature. Among limitations in actual knowledge is the fact that too many modalities of thermal interventions on sleep are understudied and, in addition, the effect of thermal interventions on sleep may differ between normal and sleep-disturbed subjects.

Changing Temperatures

Ambient temperature, especially in combination with high humidity, is important for both the quantity and quality of sleep.[52] When sleep occurs in a warm environmental temperature (31 to 38°C), duration of wakefulness increases, and at the opposite, duration of REM and NREM sleep decreases.[15,17,51,52] Also, cold exposure (21°C) induced more awakenings, less time in sleep stage 2, and less TST but did not affect the duration of the other sleep stages. Marked thermoregulatory effects were induced under such manipulations.[15] The decrease in CBT observed during the night episode was larger at 21°C compared with the thermoneutral 29°C condition. During REM sleep, forehead temperature and oxygen consumption increased and feet temperature decreased compared with SWS with cold exposure. Therefore cold-exposed humans may not exhibit a complete inhibition of thermoregulation during REM sleep as has been observed in small mammals.

When ambient temperature was gradually decreased during sleep, an earlier CBT nadir and an advanced peak for REM sleep propensity was obtained.[56] Duration of sleep stage 4 increased when the normal nocturnal decrease of CBT was augmented by a constant and mild reduction in ambient temperature, despite decreased sleep efficiency.[56] Similarly, after a 2°C reduction in ambient temperature during sleep, it was observed that the increase in SWS occurred simultaneously with the rise in slow wave activity (SWA; EEG power density ~1 to 4 Hz) without any change in sleep efficiency or reduced amount of REM sleep.[57] The thermal manipulation reduced not only leg skin temperature but also CBT and heart rate. Taken together, the augmentation of heat loss leading to reduced CBT during sleep seems to be the crucial variable for increased SWS.

In humans, body heat content and hence CBT can also be effectively manipulated by body immersion in warm or cold bathes. For instance, as a result of rapid conductive heat loss in a cold bath, CBT decreases faster compared with the drop observed in air at the same temperature. Rewarming of the cool shell after cool bathing leads to a characteristic after-drop in CBT.[58] Several studies showed effects of positive heat load on sleep,[15,51,52] but no study examined effects on sleep after a cold bath. In general, passive body heating (40° to 43°C for 30 to 90 minutes; CBT increase of 1.4° to 2.6°C) has a positive effect on many aspects of sleep in healthy young adults and in older and sleep-disturbed subjects. It was found that warm bathing in the evening shortened sleep onset latency, enhanced SWS duration, and sometimes reduced REM sleep duration. The increase in SWS, however, is not dependent on a reduction in REM sleep. Bathing performed in the morning or early afternoon had no effect on sleep architecture.[15,51] In principal, actual levels of CBT at sleep onset or the decline in CBT afterward could be related to the amount of SWS after warm bathing.[15,51] Additionally, a phase delay of the CBT nadir during the night sleep episode has been described after evening hot bathing, which correlates with increased SWS.[59] All these CBT characteristics could be correlated because directly after a positive heat load the velocity of CBT decline is larger, the CBT level is elevated before sleep onset, and the overt CBT nadir during sleep may be delayed. However, phase shifting effects of passive heat loads in humans have not been studied systematically. Other variables than CBT, such as skin

temperature, may play a role. The available findings are inconsistent because of the diversity in study designs and methodology and the low statistical power of many studies. In a study in which hot full-body and hot foot bathing were performed 35 minutes before lights off,[60] CBT increased by about 1°C only during full-body bathing. Both conditions, however, increased mean skin temperatures and reduced sleep onset latency and movement during sleep. These findings indicate that elevated skin temperature is crucial for a rapid onset of sleep, but not changes in CBT. Older sleep-disturbed subjects responded to hot foot bathing with slightly reduced sleep onset latency to stage 1 (SOL1) and significantly decreased wakefulness in the second NREM sleep period.[61] In these older subjects not only DPG but also CBT were elevated after hot foot bathing during the first hour of sleep. However, the same authors recently reported that warming the feet may improve sleep only for those who have cold feet.[62]

Of clinical relevance are the data describing that REM sleep duration and TST are reduced when electric heat blankets are used throughout the night,[63] suggesting that heat load exerted through the blanket is too strong an intervention and disturbs rather than supports sleep. In a series of experiments with a thermal suit, the effects of small changes in skin temperatures of only 0.4° to 2°C within the thermal comfort zone, without significantly altering CBT, were investigated on several sleep parameters.[64] It was demonstrated that intermittent elevation in skin temperature during the sleep episode suppressed nocturnal awakenings and triggered shifts to deeper sleep in young and older healthy subjects and in insomniac patients.[64] Whole-night studies are needed to confirm that a sleep depth–enhancing effect of mild skin warming can indeed be sustained. Nevertheless, these findings emphasize the importance of skin temperatures, primarily proximal skin temperatures (including the trunk), but also more distal skin region temperatures such as of the legs and arms, for these effects. Other studies revealed that subtle skin temperature warming was associated with a faster onset of sleep in young and older subjects and in older insomniac and narcoleptic patients.[64-66] At present, it cannot be concluded which thermophysiologic correlate represents the causal factor to increase SWS and reduce sleep onset latency. Nevertheless, heat load before sleep seems to increase the duration of SWS.

Intense exercise is a manipulation that can also raise CBT by 2°C or more. Subsequently, CBT declines as a result of the thermoregulatory heat loss drive through increased vasodilation and sweating. A number of reproducible changes on sleep have been identified after exercise in the evening: shortened sleep onset latency, increased TST and SWS, longer REM sleep onset latency, and less REM sleep.[67,68] Exercise exhibits negative effects on sleep when performed close to sleep onset; the optimal temporal positioning of physical activity is thought to be 4 to 8 hours before bedtime.[68] Chronic exercise studies have not provided much stringent evidence of a sleep-promoting effect. Conversely, with reduced exercise load in trained athletes, SWS and REM sleep onset latency were reduced and REM sleep duration and sleep onset latency were increased.[69] Taken together, after intense exercise, sleep appears to commence faster and is deeper.

In conclusion, warm distal skin temperatures induced either by endogenous circadian heat loss regulation in the evening, homeostatic downregulation of CBT after passive and active heat load, or selective skin warming predispose to a rapid onset of sleep. More sophisticated studies with respect to skin regions are necessary to show whether warming of shoulder, stomach, legs, hands, or feet, for example, exhibits the strongest effects on sleep initiation and sleep architecture. The increase in skin temperatures could be the causal factor for the acceleration of sleep onset and the increase of SWS. Further studies must investigate the optimal time interval between thermal intervention and bedtime and which physiologic mechanisms are involved in the observed effects. It is possible that thermal afferents provide a signal for the sleep-inducing brain regions in the hypothalamus.[31,70]

Changing Sleep Pressure

The studies of sleep deprivation effects on the thermoregulatory system cannot be understood without considering the circadian time at which the deprivation occurs. All thermophysiologic variables undergo significant circadian changes. Additionally, the effects of an experimental overnight sleep deprivation on the thermoregulatory system have to be controlled for changes in body position, locomotor activity, food intake, and light intensity. Using the CR protocol, it was shown that 40-hour total sleep deprivation does not change CBT, distal and proximal skin temperatures, heart rate, and energy expenditure despite the huge increase in sleepiness.[8] A comparison with a sleep pressure–reducing protocol, including regularly scheduled naps, provided evidence that changes in sleep pressure do not influence the thermoregulatory system.[12] Additionally, the nocturnal 8-hour sleep episodes before and after the two protocols revealed that CBT and distal and proximal skin temperatures did not differ even though a large difference in SWA was observed.[12] Taken together the circadian modulation of sleepiness is primarily related to the circadian regulation of distal vasodilation and hence to heat loss and circadian CBT reduction, whereas the homeostatic regulated increase of sleep pressure does not influence the thermoregulatory system,[19] contrary to earlier suggestions.[71] To be more conclusive, longer sleep deprivations may need to be performed to test whether the thermoregulatory system remains independent of sleep pressure.

INTERVENTION STUDIES IN RODENTS

Changing Temperature

The main interventions applied in rodent studies are manipulation of ambient temperature and brain temperature. In the rat, a general decrease in the daily percentage of REM sleep was seen when ambient temperature decreased,[42,72] indicating that REM sleep is very sensitive to changes in temperature and incompatible with low ambient temperature in the rat. Djungarian hamsters enter REM sleep more easily when brain temperature is relatively low,[73] but probably also in this species REM will disappear first when ambient temperature is lowered. In this context, low ambient temperatures are applied as a tool to investigate REM sleep regulatory mechanisms.[74,75]

In general the impression exists that increasing ambient temperature increases sleep pressure. In rats, when ambient temperature was increased to 33° to 35°C for 3 hours, resulting in a brain temperature of approximately 40°C, subsequent NREM sleep displayed more slow waves than in sleep-matched controls.[76] The amount of REM sleep did not change

compared with controls, and brain temperature was significantly decreased in the first 5 hours of recovery. Under these conditions animals slept less during the heating compared with baseline, indicating that too high ambient temperatures override sleep demand.

In two separate experiments in rats in which ambient temperature was increased to 30° to 32°C for 24 hours, cortical brain temperature was significantly increased by 0.3° to 1.0°C but hypothalamic temperature did not change.[77,78] This treatment resulted in one case in increased NREM sleep and in both experiments in an increase in SWA in NREM sleep in the dark period. These data indicate that changes in sleep can be induced by increasing ambient temperature without changing hypothalamic temperature.

Another approach is heating the POAH, increasing hypothalamic brain temperature locally, without changing ambient temperature. This approach resulted in increased SWA and NREM sleep during 1 hour of warming (1.0°C above baseline) in cats.[79] One hour of cooling (2.0°C below baseline) did not elicit a response. The data suggest that an acute increase in ambient or brain temperature (0.3° to 1.0°C) can increase NREM sleep and possibly increase the occurrence of SWA in the NREM sleep EEG.

Changing Sleep Pressure

During sleep deprivation, brain temperature is higher compared with baseline, and subsequent recovery is characterized by a decrease in brain temperature below baseline and an increase in NREM sleep and in SWA in NREM sleep.[41,78,80,81] This result was interpreted as a heat load incurred during the sleep deprivation that was subsequently recovered by increasing NREM sleep and SWA.[71] One of the clear results obtained from these experiments is a negative correlation between the amount of NREM sleep and the level of brain temperature.[73,80] There is no significant correlation, however, between SWA in NREM sleep and brain temperature,[73,80] ruling out the possibility that the depth of sleep determines brain temperature directly. Moreover, in Djungarian hamsters well adapted to a short winter photoperiod with a brain temperature 1°C below summer photoperiod brain temperature, recovery sleep after sleep deprivation is accompanied by an increase in brain temperature.[73] This is in contrast to the long photoperiod during which sleep deprivation is followed by a decrease in brain temperature.[41,73] A correlation between SWA and brain temperature, combining these data, supported the notion that brain temperature after sleep deprivation is set to the same temperature in both conditions,[73] suggesting that there may be an optimal temperature for high-amplitude slow waves in NREM sleep during recovery.

Two experiments in rats, in which ambient temperature was raised to 32°C during a sleep deprivation of 2.5 hours[78] or 3 hours,[81] did not result in similar outcomes. Short-lasting increases in SWA and NREM sleep were observed after a 2.5-hour sleep deprivation,[78] but not after a 3-hour sleep deprivation.[81] In contrast, a short-lasting increase in REM sleep was observed after a 3-hour sleep deprivation,[81] but not after a 2.5-hour sleep deprivation.[78] It can be questioned whether consistent results can be obtained in the rat with these short sleep deprivation durations. Probably a more systematic approach of scanning different ambient temperatures with longer sleep deprivations is needed to resolve these differences.

Brain Temperature, Electroencephalogram, and Thermosensitive Neurons

The EEG is influenced by changes in brain temperature as well. From analysis of the EEG of the Djungarian hamster during spontaneous entrance into the hypothermic state of torpor (see Hibernation, later) and from experiments in which either rats, cats, or humans were cooled, it was found that the amplitude and frequency of the EEG changes when brain temperature decreases. The amplitude becomes smaller, and prominent frequencies in the EEG slow down with decreasing temperature.[82] Recently the slowing down of the EEG was confirmed in rats in which hypothermia was induced pharmacologically by inhibiting neurons in the central nervous pathways for thermoregulatory cold defense.[83] This relation between EEG frequency and brain temperature was shown to follow a Q_{10} of approximately 2.5,[84] which means that the frequency became 2.5 times slower when brain temperature decreased by 10°C. Under influence of euthermic changes this effect is relatively small, but it can be significant even for frequencies below 5 Hz.[82] Faster frequencies like the theta rhythm (6 to 9 Hz) in rodents[82,85] and frequencies above 10 Hz[82] are significantly influenced by these daily changes in brain temperature.

Measuring the electrical activity of neurons in the POAH revealed the activity of two distinct types of neurons that either increase or decrease firing rate when brain temperature increases. The latter are called *cold-sensitive neurons,* whereas the first group is called *warm sensitive*. A biochemical process (i.e., neuronal firing) that slows down when temperature is increased is quite unique and therefore cold sensitive neurons, when observed, can be considered to be genuine. In contrast, a biochemical process that speeds up when temperature is increased is quite normal and was theoretically explained at the end of the nineteenth century.[82] Many processes, ranging from the firing rate of SCN neurons,[86] to the frequency of prominent EEG waves,[82] to muscle contraction,[87] double or triple when temperature is increased by 10°C ($2 < Q_{10} < 3$).

To identify warm-sensitive neurons, two criteria are applied in the literature. The first determines that an increase in firing rate needs to be more than double when temperature is increased by 10°C ($Q_{10} > 2$).[88] The second says that the increase in firing rate needs to be more than 0.8 impulses per second per 1°C of warming.[89] Both are insufficient. The criterion of a Q_{10} above 2 ignores the fact that most biochemical processes have a Q_{10} somewhere between 2 and 3. Therefore, a Q_{10} of at least 3 needs to be reached before one can be relatively sure that the change in firing rate can be distinguished from the passive biochemical response of the temperature insensitive neurons. With the second criterion, fast firing neurons have a relatively large chance of being included even when they follow the passive biochemical Q_{10} rule of doubling firing rate when temperature is increased by 10°C. Nevertheless there are genuine warm-sensitive neurons in the POAH[90] and other brain areas, such as the diagonal band.[91]

The firing rate of warm- and cold-sensitive neurons in the POAH is known to be vigilance state dependent. Most warm-sensitive neurons increase their activity at the onset of NREM sleep. On the other hand, most cold-sensitive neurons are more active during waking.[90,91] Those results emphasize the importance of simultaneous polysomnographic recordings to be able to disentangle the vigilance state–related changes in firing rate from temperature-related changes.[92] Noradrenergic

afferents from sleep-wake regulatory centers like the locus coeruleus and the lateral tegmental system are also involved in the change in firing rate observed in the POAH.[93] The changes in firing rate of the ensemble of neurons are thought to shape the sleep-wake response to thermoregulatory demands encountered by the animal.

HIBERNATION

The hypothermic state observed during the hibernation season may be a valuable model to investigate the relationship between thermoregulation and sleep and may generate relevant data to the understanding of human physiology. Most hibernating mammals are small and weigh between 10 and 1000 g.[94] During the hibernation season these animals spend a considerable amount of time in a torpid state with body temperature below euthermia (i.e., <30° to 32°C). This sustained hypothermic state is entered voluntarily and can be terminated by the animal. Body temperature can drop by more than 35°C,[95] and the metabolic rate is only a fraction of that during normothermia.[96]

In hibernators the torpid state can last for several days or weeks. A group of smaller mammals with body weights between 5 and 50 g display daily torpor in which body temperature is dropped for a couple of hours during the rest phase but returns to euthermia during the active phase.[94] Both hibernation and daily torpor result in substantial energy savings[97,98] and are therefore considered to be adaptive mechanisms that permit the conservation of energy during unfavorable environmental conditions.

Whether circadian rhythms continue during hibernation is unclear. In hibernating ground squirrels, rest-activity rhythms only slowly reappeared a couple of days after termination of torpor bouts, and this correlated with the number of vasopressin-containing neurons in the SCN.[99] Also, sleep-wake distribution in hibernating ground squirrels did not show a circadian modulation.[100] In contrast, responses to, for instance sleep deprivation, temperature changes, and other homeostatic regulatory processes, are still functioning.[101-105] Under certain circumstances a small circadian modulation of body temperature can be observed during hibernation[106]; however, it is possible to conclude that the contribution of the circadian clock is reduced, whereas homeostatic regulation seems to be virtually identical to its regulation outside the hibernation season.

THERMOREGULATION AND METABOLIC RATE REDUCTION

The mechanism by which the animals reach the reduction in metabolic rate is still controversial. The traditional view was that metabolic rate falls as body temperature decreases at torpor entry. The Q_{10} of metabolic rate between euthermia and torpor is often close to 2, which is typical for biochemical processes[82]; therefore the reduction in metabolic rate seems to be explained solely by the temperature effect on biochemical processes in the body.[107,108]

However, Q_{10} values above 3 for metabolic rate have been observed during torpor entry and during torpor at relatively high temperatures. It was proposed that an additional phy siologic inhibition must be involved in the reduction of metabolic rate.[109,110] Metabolic rate is downregulated before torpor entry, and the decrease in body temperature is the consequence and not the cause of the metabolic rate reduction.[111-113] As an alternative hypothesis it was proposed that metabolic rate is a function of the difference between ambient temperature and body temperature, similar to that during euthermia.[111] Because this difference is generally very small during torpor, metabolic rate is equally reduced.

Inhibition of metabolic rate during torpor may be caused by reduced pH, which slows down metabolic processes.[114] In hibernating ground squirrels the respiratory quotient drops during entrance into torpor and rises during subsequent arousal, suggesting CO_2 storage, which may result in decreased pH. In contrast, in Djungarian hamsters, who display daily torpor, respiratory quotient increases during entrance into torpor and decreases before emergence from torpor. Changes in enzyme activity are other candidates for metabolic rate reduction. Mitochondrial respiration is reduced by 50% during torpor in hibernating ground squirrels compared with euthermic individuals.

The previous data support the notion that the mechanism of metabolic rate reduction differs between hibernators and animals that use daily torpor.[114] The reduction in metabolic rate in animals who display daily torpor is largely determined by the decrease in body temperature, whereas hibernators seem to apply some kind of extra reduction in metabolic rate.

For some essential but unknown reason, deep torpor in hibernators is interrupted on a regular basis by short (<24 hours) euthermic periods.[115,116] Above 0°C body temperature, torpor bout length correlates negatively with ambient temperature,[117] and metabolic rate depends on ambient temperature and therefore body temperature. The timing of the euthermic period (also called *arousal*) correlates with the ambient temperature and the metabolic rate of the animal during the preceding hibernation bout.

More recent experiments indicate that the arousal is probably not induced by a process accumulating in the course of a multiday torpor bout but that there exists a separate arousal process of which the onset is more susceptible at higher body or ambient temperatures.[118] Although the mechanism of the arousal is not known, it appears to be totally endogenous in its origin.

The question of the reason for the arousal is important. The function of hibernation is energy conservation, and energy expenditure during the hibernation season is reduced to less than 15% of that the animals would have expended if they remained euthermic throughout the winter season.[96] However, there is still room for improvement because the energetic costs of the periodic arousals constitute 64% to 90% of the total energy expenditure during the hibernation season.[96,119,120] It has been proposed that arousals are required to eliminate metabolic waste products, to replenish blood glucose levels, or to restore cellular electrolyte balance. All these hypotheses have not survived critical experimental testing.[121] Based on EEG observations it was proposed that animals terminate torpor and return to euthermia to restore a sleep debt.[122,123] NREM sleep was deepest at the beginning of a euthermic period, and most of the euthermic period was spent in sleep. The putative restorative function of NREM sleep was thought to be incompatible with torpor.

Torpor and Sleep

Animals in torpor appear to be sleeping. They remain in their nest in a sleeplike posture with elevated arousal thresholds.

Based on several behavioral and physiologic findings it is generally accepted that torpor has evolved as an extension of sleep. When the animals enter torpor, analysis of the EEG shows that rodents are mainly in NREM sleep and that REM sleep is reduced or not present.[124,125] Recordings of neuronal activity in the hypothalamus of hibernators with brain temperatures of 10° to 20°C indicate that these animals keep alternating between long NREM sleep bouts and short waking bouts.[126] Below 10°C it is not possible to determine vigilance states with electrophysiologic methods.[126] These findings supported the hypothesis that NREM sleep is an adaptive behavior for energy conservation, which function is strengthened during torpor.[12,98,127]

However, when the animals subsequently emerge from torpor they immediately enter deep NREM sleep, irrespective of whether animals emerge from deep hibernation[122,123] or daily torpor.[125] This observation suggests that during deep torpor the function of sleep cannot be fulfilled completely and animals have to return to euthermia to recover from a sleep deprivation incurred during the hypothermic state. In the Djungarian hamster the hypothesis was confirmed by combining daily torpor with sleep deprivation experiments.[101] In contrast, similar experiments in hibernating ground squirrels resulted in rejection of the sleep deprivation hypothesis.[103,104] Recently, after recovering from pharmacologically induced hypothermia in rats, the increase in deep sleep was also observed.[83]

There appear to be fundamental differences between animals that display daily torpor and hibernating animals that display multiple-day torpor bouts. The mechanisms of metabolic rate reduction seems to differ.[114] Another fundamental difference between the two groups of animals is the effect of torpor on subsequent sleep. In hibernating animals who display deep torpor similar to ground squirrels or hamsters, regulation of temperature and sleep seems to be reduced to a level that currently cannot be measured reliably. In contrast, animals who display daily torpor appear to make use of the same processes, although applied in an extreme way, which reduces body temperature at the onset of sleep in humans. This latter similarity may provide an opportunity to investigate the relationship between sleep and body temperature with much larger variability.

CLINICAL PEARL

Humans with cold hands and feet are predisposed to difficulties initiating sleep. Better knowledge about sleep thermophysiology may help to develop evidence-based thermal interventions to alleviate sleep disturbances and to manage certain aspects of sleep and alertness problems in the general population. Therefore to assess dysfunctional thermoregulation, clinicians can use objective skin temperature measurements (e.g., wrist and instep seem promising sites for gender-related cold sensation assessment in relation to sleep) in conjunction with patient self-reports about thermal discomfort (e.g., sensation of cold hands and feet) to diagnose, monitor, and advise patients on their sleep disturbances or thermal discomfort–related complaints.

SUMMARY

The human sleep-wake cycle is tightly coupled to the circadian time course of core body temperature. The evening increase in heat loss through distal skin regions and reduction in heat production is associated with sleepiness and the ease to fall asleep, whereas the homeostatic increase in sleep pressure does not influence the thermoregulatory system. After sleep initiation, ultradian NREM-REM sleep cycle fluctuations seem to have minor thermoregulatory functions, especially in humans.

From experimental data it can be concluded that mild warming can increase sleep propensity, sleep consolidation, and the duration of SWS. More reproducible systematic investigations applying temperature levels within the thermoneutral zone on different skin regions are needed to develop applicable thermal therapeutic strategies for sleep disturbances.

The preoptic anterior hypothalamus integrates input from brain areas involved in circadian, temperature, and sleep-wake regulation and in turn influences vigilance states and body temperature in response to that input.

In animals, the torpid state may be a valuable model to investigate the relationship between thermoregulation and sleep. During daily torpor, similar physiologic processes occur as during normal entrance into sleep, but this is observed in a more extreme way, providing an excellent opportunity to investigate these processes in more detail.

Selected Readings

Colbert RW, Holley CT, Stone LH, et al. The recovery of hibernating hearts lies on a spectrum: from bears in nature to patients with coronary artery disease. *J Cardiovasc Transl Res* 2015;[Epub ahead of print].

Deboer T. Brain temperature dependent changes in the electroencephalogram power spectrum of humans and animals. *J Sleep Res* 1998;**7**(4):254–62.

Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci* 1995;**15**(5 Pt 1):3526–38.

Geiser F. Metabolic rate and body temperature reduction during hibernation and daily torpor. *Annu Rev Physiol* 2004;**66**:239–74.

Horne JA. Why sleep? *Biologist* 2002;**49**(5):213–6.

Jacquot CM, Schellen L, Kingma BR, et al. Influence of thermophysiology on thermal behavior: the essentials of categorization. *Physiol Behav* 2014;**128**:180–7.

Kräuchi K. The thermophysiological cascade leading to sleep initiation in relation to phase of entrainment. *Sleep Med Rev* 2007;**11**(6):439–51.

Krauchi K, Deboer T. The interrelationship between sleep regulation and thermoregulation. *Front Biosci (Landmark Ed)* 2010;**15**:604–25.

Mekjavic IB, Eiken O. Contribution of thermal and nonthermal factors to the regulation of body temperature in humans. *J Appl Physiol* 2006;**100**:2065–72.

Parmeggiani PL. REM sleep related increase in brain temperature: a physiologic problem. *Arch Ital Biol* 2007;**145**:13–21.

Siegel JM. Do all animal sleep? *Trends Neurosci* 2008;**31**:208–13.

Van Someren EJW. Mechanisms and functions of coupling between sleep and temperature rhythms. *Prog Brain Res* 2006;**153**:309–24.

A complete reference list can be found online at ExpertConsult.com.

Memory Processing in Relation to Sleep

Philippe Peigneux; Stuart Fogel; Carlyle Smith

Chapter 22

Chapter Highlights

- Sleep contributes to memory consolidation in the long term.
- Sleep stages and specific sleep features (e.g., brain oscillatory activity, spindles) exert different functions in memory consolidation processing.
- Disturbed and reduced sleep affect cognitive functions, especially learning and memory.

In 1867, Hervey de Saint Denys, fascinated by his dreams since the age of 14, published *Les Rêves et les Moyens de les Diriger*.[1] In his book, he reported a series of ingenious experiments showing that experienced events are incorporated into our dreams, in which they can be combined to create original associations between "memory images" of the past. Hence he opposed the idea that sleep may be a sudden drop in a state of cognitive nonbeing in which our resting brain is disconnected. On the contrary, he claimed that "sleep without dreams cannot exist, just as wake without mentation does not exist." Besides dreaming activity however, already addressed in this book, we now know that the sleeping brain houses a great variety of cognitive processes, including, to name a few, the ongoing processing of external stimuli, the revival of past experiences, and last, but not least, the consolidation of new information into memory. Interestingly, recognition that persistence of mental activity in the sleeper may be an integral part of the physiologic processes that subserve memory consolidation only arose in the last quarter of the twentieth century. It is now widely accepted that the sleeping brain is highly active and dynamic, but until recently it was not well understood to what end this heightened activity may serve, beyond the initiation and maintenance of sleep itself.

This chapter aims at introducing issues surrounding the role that sleep may play in memory consolidation, focusing on studies in humans. Reviews in which relationships between sleep and memory in animals are addressed can be found elsewhere.[2-4] We outline here key findings and milestones in probing the sleep for memory hypothesis, as well as ongoing debates and thoughtful questions that remain to be solved. First, we briefly introduce the concepts of memory consolidation and memory systems in humans. Indeed, previous chapters have shown that sleep is a multidimensional state of vigilance, composed of rapid eye movement (REM) and non–rapid eye movement (NREM) sleep states characterized by distinctive features and relying on specific neuroanatomic substrates.[4-5] It is these biologic markers of the various sleep states, rather than the categorical stages of sleep per se, that have been the focus of much of the recent investigation of the role of sleep for memory and are discussed in the following sections. Likewise, it must be kept in mind that memory is not a unitary phenomenon, both from cognitive and neurophysiologic perspectives. Rather, *memory* should be seen as a generic concept for information storage, encompassing a series of specific subdomains. Consequently, the interaction between multidimensional states of sleep and distinctive memory systems makes it logical that not all sleep manipulations will affect performance to the same extent, depending on the nature and demands of the memory task. Hereafter, we examine the growing number of behavioral studies that have enlightened our understanding of the role played by sleep episodes in memory consolidation. Besides, the rapid evolution of neurophysiologic techniques and analytical approaches now allow scrutiny of the complex relations between cognitive processes and their underlying neural substrates. Therefore it is timely to focus part of this chapter on the neurophysiologic mechanisms acting in sleep to support, or at least favor, memory consolidation processes. The relationship between consolidation of newly formed memories and sleep is a crucial issue because memory is at the root of most of our daily behaviors, such as simple skill acquisition (e.g., typewriting), sophisticated operational procedures (e.g., using computer-based systems), or keeping track of personal events and relationships, but is also an integral component of mental health therapeutic interventions.[6]

MEMORY SYSTEMS AND MEMORY CONSOLIDATION

Humans are able to learn, store, and remember various types of information in different ways and for variable periods of time, from conscious acquisition strategies to incidental detection of environmental events. Cerebral damage can selectively alter some of these processes while leaving others undisturbed. These simple observations have led to the proposal that memory is not a unitary phenomenon, but rather should be seen as a complex construct of more or less specialized memory subdomains that interact in ways that remain to be elucidated. Long-term memories in humans may further belong to multiple systems, primarily delineated between declarative and nondeclarative memories (Figure 22-1). Distinguishing features of declarative memory are that information is easily accessible to verbal description and that encoding or retrieval is usually carried out explicitly; that is, the subject is aware that the stored information exists and is being accessed. Declarative memory further comprises semantic and episodic

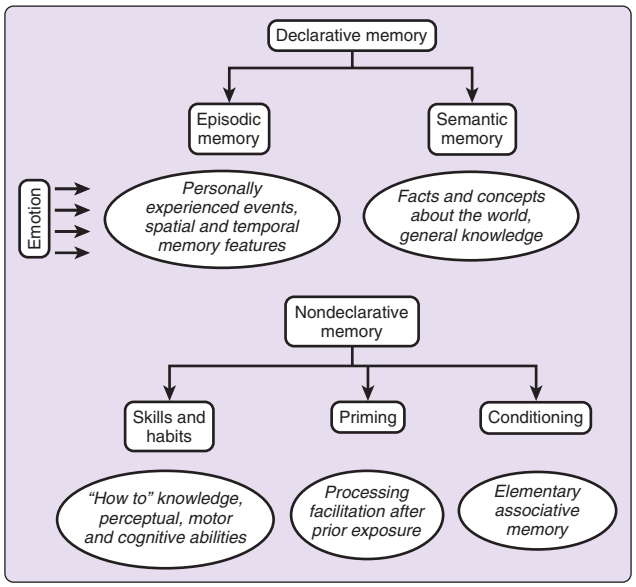


Figure 22-1 Schematic organization of long-term memory systems.

memory components. Semantic memory is the receptacle for our general knowledge about the world, regardless of the spatiotemporal context of knowledge acquisition. Conversely, episodic memory refers to the system that stores events and information along with their contextual location in time and space. On the other hand, distinctive features of nondeclarative memories are that they are not easily accessible to verbal description and can be acquired and reexpressed implicitly. It means that our behavioral performance can be affected by the new memory even if we are not necessarily consciously aware that new information has been encoded or is retrieved. Importantly, memory abilities aggregated under the nondeclarative label also include skills, habits, priming, and conditioning. Although grouped under the nondeclarative label, these various processes are subtended by distinct neuroanatomic substrates both in humans and animals, further suggesting their relative independence.

In addition to the transverse division between long-term memory subsystems proposed by these influential models, it is important to emphasize that newly acquired information is not immediately stored at the time of learning in its final form, if such a stable state exists. Rather, memories undergo a series of transformations over time periods ranging from hours to days, maybe even years, during which they are gradually incorporated into preexisting mnemonic representations[7-9] or are subjected to forgetting.[10] This dynamic longitudinal process refers to the concept of *memory consolidation*, which can be defined as the time-dependent process that converts labile memory traces into more permanent or enhanced forms.[11] Eventually, time-dependent processes of consolidation and the ensuing robust memory trace will enduringly adjust behavioral responses to the recent environmental changes, thereby expanding the organism's behavioral repertoire.[12,13]

In this framework, scientific evidence suggests that sleep and associated processes of brain plasticity[4] are major players in time-dependent processes of memory consolidation, acting as key constituents in the chain of transformations that help integrate information for the long term. Furthermore, these studies have suggested that the electroencephalogram (EEG) features of sleep and their respective sleep states that they characterize may have distinct memory-related functions, which has been interpreted in two different but nonexclusive manners. According to the *dual-process hypothesis*, REM and NREM sleep act differently on memory traces depending on the memory system or process to which they belong. For instance, it has been proposed that slow wave sleep (SWS) facilitates consolidation of declarative and spatial memories, whereas REM sleep facilitates consolidation of nondeclarative memories[14,15] (but see an alternative interpretation of sleep and declarative memory consistent with Tulving's serial-dependent-independent [SPI] model,[16] or an alternative explanation of the relationship between sleep states and nondeclarative memory[17]). Other models hypothesized that memory processing during sleep takes place in a sequential manner whereby particular transitions from one sleep state to another each handle particular aspects of memory consolidation.[18-21] Both approaches assume that it is sleep on the first posttraining night that is important for memory consolidation. Notwithstanding, it would be inappropriate to claim that *only* sleep may achieve the necessary conditions to consolidate novel memories in the nervous system because both sleeplike cognitive and neural processes of memory consolidation have also been observed during wakefulness.[22-24] The precise nature of the relationship between online and off-line consolidation processes remains to be fully elucidated.

MAIN METHODS FOR STUDYING THE ROLE OF SLEEP FOR MEMORY CONSOLIDATION

Three broad classes of experimental approaches have been used to test the hypothesis that sleep exerts a favorable or promoting effect on memory consolidation: (1) effects of posttraining sleep deprivation on memory consolidation and on the reorganization of the neural substrates of long-term memories, (2) effects of learning on posttraining sleep and reexpression of behavior-specific neural patterns during posttraining sleep, and (3) effects of within-sleep stimulation on sleep patterns and overnight memories.

Posttraining Sleep Deprivation

The first and likely earliest type of investigation has probed the putatively detrimental effect of sleep deprivation on the night following learning, based on the assumption that memory performance over the long term will be better if participants are allowed to sleep after learning, compared with sleep-deprived subjects. In classic experimental procedures, subjects learn new material. Afterward, some participants are allowed to sleep normally, whereas others either do not sleep at all (total sleep deprivation), or awakened at the onset of occurrences of the sleep stage under study (selective sleep deprivation), are kept awake during the period of the night in which the sleep stage is predominant (partial sleep deprivation), or have a shortened sleep duration (sleep restriction). Finally, prenight and postnight memory measures are compared between sleeping and sleep-deprived subgroups, either the next day or several days later. This approach was first used more than 80 years ago[25] and revealed that the normal forgetting curve for newly learned verbal material was significantly dampened by the presence of an intermediate period of sleep.

Notwithstanding, investigators did not interpret this effect as a specific role of sleep per se in memory processing but merely as a protective role that sleep may have against "interference, inhibition, or obliteration of the old by the new."[25] However, this hypothesis of a purely passive role for sleep in memory processes has been challenged by selective sleep deprivation studies attributing a specific role to REM sleep in memory storage and consolidation, both in human and animal species.[26] Also, it was demonstrated that memory over an interval with relatively high amounts of SWS was superior to memory over an interval with relatively high amounts of REM sleep.[27] This seemingly apparent contradiction was resolved later with the demonstration that recall of paired-associate lists was significantly better after sleep than wakefulness in the SWS-rich first part of the night only, whereas consolidation of mirror-tracing skills specifically benefited from sleep in the REM-rich second part of the night.[14] Other experimental manipulations inferred a specific role for stage 2 sleep in the second part of the night for the consolidation of motor memories.[28,29] Hence, sleep deprivation studies have suggested that all stages of human sleep (REM, SWS, and stage 2 sleep) might be actively involved in distinctive ways in learning and memory consolidation processes.[30,31]

In the recent past, brain imaging investigations have complemented these findings using neuroimaging techniques, especially functional magnetic resonance imaging (fMRI). As in behavioral paradigms, subjects are trained to the task and then either sleep deprived or allowed to sleep on the following night. A few days later, cerebral activity is recorded during memory retrieval and contrasted between the two posttraining sleep conditions. One or two additional nights of regular sleep are usually allowed after the posttraining night before testing in the scanner to avoid different arousal states that may confound neural activity associated with memory retrieval between sleeping and sleep-deprived subjects. For similar reasons day (wake) versus night (sleep) comparisons[32] or daytime napping protocols[33,34] have been used. These neuroimaging studies have yielded two important contributions. First, they have demonstrated that sleep deprivation during the posttraining night eventually impedes the reorganization and optimization of the cerebral activity subtending delayed retrieval of consolidated memories during wakefulness.[35-38] Second, they have shown that sleep-dependent changes in memory-related brain activity patterns may be present even in cases in which similarity in behavioral performance between posttraining sleep conditions suggests an absence of sleep-related effect on memory.[36,38,39] The latter results further indicate that long-term memory performance can be achieved using different cerebral strategies initiated as a function of the status of sleep during the posttraining night.

Posttraining Sleep Modifications

This second type of investigation stemmed from the reasoning that if newly acquired information underwent an ongoing process of consolidation during sleep, then this should be reflected in neural and physiologic features of posttraining sleep.[40-72] These changes in the characteristics of sleep provide a window into the processes that underlie sleep-dependent memory consolidation.

Numerous EEG studies have demonstrated that both the architecture of sleep and distinctive features of sleep stages can be affected by prior learning experience. For instance, postlearning sleep modifications have been observed by looking at absolute or proportional (i.e., relative to total sleep time) increases in the duration of REM sleep,[21,30,42-44] stage 2 sleep,[45] and SWS[46] episodes. Other studies have reported increased density of rapid eye movements[29,47] and stage 2 spindle activity in the sigma frequency band,[48-55] as well as increases in REM sleep theta power.[56] Many found relationships between quantitative parameters of sleep and overnight performance improvements[21,44,48-50,53-55] or levels of performance at the end of learning,[57] suggesting a close link between changes in sleep physiology and memory consolidation. In support of the double-step hypothesis described previously, others studies demonstrated relationships between performance changes and the organization of NREM-REM sleep cycles.[58,59] Hence these investigations, mostly based on noninvasive electrophysiologic techniques, have consistently demonstrated that prior learning during the day influences the physiology of sleep.

Noninvasive neuroimaging studies, initially using positron emission tomography (PET) measurements and more recently fMRI studies, have taken advantage of their superior spatial resolution and whole brain coverage to complement this picture, in looking more precisely at the neural correlates of postlearning modifications during sleep. Most important, they have provided evidence that neural activity occurring during memory task practice in learning-related cerebral structures can be reexpressed or continued during both REM[63,64] and NREM[65] stages of sleep as well as during posttraining wakefulness.[22] More recently, advances in simultaneous EEG-fMRI recordings have enabled investigating functional activations in brain regions associated with sleep EEG features. These studies showed that activity in brain regions active during learning and then during subsequent sleep is time-locked to phasic events, such as sleep spindles.[66,67] These data are in close agreement with intracerebral recording studies in animals, which demonstrated neuronal reactivation during sleep,[68,69] although these seminal animal studies did not seek to establish a link to memory consolidation or behavioral relevance of this reactivation per se. In this context, a significant contribution of human neuroimaging data was to show that experience-dependent reactivations of local (hippocampal) activity during SWS correlated with overnight gains in memory performance after spatial navigation.[65] Conversely, levels of implicit procedural learning achieved before sleep correlated with the amplitude of reactivation in cortical areas during REM sleep,[64] when connectivity patterns between learning-related areas were additionally reinforced.[70] Functional MRI studies also revealed reactivation of hippocampal activity time-locked to sleep spindles following both declarative[66] and procedural[67] learning. Moreover, procedural learning-dependent reactivation recruited other task-relevant structures, including the putamen,[67] which correlated with overnight improvement. Taken together, human reactivation studies have suggested that there is neuronal replay of previous experience during sleep and that posttraining sleep activity in brain areas involved during the learning episode represent a neural signature of memory-related cognitive processes, possibly linked with phasic sleep events such as spindles. An alternate but not exclusive hypothesis is that learning at wake induces local synaptic changes which themselves induce local changes in slow wave activity (SWA), the main marker of

sleep homeostasis, and that these changes are ultimately beneficial to imprinting novel memories.[71,72]

Within-Sleep Stimulations

Finally, a third approach using more causal manipulations to demonstrate enhanced (or diminished) sleep-dependent memory consolidation has consisted of providing specific stimulations during sleep with the aim to investigate whether meaningful stimuli could be recognized or new associations formed or whether presleep learning can be modified by nonawakening stimulations.[73-94] Indeed, evidence for such phenomena would be indicative that active plastic processes are taking place during sleep.

Demonstrating an active and potentially causal role for sleep in memory consolidation processes, several studies have established that stimulations within the posttraining sleep period may enhance performance compared with a normal, unmodified posttraining night.[84-88] Indeed, presentation of nonawakening auditory stimulations during REM sleep after Morse code learning[84] or re-presentation during REM sleep of sounds heard in the background while learning a complex logic task[85] increased overnight memory performance. Importantly, the effect was only present when auditory stimulations were displayed in coincidence with the bursts of rapid eye movements that reflect phasic ponto-geniculo-occipital (PGO) activity in humans, further suggesting that the characteristics of sleep are important markers and perhaps are even tied to cellular mechanisms of memory consolidation processes but reflected at a macroscopic level. Likewise, presentation during SWS of odors that were used as contextual cues during the learning episode triggers hippocampal responses and improves overnight retention of declarative memories.[88] Reactivation of memory traces is not merely an epiphenomenon of sleep, but rather is highly specific to the previously learned task. Indeed, in subjects trained to produce two musical melodies by following sequences of movements, motor sequence accuracy was significantly more enhanced for the melody that was replayed during sleep,[89] and improvement correlated with the amount of posttraining SWS and sleep spindles. Likewise, transcranial direct current stimulation that modulates excitability in cortical areas improves declarative memory when applied during SWS,[87] especially when application of oscillating potentials at about 0.75 Hz induced slow oscillation-like potential fields that mimic the slow oscillations of deep NREM sleep.[86] Finally, artificially maintaining high levels of cortisol feedback and cholinergic tone during SWS impairs hippocampus-dependent declarative memory formation,[90-92] suggesting that the natural shift in central nervous system cholinergic tone from high levels during acquisition-related wakefulness periods to minimal levels during SWS optimizes declarative memory consolidation.[93] Conversely, preventing natural increases in cortisol during REM sleep periods appears to enhance amygdala-dependent emotional memory.[92] These studies have demonstrated the beneficial (or detrimental) effects of various stimulations and manipulations within the posttraining sleep periods on overnight gains in performance and therefore have provided stimulating evidence that sleep does not merely play a passive role in memory processing by protecting novel memories from interference. Rather, they support the hypothesis that sleep acts in a complex manner in providing optimal conditions for the consolidation of novel memories in the nervous system.

SLEEP AND DECLARATIVE MEMORY

As described previously, long-term declarative memory comprises (1) semantic and (2) episodic memory components. Experimental data indicate that the role of sleep in consolidating these two memory components may be dissociated to a certain extent and that (3) emotional variables play a modulatory role in episodic memory consolidation. These aspects are covered in the following section.

Semantic Memory

Few studies have looked at the role of sleep for consolidation of semantic information per se, although event-related potential studies have demonstrated that semantic processing of externally presented stimuli is possible during REM and stage 2 sleep, but not during SWS.[76,81,82] Also, it has been shown that semantic priming (i.e., the facilitating processing effect resulting from prior presentation of semantically related material) qualitatively differs on awakening from stage 2 and REM sleep.[95] Sleep also enhances the creation of semantically related false memories (i.e., "theme words" associated with lists of semantically related words[96]; but see Fenn and colleagues[97] for an alternative view). Furthermore, both accurate and illusory recollections have been associated with increased hippocampal activity after sleep, although behavioral effects were similar.[98] Despite evidence for residual semantic processing, attempts to create novel semantic associations using direct auditory stimulation during sleep have been unsuccessful.[75] One possible explanation for this is that transfer of novel information from hippocampus-dependent episodic memory stores to neocortical semantic, decontextualized memory representations is a gradual, slow process that may take years to complete.[7] Therefore protocols in which initial encoding, posttraining sleep periods, and retrieval are temporally close are probably not best suited to segregate the semantic component of memory from other constituents. One promising approach used to address this issue[99] is to investigate the integration of novel words that were part of either densely or sparsely populated networks of semantic memory. Results showed that sleep spindles and SWA (0.5 to 4 Hz) increased more after learning sparsely than densely semantically integrated words, suggesting the involvement of these NREM sleep parameters in the integration of new information into existing semantic networks.

In this respect, a few neuroimaging studies have investigated the cerebral correlates of declarative memory retrieval after extended periods of up to 6 months using paired-associates list of words[37] or pictures of landscapes.[100] Although these experiments were not specifically designed to probe whether the memorized material was "semanticized" at the time of retesting, neuroimaging results clearly indicated a transfer from activity in hippocampal locations, observed early after learning, toward activity in medial prefrontal cortical sites recorded 6 months later during memory retrieval.[37,100] Furthermore, total sleep deprivation on the postlearning night hindered this gradual process of consolidation. Indeed, 6 months after learning verbal material, memory retrieval more strongly activated the medial prefrontal cortical when initial encoding was followed by a night of sleep than by sleep deprivation, suggesting that sleep leads to long-lasting changes in the representation of memories at the neuroanatomic level.[37] These results may be consistent with the hypothesis that sleep

exerts an effect on the gradual semantic integration of the learned material.

Episodic Memory

The effect of sleep on episodic memory has been extensively studied using a series of declarative memory paradigms encompassing learning of verbal material such as lists of words or paired-associates words[14,15,37,48,49,55,56,86,87,101-107,108-109] and sentences or prose passages,[110] but also explicit encoding of landscapes,[100] objects locations[88] and faces,[111] or visuospatial memory[15,54] and navigation within virtual[36,38,65,112] or natural[113] environments.

Among these, most studies using partial behavioral[14,15,37,103] or pharmacologic[91] sleep deprivation have consistently found that SWS, or at least the first half of the night of sleep, which is rich in SWS, is beneficial for the consolidation of novel declarative memories, particularly when the to-be-remembered material is of future importance.[114] Moreover, the benefit of sleep to episodic recall is greater for young than for older subjects and correlated with sleep duration,[115] and long-term forgetting is correlated with sleep disruptions in this population.[116] It suggests that age-related changes in sleep contribute to deterioration of episodic memory with age. Additionally, consolidation of declarative learning was linked to increased spindle activity during posttraining stage 2 sleep,[48,49,53-55] as well as to the alteration of SWS[117] and spindles[118] in schizophrenia. Spindles are strong candidates to subserve memory consolidation processes during NREM sleep because they are thought to support neural plasticity.[119] As well, declarative learning abilities have been linked with increased spindle activity during stage 2 sleep[50,120] and periodic arousal fluctuations during NREM sleep[121] in healthy subjects. Conversely, declarative memory deficits are associated with decreased sleep spindle activity in patients with Alzheimer disease[122] and with NREM sleep duration and number of cycles in patients with chronic nonrestorative sleep.[123] On a different note, others reported that overnight performance gains on declarative verbal memory primarily depend on preserved organization of sleep cycles (i.e., sleep continuity), rather than on the integrity of a specific sleep stage per se.[59] Also, a more recent study found that REM sleep deprivation specifically impairs recall of spatial and temporal features of memories, as well as the subject's confidence in his or her own remembering,[107] these parameters being considered as genuine components of episodic memory, as opposed to general recall, which may partially rely on semantic, decontextualized memories.[16] Accordingly, it has been reported that consolidation during sleep enhances explicit recollection in recognition memory[124] and strengthens the original temporal sequence structure for lists of triplet words.[125] Taken together, these studies provide compelling evidence that sleep is important for the formation of novel episodic experiences and the preservation, integration, and recollection of episodic memory. When sleep is adversely affected by age, by psychiatric and neurodegenerative conditions, or more simply by sleep disruption or deprivation, episodic memory is impaired.

Emotion in Episodic Memory

The role of emotional variables in sleep-dependent processes has been a focus of recent interest.[92,110,126-135] Emotion can be seen as an important contextual cue in retention of episodic memories. Notwithstanding, emotional material was better recalled after REM than NREM sleep[110] and was altered after sleep deprivation,[126] although more for the emotional content than the context of the information.[129] Others have found that emotional memories were preserved, or at least less disrupted than neutral memories, after total sleep deprivation.[92,131] These latter results suggest that consolidation of emotional memories also occurs efficiently during wake periods, an effect that may be explained by the important ecologic value of rapidly acquiring emotional stimulus-response associations. Moreover, a lack of behavioral effect does not guarantee that sleep was without any consequence on memory consolidation processes because the underlying patterns of brain activity at retrieval were effectively altered by sleep deprivation on the posttraining night.[36,38,131]

From another perspective, it has been proposed that sleep may play an instrumental role in dissociating memories from their emotional context,[132] contributing to the regulation and integration of emotional experiences. Accordingly, mood-dependent memory effects (i.e., better recall for neutral material in the same than in a different mood during learning) were attenuated when tested after 3 days of normal sleep, but not when sleep deprivation took place on the postlearning night.[133] However, mood-dependent memory effects were equally preserved after 4 hours of sleep dominated either by NREM or REM sleep,[134] or after a night of sleep,[135] suggesting that emotional decontextualization of neutral memories takes place over several nights of sleep, the process being initiated on the first posttraining night. Further studies are needed to ascertain whether this effect is specifically emotional in nature or merely related to contextual information in general.[136]

SLEEP AND NONDECLARATIVE MEMORIES

As previously mentioned, memory abilities aggregated under the nondeclarative category comprise various heterogeneous subtypes. These may be relatively independent from both cognitive and neuroanatomic standpoints but are collectively considered procedures for "how" to perform various tasks that are often divorced from explicit knowledge of the experiences where these skills were acquired. Skills and habits that refer to the gradual acquisition of novel perceptual, motor, and cognitive abilities through repeated practice (e.g., discriminating figures, playing piano, riding a bicycle, or detecting environmental regularities) have been the most widely investigated in relation to sleep. In this respect, numerous studies have found that posttraining sleep boosts acquisition levels on nonverbal motor,[32,42,51,137-144] perceptual,[61,145-154] and perceptual-motor[28,35,57,63,64,70,72,155-163] procedural learning tasks. Sleep was also shown to reinforce proactive interference effects in motor learning in children,[158] a population in whom procedural memory effects are typically not observed using direct measures.[159] Additionally, a few studies have found that pharmacologic enhancement (e.g., using nonbenzodiazepine hypnotics) of sleep spindles restores sleep-dependent motor memory consolidation in schizophrenia.[160] Also, sleep particularly enhances procedural memory performance in brain-damaged individuals,[162] but not in patients suffering from degenerative Parkinson disease.[163] Age-related changes in sleep spindles may also underlie age-related reduction in striatal activation and the related deficits in motor memory consolidation.[33] In the following section we further describe the

role of sleep on (1) perceptual, (2) motor, and (3) perceptual-motor learning and on (4) priming.

Sleep and Perceptual Learning

One of the most consistent findings in the literature is the prominent role of sleep in the development of visual discrimination abilities. Most studies have used the texture discrimination task,[164] in which learning is retinotopic, that is, specific to the trained visual quadrant.[145,149-154] It was initially found that selective REM sleep deprivation, but not SWS deprivation, abolishes overnight performance improvement during visual perceptual learning, whereas no or only feeble improvements occur over episodes of wakefulness.[164] However, another study using the same task found that improvement in visual discrimination skills was mostly disrupted by early sleep deprivation (i.e., rich in SWS) and even more so by total sleep deprivation.[145] This apparent contradiction was partially resolved with the finding that overnight improvement was a direct function of both the amount of SWS in the first fourth of the night and the amount of REM sleep in the last fourth of the night,[21] suggesting that SWS and the predominance of slow—frequency EEGs prompt memory formation, which is possibly, but not necessarily, consolidated during REM sleep. There is further evidence of the complementary role of sleep in the off-line (i.e., occurring outside of actual practice) processes of consolidation for visual perceptual learning. Indeed, others have found that repeated practice on the task within the same day does not lead to any improvement, and can even result in performance deterioration for the trained visual quadrant, unless there is an intervening sleep episode.[152] But most important, they demonstrated that the duration of the sleep episode and its constituent phases are crucial in this process because 30-minute daytime naps merely discontinued performance deterioration over repeated sessions,[152] 60-minutes naps reverted performance to its original level,[152] and 90-minutes naps yielded improvement in discrimination performance.[151] The main differences were more time spent in NREM sleep in the 60-minute versus the 30-minute nap and the occurrence of REM sleep in the 90-minute nap. Other studies confirm the importance of posttraining sleep in the consolidation of coarse visual discrimination[148] as well as the effect of visual adaptation paradigms on subsequent sleep parameters,[61] thus demonstrating the importance of both REM and NREM sleep states for consolidation of procedural abilities in the visual system.

Interestingly, visual perceptual skill performance transfer to the untrained eye also tends to generalize to the untrained visual quadrant following sleep, suggesting that sleep contributes to the generalization of perceptual skill learning beyond the primary visual cortex.[165] However, sleep-dependent improvements cannot yet be fully generalized across modalities because contradictory results have been reported in the auditory domain.[146,147,166-169] Regarding more sophisticated auditory discrimination abilities, however, it has been shown that integration of newly learned spoken word forms, which must be discriminated from similar-sounding entries during auditory word recognition, requires an incubation-like period containing sleep.[170]

Sleep and Motor Learning

The time-dependent evolution of motor learning has been mostly investigated using a simple sequential thumb-to-fingers opposition task.[171] Skilled performance on this task is acquired across an initial, within-session fast learning phase, followed by a slow phase of consolidation and optimization that can extend for several weeks of repeated practice.[171] Using the same task or a keyboard-presses variant, others subsequently found that posttraining sleep significantly enhances performance in the absence of further practice compared with the same amount of time elapsed awake.[51,138,140,141] However, contrary to the observations made using perceptual visual discrimination tasks described in the prior section, it cannot be claimed here that posttraining time spent awake prevents the formation of long-term motor memories because performance merely stabilizes at the level achieved at the end of learning and does not deteriorate, but rather modestly improves over repeated practice sessions. Still, a transitory boost in performance is observed 5 to 30 minutes after the end of learning,[172] which disappears if tested without an intermediate sleep period more than 4 hours later.[51,51,172] Interestingly, performance improvement over the 5- to 30-minute boost predicts performance levels after a night of sleep.[172] This precocious posttraining period appears important for the initial stabilization of motor memories[22] because it has been shown that learning another sequence during that time can interfere with the initial sequence,[143,173] unless a nap is allowed.[143] When a night of sleep occurs after initial training, off-line gains in performance and increased activation in the putamen are observed using fMRI 12 hours later compared with an equivalent period of wake.[32,137] Sleep spindle density, particularly for fast frequency (~13 to 15 Hz) spindles at parietal regions, is correlated with both sleep-dependent behavioral changes[174] and neural changes that occur.[175] With age, the beneficial impact of sleep on motor sequence consolidation is reduced, along with an age-related reduction in the activation of the putamen, in turn correlated with spindle density.[33] Noteworthy, a role of sleep for motor memory consolidation extends beyond a genuine motor component because training-related changes in sleep architecture are observed after practice of a sequence of finger movements but not after random key presses,[139] as well as after acquiring new and complex motor patterns such as trampolining[42,144] but not after the familiar and well-known motor activities of soccer or dancing.[144] These latter results are in line with the demonstration that sleep provides maximal benefit for motor skill procedures that proved to be most difficult during learning.[142] Motor learning-related changes in posttraining sleep parameters were mostly observed during stage 2 sleep,[51,139,140] although others have reported links with REM sleep[144,176] or both SWS and REM, resulting in a lengthening of the ultradian sleep cycle.[42] Taken together, these results suggest that the sleep may strengthen the initially labile memory trace formed during learning, possibly through a reactivation process. This hypothesis was tested using combined EEG-fMRI during sleep following acquisition of a motor sequence task.[33] Results revealed that sleep spindle activity in the same brain regions that were active during practice correlated with overnight improvement in performance. Moreover, slow and fast spindles were associated with activation in hippocampal and striatal regions, respectively, suggesting that different types of spindles may serve to consolidate different aspects of the memory trace. Indeed, hippocampus and striatum have been found to interact whereby a competitive interplay between

these structures was correlated with improved performance at retest in participants who slept after learning compared with sleep-deprived participants.[177] Further studies are needed to delineate precisely the role and benefit of sleep on various types of motor learning and the neurophysiologic mechanisms involved.

Sleep and Perceptual-Motor Learning

The motor learning tasks described previously have definite features. They are self-initiated, and acquisition is initially carried out in an explicit, almost declarative manner because the motor sequence of movements to be generated is already known and can even be verbalized. In this respect, motor procedural learning reflects the optimization of predefined motor forms possibly created with the contribution of episodic memory processes. By contrast, perceptual-motor procedural learning entails a motor performance triggered by external stimulations, but the organization of the material to be learned is not necessarily obvious to the subject, although it affects their performance. These features should allow us to distinguish the role of sleep for consolidation of covertly versus overtly formed memories. We review this aspect at the end of this section after reviewing perceptual-motor studies.

It was initially found that performance improvement on the pursuit rotor task was blocked by total sleep deprivation and by sleep deprivation of the second part of the night, but not by selective REM sleep deprivation.[28] Task improvement was also associated with increased sleep spindle activity,[57,178] together suggesting that consolidation of motor adaptive memories is mostly dependent on stage 2 sleep. Functional MRI additionally showed that posttraining total sleep deprivation hampered both performance improvement and the reorganization of brain activity on a visuomotor pursuit task with hidden regularities in the target's trajectory.[35] However, mirror-tracing skills were reported to improve more during the late part of the night[14] and to be associated with REM sleep increases in well-performing individuals,[56] suggesting that this latter task is rather REM sleep dependent. Still, performance improvement on mirror tracing was also found to occur after naps dominated by stage 2 sleep[155] and to correlate with stage 2 sleep spindle activity. This apparent discrepancy between sleep stages and task associations within the perceptual-motor learning domain might be explained by differences in initial skill levels of participants. Indeed, an association was reported between performance improvement on the pursuit of rotor task and stage 2 spindle activity in highly skilled subjects, whereas a similar relationship was established with REM density in lower skilled subjects,[57] in line with the proposal that motor skills tasks involving REM and stage 2 sleep might be dependent on two separate but overlapping neural systems.[29]

This interpretation may be consistent with neuroimaging data that have established a relationship between posttraining REM sleep activity and the consolidation of higher order perceptual-motor cognitive skills. Indeed, subcortical and neocortical areas already activated during practice on a probabilistic sequence learning task[179] (i.e., a paradigm of implicit sequence learning) were reactivated during posttraining REM sleep in subjects previously trained to the task.[63,64,70] It was further demonstrated that these reactivations were not merely activity dependent but rather occurred specifically because a rules-based sequence was implicitly learned during prior practice,[64] supporting the hypothesis that REM sleep might be deeply involved in the reprocessing and optimization of the higher order information contained in the material to be learned (as in the mirror-tracing and Tower of Hanoi tasks), as opposed to its motor component alone.

Notwithstanding, further studies using sequence-learning tasks have raised several issues, and some of these remain to be resolved. Indeed, it has been claimed that sleep is beneficial for sequence learning only when acquisition of the sequential regularities practiced during the task was explicit, with time alone being sufficient for the consolidation of implicitly acquired sequences.[24,180] These results appear to contradict the above findings of REM sleep dependency for higher order probabilistic sequence learning, in which learning was undoubtedly implicit.[64,179,181] However, one overlooked difference between these studies and those claiming a sleep dependency exclusively for explicit sequential material is that the latter have used deterministic, repeated sequences, whereas the probabilistic sequences used in the former studies are much more ambiguous. Accordingly, neural responses to nonpredictable elements in the probabilistic serial reaction time were attenuated over sleep, suggesting better integration of the learned structure.[39] It is therefore possible that sleep (and especially REM sleep) mostly supports the consolidation of implicitly acquired complex relationships. This interpretation may be in line with the proposal that different aspects of a procedural memory are processed separately during consolidation; for example, that the movement sequence (e.g., a repeated, deterministic sequence) improves over daytime wakefulness periods independently of sleep whereas its goal (e.g., the complex, abstract rules for items succession in a probabilistic sequence) improves after a night of sleep.[182] How and whether off-line processes of memory consolidation do actually benefit from posttraining sleep may also be a function of the circadian moment of the day when the material is acquired,[157] as well as the nature of the complex interactions between declarative and procedural memory systems.[23] The complexity of these interactions is further illustrated by the demonstration that learning-related cerebral responses in cerebral structures linked to procedural (i.e., striatum) and declarative (i.e., hippocampus) memory systems are both correlated to an overnight gain in performance in an implicit oculomotor sequence learning task[183] and the finger-tapping task.[177]

Sleep and Priming

Perceptual priming refers to the facilitation or bias in the processing of a stimulus as a function of a recent encounter with that stimulus.[184] The few studies that have investigated a role for sleep in consolidating the memory representations subtending priming[15,185,186] or priminglike[185,187] effects have yielded discrepant results.[186,187] Although studies have found that intervening deprivation of sleep, especially REM sleep, alters priming effects in word stem completion[15] as well as face processing[185,187] and enhances reactivity for emotional pictures,[188] another study failed to disclose sleep-dependent effects using better controlled tachistoscopic identification of drawings,[186] and total sleep deprivation was found to affect priminglike performance in the right hemisphere only, suggesting interhemispheric differences in sleep-dependent processes of memory consolidation.[187]

SLEEP-DEPENDENT MECHANISMS OF BRAIN PLASTICITY AND MEMORY CONSOLIDATION

In this section, we provide an overview of specific mechanisms viewed as particularly important to support sleep stage–related processes of brain plasticity and memory consolidation: (1) PGO waves, (2) hippocampal rhythms, and (3) sleep spindles.

Ponto-Geniculo-Occipital Waves

PGO waves are prominent phasic bioelectrical potentials, closely related to rapid eye movements, that occur in isolation or in bursts during the transition from NREM to REM sleep or during REM sleep itself. PGO waves are a fundamental process of REM sleep in animals, playing a significant role in central nervous system maturation.[189] In humans, intracerebral recordings in epileptic patients[190] and noninvasive PET,[191] fMRI,[192] and magnetoencephalography (MEG)[193] scanning in healthy volunteers indicate that the rapid eye movements observed during REM sleep are generated by mechanisms similar or identical to PGO waves in animals. Most important, animal data indicate that PGO activity during REM sleep is associated with learning and memory consolidation,[189] suggesting that activation of this generator during REM sleep may represent one of the natural physiologic processes of memory, possibly through the synchronization of fast oscillations that would convey experience-dependent information in thalamocortical and intracortical circuits.[194]

Despite the evidence from animal studies, a direct demonstration of the association between PGO activity and memory consolidation during REM sleep in humans has yet to be done. Nonetheless, the hypothesis is supported by studies showing an increase in the density of rapid eye movements during REM sleep following procedural learning[29,195] and intensive learning periods,[47] and correlations between declarative memory performance[196] or retention levels after learning a Morse code[44] and rapid eye movements during posttraining REM sleep.[44] Also, presenting sounds in the background while learning a complex logic task at wake enhanced next-morning performance when the same sounds were presented again during REM sleep. Most interestingly, however, enhancement was found only when sounds were coincident with the bursts of posttraining rapid eye movements that reflect PGO activity,[85] further suggesting its association with memory consolidation processes. Also, it has been proposed that during human phasic REM sleep, propagation of PGO activity in the parahippocampal and hippocampal areas is linked with verbal learning performance and mnemonic retention values.[197]

Hippocampal Rhythms

The theta rhythm (i.e., regular sinusoidal oscillations in the frequency range of 4 to 7 Hz recorded in the hippocampal EEG) constitutes a prominent signature of REM sleep in mammals, including humans.[198] Theta represents the "online" state of the hippocampus, believed to be critical for temporal coding and decoding of active neuronal ensembles and the modification of synaptic weights.[198] Additionally, population synchrony of pyramidal cells is maximal during quiet wakefulness and SWS associated with sharp waves (i.e., sharp waves of SWS are the consequence of synchronous discharge of bursting CA3 pyramidal neurons) and fast ripples (140 to 200 Hz). Sharp waves and ripples during SWS constitute good candidates to induce neuronal plasticity.[199] Hence, the alternation between both REM sleep and active wake theta activity and SWS and quiet wakefulness sharp waves and ripples could contribute to brain plasticity. According to the two-stage model of memory formation,[199] neocortical information activates the entorhinal input, which causes synaptic changes to occur in the hippocampal CA3 system during learning associated with active waking and REM sleep theta rhythmic activity in the hippocampus. In the subsequent non-theta state (i.e., SWS but also possibly quiet wakefulness), previously activated neurons are reactivated during sharp waves bursts, and the memory representation transiently stored in the CA3 region can be transferred to neocortical targets for the long term.

Also, animal[52] and human studies have shown increases in theta rhythms during REM sleep following learning in declarative memory tasks,[56] as well as correlations between theta during REM sleep and memory facilitation for emotional tasks.[200] Human brain imaging studies have demonstrated after spatial navigation in a virtual town the experience-dependent reactivation of hippocampal activity during SWS, but not during REM sleep.[65] Additionally, intracerebral recordings in humans suggest that slow EEG frequencies in the hippocampus during NREM sleep contribute to the consolidation of spatial memories.[201] Similarly, an odor-cued activation in hippocampus-related memory areas was observed during SWS,[88] eventually leading to overnight performance improvement. These studies also found long-term transfer of hippocampal memories toward neocortical stores,[37,100] an effect disturbed by sleep deprivation on the posttraining night.[37] In parallel, neuroimaging data have suggested the off-line persistence of memory-related cerebral activity during active wakefulness and its dynamic evolution in the hippocampus.[22] Although this model is well supported by the previous data and may account for the off-line processing of declarative material, the fact that reactivations have been observed during both posttraining REM sleep[63,64] and wakefulness[22] after procedural, nonhippocampal learning suggests that other routes for consolidation exist, which should likewise be investigated.

Sleep Spindles and Slow Waves

Traditionally, the sleep spindle has been defined as the presence of rhythmic, about 12- to 14-Hz, activity lasting a minimum of about 0.5 second and displaying an increasing, then decreasing amplitude envelope. This definition has expanded to 12 to 16 Hz and includes both slow (~11 to 13.5 Hz) and fast (~13.5 to 16 Hz) spindles, which may reflect two separate spindle generators with different brain topographies,[202,203] and modern automated detection methods are capable of reliably identifying shorter duration spindles, even during SWS. Although spindles occur most frequently in stage 2, they also appear to a lesser extent in SWS but are obscured by delta activity. Spindles are considered to be caused by the intrinsic properties and the connectivity patterns of the thalamic neurons and are regulated by slow oscillations.[204-206] Spindle generation seems an ideal mechanism for neural plasticity.[119] Therefore spindles may play an important role in the processes of memory consolidation during sleep.

Accordingly, several studies have reported links between the consolidation of verbal declarative memory and increases

in spindle density during nocturnal and diurnal sleep that follows learning.[48-50,53] Verbal memory association with spindle density increased in the left frontocentral scalp location at night, whereas memory for faces did not elicit this effect.[53] Similarly, postlearning increases are observed during daytime napping in low sigma frequency spectral power (11.25 to 13.75 Hz), particularly in the left frontal scalp region. Additionally, these increases are positively correlated with learning performance for difficult word associations, but not easy word associations,[49] and the integration of new verbal learning with existing knowledge is correlated with sleep spindles.[207] Recent studies have attempted to causally manipulate spindles to investigate memory enhancement processes. One such study observed that increasing sleep spindle density improved verbal memory.[160] Functional neuroimaging investigation found hippocampal reactivation during sleep following declarative learning, time-locked to the onset of sleep spindles, modulated by spindle amplitude and correlated with memory performance.[66] Taken together, these studies suggest that sleep spindles are important for the overnight consolidation of declarative memories and that the occurrence of sleep spindles in brain structures recruited during initial learning contributes to reactivation and consolidation processes.

Similarly for procedural memory, posttraining stage 2 sleep deprivation impaired memory for a perceptual-motor task,[28] and the amount of stage 2 sleep correlated with learning progress on the finger-tapping task.[138,140] Also, intensive training on perceptual-motor learning tasks results in marked increases in number and density of spindles during subsequent stage 2 sleep,[56] as well as an increase in average spindle duration.[56] Posttraining increases in slow sigma power (12 to 14 Hz) have been observed in frontal and occipital regions, with no changes in high-frequency sigma.[56] Napping studies using the same task provided similar results: subjects that take naps regularly showed positive correlations between postnap performance and stage 2 spindle density.[208] In this case, low-frequency spindle activity (12 to 14 Hz) was correlated with performance at frontal sites, whereas high-frequency spindle activity (14 to 16 Hz) was correlated with performance at central and parietal sites. Moreover, subjects who did not nap on a regular basis did not benefit from the nap.[208] In another study using a longer (60 to 90 minute) nap after motor performance for the finger-tapping task, it was found that subjects with the most significant increases in motor performance also had the largest increases in stage 2 sleep, with a significant correlation between spindle density and postnap performance that was confined to the learning hemisphere.[51] Finally, a marked increase in stage 2 spindle densities was observed following acquisition of the finger-tapping task, but not after a control motor task, indicating that the changes were not a result of general motor activity.[139] Neuroimaging studies also revealed that similar to declarative memory, activation time-locked to sleep spindles in brain regions (e.g., hippocampus, putamen) recruited during practice of a motor procedural task was correlated with overnight gains in performance.[67] The fact that spindles have been related to consolidation both for declarative and nondeclarative memories indicates their prominent role in sleep-dependent memory processes, although it remains to be investigated whether spindles subserve the same process across declarative and procedural memory.

Finally, spindle oscillations may not act in isolation because they are grouped and regulated by slow waves[206] that are thought important for memory consolidation and synaptic plasticity, as discussed previously. Indeed, experience-dependent regional increases in delta activity have been observed during NREM sleep,[72] suggesting the existence of local homeostatic mechanisms for memory consolidation.[209] Furthermore, coherence was found to increase after learning in the depolarizing phase of the slow oscillations below the 1-Hz frequency.[206] Conversely, transcranial direct current stimulation, which may contribute to modulation of cortical excitability, was found to improve the overnight retention of declarative memories when applied during NREM sleep at a slow rhythm whose frequency (<1 Hz) approximates slow oscillations.[86] Further, reducing SWA during NREM sleep with transcranial direct current stimulation reduces verbal declarative memory retention,[210] suggesting a causal relationship between slow waves and declarative memory. Taken together, available data suggest that slow oscillations during NREM sleep may play a facilitating role in neuronal plasticity and the ongoing transformations and consolidation of memory traces, either in concert with spindles or perhaps contributing in some unique way to memory consolidation processes. Further study is required to better dissociate the unique roles that these two interrelated aspects of non-REM sleep may have in memory consolidation.

CONCLUSIONS

Although elusive at times, studies supporting the proposal that sleep is an integral component in the off-line processes that subtend memory consolidation have now substantially flourished. Still, when compared with other domains of cognition, the field obviously remains underdeveloped. Indeed, there is an urgent need of replication and validation of studies to confirm or disprove a series of hypotheses regarding the type of memories that can benefit from sleep and in which circumstances it can occur. Furthermore, the mechanisms that support these processes remain to be fully elucidated or fully understood. Most important, it remains to be understood how sleep disorders and the numerous pathologies accompanied by sleep disturbances affect cognitive processes and especially learning and memory in humans.

CLINICAL PEARL

In the long term, patients with inadequate REM sleep may be expected to experience more difficulty learning novel cognitive procedures than same-age individuals with healthy sleep. Patients exhibiting little or impaired SWS may be impaired in declarative learning, for example, in memorizing large amounts of factual material. Those with specific stage 2 sleep disturbances (e.g., reduced spindles) may be expected to have trouble refining and performing reasonably simple motor skill tasks and exhibit declarative memory deficits. Finally, those patients with poor sleep quality or pathologies affecting all sleep stages would be expected to have deteriorated long-term memory performance pervasively on all types of memory. Because sleep is only a part of memory consolidation processes, however, these deficits may manifest in a subtle manner and be compensated to a certain extent by alternate consolidation strategies. In-depth clinical investigation of memory problems should always be performed by a trained neuropsychologist.

SUMMARY

Although the utmost importance of sleep for the quality of everyday life is empirically recognized, the functions of sleep have long remained shrouded in mystery. Beyond its putative physiologic functions, there is now growing evidence that sleep plays a prominent role in brain plasticity and memory consolidation processes. According to this proposal, memory traces formed during a learning episode are not immediately stored in their ultimate form. Rather, they initially remain in a labile, fragile state during which they can be easily disrupted or interrupted. Over time and especially during sleep, memory traces subsequently undergo a series of transformations during which they will be consolidated and fully integrated into long-term memory. In this chapter, we present experimental data that provide support for the hypothesis that sleep exerts a promoting effect on plastic processes of memory consolidation. Several sources of support for this hypothesis are described in this chapter, including studies that have assessed (1) the effects of posttraining sleep deprivation on memory consolidation and on the reorganization of the neural substrates of long-term memories; (2) the effects of learning on posttraining sleep and reexpression of behavior-specific neural patterns during posttraining sleep; and (3) the effects of within-sleep stimulation on sleep patterns and overnight memories. Despite advances that have refined the understanding of the relationships between sleep and cognitive processes, the underlying mechanisms still remain to be fully elucidated. Further steps are now required to understand how sleep disorders and pathologies accompanied by sleep disturbances affect cognitive functions, especially learning and memory consolidation in humans, eventually leading to remedial interventions.

Selected Readings

Fogel S, Smith C. The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci Biobehav Rev* 2011;**35**:1154–65.

Maquet P, Smith C, Stickgold R, editors. *Sleep and brain plasticity*. Oxford, UK: Oxford University Press; 2003.

Peigneux P. Neuroimaging studies of sleep and memory in humans. In: Abel T, Benca R, Meerlo P, editors. *Sleep, neuronal plasticity and brain function.* Springer series in computational neuroscience, vol. 12. Marlton: Springer; 2015. p. 225–43.

Peigneux P, Urbain C, Schmitz R. Sleep and the brain. In: Espie C, Morin C, editors. *Oxford handbook of sleep and sleep disorders*. New York: Oxford University Press; 2011. p. 11–37.

Rasch B, Born J. About sleep's role in memory. *Physiol Rev* 2013;**93**: 681–766.

A complete reference list can be found online at ExpertConsult.com.

Sensory and Motor Processing During Sleep and Wakefulness

Chapter 23

John H. Peever; Barry J. Sessle

Chapter Highlights

- Sensory and motor processing are reduced during sleep, and their integration is differentially affected during different sleep states. For example, respiratory reflexes are reduced during non–rapid eye movement (NREM) sleep but are virtually abolished during rapid eye movement (REM) sleep. Such observations suggest that discrete NREM and REM sleep mechanisms modulate and control sensory and motor processes in a sleep state–dependent manner.
- Changes in sensory processing during sleep affect the somatosensory pathways that transduce and relay nociceptive signals to and within the central nervous system. Sensory mechanisms are affected by sleep, and sleep is in turn affected by nociceptive processes. The interaction between pain and sleep is common in patients suffering from chronic pain such as cancer.
- Abnormal sensory and motor functions during sleep underlie common sleep disorders. For example, reduced chemical control of breathing during sleep plays a role in congenital central hypoventilation syndrome, and pathologic motor control during REM sleep underlies REM sleep behavior disorder. Abnormal motor control and sensory processes also are associated with periodic limb movement disorder, narcolepsy, and sleep apnea.

Sleep has a profound impact on the mechanisms by which the central nervous system (CNS) transduces and expresses sensory and motor commands. Transmission of sensory information to the CNS is markedly attenuated during sleep, which enables sleep continuity; in certain circumstances, however, some sensory inputs such as pain may influence sleep patterns. Sleep mechanisms not only modulate the pathways that relay sensory commands to the central nervous system but also affect the manner in which regulatory brain centers such as the thalamocortical circuits relay afferent inflow to appropriate sensory and motor control centers, including somatic motoneurons. The neurocircuits that generate sleep-wake states also affect the fundamental integration and expression of sensorimotor processes.

This chapter comprises three main topics: (1) pathways and mechanisms of sensory processes, especially those related to pain; (2) the integration of sensory and motor processes; and (3) respiratory reflexes. As appropriate, each of these various functions is considered during both sleep and wakefulness, to highlight the contributions of changes in sensory and motor processes and their integration during sleep to both normal and abnormal physiology. Also discussed are the mechanisms by which sleep influences the central processing of nociceptive information and how pain itself may affect sleep. Although all sensory and motor systems are affected by sleep, a primary focus in this context is on how sleep affects the sensorimotor processes of spinal reflexes and the respiratory responses to chemical (e.g., CO_2) and mechanical stimuli (e.g., airway pressure). An appreciation of these physiologic systems is important for the clinician to realize that sleep-related changes in sensory-motor processing and their integration have a direct impact on human health and disease and can underlie common and serious sleep-related disorders.

MODULATION OF SENSORY PROCESSES DURING SLEEP AND WAKEFULNESS

Sensory Pathways and Mechanisms

The body's peripheral tissues—skin, mucosa, muscles, and joints—are supplied by small-, medium-, and large-diameter primary afferent nerve fibers. The medium- and large-diameter afferents are myelinated and comprise A beta and A delta afferents, most of which terminate in the tissues as sense organs (receptors) detecting tactile or proprioceptive stimuli applied to the tissues. However, some of the A delta afferents, along with unmyelinated C-fiber afferents, terminate in the tissues as free nerve endings, many of which act as nociceptors—that is, they are the sense organs that respond to noxious stimulation of the peripheral tissues; other A delta or C-fiber afferent terminals function as thermoreceptors responding to warming or cooling stimuli. The activation of nociceptors may result in the elicitation of action potentials in the A delta or C-fiber afferents, and these action potentials are conducted along the afferents into the CNS and thereby provide the brain with sensory-discriminative information about the location, quality, intensity, and duration of the noxious stimulus.[1-4]

In some cases, a prolonged increase in excitability of the nociceptors may occur after tissue injury or inflammation, to the stage at which they become more responsive to noxious

stimulation or even start responding to stimuli that normally are innocuous. Thus this process of "peripheral sensitization" may contribute to the hyperalgesia (increased sensitivity to a stimulus that is normally painful) and allodynia (pain resulting from a stimulus that typically does not provoke pain) that occur in certain pain conditions. Numerous chemical mediators are involved in peripheral sensitization of nociceptive endings, as well as in their activation by noxious stimuli, including several that also have actions in the CNS, such as glutamate, serotonin (5-hydroxytryptamine [5-HT]), and norepinephrine.[5-7]

The primary afferents in spinal nerves supplying skin, viscera, and musculoskeletal tissues of the limbs, trunk, and neck project to the spinal cord and synapse on second-order neurons that are predominantly located in the spinal dorsal horn and the dorsal column nuclei.[3,4] The afferents innervating orofacial tissues project mainly to the trigeminal brainstem sensory nuclear complex (V-BSNC). This structure is subdivided into the main or principal sensory nucleus and the spinal tract nucleus, which from rostral to caudal comprises three subnuclei: oralis, interpolaris, and caudalis.[2,8,9] The subnucleus caudalis has many anatomic and physiologic features that are analogous to those of the spinal dorsal horn.[8] Another brainstem region receiving primary afferent inputs is the solitary tract nucleus (NTS); it contains second-order neurons on which respiratory and alimentary tract afferents, taste afferents, and cardiovascular and chemoreceptor and baroreceptor afferents synapse, and it plays an important role in autonomic (e.g., respiratory, cardiovascular) functions, taste, and reflexes evoked from the respiratory and alimentary tracts (e.g., cough, swallow).

The spinal primary afferents conducting actions potentials evoked by tactile or proprioceptive stimuli terminate principally in the spinal dorsal horn and dorsal column nuclei, where they may activate second-order nonnociceptive neurons (e.g., low-threshold mechanoreceptive [LTM] neurons). Analogous trigeminal (V) afferents terminate at all levels of the V-BSNC. The primary afferents carrying thermoreceptive or nociceptive information predominantly terminate in the spinal dorsal horn (Figure 23-1) and subnucleus caudalis (Figure 23-2). The main excitatory neurotransmitters or neuromodulators released from the central terminals of the nociceptive afferents are glutamate and substance P.[4,8] In addition to LTM and thermoreceptive neurons, there are nociceptive neurons that are of two main types: nociceptive-specific (NS) neurons, which are excited only by noxious stimuli (e.g., pinch, heat) applied to a localized receptive field in peripheral tissues (e.g., skin), and wide-dynamic-range (WDR) neurons, which are excited by nonnoxious (e.g., tactile) stimuli as well as by noxious stimuli.

Nociceptive afferent inputs relayed toward many neurons in the spinal dorsal horn and subnucleus caudalis appear to derive exclusively from superficial (e.g., cutaneous) tissues and endow these neurons with coding properties, suggesting that they are critical neural elements for the detection and discrimination of superficial pain. However, nociceptive information from deep tissues (e.g., muscle, joint, viscera) is processed predominantly by subsets of these cutaneous nociceptive spinal dorsal horn or caudalis neurons that receive extensive convergent afferent inputs from these tissues. These convergence patterns appear to underlie the CNS mechanisms contributing to deep pain and also may explain the poor

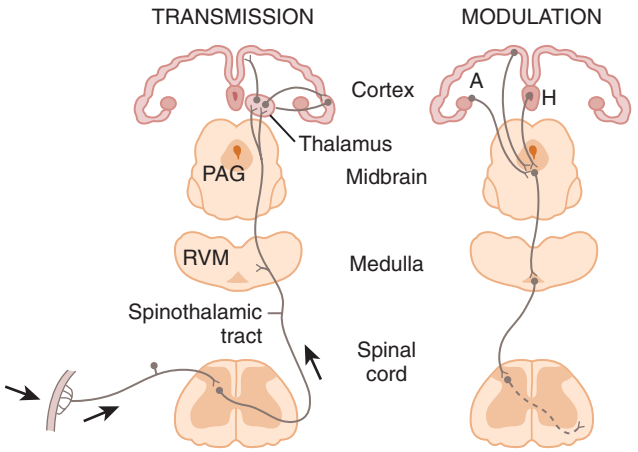


Figure 23-1 Major nociceptive pathways from spinal tissues. This schematic shows the ascending and descending pathways that transmit and modulate nociceptive signals from spinally innervated tissues. Nociceptive afferents synapse onto spinothalamic neurons in the spinal dorsal horn, which in turn relay afferent signals to the thalamus and then to the cortex (e.g., anterior and posterior cingulate cortex). Ascending spinothalamic neuronal inputs also are modulated by descending pathways located in higher brain centers such as the cortex, amygdala (A), and hypothalamus (H). Such descending pathways converge in the periaqueductal gray (PAG) and then in the rostral ventral medulla (RVM) before synapsing onto spinothalamic neurons. (From Price DD. *Psychological mechanisms of pain and analgesia.* Seattle: IASP Press; 1999, p. 250–71.)

localization, spread, and referral of pain that are typical of pain conditions involving deep tissues. Some of the convergent afferent inputs also may be "unmasked" in pathophysiologic situations and become more effective in exciting the nociceptive neurons.

Neuroplastic changes can be manifested in caudalis and spinal dorsal horn (and thalamocortical) nociceptive neurons as a result of nociceptive afferent inputs evoked by injury or inflammation.[4,8] This neuroplasticity results from the release from the nociceptive afferent terminals of neurochemicals, such as glutamate, that act by way of *N*-methyl-D-aspartate receptor mechanisms to induce a cascade of intracellular events in the nociceptive neurons. These events can result in an increase in neuronal excitability, reflecting what has been termed a "central sensitization" of the nociceptive neurons. Central sensitization has been documented in both acute and chronic pain models and involves glutamatergic mechanisms as well as nonneuronal (i.e., glial cells) processes. The neuroplastic alterations in the nociceptive neuronal properties, and in the reflex neuromuscular changes and other behavioral activities that may be induced by injury or inflammation, represent mechanisms that along with peripheral sensitization (discussed previously) can explain the allodynia and hyperalgesia as well as pain spread and referral that characterize several pain conditions.

Some neurons in the spinal dorsal horn, dorsal column nuclei, NTS, and the V-BSNC give rise to axons that ramify within that structure and serve to modulate the activity of other neurons. Nevertheless, many neurons in these various structures project to other spinal cord or brainstem regions including the reticular formation, raphe nuclei, and spinal ventral horn or cranial nerve motor nuclei. Such connections provide the central circuitry underlying autonomic and muscle

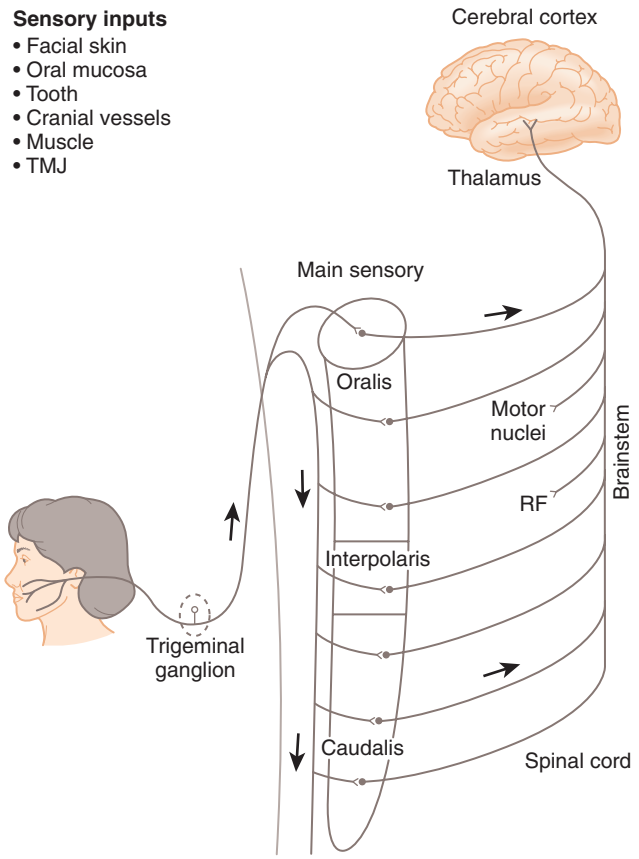


Figure 23-2 Major somatosensory pathway from the face and mouth. Trigeminal primary afferents project by way of the trigeminal ganglion to second-order neurons in the trigeminal brainstem sensory nuclear complex. These neurons may project to neurons in higher levels of the brain (e.g., thalamus) or in brainstem regions such as cranial motor pools or the reticular formation (RF). *Not shown* are the projections of some cervical nerves and cranial nerve VII, X, and XII afferents to the trigeminal complex and the projection of many VII, IX, and X afferents to the solitary tract nucleus. TMJ, Temporomandibular joint. (From Sessle BJ. Acute and chronic craniofacial pain: brainstem mechanisms of nociceptive transmission and neuroplasticity, and their clinical correlates. *Crit Rev Oral Biol Med* 2000;11:57–91.)

reflex responses to peripheral stimuli or provide sensory inputs to neurons in some of these structures that contribute to descending modulatory systems (discussed hereafter). In view of the role that the raphe and reticular formation areas play in wakefulness and sleep, projections to these areas also may provide neural substrates by which noxious stimulation may influence sleep and consciousness.[10]

Many neurons in the spinal dorsal horn, dorsal column nuclei, NTS, and V-BSNC also (or instead) project to the contralateral thalamus (see Figures 23-1 and 23-2), either directly or along multisynaptic paths. The main regions of the thalamus that receive and relay this sensory information are the ventrobasal complex (or ventroposterior nucleus in the primate) as well as the medial thalamus and the posterior nuclear group.[3,11,12] These thalamic regions contain LTM and thermoreceptive neurons that project to the overlying somatosensory cerebral cortex, where their relayed signals are processed by analogous neurons to provide for the detection and localization of tactile and nonnoxious thermal stimuli. In addition, NS and WDR neurons are present in these thalamic regions, and most have spatiotemporal coding properties and connections to NS and WDR neurons in the overlying somatosensory cortex that indicate a role for them in defining the spatiotemporal features of peripheral stimuli and hence in the sensory-discriminative dimension of pain. By contrast, the spatiotemporal coding properties of most nociceptive neurons in the medial thalamic nuclei and posterior nuclear group and their connections to areas such as the anterior cingulate cortex are more suggestive of a role in the motivational or affective dimensions of pain. These features generally are consistent with brain imaging findings in humans that noxious stimuli activate several cortical regions, including the somatosensory cortex and the anterior cingulate cortex.[13,14] It also is noteworthy that neurons at brain higher levels (e.g., somatosensory thalamus, sensorimotor cortex) also are subject to neuroplastic changes reflecting central sensitization.

Modulation of Sensory Processes Including Those Related to Pain

In addition to its role in transferring to the cortex neural signals used for sensory, affective, and motivational functions, the thalamus also is the site of modulatory influences involved in the state of consciousness. The thalamocortical transfer of sensory information is subject to modulation or "gating" as a result of facilitatory and inhibitory processes exerted by local neural circuits or inputs to the thalamus and cortex from other CNS regions such as the reticular formation.[15-17] These gating processes are especially apparent during changes in behavioral state and consciousness, such as during non–rapid eye movement (NREM) sleep. For example, there is marked attenuation during NREM sleep, which may function to maintain sleep continuity.

Nonetheless, modification of somatosensory transmission also can occur at spinal and brainstem levels and, in addition, may be operational in various degrees during different behavioral states and indeed may contribute to specific features of these states—for example, the hypotonia of rapid eye movement (REM) sleep. These modulatory mechanisms may involve neural circuits within the spinal dorsal horn, V-BSNC, and NTS and adjacent regions, as well as inputs from primary afferents and descending inputs from reticular formation, raphe nuclei, locus coeruleus (LC), and cerebral cortex, to name a few. Various neurochemicals are used by these circuits and inputs, including gamma-aminobutyric acid (GABA), norepinephrine, 5-HT, and opioids.

In the case of nociceptive transmission, the variety of inputs and interconnections in spinal dorsal horn and subnucleuscaudalis noted earlier provide the basis for considerable interaction between the various afferent inputs derived from peripheral tissues (e.g., so-called segmental or afferent inhibition) or from intrinsic brain regions (e.g., descending inhibition). Examples are the interneuronal system within the substantiagelatinosa of the spinal dorsal horn and subnucleuscaudalis, and the descending inputs to these structures from the periaqueductal gray/raphe nuclei, reticular formation, cerebral cortex, and several other brain centers (see Figure 23-1). Several neurochemicals, including GABA, 5-HT, norepinephrine, dopamine, hypocretin, cholecystokinin, prolactin, melatonin, and opioids (e.g., enkephalins), provide a neurochemical substrate by which many of the afferent and descending inputs can exert their modulatory

actions on nociceptive transmission. Of note, many of these neurochemicals and descending influences also are involved in sleep mechanisms.[15,16] Inhibitory influences exerted by many of these inputs on nociceptive neurons have been implicated as intrinsic mechanisms contributing to the analgesic effects of several procedures used to control pain, including deep brain stimulation, acupuncture, and opiate-related (e.g., morphine) and 5-HT agonist (e.g., amitriptyline) drugs. Some, instead, have a role in facilitating nociceptive transmission (e.g., in the central sensitization process noted earlier).[4,8]

Processing Related to Pain during Sleep and Wakefulness

As noted previously, sensory transmission through the spinal cord and brainstem, as well as thalamus and cerebral cortex, may be modulated during sleep and thereby may contribute, for example, to the decreased responsiveness to external stimuli during NREM sleep to favor sleep continuity. The modulatory influences on nociceptive neurons of behavioral factors, including state of alertness, attention, and distraction, are just some examples in which the higher brain centers involved in these states give rise to descending influences operating at these levels and thereby contribute to the influence of these behavioral factors on pain. Unfortunately, although much has been learned independently about pain and sleep, investigation of how sleep affects pain, and vice versa, has been limited (see Chapter 133).

Available evidence indicates that sensory processing, including that related to pain, is reduced in sleep states compared with wakefulness, but the underlying processes are still unclear. Nonetheless, with the advent of studies in chronic (awake and sleeping) animal preparations as well as more conventional anesthetized preparations, as well as some correlated studies in humans, it has been shown that somatosensory transmission through several ascending pathways originating in the spinal dorsal horn is tonically diminished during REM sleep. Apparently, however, the gating process is quite complex during REM sleep, because modulation of somatosensory transmission may be different between phasic and tonic REM sleep.[18] Other evidence of complex gating during NREM sleep includes findings that sensory transmission through some ascending tracts may be attenuated in some stages of NREM sleep and that this is reflected in reduced thalamic and cortical activity in these stages.[16-20] Recent studies also suggest that the modulation of spinal sensory transmission during REM sleep appears to involve presynaptic inhibition as well as postsynaptic inhibitory mechanisms, supporting earlier findings that presynaptic regulatory processes are important in REM sleep and contribute to the reduction in motor activity during this sleep phase (discussed further on). In addition, the presynaptic inhibitory processes in the spinal cord utilize GABA, and the postsynaptic inhibition utilizes glycine. Several other neurochemicals released from neurons in higher brain centers also have been implicated in the sleep-dependent modulation of spinal cord neurons or their thalamic targets; these include 5-HT, norepinephrine, acetylcholine, and hypocretin.[16,18-20]

A complex gating mechanism also has been documented in the rostral components of the V-BSNC that is state-dependent; for example, the activity of many rostral V-BSNC neurons and their responses to sensory inputs, including those evoked by noxious (tooth pulp) stimuli, may be presynaptically and postsynaptically inhibited during REM sleep through, respectively, GABA and glycine-based modulatory processes.[16,18-20] Comparable investigations, however, have not been made in more caudal V-BSNC nociceptive (e.g., caudalis) neurons, which as noted previously have many features analogous to those of spinal dorsal horn nociceptive neurons and play crucial roles in trigeminal nociceptive transmission; thus, if and how sleep stages affect their properties constitute an important area of future study. Of note, the state-dependent processes on spinal and V-BSNC nociceptive transmission that have been revealed have been studied only in acute sleep models, and it is unclear what neural modulatory processes are involved in the more clinically challenging sleep-pain interactions that may occur in chronic pain conditions (discussed later). Furthermore, if nociceptive afferents inputs do reach cortical levels and result in the conscious feeling of pain, then sleep may be disrupted to alert the subject to the noxious event.

Clinical Correlates

When healthy subjects are sleeping in conditions favoring good sleep quality (e.g., a quiet, comfortable environment), low-intensity stimuli may have little or no effect on sleep quality, whereas a loud noise or a sudden pain attack during sleep can produce awakening that may potentially trigger anxiety or concern that impedes subsequent sleep.[21] Nonetheless, several other factors in addition to pain can influence sleep quality (e.g., past experiences, psychological variables, anxiety, mood, life style, health status, and any concomitant pain-sleep interactions).

In the case of so-called pain patients, they may experience long delays in falling asleep, as well as sleep-stage shifts, frequent sleep arousals, and undesirable body movements; furthermore, some analgesic drugs (e.g., opioids) may affect sleep patterns (see Chapter 24). Although the majority of patients with chronic pain report that pain occurred before or at the onset of poor sleep, suggesting that pain may have a direct effect on sleep quality, studies using experimental noxious stimuli have revealed only minor sleep disturbance in healthy subjects.[22,23] However, these studies have elicited acute pain, and it remains unclear the mechanisms by which chronic pain may negatively influence sleep quality. Nonetheless, as noted earlier, chronic pain can be associated with neuroplastic changes in spinal dorsal horn, V-BSNC, and thalamocortical relays, so these or other pain-related changes may potentially have effects on sleep; however, how these changes influence the brainstem and thalamocortical circuits involved in sleep is largely unknown.[16]

The reverse situation, that poor sleep might be a significant cause of pain, also is unclear. According to some reports, experimental sleep deprivation or fragmentation can lead to pain, and restoration of sleep quality (continuity and duration) can reduce this pain in humans. Likewise, in animals, sleep limitation leads to a reduction in nociceptive threshold during wakefulness that is reversed once sleep is restored. Some studies, however, suggest that other factors (e.g., fatigue, mood changes, cognitive impairment) may be involved,[24] thereby raising doubts about whether poor sleep is indeed a dominant or a pathognomonic cause of pain. Accordingly, further investigations into the relationship between chronic pain and poor sleep quality are needed.

MODULATION OF SENSORIMOTOR PROCESSES DURING SLEEP AND WAKEFULNESS

Sleep mechanisms markedly affect sensorimotor functions. Sleep not only suppresses basal muscle tone but also attenuates and, in some cases, even abolishes motor reflexes (see later section on Specific Reflexes). Sensory and motor processes and their integration are also differentially affected by prevailing behavioral states. For example, during wakefulness, chemical and mechanical stimulation of laryngeal tissues evokes coughing; during sleep, however, the same stimulus elicits only a brief expiration. Such observations suggest not only that sensorimotor processes are suppressed by sleep but also that sleep mechanisms themselves control the integration and expression of sensorimotor processes.

Sensorimotor Pathways and Mechanisms

The afferent inputs that provide access to the spinal cord, brainstem, thalamus, and cerebral cortex are involved not only in perceptual processes but also in sensorimotor integration and control.[1,9,25,26] The neuromuscular system is reflexively influenced by receptors that signal pain, touch, joint position, and muscle stretch or tension. In the case of muscles of the trunk, neck, and limbs, afferents from spinally innervated tissues enter the dorsal root and project into the spinal cord, where they can excite or inhibit motoneurons that innervate skeletal muscles. Motoneurons that innervate craniofacial muscles (e.g., jaw, soft palate, laryngeal, tongue) reside in motoneuron pools in the brainstem, most notably the trigeminal, facial, ambiguus, and hypoglossal nuclei.

One source of afferent input that reflexively influences motoneurons is derived from the muscle spindle, a stretch-sensitive receptor signaling muscle length. The group Ia primary afferents innervating muscle spindles monosynaptically excite motoneurons to produce contraction of the stretched muscle. This monosynaptic circuitry is the neural substrate for the H-reflex and jaw-closing reflex. Muscle spindle afferents are a fundamental motor control mechanism in limb, neck, and trunk muscles; however, because there are insignificant numbers of muscle afferents in most craniofacial muscles (e.g., anterior digastric, facial, pharyngolaryngeal muscles), other receptor systems, such as the receptors in tooth-supporting tissues and in the pharyngeal and laryngeal mucosa, play a significant role in the control of these muscles.[1,9,25]

Some afferent inputs trigger either excitatory or inhibitory reflexes by affecting the activity of interneurons, which are located in the spinal dorsal horn, V-BSNC, or NTS, or within motor pools themselves. Both interneurons and motoneurons also are modulated by the descending neural systems that regulate somatosensory transmission such as the reticular formation, limbic system, lateral hypothalamus, basal ganglia, LC, red nucleus, cerebellum, and sensorimotor cerebral cortex.[25,26] Some of these systems also are involved in the initiation and guidance of movements such as the sensorimotor cortex, which contributes to the learning of novel sensorimotor skills through neuroplastic changes in sensory and motor representations in the sensorimotor cortex. Others are involved in the control and guidance of movements and their integration with other sensorimotor functions (e.g., basal ganglia; spinal or brainstem pattern generators for locomotion, chewing, or swallowing), and yet others in the regulation of sleep-wake states themselves (e.g., periaqueductal gray, lateral pontine tegmentum, lateral hypothalamus).[27]

Processing of Somatic Reflexes during Sleep and Wakefulness

Sleep influences the processing of sensory-motor responses such as the withdrawal reflex, jaw-closing reflex, jaw-opening reflex, and other H-reflexes. Because changes in somatic reflex responses are linked to movement disorders, such as PLMD and restless legs syndrome (RLS),[28] it is important to consider how such reflexes normally are affected by sleep and to identify mechanisms by which these reflexes are modulated during sleep, as reviewed next. Also discussed is how H-reflexes are affected during periods of cataplexy, because such changes provide valuable insight into understanding how brain function is affected by narcolepsy.

The withdrawal reflex (also known as the flexion reflex) is a polysynaptic multisegmental spinal reflex that triggers withdrawal (flexion) of the stimulated limb. It functions to remove the affected limb from potentially noxious stimuli that could cause tissue damage. The withdrawal reflex is composed of two excitatory responses. The first component, termed RII, is considered a tactile response and is characterized by muscle activation (e.g., biceps femoris muscle) that occurs 40 to 60 ms after nerve stimulation (e.g., sural nerve). The second component, termed RIII, occurs 85 to 120 ms after stimulation and is considered a nociceptive polysynaptic reflex.[29] The first and second components probably represent, respectively, the activation of A-beta and A-delta cutaneous afferent fibers. Because a robust correlation has been found between the RIII threshold and subjective pain thresholds, the primary focus here is on the effects of sleep on this component of the reflex.

Sleep has a marked impact on the expression of the withdrawal reflex in humans. Compared with wakefulness, a significant increase in the stimulus intensity is required to activate the reflex during both NREM and REM sleep.[30] The polysynaptic jaw-opening reflex (a reflex analogous to the withdrawal reflex) also is suppressed during NREM sleep in monkeys; reduced trigeminal motoneuron excitability is one factor that may cause suppression of this reflex during NREM sleep.[10,31] Although stimulus thresholds are elevated during sleep, the withdrawal reflex, like other somatic reflexes, is most reliably activated and most stable during NREM sleep. An increase in the latency to the RIII component also is observed during both NREM and REM sleep,[30] suggesting that sleep reduces nociceptive reflex excitability and acts to filter sensory inputs in order to preserve sleep continuity. However, there is an increase in both the magnitude and duration of the RIII reflex during REM sleep. This second finding is paradoxical, because monosynaptic reflexes are maximally suppressed during this state,[32] and somatic motoneurons are hyperpolarized during REM sleep in cats.[33] The mechanism(s) underlying RIII reflex facilitation during REM sleep is unknown.

Unlike the RIII nociceptive component, the RII tactile reflex is absent during sleep.[30] This observation is physiologically important because it suggests that sleep differentially affects the individual components of the sensory pathways that mediate this two-part reflex, which implies that sleep mechanisms do not affect all physiologic systems equally. This concept is of clinical relevance because it suggests that sleep suppresses tactile responses (i.e., RII), which if activated by body repositioning, for example, could cause unwanted arousal,

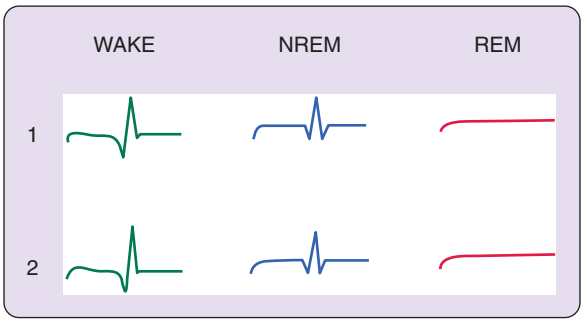


Figure 23-3 Magnitude of the H-reflex during sleep and wakefulness. The H-reflex of the calf muscle was triggered by electrical stimulation of the tibial nerve in a 22-year-old man during wakefulness, NREM sleep, and REM sleep, with reflex responses obtained on two separate measurements (1 and 2). Relative to that in wakefulness (WAKE), the H-reflex is depressed in NREM sleep and abolished in REM sleep. (From Shimizu A, Yamada, Y Yamamoto J, et al. Pathways of descending influence on H reflex during sleep. *Electroencephalogr Clin Neurophysiol* 1966;20:337–47.)

whereas nociceptive reflexes (i.e., RIII) and their motor component remain active during sleep. Preservation of this protective reflex would therefore ensure that painful stimuli elicit appropriate motor responses and if necessary arousal from sleep.

Sleep also has powerful effects on the expression of monosynaptic reflexes such as the masseteric jaw-closing reflex and the classic H-reflex. In humans, a progressive reduction in the magnitude of the H-reflex amplitude is seen during stages 1 to 4 of NREM sleep, along with a near-complete loss of the reflex during REM sleep[32] (Figure 23-3). Animal studies also show that the stimulus intensity needed to trigger the masseteric reflex is increased in sleep and that reflex responses are reduced during NREM sleep, minimal during tonic REM sleep, and maximally suppressed during periods of active REM sleep (i.e., during rapid eye movements). Because trigeminal motoneurons, which ultimately mediate the masseteric reflex, are inhibited by GABA and glycine during both NREM and REM sleep,[31] such inhibitory mechanisms may act to shunt the excitatory afferent inputs onto motoneurons during sleep. The available evidence, however, suggests that presynaptic inhibition of Ia afferents onto spinal motoneurons also may function to suppress monosynaptic spinal reflexes during REM sleep.[34] Therefore descending inhibitory inputs acting on both motoneurons and Ia afferents contribute to reduce monosynaptic reflexes during sleep.

Clinical Correlates

Sleep Bruxism

Sleep bruxism, classified as a sleep-related movement disorder, is characterized by tooth grinding and jaw clenching that can lead to destruction of teeth or dental restorations and cause jaw and head pain (see Chapters 144 and 145).[35,36] It occurs in 8% to 10% of the adult population and is characterized by rhythmic episodes of involuntary jaw muscle activity occurring primarily during NREM sleep and less frequently in REM sleep. Sleep bruxism is characterized by phasic contractions of jaw muscles, with three or more muscle bursts, although it also can manifest as tonic jaw muscle activity lasting 2 seconds or more or a mixture of the phasic and tonic activities. Because rhythmic jaw activity also occurs in most healthy subjects during sleep, sleep bruxism may represent an extreme manifestation of this normal pattern. Although the etiology and pathophysiology of sleep bruxism are not fully understood, abnormal sensorimotor processing within the CNS level, rather than an abnormal processing of peripheral sensory feedback (e.g., from the teeth), is recognized to underlie the genesis of sleep bruxism.[10]

Brainstem structures (e.g., reticular formation, LC, raphe nuclei) that influence sleep-wake control and sensorimotor processes also influence the mechanisms underlying sleep bruxism, and involve many of the same neurochemicals (e.g., 5-HT, GABA, acetylcholine, dopamine). In addition, the rhythmic movements appear secondary to CNS events related to "arousals," with increases in autonomic (cardiac and respiratory) and brain activity reflecting reactivation of the reticular formation arousal system preceding the movements.[35] The activity of noradrenergic cells in the LC also may contribute to the pathogenesis of sleep bruxism (see Chapter 144).

Descending cortical pathways, producing rhythmic masticatory movements during wakefulness, also may be involved in bruxism. Recent findings comparing cortically induced jaw movements in the awake and sleep states of awake versus sleeping primates indicate that cortical influences are suppressed during NREM sleep.[10] These findings suggest that corticobulbar excitatory influences are deactivated during sleep to preserve sleep continuity, providing indirect support for the importance of brainstem structures in the genesis of sleep bruxism.

Periodic Limb Movement Disorder and Restless Legs Syndrome

RLS is a common neurologic disorder that results from pathologic sensory and motor processing (see Chapter 95). It is characterized by unpleasant, uncomfortable, or painful sensations in the limbs, most often the legs, that lessen or disappear following limb movement.[36] Symptoms of RLS worsen during the night and can cause sleep disturbances that underlie the excessive daytime sleepiness and somnolence that typify RLS. Most affected patients also experience periodic limb movements (PLMs), which occur mainly during NREM sleep. PLMD is a sleep disorder characterized by the occurrence of brief (0.5 to 5 seconds), repetitive limb movements, typically involving the legs, every 5 to 90 seconds, arising periodically throughout the night. Common physiologic mechanisms probably underlie both RLS and PLMD, because 80% of patients with RLS also experience PLMs, and both RLS and PLMD are treated with dopaminergic drugs.[28]

Abnormal sensorimotor processing contributes to both RLS and PLMD. For example, marked hyperexcitability of both monosynaptic and polysynaptic reflexes is a feature of these two disorders.[28] The threshold required to activate the withdrawal reflex is lower and the response magnitude is larger during sleep in patients with RLS and PLMD than in healthy control subjects.[28] Changes in the response attributes of the soleus H-reflex in patients with RLS and PLMD also are seen—specifically, an increased late facilitation and a decreased late inhibition of the reflex response.[37] Reduced inhibition of either spinal motoneurons or presynaptic mechanisms may underlie the facilitation of motor responses and could explain why patients with RLS and PLMD exhibit hyperexcitability of spinal reflexes. Brain imaging studies show changes in thalamic, cerebellar, and pontine activity in both patient populations,[38] suggesting that these regions may

affect spinal reflex activity and hence contribute to RLS and PLMD.

Narcolepsy

Cataplexy is characterized by a sudden loss of skeletal muscle tone (i.e., atonia) despite maintenance of consciousness, and it is a pathognomonic symptom of narcolepsy. It generally is triggered by strong positive emotions (see Chapters 89 and 90). Because motor atonia is a defining feature of both cataplexy and REM sleep, similar neurocircuits may mediate both motor phenomena. In human and canine narcolepsy, the atonia of cataplexy can last from several seconds to several minutes. In both dogs and humans, the monosynaptic H-reflex is either absent or minimal during cataplectic attacks.[39] This observation demonstrates that the "wakefulness drive," which has been proposed to cause loss of muscle tone in sleep,[40] does not mediate motor suppression during cataplexy because wakefulness is preserved during this state. Loss of excitatory noradrenergic drives onto motoneurons may mediate the loss of muscle tone and the H-reflex during cataplexy, because noradrenergic cells in the LC project to motoneurons, and they cease firing during cataplectic attacks in narcoleptic dogs[41] (Figure 23-4). Furthermore, because LC neurons reduce their discharge activity in NREM sleep and virtually stop firing during REM sleep,[41] loss of noradrenergic drives onto motoneurons also may explain why motor reflexes and muscle tone are reduced in NREM and REM sleep.[42]

RESPIRATORY REFLEXES DURING SLEEP AND WAKEFULNESS

Sleep not only affects the processing of spinal and craniofacial sensorimotor activities but also changes how the respiratory system responds to both mechanical and chemical stimuli (see Chapters 15 to 17). The focus of this section is on sensory-motor integration in upper airway and craniofacial muscles, because sleep-dependent changes in these reflexes can cause respiratory instability during sleep. Airway reflexes, including coughing, swallowing, laryngeal closure, apnea, and the negative-pressure reflex, function to protect the airway from inhalation of inappropriate substances and to preserve airway patency during vulnerable periods such as during anesthesia and sleep. A multitude of afferent nerve endings in the upper alimentary and airway mucosa and muscles that detect changes in muscle tone, pressure, airflow, temperature, and chemical status (e.g., acidic fluids) are able to respond to and thus trigger appropriate respiratory reflexes. Presented next is a summary of how two important airway reflexes are affected by sleep; the potential mechanisms underlying state-dependent regulation of such reflexes are then reviewed.

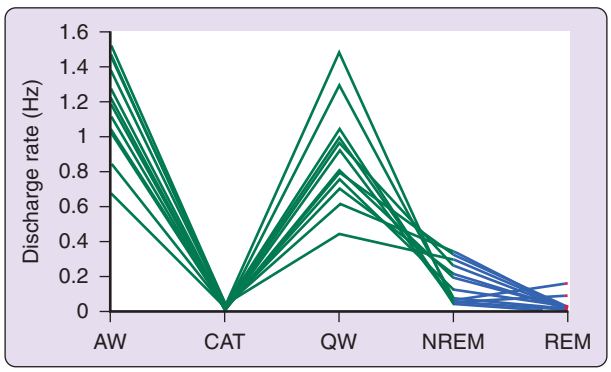


Figure 23-4 Discharge activity of putative noradrenergic cells in the locus coeruleus of narcoleptic dogs during wakefulness, sleep, and cataplexy. The discharge rate for locus coeruleus neurons is tightly correlated with behavioral state. Cell activity is maximal during active and quiet wakefulness (AW and QW, respectively), reduced during NREM sleep, and minimal or absent when motor hypotonia/atonia is present, that is, during REM sleep and cataplexy (CAT). (From Wu MF, Gulyani SA, Yau E, et al. Locus coeruleus neurons: cessation of activity during cataplexy. *Neuroscience* 1999;91:1389–99.)

Specific Reflexes

Airway Negative-Pressure Reflex

One of the most clinically relevant reflexes is the airway response to negative pressure. The airway negative-pressure reflex is characterized by increased upper airway muscle tone in response to the negative suction pressures generated by diaphragmatic contraction. During normal breathing, the reflex is inactive because airway pressure is below the stimulus threshold required to trigger it; however, during sleep, particularly REM sleep, reductions in airway muscle tone cause airway narrowing and increased resistance, which increases negative pressure, and this triggers the pressure reflex. The primary function of the negative-pressure response is to increase upper airway muscle tone when airway pressures threaten to occlude the airspace. This reflex response plays an important role in obstructive sleep apnea (OSA) (see the following Clinical Correlates section and various chapters in Section 14).

The negative-pressure reflex is typified by short-latency activation of a variety of craniofacial and pharyngeal muscles (e.g., tensor palatini and genioglossus) when airway pressure drops below a variable threshold. Studies using resistive loading, which approximates the effects of sleep on airway pressure, show that the negative-pressure reflex is reduced in sleep and indeed often is absent, except in the presence of hypercapnia or hypoxia.[43] Studies in both animals and humans show that the magnitude of pharyngeal dilator activation is reduced during NREM sleep and further reduced during REM sleep[44] (Figure 23-5).

The mechanisms mediating the airway negative-pressure reflex are not fully understood (see Chapter 15). Changes in pressure probably are detected by mechanoreceptors located within both airway mucosa and airway muscles themselves; whether both mechanisms operate at all levels of the airway is unknown. Trigeminal nerve afferents play a particularly important role in detecting upper airway pressure changes because not only do they respond to pressure changes, but the pressure reflex also is diminished when regions of the airspace innervated by trigeminal afferents are locally anesthetized.[45] Reductions in the negative-pressure reflex during sleep could be mediated, at least in part, by changes in the excitability of trigeminal pathways. Neuronal activity in the rostral V-BSNC is reduced during REM sleep[18]; because some of these neurons relay their excitatory signals to the respiratory centers and motoneurons that trigger reflex expression, reductions in their activity could function to reduce reflex responses in sleep, particularly during REM sleep.

Changes in airway pressure also are detected by mechanoreceptors supplied by the superior laryngeal nerve; such afferent signals are relayed to the NTS before reaching respiratory centers. Because excitatory (e.g., serotonergic raphe neurons,

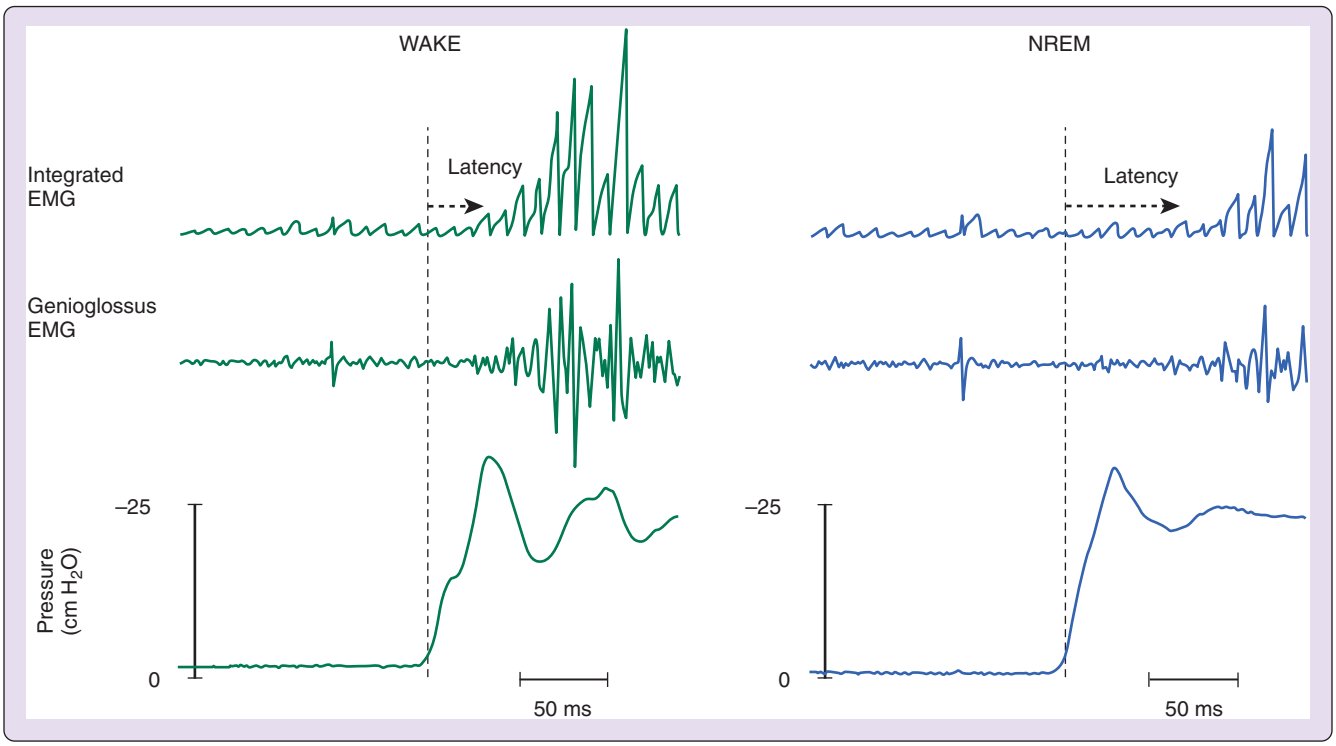


Figure 23-5 The airway negative-pressure reflex is suppressed during sleep, as seen in this typical example of an electromyogram (EMG) recorded in a human subject during wakefulness and sleep. *Top two traces,* Integrated and raw EMG activity recorded from the genioglossus muscle; *bottom trace,* changes in airway pressure. During waking, negative airway pressure (–25 cm H_2O) triggers a short latency activation of genioglossus muscle tone; during NREM sleep, however, the reflex magnitude is suppressed, and a significant increase in response latency is evident. (From Horner RL, Innes JA, Morrell MJ, et al. The effect of sleep on reflex genioglossus muscle activation by stimuli of negative airway pressure in humans. *J Physiol* 1994;476:141–51.)

hypocretin cells) and inhibitory (GABAergic in the ventrolateral preoptic nucleus) components of the sleep-generating circuitry also project to and synapse onto NTS neurons, reduced excitation and increased inhibition of NTS neurons could act to reduce the negative-pressure reflex during sleep.[27] In addition, GABAergic and glycinergic inhibition of upper airway motoneurons during NREM and REM sleep[31] may function to shunt the excitatory signals arising from pressure afferents; this effect also would limit the expression of the negative-pressure reflex during sleep.

Laryngeal and Bronchopulmonary Reflexes

Chemoreceptors in the nose, mouth, pharynx, larynx, and lower airways (e.g., trachea) detect various chemical stimuli that elicit protective upper airway reflexes. Stimulation of the larynx with acidic liquids or mechanical forces causes the laryngeal reflex, which during wakefulness in adults triggers swallowing or coughing. During NREM and REM sleep, however, the laryngeal reflex is not simply suppressed (and the stimulus threshold increased); it is transformed into a fundamentally different motor response. Compared with the waking reflex, which triggers coughing, laryngeal stimulation during sleep often initiates apnea and bradycardia[46] (Figure 23-6). Activation of bronchopulmonary receptors also elicits different respiratory motor responses during waking and sleep. Mechanical activation of the bronchopulmonary receptors that detect changes in airflow and/or lung stretch triggers the cough reflex during wakefulness but elicits reflexive apnea during both NREM and REM sleep.[47] Such an observation implies that sleep circuits inhibit the expression of these reflex responses or that activation of wake-generating circuits is required to trigger these reflexes.

The mechanisms responsible for reversing laryngeal and bronchopulmonary reflexes during sleep have not been determined. This type of response reversal, however, previously has been identified and studied in other motor pathways. During wakefulness, auditory stimuli (e.g., sudden loud noises) or skeletal muscle stimulation (e.g., muscle stretch) cause reflexive activation of somatic motoneurons with subsequent facilitation of skeletal muscle tone; during REM sleep, however, identical stimuli, albeit at greater intensities, cause the opposite effect—that is, motoneuron inhibition and reduced muscle tone. This type of state-dependent response reversal is mediated, at least in part, by the pontomescencephalic reticular formation (PMRF) that participates in sleep-wake regulation.[48] Stimulation of the PMRF causes excitatory postsynaptic potentials (i.e., excitation) in trigeminal motoneurons when the jaw-closing masseteric reflex is evoked during waking, but PMRF stimulation during REM sleep triggers inhibitory postsynaptic potentials (i.e., inhibition) in motoneurons and reduced masseteric reflex output.[48] The PMRF may therefore function to gate sensory-motor reflexes during different sleep-wake states; during wakefulness, it allows sensory stimuli to produce appropriate motor facilitation such as coughing, whereas during sleep, when motor activation could elicit inappropriate arousals, such stimuli trigger motor

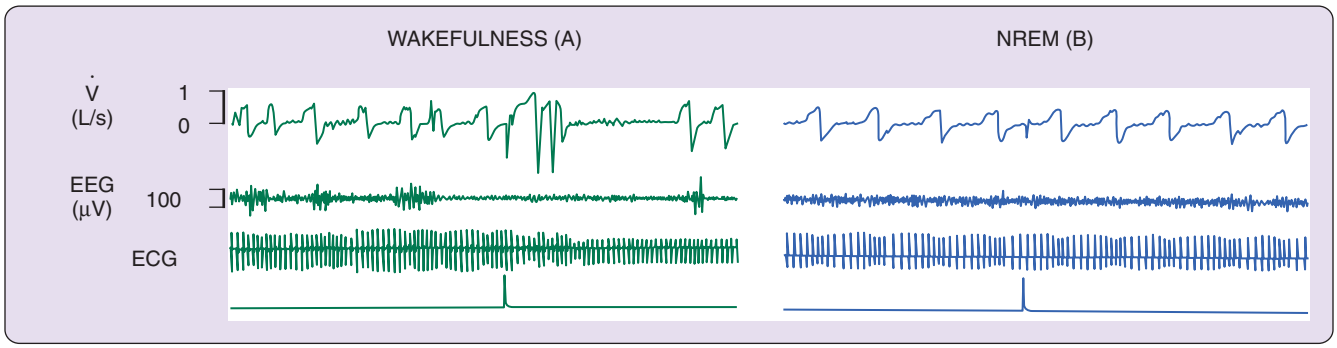


Figure 23-6 Laryngeal stimulation triggers the cough reflex during wakefulness but not during sleep, as seen in this example depicting the effects of such stimulation on respiratory activity during wakefulness **(A)** and sleep **(B)** in a dog. Airflow ($\dot{V}$) data are tracked on the *top trace,* with inspirations shown as upward spikes and expirations as downward spikes. The *upper middle trace* is the concomitant EEG; the *lower middle trace* is heart rate on the ECG).The *bottom trace* indicates when water was injected into the larynx. Injection of 0.2 mL of water into the laryngeal region of the trachea caused immediate expiration followed by the cough reflex **(A)**. Injection of the same volume of water during NREM sleep produced only a brief expiration; it did not trigger the cough reflex **(B)**. ECG, Electrocardiogram; EEG, electroencephalogram; $\dot{V}$, Ventilation. (From Sullivan CE, Murphy E, Kozar LF, et al. Waking and ventilatory responses to laryngeal stimulation in sleeping dogs. *J Appl Physiol* 1978;45:681–9.)

inhibition (e.g., apnea). Whether this type of response reversal underlies state-dependent changes in laryngeal and bronchopulmonary reflexes is unknown.

Processing of Chemoreflexes during Sleep and Wakefulness

During wakefulness, respiratory control mechanisms are acutely sensitive to changes in blood and tissue carbon dioxide (CO_2) levels, with subtle increases in CO_2 levels causing potent increases in ventilation. However, sleep has profound effects on the respiratory sensitivity to CO_2, which influences the control of breathing during sleep. During sleep, ventilation decreases even though CO_2 levels increase. The decrease in ventilation is paradoxical in that increased CO_2 levels trigger potent increases in ventilation during waking. These changes are of physiologic importance because during waking, an increase of 1 mm Hg in the partial pressure of CO_2 (P_{CO_2}) causes a 20% to 30% increase in ventilation; during sleep, however, CO_2 levels can increase by 2 to 6 mm Hg without affecting breathing. Therefore, during sleep there is a change in the relationship between ventilation and the mechanisms that detect CO_2 levels; this change reflects the fact that sleep dramatically affects the processing of respiratory chemoreflexes.

The hypercapnic ventilatory response undergoes two important changes during sleep. First, the P_{CO_2} threshold that normally triggers increased ventilation during wakefulness is significantly increased during sleep. This sleep-related change indicates that higher CO_2 levels are required to increase ventilation during sleep, which could partially explain why ventilation decreases below waking levels during NREM sleep. Second, the sensitivity of the ventilatory response to CO_2 during sleep is significantly reduced from that in awake states. For example, decreases of 25% to 75% have been reported in the hypercapnic ventilatory response during NREM sleep compared with waking.

REM sleep also has a profound impact on the ventilatory sensitivity to CO_2. Compared with NREM sleep, there is a further reduction in the hypercapnic ventilatory response during REM sleep. In fact, CO_2 sensitivity often is suppressed to such a degree that high levels of CO_2 have negligible effects on breathing during REM sleep. In dogs, even high levels of CO_2 have minimal effects on breathing, and despite potent CO_2 stimulation, the irregular breathing patterns (e.g., rapid fluctuations in tidal volume and breathing frequency) that predominate during REM sleep still persist.[49] This is a notable observation because such levels of CO_2 normally would cause breathing to become deep and regular during waking and NREM sleep. The persistence of irregular breathing during REM sleep indicates that REM phenomena override the sensory control of breathing.

CO_2 sensitivity is different during tonic (i.e., no eye movements) versus phasic (i.e., eye movements) REM sleep. Compared with either waking or NREM sleep, the ventilatory response to CO_2 is potently suppressed during periods of phasic REM sleep. By contrast, the CO_2 response during tonic REM sleep is no different than during NREM sleep; however, CO_2 sensitivity remains below waking levels during this state.[50] This important observation underscores the fact that CO_2 sensitivity is differentially affected by NREM and REM sleep and also is subject to intrastate variability. Such findings suggest that behavioral states have a marked impact on the sensory mechanisms by which the CNS detects and responds to changes in CO_2.

Sleep also affects the manner by which different respiratory muscles respond to CO_2. In both humans and animals, hypercapnia causes marked increases in both diaphragmatic and upper airway (e.g., the genioglossus muscle) muscle activity; compared with waking, however, a marked reduction in levels of muscle activity is seen during both NREM and REM sleep. Although hypercapnia increases diaphragmatic muscle tone during NREM and REM sleep, the sensitivity of genioglossus upper airway muscle tone is severely blunted during REM sleep.[51] Therefore, although hypercapnia can trigger increased diaphragm activity in REM sleep, it has comparatively minimal effects on recruiting upper airway muscles whose activation is required for keeping the airspace open and for overcoming OSA (see the following Clinical Correlates section).

Sleep-dependent suppression of respiratory muscle function could limit the expression of the hypercapnic ventilatory response. Although respiratory muscle tone is reduced in

NREM sleep, it is unlikely that this change explains reduced CO_2 sensitivity, because hypercapnic exposure during this state increases ventilation and respiratory muscle tone (e.g., diaphragm and genioglossus) to within waking levels. During REM sleep, however, upper airway muscle tone is minimal and virtually unresponsive to hypercapnia; these effects could mask the ventilatory response to CO_2. Bypassing the upper airway by means of chronic tracheotomy can prevent the stereotypical loss of the ventilatory responses to CO_2 during REM sleep in cats,[52] which suggests that loss of upper airway muscle tone is, at least in part, responsible for the suppression of CO_2 sensitivity during REM sleep. Nevertheless, it is unlikely that REM sleep muscle atonia fully accounts for suppression of the hypercapnic ventilatory response, because CO_2 responses are minimal or absent during phasic REM sleep, when skeletal muscles are transiently activated, and CO_2 responses are present during tonic REM sleep, when skeletal muscle tone is lowest.[50]

Two mechanisms by which the nervous system detects changes in CO_2 levels are recognized. First, sensory cells (glomus cells) in the carotid body chemoreceptors respond to changes in arterial P_{CO_2} (and O_2). It is unknown how this sensing mechanism contributes to sleep-related changes in CO_2 sensitivity; however, because glomus cells relay their afferent signals to the brain via the NTS, sleep mechanisms that affect the activity of NTS neurons presumably could affect CO_2 sensitivity. Second, intrinsic CO_2-sensing mechanisms also are located within the brain itself. These CO_2 sensors, termed central chemoreceptors, consist of neurons that respond to local tissue changes in CO_2, which in turn trigger ventilation. The primary areas that detect changes in CO_2 are the NTS, LC, raphe nuclei, and retrotrapezoid nucleus (RTN).[53] Hypocretin/orexin neurons in the lateral hypothalamus, which form part of the neurocircuitry underlying arousal, also respond to and detect changes in CO_2. Focal acidification (i.e., increased CO_2) of hypocretin neurons causes them to depolarize and increase their firing frequency.[54] Genetic deletion of hypocretin neurons reduces the ventilatory response to CO_2 in awake mice and promotes respiratory instability during NREM and REM sleep,[55] suggesting that hypocretin neurons act as central CO_2 sensors that function to maintain normal respiratory reflexes during both sleep and waking.

Animal research demonstrates that some CO_2 sensors are able to detect changes in CO_2 only during waking, whereas others are able to detect CO_2 changes *only* during sleep. For example, local increases in CO_2 within the retrotrapezoid nucleus activates breathing only during waking; the same stimulus has no effect on breathing during sleep[53] (Figure 23-7). Although CO_2 stimulation of the NTS triggers increased ventilation during both waking and sleep, the response is significantly larger during waking than in sleep. Therefore the RTN and NTS could serve as the primary CO_2 sensors during wakefulness; loss of their CO_2 sensitivity during sleep may explain why the hypercapnic ventilatory response is suppressed during sleep. Although CO_2 sensitivity is reduced in sleep, it is not abolished; therefore other CO_2 sensors must continue to monitor tissue CO_2 levels during sleep. The serotoninergic raphe nucleus is one potential candidate, because local increases in CO_2 levels in this brain region increase breathing only during sleep; it does not affect

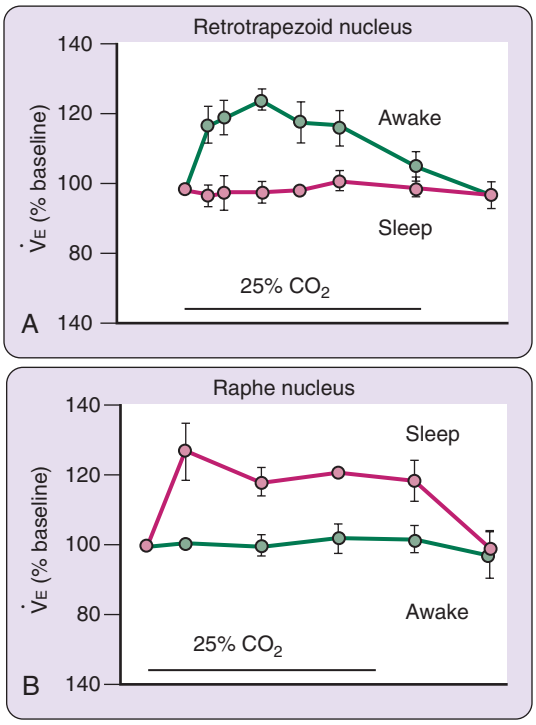


Figure 23-7 Some CO_2 sensors function only during wakefulness, whereas others function only during sleep. **A** and **B,** Plots of data for changes in $\dot{V}_E$ triggered by activation of CO_2 sensors in the retrotrapezoid nucleus RTN **(A)** or raphe nucleus **(B)** during wakefulness (*green line*) and sleep (*red line*) in behaving rats. To activate CO_2 sensors, 25% CO_2 (in saline) was perfused directly into either the RTN or raphe nuclei; this CO_2 concentration has physiologically relevant effects on CO_2 sensors. Activation of CO_2 sensors in the RTN increases ventilation only during waking; it has no effect during sleep (see **A**). Conversely, activation of CO_2 sensors in the raphe nucleus increases ventilation only during sleep; ventilation is unaffected during wakefulness (see **B**). (From Nattie EE. Central chemosensitivity, sleep, and wakefulness. *Respir Physiol* 2001;129:257–68.)

breathing during wakefulness (see Figure 23-7).[53] The raphe nucleus may therefore be the primary CO_2 sensor during sleep. This concept is supported by the fact that mice lacking functional serotonin neurons have abnormal arousal responses to CO_2 stimulation during sleep.

The neurocircuits that regulate sleep may also influence the ability of CO_2 sensors to transmit their excitatory drives to respiratory centers, thus limiting the expression of the hypercapnic ventilatory response during sleep. For example, because monoaminergic (e.g., noradrenergic in the LC; see Figure 23-4) and hypocretin neurons reduce their discharge activity during sleep and because these neurons also project to CO_2 sensors (e.g., RTN), then reduced excitatory drive from these sites would limit the ability of CO_2 sensors to relay their afferent signals to respiratory motor pools and to the respiratory centers that trigger the CO_2 response. Conversely, increased inhibitory drives from the GABAergic neurons in the ventrolateral and median preoptic nuclei that regulate NREM sleep could function to suppress CO_2 sensitivity during this state, because such inhibitory regions also project to CO_2 sensors and respiratory centers. Therefore both reduced excitation and increased inhibition during sleep may curtail CO_2–sensing mechanisms such that the hypercapnic ventilatory response is reduced in sleep.

Clinical Correlates

Although breathing continues without respite during wakefulness, it becomes fragile during sleep, and in approximately 2% to 5% of the adult population, it is punctuated by transient apneic episodes, which can manifest clinically as sleep-disordered breathing (see Section 14). OSA is the most common and serious respiratory-related sleep disorder. Its prevalence is growing, in keeping with its recognized link with obesity and the current obesity epidemic. A defining feature of OSA is that it occurs exclusively during sleep—patients with OSA breathe normally while awake but not while asleep. Changes in both chemical and mechanical processing during sleep account for the drastic changes in respiratory control that contribute to OSA. Reduced upper airway muscle tone during sleep, particularly REM sleep, is the primary cause of OSA; it causes either airway narrowing or complete airway obstruction, both of which lead to hypoventilation and subsequent asphyxia. Under waking conditions, these respiratory stimuli (i.e., airway narrowing and asphyxia) trigger an increase in upper airway muscle tone, which reopens the airspace by activating both the negative-pressure reflex and the hypercapnic ventilatory responses. Because these reflexes are suppressed during sleep, however, they are only partially reactivated, so airway muscle tone does not cause airway reopening. Accordingly, the airway remains occluded, and asphyxia worsens until it finally triggers awakening from sleep. Although wakefulness reinstates and reactivates the respiratory reflexes that increase muscle tone and thus causes airway reopening, it also causes hyperventilation, which in turn leads to hypocapnia. Hypocapnia then reflexively triggers a brief apneic episode, coincident with reentrance into sleep; this timing is problematic because sleep reduces the sensitivity of the mechanoreflexes and chemoreflexes needed to overcome apnea. Therefore apnea continues until the accompanying asphyxia produces arousal from sleep. This vicious circle occurs repeatedly throughout the night and underlies the sleep fragmentation, excessive daytime sleepiness, and related comorbid conditions (e.g., hypertension) that typify OSA.

CLINICAL PEARL

Determining how sleep controls sensory and motor processes is of notable importance in sleep medicine because some sleep disorders result from disturbances in sensory or motor control. As indicated by the available evidence, nociceptive transmission in the CNS is affected by sleep, and acute pain causes some sleep disturbance, but the extent and mechanisms by which chronic pain influences sleep, and vice versa, are largely unexplored. Clinicians need to be mindful that there is no simple relationship between chronic pain and sleep quality. Abnormalities in motor control during sleep underlie most of the major sleep disorders, including OSA, narcolepsy-cataplexy, RLS and PLMD, and sleep bruxism. Identifying the underlying mechanisms of motor regulation during sleep is a prerequisite for developing more rational treatments for such sleep disorders.

SUMMARY

Sensory and motor processing are reduced during sleep, and the integration of these processes is differentially affected during different sleep states, suggesting that discrete NREM and REM sleep mechanisms modulate and control sensory and motor processes in a sleep state–dependent manner. Changes in sensory processing during sleep affect the somatosensory pathways that transduce and relay nociceptive signals to and within the CNS. Sensory mechanisms are affected by sleep, and sleep is in turn affected by nociceptive processes. The interaction between pain and sleep is common in patients suffering from chronic pain such as that associated with cancer. Abnormal sensory and motor functions during sleep underlie common sleep disorders. For example, reduced chemical control of breathing during sleep plays a role in congenital central hypoventilation syndrome, and pathologic motor control during REM sleep underlies REM sleep behavior disorder.

Selected Readings

Brooks PL, Peever JH. Glycinergic and GABA(A)-mediated inhibition of somatic motoneurons does not mediate rapid eye movement sleep motor atonia. *J Neurosci* 2008;**28**:3535–45.

Burke PG, Kanbar R, Basting TM, et al. State-dependent control of breathing by the retrotrapezoid nucleus. *J Physiol* 2015;**593**:2909–26.

Dauvilliers Y, Siegel JM, Lopez R, et al. Cataplexy–clinical aspects, pathophysiology and management strategy. *Nat Rev Neurol* 2014;**10**:386–95. [Epub 2014 Jun 3].

Lavigne G, Khoury S, Laverdure-Dupont D, et al. Tools and methodological issues in the investigation of sleep and pain interactions. In: Lavigne G, Sessle BJ, Choinière M, Soja PJ, editors. *Sleep and pain.* Seattle: IASP Press; 2007. p. 235–63.

Massimini M, Ferrarelli F, Huber R, et al. Breakdown of cortical effective connectivity during sleep. *Science* 2005;**309**:2228–32.

Soja PJ. Modulation of prethalamic sensory inflow during sleep versus wakefulness. In: Lavigne G, Sessle BJ, Choinière M, Soja PJ, editors. *Sleep and pain.* Seattle: IASP Press; 2007. p. 45–76.

Sullivan CE, Murphy E, Kozar LF, Phillipson EA. Waking and ventilatory responses to laryngeal stimulation in sleeping dogs. *J Appl Physiol* 1978; **45**:681–9.

Sullivan CE, Murphy E, Kozar LF, Phillipson EA. Ventilatory responses to CO_2 and lung inflation in tonic versus phasic REM sleep. *J Appl Physiol* 1979;**47**:1305–10.

A complete reference list can be found online at ExpertConsult.com.

Chapter
24

Opiate Action on Sleep and Breathing

Ralph Lydic; John C. Keifer; Helen A. Baghdoyan; Robert Craft; Chelsea Angel

Chapter Highlights

- Opiates are not commonly prescribed for primary disorders of sleep, but with widespread use and increasing misuse of this class of drugs, opiate-related problems of respiratory depression, addiction, neuropsychiatric disorders, and optimal versus inadequate pain control, will be encountered with increasing frequency in sleep medicine clinics.
- Sleep disruption caused by opiates is consistently documented by studies using objective, polysomnographic measures. The effects of opiates on sleep are complex and vary as a function of the particular opiate administered, dose, and acute versus chronic administration. In the absence of data demonstrating significant, deleterious effects of opiate-induced sleep disruption, a prevailing view is to focus on the positive role that opiates play in acute pain management.
- Wakefulness is blunted by opiates, and chronic administration of opiates to manage chronic pain causes a dose-dependent torpor that is characterized by decreased physical and mental vitality. Drugs currently deployed to counter opiate-induced torpor and sedation are used off-label.
- The chapter concludes by emphasizing gaps in knowledge as opportunities for research on how to achieve both pain reduction and a restorative, sleep-like state that is as similar as possible to normal, nondrugged sleep.

Opiates have been briefly considered in previous editions of this book. Opiates can be useful for treating restless legs syndrome in some patients,[1] but these agents have the unwanted side effects of sleep disruption[2,3] and exacerbation of sleep-disordered breathing,[4] in concert with significant abuse potential.[4] Therefore, opiates are indeed relevant for practitioners of sleep medicine. The complex relationship among opiates, pain, and sleep means that sleep medicine clinicians will increasingly be confronted by problems of addiction, neuropsychiatric disorders, and pain. For example, deaths caused by overdose of prescription opiates have been described as an ongoing public health crisis.[5] Insomnia is the most common presenting condition in sleep medicine and often is a comorbid feature of posttraumatic stress disorder and opiate dependence.[6] This chapter updates and expands previous reviews of opiates and sleep,[7-9] and confirms that polysomnographic data repeatedly demonstrate that opiates disrupt sleep. This confirmed association should not be construed as indicating that the relationship between sleep and opiates has been thoroughly studied. Reviews of the literature reveal few objective polysomnographic studies that quantitatively describe how opiates, independent of pain, illness, or addiction, influence the temporal organization of sleep. This issue is directly relevant for a scientific approach to sleep disorders because although self-assessment of sleep quality is clinically valuable,[10] the data obtained using such assessments can differ significantly from measures provided by objective polysomnography.[11] To date, no polysomnographic studies in humans characterizing how the effects of opiates on sleep vary as a function of age, sex, body habitus, or race have been published. Furthermore, human polysomnographic data systematically comparing effects on sleep across multiple opiates, or across the variety of pain modalities are lacking. These gaps in knowledge represent important opportunities for sleep research. Studies also are needed to address the paradox that opiate-induced sleep disruption can contribute to increases in the doses of opiates that are needed to achieve pain relief, and to examine the putative role of opiates in causing hyperalgesia.[12]

This chapter is organized into four main topics: (1) the historical context for ongoing social and medical issues concerning opiates, (2) the mechanisms of sleep disruption caused by opiates, (3) the clinical relevance of opiate-induced respiratory depression, and (4) emerging opportunities for research on opiates and sleep. Not included here is the important topic of acute pain management by anesthesiologists in the perioperative setting. For more information on sleep and pain interactions, see Chapters 23 and 133.

HISTORICAL CONTEXT FOR ONGOING SOCIAL-MEDICAL ISSUES CONCERNING OPIATES: OPIOPHILIA AND OPIOPHOBIA

Efforts to promote patient access to pain management[13] and consensus statements on safe and effective treatment guidelines[14] have not resolved the longstanding debate regarding legitimate prescriptions for opiates, with both pros ("opiophilia") and cons ("opiophobia") well described.[15-17] Calls for a "balance between treating legitimate pain patients and mitigating opiate abuse, overdoses, and related deaths" reflect this

reality.[18] The increase in prescriptions for opiate analgesics[19,20] means that practitioners of sleep medicine commonly will confront the challenge of how to preserve the human right of pain relief while protecting patients against opiate side effects. The difficulty of this challenge is emphasized by the lack of consensus regarding prevalence and incidence of addiction associated with long-term use of opiates for chronic pain.[21-23]

Data selected from a large body of literature on substance abuse provide statistics showing that analgesic-induced overdose fatalities in New York City increased almost 700% during the interval from 1990 to 2006.[24] In September 2014, the National Institutes of Health (NIH) Office of Disease Prevention sponsored a workshop entitled "The Role of Opioids in the Treatment of Chronic Pain."[25] The NIH announcement also noted that chronic pain is a major public health problem affecting about 20% to 30% of the world's population. This statistic is emphasized by data from the Centers for Disease Control and Prevention showing that the problem of prescription opiate abuse is exacerbated by a 300% increase over the past 20 years in clinical prescriptions written for opiates. The number of deaths due to prescription opiates has exceeded the total number of deaths caused by cocaine and heroin.[25] Since 2005 in the state of Massachusetts, annual deaths attributed to opiate overdose have outnumbered those caused by motor vehicle accidents.[19] This crisis continues to receive extensive media coverage in North America and also is a concern in other parts of the world experiencing increased access to opiates.[26]

Figure 24-1 illustrates opium-induced states of torpor as a historical predecessor of contemporary opiophobia. During the opium wars of the early 19th century, the British foisted upon China the use of opium, with the drug serving as a form of currency intended to reduce a British trade deficit.[27] The Chinese administration fought this and sought to ban the sale of opium. The Chinese leadership understood that the state of torpor caused by highly addictive opium was a threat to the Chinese economic and social structure.

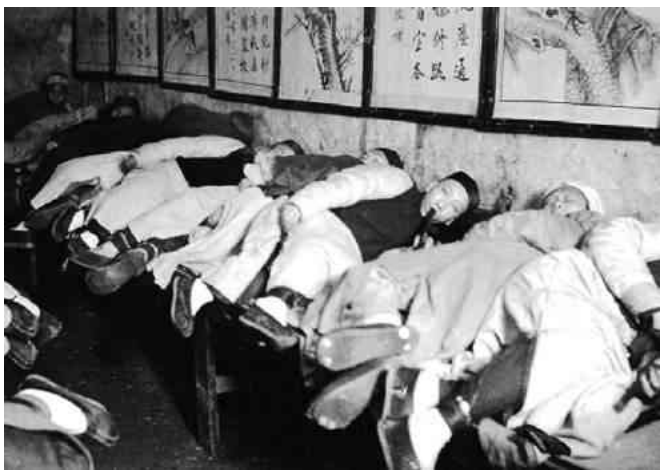


Figure 24-1 Photograph from the late 1890s showing persons in opium-induced torpor, characterized by diminished physical and mental activity, eyelid closure, and recumbency. This constellation of traits accounts for the mislabeling of opioid-induced torpor as sleep. One neuronal substrate for the opioid-induced blunting of wakefulness is suggested by the finding that mu opioid agonists depress the discharge of arousal-promoting hypocretin-containing neurons. (Li Y, van den Pol AN. μ-opioid receptor-mediated depression of the hypothalamic hypocretin/orexin arousal system. *J Neurosci* 2008; 28:2814–9.)

SLEEP MEDICINE, PAIN MEDICINE, AND HEALTH POLICY

The association between sleep disorders medicine and pain medicine has been slow to develop. More than 10 years ago the National Sleep Disorders Research Plan[28] documented the importance of achieving an understanding of the mechanisms by which opiates disrupt sleep. Opiates remain a key component of acute and chronic pain management[29,30] despite slow progress in overcoming unwanted side effects.[31] The U.S. Congress declared the years 2001 to 2010 as the Decade of Pain Control and Research (Public Law 106-386-OCT). The Pain Care Policy Acts of 2003 (HR 1863) and 2005 (HR 1020) were bills calling for establishment of a Center for Pain and Palliative Care Research within the NIH. These mandates were encouraged by data on the prevalence of clinically significant pain and by evidence that pain often is inadequately managed. For example, adequate postsurgical relief of acute pain is provided to only approximately 25% of patients.[32,33] Some studies report that only 40% of patients with chronic pain can obtain adequate relief.[33] More recent studies also question the efficacy of long-term opiate administration for treating nonmalignant chronic pain.[34] In 2001, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), since renamed "The Joint Commission," mandated that health care organizations comply with the agency's Pain Assessment and Management Standards in order to receive accreditation.[33] Compliance includes documenting the efficacy of pain treatment plans and educating patients and their families, as well as health care providers, about pain.[35] The 2010 Patient Protection and Affordable Care Act mandated the U.S. Department of Health and Human Services to "increase the recognition of pain as a significant public health problem in the United States."[13] In 2011 a report from the Institute of Medicine indicated that at least 116 million adults in the United States suffer from chronic pain resulting in an economic cost of up to $635 billion, exceeding the combined costs of cancer, diabetes, and heart disease.[13] Thus, nearly 40% of the U.S. population experiences chronic pain, a disease prevalence that is significantly greater than that for sleep disorders, affecting 50 to 70 million people.[36]

OPIATES AND OPIOID PHARMACOLOGY

The nomenclature used in this chapter is that recommended by the International Union of Basic and Clinical Pharmacology (IUPHAR), which distinguishes between exogenous opiate drugs and receptors that are activated by endogenous opioid peptides[37] including enkephalins, beta-endorphins, and dynorphins.[38] The IUPHAR review, summarizes nomenclature for the four opioid peptide receptors (mu, delta, kappa, and nociceptin/orphanin FQ), their nonapproved monikers, and the presumed endogenous ligands for each opioid receptor.[37] The mu opioid receptor was named for morphine, and the clinical management of both acute and chronic pain commonly involves administration of mu opiates. Mu opioid receptor activation closes voltage-gated calcium channels and opens potassium channels, thereby hyperpolarizing neurons

and decreasing neuronal excitability.[39] Opioid receptors are part of the family of G protein–coupled receptors, and the complexity of opioid signaling is illustrated by the identification of multiple splice variants of the mu opioid receptor throughout brain and spinal cord.[40,41] In addition, distinctly different opioid receptors can combine to form dimers and/or oligomers with unique functional roles.[42] Opiates also interact with multiple receptor systems. Adenosine is a sleep-promoting molecule, and the potential relevance of the foregoing details for sleep is demonstrated by evidence of interaction between mu opioid and adenosine A_1 receptors. For example, mu opiates enhance G protein expression in sleep-regulating regions of the brainstem, where opiates and adenosinergic agonists can additively activate G proteins.[43]

Despite the incontrovertible evidence of unwanted opiate side effects, it is important to acknowledge that opiates make a significant, positive contribution to the management of pain. Sleep disorders medicine and pain management are linked by a shared focus on states of consciousness.[32] States of pain are composed of complex psychophysiologic events that include nonnociceptive variables such as alterations in affect, cognition, and autonomic physiology,[44] each of which varies as a function of sleep and wakefulness. Finally, the problem of polypharmacy clearly complicates efforts to delineate the effects of opiates on sleep and wakefulness. Clinical efforts to manage cancer pain can involve a complex array of pure mu agonists, partial agonists such as buprenorphine, and opiates with multiple mechanisms of action.[45] At present, studies characterizing opiate effects on sleep-wake states using polysomnography are not available for a majority of 20 opioid medications considered in the context of managing cancer pain.[45] Readers are referred to cited references[46-48] for reviews of opiates and opioid pharmacology.

Systemically administered opiates alter cell excitability throughout the nervous system (Figure 24-2). These anatomically distributed effects cause diverse alterations in physiologic traits including cortical electroencephalogram (EEG) activity, breathing, and motor control, as well as in more global behavioral states. At present, it is not possible to selectively achieve the desired opiate effect of diminishing states of pain and anxiety without causing the unwanted side effects of disrupting breathing and states of sleep and wakefulness. The subsequent sections emphasize the persisting lack of data regarding unwanted opiate side effects.

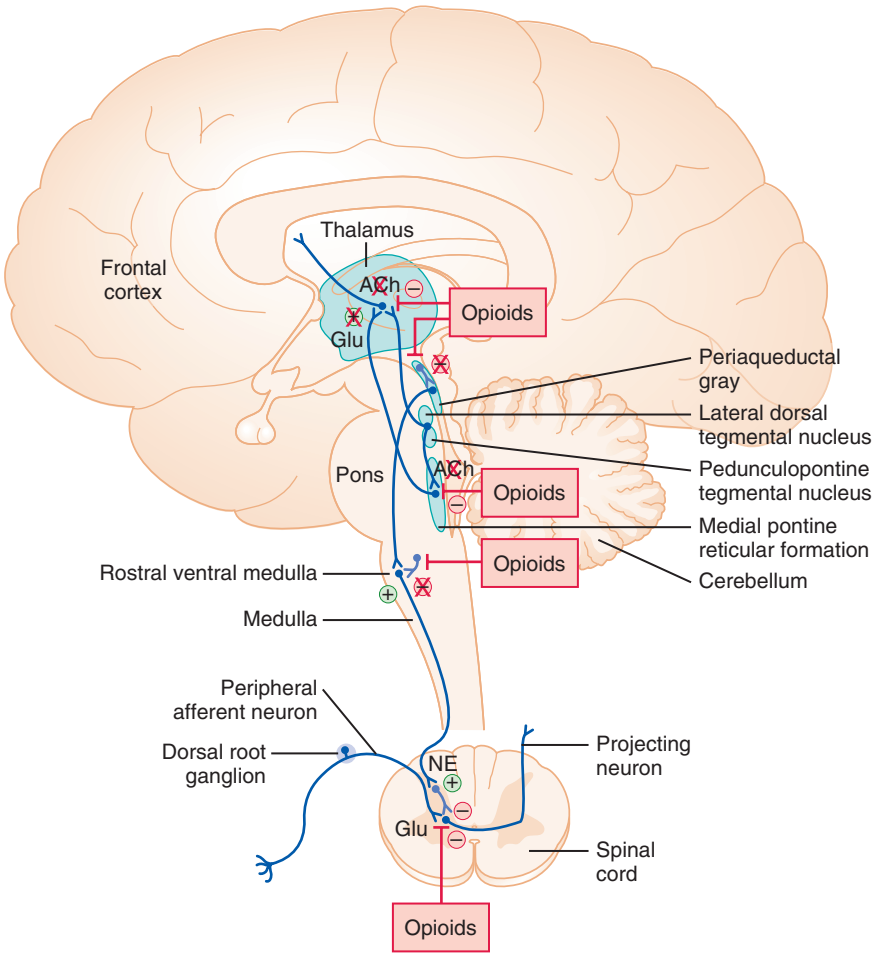


Figure 24-2 Schematic view of the brain and spinal cord. The figure emphasizes the wide anatomic distribution of sites where opioids have been shown to decrease neuronal excitability and/or neurotransmitter release. ACh, Acetylcholine; GLU, glutamate; NE, norepinephrine. (Modified from Brown EN, Lydic R, Schiff ND. General anesthesia, sleep, and coma. *N Engl J Med* 2010;363:2638–50.)

Opiates Disrupt the Temporal Organization of Sleep and Wakefulness

Sleep disruption has long been recognized as a major complaint of patients experiencing pain.[49] Unfortunately, opiates used in the clinical management of acute pain also can cause sleep disruption. This paradoxical effect associated with systemic administration of opiates has been consistently documented in studies using objective, polysomnographic measures of sleep (Table 24-1). Considered as a group, the studies summarized in Table 24-1 distinguish between disruptions in sleep that are caused by opiates and sleep changes that are caused by pain, disease, or addiction. Administering clinically relevant doses of opiates to otherwise healthy people increases light (stage N1) non–rapid eye movement (NREM) sleep, decreases deeper NREM sleep (stages N3, N4), and inhibits rapid eye movement (REM) sleep.[50-55] Sleep and EEG data recorded from rodents fit well with clinical evidence showing that opiates increase EEG power in the delta (0.35 to 3.5 Hz) and theta (3.5 to 8 Hz) frequency ranges.[56,57] Many studies have shown that even in pain-free subjects, sleep disruption promotes hyperalgesia.[58-61] The pronociceptive effects of sleep disruption are likely to be mediated by many neurochemicals in multiple brain regions. Available evidence indicates that gamma-aminobutyric acid (GABA)-ergic transmission in the pontine reticular formation contributes to hyperalgesia caused by sleep disruption.[62]

The effects on sleep of acute opiate administration in people who are pain-free are complex and also vary as a function of the particular opiate administered, dose, and route of administration. Remifentanil, for example, is a synthetic opioid that is delivered by intravenous infusion, rapidly reaches a steady state level in plasma, and has an elimination half-life of less than 4 minutes.[55] When delivered to pain-free subjects, remifentanil disrupted the temporal organization of sleep and inhibited REM sleep.[54,55] Buprenorphine, an agonist at mu opioid receptors and an antagonist at kappa opioid receptors, was discovered in 1966, and its role in the search for addiction therapeutics has been reviewed elsewhere.[63] Buprenorphine is advocated by some researchers as an agent for treating opiate addiction.[64] Survey data also indicate a large increase in buprenorphine use for nontherapeutic purposes.[65] The first study[66] showing that an antinociceptive dose of buprenorphine significantly disrupts sleep was published as recently as 2011. Systemic administration of buprenorphine to rats delayed sleep onset (Figure 24-3) and caused a significant increase in wakefulness and a significant decrease in both NREM sleep and REM sleep (Figure 24-4, *A-E*).[66] To date, no polysomnographic data that quantify the effects of buprenorphine on the sleep of healthy, pain-free, nonaddicted

Table 24-1 Polysomnographic Characterization of Opiate-Induced Disruption of Sleep and Wakefulness*

Study	Opiate	Trade Name(s)	No. of Patients (Female/ Male)	Species, Disposition	Wake Time	Light NREM Sleep Duration	Deep NREM Sleep Duration	REM Sleep Duration	Total Sleep Time	Latency to Onset
Shaw et al., 2005[50]	Morphine	Avinza, Kadian, MS Contin, Roxanol, Roxanol-T	7 (2/5)	Human 24–28 yo, healthy	—	↑	↓	↓	—	—
Bernards et al., 2009[55]	Remifentanil	Ultiva	19 (8/11)	Human 38–62 yo, moderate OSA	—	↑	NS	↓	↓	—
Axelin et al., 2010†	Oxycodone	Tylox, Percodan, OxyContin	18 (7/11)	Human 28–32 weeks, preterm	NS	↑ (NREM only)	↑ (NREM only)	↓	NS	—
Dimsdale et al., 2007[53]	Methadone	Methadose, Dolophine	42 (25/17)	Human 18–60 yo, healthy	—	↑	↓	NS	NS	—
Gauthier et al., 2011[66]	Buprenorphine	Buprenex, Suboxone, Subutex	26 (0/26)	Sprague Dawley rat, adult, healthy	↑	↓ (NREM only)	↓ (NREM only)	↓	↓	↓
Williams et al., 2012‡	Butorphanol	Stadol	6 (0/6)	Horse, adult, healthy	↑	↓ (SWS only)	↓ (SWS only)	↓	—	↓

*These studies are unique in that they do not include subjects with the potential confounding factors of surgery, pain, drug addiction, or other comorbid conditions.

†Axelin A, Kirjavainen J, Salantera S, Lehtonen L. Effects of pain management on sleep in preterm infants. *Eur J Pain* 2010;14:752–8.

‡Williams DC, Aleman M, Tharp B, et al. Qualitative and quantitative characteristics of the electroencephalogram in normal horses after sedation. *J Vet Intern Med* 2012;26:645–53.

NS, Not significant; OSA, obstructive sleep apnea; SWS, slow wave sleep; yo, years old.

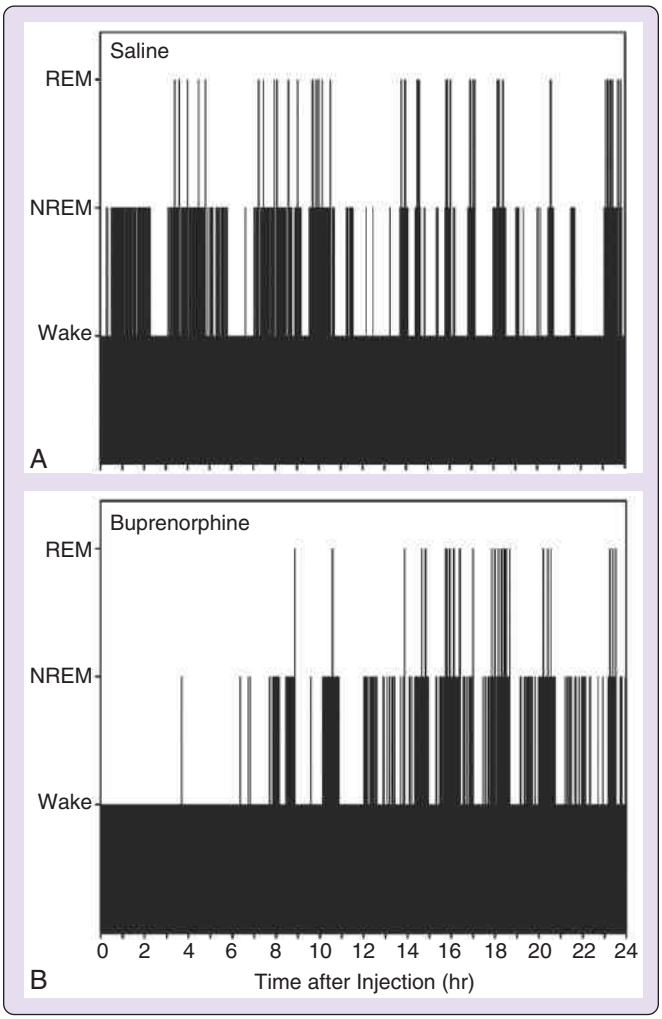


Figure 24-3 States of wakefulness (Wake), NREM sleep, and REM sleep in a Sprague Dawley rat.The states are plotted as a function of time after systemic administration of saline (**A**) or buprenorphine (**B**) at time point zero. Buprenorphine decreased both NREM sleep and REM sleep. (From Gauthier EA, Guzick SE, Brummett CM, et al. Buprenorphine disrupts sleep and decreases adenosine concentrations in sleep-regulating brain regions of Sprague Dawley rat. *Anesthesiology* 2011;115:743–53.)

humans are available. The ideal clinical goal is to achieve both pain reduction and a restorative, sleep-like state that is as similar to normal, nondrugged sleep as possible. The potential to achieve this dual goal is suggested by the results of preclinical studies showing that buprenorphine-induced disruption in the temporal organization of sleep can be diminished by coadministering the sedative-hypnotic eszopiclone (Figure 24-4, *F-J*).[66]

Opiates Promote Torpor-like States

Tissue injury and some pain-producing diseases are accompanied by inflammation. The signs of inflammation were recognized in the First century by Celsus to be heat, pain, redness, and swelling (*calor*, *dolor*, *rubor*, and *tumor*).[67] The use of opiates to provide pain relief adds a fifth cardinal sign of "torpor." The clinical manifestations of torpor are expressed as lethargy and a reduced ability to sustain physical and mental activity. (Additional information on sleep and torpor is provided in Chapter 21.) Opiates blunt wakefulness and slow cortical EEG activity.[68-72] Systemic delivery of morphine to rodents significantly increases low-frequency EEG power.[73,74] EEG slow wave activity is characteristic of the postmorphine state and is accompanied by behavioral torpor and muscle rigidity.[72,74] The action of opiates to blunt wakefulness is relevant because the duration and quality of wakefulness significantly modulate subsequent sleep[75,76] and global brain states.[77] The torporlike state experienced by some patients receiving opiates for pain on an acute basis, in whom a rapid return to health is anticipated, is commonly regarded as an equitable exchange for relief from pain. For patients experiencing chronic pain, however, the decreased physical and mental vitality characteristic of torpor can significantly diminish the quality of life. Inflammation alone can produce fatigue,[78] which may interact with the sedating effects of opiates to induce a state of torpor. Studies on the neurobehavioral effects of opiates are available,[79-81] but the psychophysical characteristics of torpor associated with chronic administration of opiates remain poorly understood. A 1-year retrospective chart review of findings in adult patients with nonmalignant pain who were identified by the Epworth Sleepiness Scale as experiencing opiate-induced sedation suggested that treatment with modafinil diminished sedation.[82] By contrast, a retrospective examination of the 1996 to 2004 MEDLINE database provided limited support for the view that methylphenidate, donepezil, and modafinil can offset the torporlike state caused by opiates.[83] The study authors emphasized the limitations of their review[83] and provided a useful summary (Table 24-2) that can inform the design of prospective clinical studies aiming to achieve opiate-induced pain relief while minimizing a torpor-like blunting of wakefulness. The drugs shown in Table 24-2 include a three-pronged approach of enhancing hypocretinergic transmission (modafinil), decreasing the degradation of acetylcholine (donepezil), and enhancing sympathomimetics (methylphenidate and dextroamphetamine). For all of the listed drugs, recommended uses are off-label.

CLINICAL RELEVANCE OF OPIATE-INDUCED RESPIRATORY DEPRESSION

Although many nonanalgesic effects of opiates have been documented,[84,85] the most notorious side effect is opiate-induced respiratory depression. This section focuses on the intersection of opiate-induced respiratory depression and sleep medicine and on how the overlap affects specific patient populations. For a more comprehensive review of opiate-induced respiratory depression, readers are referred elsewhere.[86-95]

A complex association has been recognized between sleep-disordered breathing, being overweight or obese, and pain. For example, sleep-disordered breathing is common in overweight or obese persons.[96] Obesity itself is associated with increased pain,[97,98] and among patients with sleep-disordered breathing, increased reports of pain are typical.[99] The neurochemical mechanisms modulating the foregoing associations remain unknown. Preclinical studies show that an adenosine A_1 receptor agonist is antinociceptive only in the presence of the satiety factor leptin.[100] The hypothesis that nocturnal (or recurrent) hypoxia somehow causes pain is currently being studied. When nocturnal hypoxemia was examined independently of sleep fragmentation in patients with obstructive sleep apnea (OSA), it was found that pain reported by the participants on awakening significantly increased.[99] These findings, along with the discovery that use of continuous

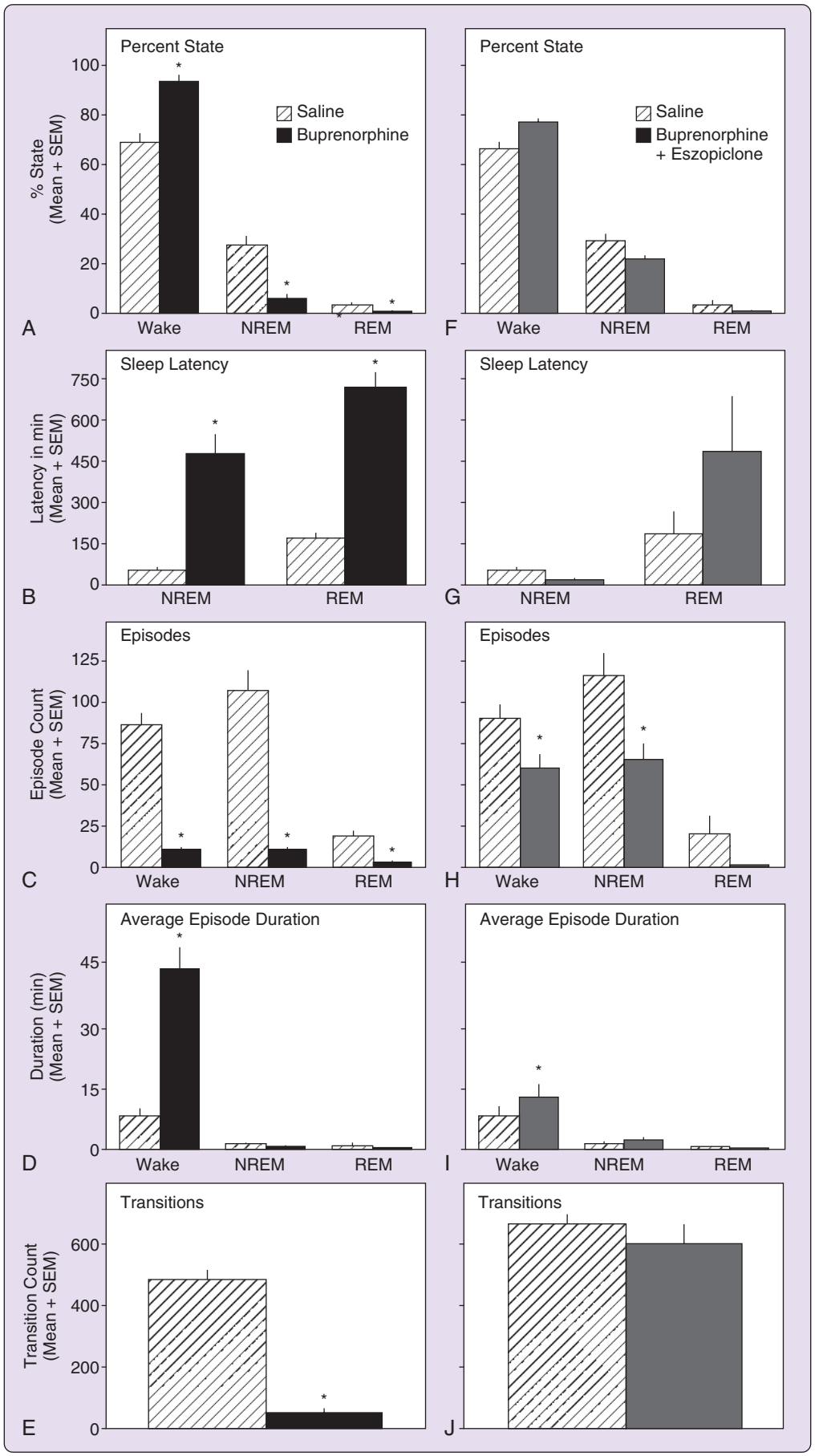


Figure 24-4 Disruption of sleep and wakefulness by buprenorphine (**A** to **E**) and reductions in the sleep-wake effects of buprenorphine by coadministration of the sedative-hypnotic eszopiclone (**F** to **J**). Frames **A** through **E** show group data from seven animals that received either saline (control) or buprenorphine. Wakefulness was increased, and NREM sleep and REM sleep were decreased as quantified for the percentage of time spent in each state (**A**) during a 24-hour recording and as quantified by the number of episodes of each state (**C**). Frames **F** through **J** show that coadministration of eszopiclone diminished sleep disruption caused by buprenorphine. (From Gauthier EA, Guzick SE, Brummett CM, et al. Buprenorphine disrupts sleep and decreases adenosine concentrations in sleep-regulating brain regions of Sprague Dawley rat. *Anesthesiology* 2011;115:743–53.)

Table 24-2 Medications Used to Counter States of Sedation and Torpor Caused by Opiates

Medication	FDA Indication	Typical Use(s)	Potential Adverse Effects	Dosing Range*	Advantages	Disadvantages
Methylphenidate	CNS stimulant	Narcolepsy, ADHD	Agitation, anxiety, psychosis, anorexia, hypertension, tremor, tachycardia, diaphoresis	10–15 mg/day	Most clinical trial support, rapid onset in responders, inexpensive	Significant adverse effects, schedule II prescribing restrictions, abuse potential, diversion potential, stigma associated with use
Dextroamphetamine	CNS stimulant	Narcolepsy, ADHD	Agitation, anxiety, psychosis, anorexia, hypertension, tremor, tachycardia, diaphoresis	5–30 mg/day	Few compared with methylphenidate	Similar to methylphenidate
Donepezil	ACh esterase inhibitor	Alzheimer's disease, dementia	Nausea, vomiting, diarrhea, headache	2.5–15 mg/day	Generally well tolerated, once-daily dosing, nonscheduled, promotility effect may relieve constipation, safe in patients with cognitive dysfunction, no abuse potential	Limited clinical trial support, expensive
Modafinil	CNS stimulant	Narcolepsy, shift work sleep disorder	Agitation, anxiety, dizziness, psychosis, diarrhea, nausea, hypertension, tachycardia	100–600 mg/day	Generally well tolerated, schedule IV some prescribing restrictions, low abuse potential	Limited clinical trial support, expensive, multiple CYP450 enzyme system effects
Caffeine	CNS stimulant	Headache	Agitation, insomnia, GI upset, nervousness, restlessness	Unknown	Readily available, generally well tolerated, inexpensive	No data to support efficacy, variable patient tolerance

*Dosing range represents doses described in trials of opiate-induced sedation.
ADHD, Attention deficit–hyperactivity disorder; CNS, central nervous system; GI, gastrointestinal.
Modified from Reissig JE, Rybarczyk AM. Pharmacologic treatment of opioid-induced sedation in chronic pain. *Ann Pharmacother* 2005;39:727–31.

positive airway pressure decreased the pain experienced by patients with severe OSA,[101] are consistent with the interpretation that intermittent hypoxia may be one source of pain experienced by patients with OSA. Conversely, subsequent work has shown that both adult and pediatric patients with recurrent hypoxia and severe OSA require less opiate administration postoperatively to obtain adequate analgesia.[102,103] On the basis of the available literature, it is not possible to unequivocally resolve the association between sleep-disordered breathing and pain.[99,102] Mu opioid receptor upregulation occurs in the developing rat in response to intermittent hypoxia.[104] Humans with high levels of endogenous opioids display less analgesia when administered morphine.[105] Thus hypoxia-associated pain and decreased opioid requirement may not always be mutually exclusive.[106]

During both REM sleep and NREM sleep, the ventilatory response to carbon dioxide is reduced, minute ventilation decreases, and a mild to moderate hypoxemia prevails.[90] Similarly, opiates induce central nervous system depression by blunting the chemoreceptive response to both carbon dioxide and hypoxia, prolonging exhalation time, and suppressing tidal volume, while also increasing upper airway resistance.[90,107-109] Thus, sleep and opiates both impair breathing by reducing respiratory rate and decreasing tonic respiratory drive.[90,108] These additive effects contribute to opiate-induced respiratory depression.

Opiates and Sleep-Disordered Breathing

Sleep-disordered breathing is a term that encompasses OSA, central sleep apnea, and upper airway resistance syndrome (snoring).[107] In the general population, men are twice as likely to be diagnosed with OSA as women are, with middle-aged obese males identified as the most susceptible.[91,110–112] Many persons with OSA remain undiagnosed preoperatively.[112] This is problematic because these patients are at increased risk for upper airway collapse,[110] and because accumulating evidence suggests that administration of mu opioid analgesics in this population has been associated with an increased risk of respiratory complications in the postoperative period.[102] The evidence also points to an increase in endogenous opioids in the cerebrospinal fluid of patients with sleep apnea syndrome, which can in turn increase sensitivity to opiates.[113] Additionally, a dose-dependent relationship has been confirmed between chronic opiate use and sleep-disordered breathing regardless of the presence or absence of a preexisting condition.[9,90,91,114-118] Thus, a preoperative condition of sleep-disordered breathing can intensify opiate-induced respiratory depression, and opiates alone can lead to sleep-disordered breathing.[109] Patients with pure OSA who also are being treated with opiates are at increased risk for development of complex sleep apnea, characterized by central as well as obstructive apneic events.[119–121]

Occurrence of central apneas during NREM sleep is the most common effect of chronic opiate use on breathing and sleep.[87,109,122,123] Seventy percent of patients chronically using opiates developed ataxic breathing and central apneic episodes during NREM sleep, compared with 5% of control subjects.[91] By contrast, central sleep apnea was reported in 30% of patients undergoing methadone maintenance therapy.[124] Data on the occurrence of sleep apneas after surgery also are available: An increase in the apnea index has been found to be most severe in the first 24 to 48 hours postoperatively.[110,125] Understanding the degree and timing of postoperative changes in sleep-disordered breathing and sleep architecture can be expected to help develop strategies for postoperative monitoring and treatment.[110] A recent innovation is development of a new mode of minute ventilation–targeted servoventilation for central apneas associated with opiate use. This approach was significantly more effective than bilevel ventilation with back-up respiratory rate (bilevel-ST) in decreasing both the apnea-hypopnea index and the central apnea index.[126]

Considered next are specific populations known to be vulnerable to opiate-induced respiratory depression.

Obesity and Opiate-Induced Respiratory Depression

Obesity has become a worldwide epidemic.[127] In the United States, the incidence of obesity has doubled over the past 20 years, yielding frequencies three times higher than for some European countries, with the greatest relative increase in the subpopulation with a body mass index greater than 50 kg/m^2.[112,128] Paralleling this increase in obesity over the past decades has been an eightfold increase in the occurrence of OSA, with as many as 50% of morbidly obese patients experiencing sleep apnea.[111,119] Obesity functions as a comorbid condition for OSA by exacerbating airway obstruction through anatomic factors, thereby increasing the risk of life-threatening respiratory complications induced by opiates.[112,125] Affected patients need to be closely monitored postoperatively, and therapies such as continuous positive airway pressure, bilevel positive airway pressure, or auto-titrating positive airway pressure should be considered.[112]

Geriatric Population Is Vulnerable to Opiate-Induced Respiratory Depression

The geriatric community represents the fastest-growing sector of society, and older patients undergo surgery four times more frequently than other age groups.[129] A combination of age-related physiologic changes and increased risk of disease augments the potential for negative disease-drug and drug-drug interactions. Furthermore, the complex association between sleep-disordered breathing, pain, and postoperative outcome is exacerbated by the fact that the prevalence rates for both sleep-disordered breathing and pain increase with age.[130,131] Adequate postoperative analgesia in elderly patients is of particular importance because insufficient pain control is associated with unfavorable outcomes.[129] As a consequence of aging, the pharmacokinetics and pharmacodynamics of drug action change, leading to increases in mean elimination half-life and brain sensitivity with decreases in volume of distribution, clearance, and protein binding.[129] In addition, normal reduction in renal function occurring as a consequence of aging hinders the elimination of many drugs and their metabolites.[132] Simultaneously, a drug's EC_{50}, the concentration producing half of the maximal effect, linearly decreases with age.[132] The major morphine metabolite, morphine-6-glucuronide, is more potent than morphine. Accumulation of morphine-6-glucuronide, in concert with its increased potency, can lead to heightened sensitivity to opiates and increased risk of respiratory depression.[132] These effects may be especially pronounced in geriatric patients in the seventh to ninth decades of life, in whom the risk of critical ventilatory depression is between 2.8 and 8.7 times higher than that in younger patients.[132] Owing to these natural physiologic changes, it is

important to adopt the "start low and go slow" approach for titrating the opiate dose prescribed to elderly patients.[129]

Opiates Alter Breathing in Children

One of the most common procedures performed in response to pediatric OSA is adenotonsillectomy, for which postsurgical administration of opiates often results in respiratory complications similar to those that occur in adults.[103] A prospective, stratified, and blinded study found that "the postsurgical total analgesic morphine dose required for children with a history of significant recurrent hypoxemia is lower than that in children who have not experienced such hypoxemia."[103] (See also Chapters 148 and 149 on perioperative use of opiates in patients with obstructive sleep apnea.) Additionally, this study identified a twofold difference in the calculated total analgesic morphine dose required between children with severe versus mild hypoxemia.[103] These results are supported by recent preclinical studies showing that respiratory sensitivity to opiates is enhanced by recurrent hypoxic episodes.[133] Finally, similar to the geriatric population, the pediatric cohort demonstrates age-related physiologic patterns that affect the metabolism of opiates. Some of these differences in pediatric patients include a potentially underdeveloped renal pathway and *CP2D6*-dependent conversion of codeine to morphine in infants (before 6 months of age).[133] These findings further emphasize the significance of individual dose titration, encouraging clinicians to determine the level of hypoxemia in each child diagnosed with OSA before surgery.[103]

Chronic Pain

The number of patients experiencing chronic pain has been steadily on the rise, with 100 million Americans suffering from the condition as of 2011.[13] Prevalence of chronic pain is higher in patients with a body mass index of 27 or greater, women, and the elderly; the chronic pain patient population is thus likely to grow along with the aforementioned cohorts.[134,135] Fifty percent to 90% of patients with chronic pain are estimated to report poor sleep quality, including sleep disturbances, difficulty falling and staying asleep, and less restful sleep.[134,136] Additionally, chronic pain is known to alter sleep architecture.[9,58,134] Chronic pain disrupts sleep, and sleep disruption worsens pain.[9,58,93,134,136,137] This vicious circle can contribute to a diminished quality of life by inflicting emotional, physical, economic, and social distress on the patient.[9,136,138] Treatment for some forms of chronic pain may involve around-the-clock administration of long-acting opiates, supplemented by short-acting opiates for breakthrough pain.[9,93,139] It is hypothesized that by utilizing extended-release opiates, the plasma levels of the analgesic can be maintained at a constant level, thereby improving sleep by eliminating gaps in analgesic action and blocking the associated breakthrough pain.[134] A study in which patients with chronic pain were treated with extended-release morphine resulted in significant improvements to several sleep measures compared with placebo.[134] These improvements in the quality of sleep for patients with chronic pain who are on extended opiate therapy[136] are tempered by the finding that chronic opiate use has been associated with increased risk of central sleep apnea in this population.[90,109,115,134,136] Additionally, it is a common occurrence during opiate therapy for patients with chronic pain to experience a loss of treatment effect, necessitating an increase in opiate dose to obtain adequate analgesia.[140,141] Some researchers suggest that this increased need is attributable to opiate-induced hyperalgesia, in which exposure to opiates results in an increased sensitivity to pain.[140] Owing to variability in experimental methods and inconsistent results, however, tolerance and acute withdrawal cannot be ruled out as potential causes of opiate dose escalation; accordingly, further investigation is required.[141]

Substance Abuse, Opiates, and Respiratory Control

Sleep disorders are 5 to 10 times more common in substance abusers than in nonabusers.[142] Oral administration of methadone to persons addicted to heroin has become the international standard for treatment.[143] This approach has proved to be successful in keeping heroin-dependent patients in treatment programs while reducing their risk of relapse by preventing or reducing withdrawal symptoms.[143] Because of these results, enrollment in methadone maintenance programs has been on the rise, with more than 160,000 enrollees in the United States.[144] Unlike previously discussed cohorts, methadone-maintained patients have been found to develop tolerance to respiratory depression caused by the opiate over a period of 5 to 8 months, although the reduction in hypoxic drive does not fully reverse during this time.[144] Patients undergoing methadone maintenance therapy exhibited significant changes in sleep architecture compared with control subjects, including decreases in total sleep time, slow wave sleep, and REM sleep, with an increase in stage 2 sleep.[144] Similar to patients with chronic pain, methadone-maintained patients experienced hyperalgesia when pain tolerance was tested using cold pressor techniques.[145] Additionally, a weak but significant correlation between methadone blood concentration and central sleep apnea was found.[144,146] As the number of patients undergoing methadone maintenance therapy increases, the likelihood that these patients will require treatment for either acute or chronic pain also increases. This duality of diagnosis can pose a problem because drug tolerance in these patients is not limited to methadone but extends to different opiate medications, such as morphine.[145,147] These changes in pain sensitivity and tolerance require clinicians to administer a larger quantity of drug more frequently to effect adequate pain control, thereby increasing the risk of respiratory complications.[145,147] In these patients, careful titration of dosage, with concomitant close monitoring of respiratory function, is essential.

The foregoing overview illustrates that opiate-induced respiratory depression varies for subgroups of patients. The acknowledged limitations of such studies, and the ethical challenges to performing such studies, also can be read as opportunities for future research. As summarized by Orlov and associates[117]:

> [O]ur findings underscore the need for large, prospective, observational studies that incorporate validated screening tools, reliable diagnostic tests, homogeneous sample populations, standardized monitoring for operationally defined complications, and faithful recording of adverse events in order to reliably quantify risk.

EMERGING OPPORTUNITIES FOR RESEARCH ON OPIATES, SLEEP, AND PAIN

A unifying theme throughout this chapter is insufficient interaction between pain medicine and sleep medicine. For example,

the word *pain* does not appear in an otherwise excellent 2014 consensus statement[36] on how best to achieve the goals of the 2011 NIH Sleep Disorders Research Plan. In 2011 the Society for Anesthesia and Sleep Medicine (SASM)[148] was incorporated in the United States "to promote interdisciplinary communication, education and research in matters common to anesthesia and sleep."[149] This SASM mission aims to address the estimated 15 million to 20 million cases of surgical anesthesia in the United States per year[150] and the relevance of postsurgical pain disrupting sleep.[151] SASM has recently partnered with the International Anesthesia Research Society (IARS), and the IARS journal *Anesthesia and Analgesia* has established an editorial board position focused specifically on respiration and sleep.[152]

A better understanding of endogenous opioids may provide mechanistic insights into the neurobiology underlying states of hibernation[153] and opiate-induced torpor. Just as sleep and anesthesia are different states sharing some similar traits, states of torpor and states of hibernation share an intimate link with thermoregulation and energy balance. (Chapter 21 provides additional information on hibernation and seasonal triggering of torpor.) The opiate antagonist naloxone can prevent entry into and reduce the duration of hibernation.[154,155] As reviewed elsewhere,[156] hibernation-inducing factor is an 88-kDa polypeptide found in ground squirrels that binds to the delta opioid receptor. Induction of hibernation mediated by hibernation-inducing factor activation of delta opioid receptors was confirmed by the finding that the delta opioid agonist (D-Ala 2, D-Leu 5) enkephalin (DADLE) did induce hibernation.[156] Studies of hibernating, nonhuman primates also indicate an interaction among sleep, thermoregulation, and metabolic rate.[157] These findings encourage future research aiming to determine the extent to which opioid receptors modulate sleep, thermoregulation, and metabolic rate in human and nonhuman primates.

Substantial progress has been achieved in identification of the multiple brain regions and molecules regulating sleep and in elucidation of how levels of these neurotransmitters are altered by opiates.[7,158] Additional work is needed to characterize the effects of an extensive number of molecules including cannabinoids, oleamides, and resolvins/protectins,[159-161] which show promise for treating chronic pain[162,163] and altering states of sleep and wakefulness.[164,165]

CLINICAL PEARL

Opiates significantly alter states of sleep and wakefulness as well as control of breathing. The variety of opiate receptors and endogenous opioid neurotransmitters within numerous tissues emphasizes the myriad physiologic functions that are modulated by opiates. Opiates acting primarily through G protein–coupled receptors also have an impact on non–G protein–related functions, underscoring the potential synergy of opiates with a vast array of nonopioid compounds. Clinical syndromes requiring long-term narcotics are not cured but are managed. To minimize untoward side effects, changes in prescriptions for sedating or wakefulness-promoting medications must be made in concert with the prescriber of narcotic medications.

SUMMARY

Opiates are vital for the clinical management of acute and chronic pain. Avoiding the potential for addiction, respiratory depression, blunted wakefulness, and sleep disruption makes use of opiates a nexus for efforts to optimize patient care. Opiate-induced sleep disruption has been documented by studies using objective, polysomnographic measures of sleep. Opiates disrupt the temporal organization of sleep, and even in pain-free persons, sleep disruption promotes hyperalgesia. Opiates also blunt wakefulness and can produce dose-dependent states of torpor. Opiates blunt the ventilatory response to hypercarbia and hypoxia, prolong expiratory time, and suppress tidal volume, while increasing upper airway resistance. Opiate-induced respiratory depression can be synergistic with comorbid conditions commonly encountered in sleep medicine. Increased vigilance on the part of health care providers is indicated when opiates are part of a treatment plan for obese patients and for patients with obstructive and central apnea. In geriatric as well as pediatric patients, age-specific pharmacokinetics can enhance opiate potency and decrease opiate elimination. Among patients with chronic pain and with a history of substance abuse, a significantly increased incidence of sleep disorders has been documented.

Selected Readings

Atkinson TJ, Schatman ME, Fudin J. The damage done by the war on opioids: the pendulum has swung too far. *J Pain Res* 2014;**7**:265–8.

Baghdoyan HA, Lydic R. The neurochemistry of sleep and wakefulness. In: Brady ST, Albers RW, Price DL, Siegel GJ, editors. *Basic neurochemistry*. 8th ed. New York: Academic Press; 2012. p. 982–99.

Brennan MJ, Lieberman JA. Sleep disturbances in patients with chronic pain: effectively managing opioid analgesia to improve outcomes. *Curr Med Res Opin* 2009;**25**:1045–55.

Campbell ND, Lovell AM. The history of the development of buprenorphine as an addiction therapeutic. *Ann N Y Acad Sci* 2012;**1248**:124–39.

Chou R, Fanciullo GJ, Fine PG, et al. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *J Pain* 2009;**10**:113–30.

Garrity AG, Botta S, Lazar SB, et al. Dexmedetomidine-induced sedation does not mimic the neurobehavioral phenotypes of sleep in Sprague Dawley rat. *Sleep* 2015;**38**:73–84.

Harrington MD. Neurobiological studies of fatigue. *Prog Neurobiol* 2012;**99**:93–105.

Horner RL, Hughes SW, Malhotra A. State-dependent and reflex drives to the upper airway: basic physiology with clinical implications. *J Appl Physiol* 2014;**116**:325–36.

Lydic R, Baghdoyan HA. Neurochemical mechanisms mediating opioid-induced REM sleep disruption. In: Lavigne G, Sessle BJ, Choinière M, Soja PJ, editors. *Sleep and pain*. Seattle: International Association for the Study of Pain; 2007. p. 99–122.

Mason M, Cates CJ, Smith I. Effects of opioid, hypnotic and sedating medications on sleep-disordered breathing in adults with obstructive sleep apnoea. *Cochrane Database Syst Rev* 2015;(**7**):CD011090.

National Institutes of Health Office of Disease Prevention. *Pathways to prevention: the role of opioids in the treatment of chronic pain*. <https://prevention.nih.gov/programs-events/pathways-to-prevention/upcoming-workshops/opioids-chronic-pain>; 2014.

Nuckols TK, Anderson L, Popescu I, et al. Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. *Ann Intern Med* 2014;**160**:38–47.

Silverman SM. Opioid induced hyperalgesia: clinical implications for the pain practitioner. *Pain Physician* 2009;**12**:679–84.

A complete reference list can be found online at ExpertConsult.com.

Chapter
25

Pathophysiology of Sleep-Wake Disturbances After Traumatic Brain Injury

Nadia Gosselin; Christian R. Baumann

Chapter Highlights

- Sleep-wake disturbances (SWDs), particularly fatigue, excessive daytime sleepiness, and pleiosomnia (an increased sleep need per 24 hours), are very common after traumatic brain injury (TBI), but their pathophysiology is poorly understood.
- The bulk of animal studies consistently show that TBI leads to modification in sleep-wake patterns, which suggests that the TBI itself causes a proportion of SWDs.
- Brain dysfunction, in animals and humans, includes impaired neurotransmitter signaling, changes in gene expression, altered circadian rhythms, cerebral hypometabolism, neuroinflammation, and hypopituitarism; all have been identified as potential contributors to the emergence and persistence of posttraumatic SWD.
- Other factors that may occur in association with TBI in humans are pain and psychiatric comorbidity, which can independently affect sleep-wake pattern after TBI. Prolonged use of sedatives/analgesics also was found to exacerbate SWD.
- The role of sleep in cognition, neuroplasticity, and neurogenesis is now well recognized, but increasing evidence suggests that sleep and its abnormalities after TBI influence recovery, on both a clinical and a neuropathologic level. Improving the current understanding of factors involved in posttraumatic SWD could lead to more specific and efficient interventions in this clinical population.

Traumatic brain injury (TBI) is the leading cause of death and disability among young adults in industrialized countries.[1] Incidence is estimated at up to 600 per 100,000 in the general population, and a majority of victims are young men entering their most productive years. TBIs often result in short- and long-term impairments that interfere with the return to normal life. Fatigue, excessive daytime sleepiness, pleiosomnia (an increased sleep need per 24 hours), and insomnia are among the most persistent and the most disabling symptoms after TBI. Chronic SWDs are reported by at least 50% of patients with TBI, but despite their high prevalence, the emergence and evolution of SWD are still poorly understood and probably involve interaction aong multiple causal factors.

This chapter is organized into three main sections, as follows: (1) an introduction to the diagnosis and the pathophysiology of TBI; (2) a survey of recent animal models that were specifically developed to characterize posttraumatic sleep-wake patterns; and (3) an update on human neuropathology data giving insights into potential pathophysiologic causes, physiologic factors, and influences of comorbid conditions that may contribute to SWD onset or maintenance. Other important sleep-wake disorders associated with TBI, not covered in this chapter, include insomnia, narcolepsy, sleep-disordered breathing, and sleep-related movement disorders; these are discussed in Chapter 99 and other subsequent chapters.

INTRODUCTION TO TRAUMATIC BRAIN INJURY

According to the definition proposed by the Demographics and Clinical Assessment Working Group of the International and Interagency Initiative, TBI is an alteration in brain function, or other evidence of brain pathology, caused by an external force.[2] In turn, alteration in brain function is defined as one of the following signs: any period of decreased level of consciousness, any loss of memory for events immediately before (retrograde amnesia) or after the injury (posttraumatic amnesia), neurologic deficits (e.g., weakness, loss of balance, change in vision, aphasia), or any alteration in mental state at the time of the injury (e.g., confusion, disorientation). Other evidence of brain pathology may include visual, neuroradiologic, or laboratory confirmation of damage to the brain. The diagnosis of TBI involves a severity assessment.[3] Generally, mild TBI is characterized by a short loss of consciousness (<30 minutes), a Glasgow Coma Scale (GCS) score between 13 and 15, and a short period of posttraumatic amnesia (<24 hours). Moderate TBI typically is associated with a loss of consciousness of 30 minutes to 24 hours, a GCS score between 9 and 12, and posttraumatic amnesia lasting 1 to 14 days.

Severe TBI generally is characterized by a loss of consciousness of more than 24 hours, a GCS score between 3 and 8, and posttraumatic amnesia persisting during several weeks.

Pathophysiology of Traumatic Brain Injury

Cerebral damage after TBI results from primary and secondary insults.[4] Primary insults result when the application of biomechanical forces causes local lesions (i.e., skull fracture, intracranial hematoma, lacerations, and contusions) or diffuse axonal injury. Secondary insults occur in the hours or days after the initial trauma as a consequence of neuronal hypoxia caused by several dysfunctions that alter the cerebral oxygen and glucose supplies.[5] Intracranial hypertension is the most common form of secondary insult. Both primary and secondary insults lead to neurochemical changes, such as an increase in extracellular excitatory amino acids (glutamate and aspartate),[6] which provokes excessive intracellular calcium influx, leading to free radical production, mitochondrial dysfunction, apoptosis, and inflammatory response.

Brain CT scan typically is performed in patients who present with TBI, particularly those with moderate or severe TBI, and is considered a good clinical tool for detecting focal lesions[7] that are mostly localized in the ventral and polar frontal regions and in the anterior temporal lobes.[8] However, TBI often is associated with diffuse axonal injury,[9] which is difficult to quantify by conventional imaging techniques. Rapid acceleration-deceleration of the brain generates traumatic shearing forces, ultimately leading to diffuse axonal injury. Diffusion tensor imaging, which allows the measurement of microstructural white matter integrity, showed white matter injuries in the corpus callosum and the internal capsule, as well as a global white matter loss, within 2 weeks after TBI that were mostly attributed to axonal swelling.[10,11] Neuroimaging findings in the acute stage of TBI are extremely variable from one patient to another in terms of severity and localization of lesions, with mild TBI generally causing no acute white matter lesions.[12]

TBI also may lead to permanent pathologic modifications of brain structures, which can explain some of the neuropsychiatric and cognitive dysfunctions observed in this population. In patients with chronic TBI, reduced gray matter may be observed, particularly in the hippocampus, but also in the frontal and temporal cortex, the thalamus, the basal forebrain, the anterior cingulate, the caudate nucleus, and the insula.[13,14] This long-term decrease in gray matter density is more severe in patients with lower initial GCS score or in those who presented with intracranial hypertension in the intensive care unit.[15] In these TBI subgroups of more severely affected patients, white matter degeneration can be observed in all of the major fiber tracts, including the corpus callosum, the cingulum, the superior and inferior longitudinal fasciculus, the uncinate fasciculus, and the brainstem. More severe disruptions in cortical and cortical-subcortical pathways were found to correlate with poorer cognitive functioning.[10,14]

Traumatic Brain Injury and Neurodegenerative Diseases

Several studies have drawn a parallel between a history of TBI and the later development of a neurodegenerative disease. The first evidence comes from epidemiologic studies, which showed that TBI increases the risk of developing Alzheimer disease or another dementia later in life by 50%, with more severe TBI associated with greater risk.[16] The link between TBI and Alzheimer disease is particularly important. Healthy axons contain amyloid precursor protein in high concentrations, and amyloid precursor protein markedly accumulates in damaged axons, most likely through impaired axonal transport, which in general leads to the production and accumulation of toxic proteins and peptides including amyloid-beta plaques and neurofibrillary tangles after TBI.[17] Other posttraumatic processes leading to brain network dysfunction and abnormal information processing also play a role in the generation of neurologic and neuropsychological impairment after TBI.[18]

Chronic traumatic encephalopathy, or dementia pugilistica, is a tauopathy caused by repeated TBI, most often sports concussion, that has recently emerged as a postmortem entity,[19] confirming the link between TBI and subsequent risk of neurodegenerative disease. Increased levels of alpha-synuclein protein, which is known to play an important role in the development of Parkinson disease and other neurodegenerative disorders, also have been detected in brain tissue samples from persons with a history of TBI.[20] In view of the probable role of TBI in the development of later neurodegenerative disease, it is not surprising that such injuries are reported to have chronic and long-term impact on brain functioning, including sleep-wake and circadian regulation.

Behavioral and Cognitive Consequences of Traumatic Brain Injury

After moderate and severe TBI, an alteration in the level of consciousness, followed by a period of delirium, confusion, agitation, and posttraumatic amnesia, generally is apparent when patients are in the awakening stage in the intensive care unit. Neurobehavioral impairments such as impulsivity, irritability, disinhibition, mutism, and apathy may be noted.[21] With regard to neuroimaging findings, the extent of functional and cognitive deficits observed in the post-acute period is highly variable among patients with TBI and depends on several factors, such as brain injury severity, location of focal lesions, severity of diffuse axonal injury, duration of posttraumatic amnesia, age, level of education, and preexisting conditions. Despite the variability of cognitive deficits, certain deficits are common, including arousal and alertness impairments, reduced information processing speed, impaired memory, executive dysfunctions, impaired language, and reduced self-awareness. Unfortunately, these impairments persist for longer than 1 year in 50% of patients with moderate and severe TBI, also may be observed in chronic mild TBI, and may lead to dependency in activities of daily living and problems with return to work or school.[22]

Biomarkers of Outcome after Traumatic Brain Injury

Several TBI studies have searched for sensitive and reliable markers of short- and long-term neurologic and functional outcome. The first group of studies supports that genetic factors influence functional outcome after TBI. Among the most-studied polymorphisms in the TBI population is the apolipoprotein (APOE) ε4 allele that was associated with poor outcome at 6 month after injury and with greater risk of later cognitive decline.[23] Preliminary evidence also suggests that functional polymorphisms of the brain-derived neurotropic factor (BDNF) and the catechol-*O*-methyltransferase (COMT) genes, involved, respectively, in brain

plasticity regulation and dopamine modulation, influence cognitive recovery after TBI.[24,25] Moreover, BDNF polymorphisms contribute to the regulation of slow wave sleep oscillations and hence non-rapid eye movement (NREM) sleep intensity in humans possibly explaining variability in SWD after TBI.[26] Other potential genetic biomarkers may potentially be related to circadian disorders. For instance, the expression of bone morphogenetic protein 6 (BMP6) coincides with melatonin levels.[27]

The second group of studies investigated proteomic markers of outcome. The serum level of the S100B protein is among the most extensively studied marker among the TBI patient population. It is known to predict unfavorable outcome in severe TBI[28] and also is sensitive and specific for intracranial lesions in mild TBI,[29] but changes in S100B serum concentration are not specific for TBI and can be observed in several other conditions, such as infection and with orthopedic injuries. Glial fibrillary acid protein (GFAP) is a protein specifically expressed in astrocytes and is a well-known marker of central nervous system pathology. In a study that included 94 patients with mild TBI, GFAP serum level at admission predicted return to work and functional outcome.[30] A study in mice suggested that GFAP expression and microglial cell activation are increased in the reticular thalamic nucleus preceding sleep disturbance, 4 weeks after TBI.[31] No sufficient evidence, however, is available on how these proteins might affect sleep-wake behavior after TBI.

PATHOPHYSIOLOGY OF POSTTRAUMATIC SLEEP-WAKE DISTURBANCES

Surveyed in this section are recent animal models that were specifically developed to improve the current understanding of posttraumatic sleep-wake patterns. Also reviewed are potential pathophysiologic causes, physiologic factors, and influences of comorbid conditions that may contribute to SWD onset or maintenance. Figure 25-1 presents an integrative model of potential contributors to posttraumatic SWD.

Experimental Models of Traumatic Brain Injury and Studies of Sleep-Wake Behavior in Animals

So far, only few animal studies have explored posttraumatic SWD. This limitation might be explained by at least two main reasons: apart from the fact that posttraumatic SWD have not been a focus of major interest for many years, combining trauma and electrophysiologic recordings in animals is challenging. In addition to the scarcity of systematic studies in animals, the few available studies applied different TBI models that lead to different pathologic mechanisms and to different lesions at cortical and subcortical levels. A summary of findings in animal models of posttraumatic SWD is presented in Table 25-1. Models of injury included fluid percussion and weight drop with acceleration and deceleration. Both mice and rats were studied, and outcomes were evaluated using different methods (electroencephalography [EEG]/electromyography [EMG] or automated locomotor rhythm analyses) and observed within different time windows.

Fluid Percussion

One group of investigators used a midline fluid percussion injury model in male mice.[32] In this model, craniotomy is performed, and then the insult is inflicted by application of a fluid pressure pulse to the dura. It has been shown that midline fluid percussion injury leads to bilateral cortical alterations associated with direct axial movement of the lower brainstem, along with changes in blood pressure, brief respiratory arrest, increased craniocerebral pressure, decreased cerebral perfusion pressure, reduced cerebral blood flow, and increased cerebral vascular resistance, and has therefore been interpreted as a good model for human TBI.[33] Midline fluid percussion injury led to a marked increase in sleep within the first 6 hours after trauma, as assessed by a noninvasive, piezoelectric cage system.[32] In a subsequent study and by applying the same methods, these investigators could not identify persistent SWD 2 to 5 weeks after injury.[34] EEG recordings were not available.

Lateral fluid percussion injury belongs to the most established and commonly used models to evaluate mixed focal and diffuse brain injury, because it is reproducible and recapitulates many injuries observed in humans.[35] EEG/EMG recordings after lateral fluid percussion injury in mice were performed by another group.[36] The behavioral data from days 8 to 12 after trauma indicated that TBI persistently impairs the ability to sustain wakefulness. The brain-injured mice exhibited more wake to NREM, NREM to wake, and rapid eye movement (REM) to wake transitions in comparison with sham injury control animals, indicating a higher fragmentation of behavioral states. In a recent study that used a similar protocol in rats, moderate TBI led to a decreased ability to maintain consolidated wake bouts during the active phase, and this observation was made for 6, 9, 19, and 29 days after injury.[37]

Controlled Cortical Impact Model

One study performing EEG/EMG recordings within the first 3 days after controlled cortical impact injury in mice reported a reduction in wakefulness and shorter wake bouts during the active period after trauma but not in control animals.[38] This TBI model permits better control over mechanical factors because a controlled impact is delivered to the intact dura by a strictly controlled compressed air-driven metallic piston, causing deformation of the underlying cortex.[33] It leads to diffuse axonal injury and other pathologic alterations in subcortical areas, cerebellar structures, midbrain, and brainstem.

Weight Drop Model

In a recent study,[39] a closed-head mouse TBI model using a weight drop system was applied in adult mice, compared with mice subjected to sham surgery. Sleep was recorded by EEG/EMG starting 14 to 16 hours after injury. Results showed that mild TBI decreased long bouts of wakefulness in the first 24 hours after mild injury, similar to findings reported independently by other studies,[38] suggesting an altered capacity to sustain wakefulness after TBI. In addition, some studies adopted a closed-impact acceleration TBI rodent model with a physical hit to the skull with subsequent acceleration-decelerations mechanisms, which also allowed for EEG/EMG recordings (Noain et al., unpublished results). Briefly summarized, a blow from a falling metal rod onto the exposed rat skull, which is placed on a thick foam pad, led to acceleration mechanisms similar to those in human TBI and to brain compression, with subsequent widespread histopathologic changes including diffuse axonal injury, particularly in the corpus callosum, midbrain, cerebral and cerebellar

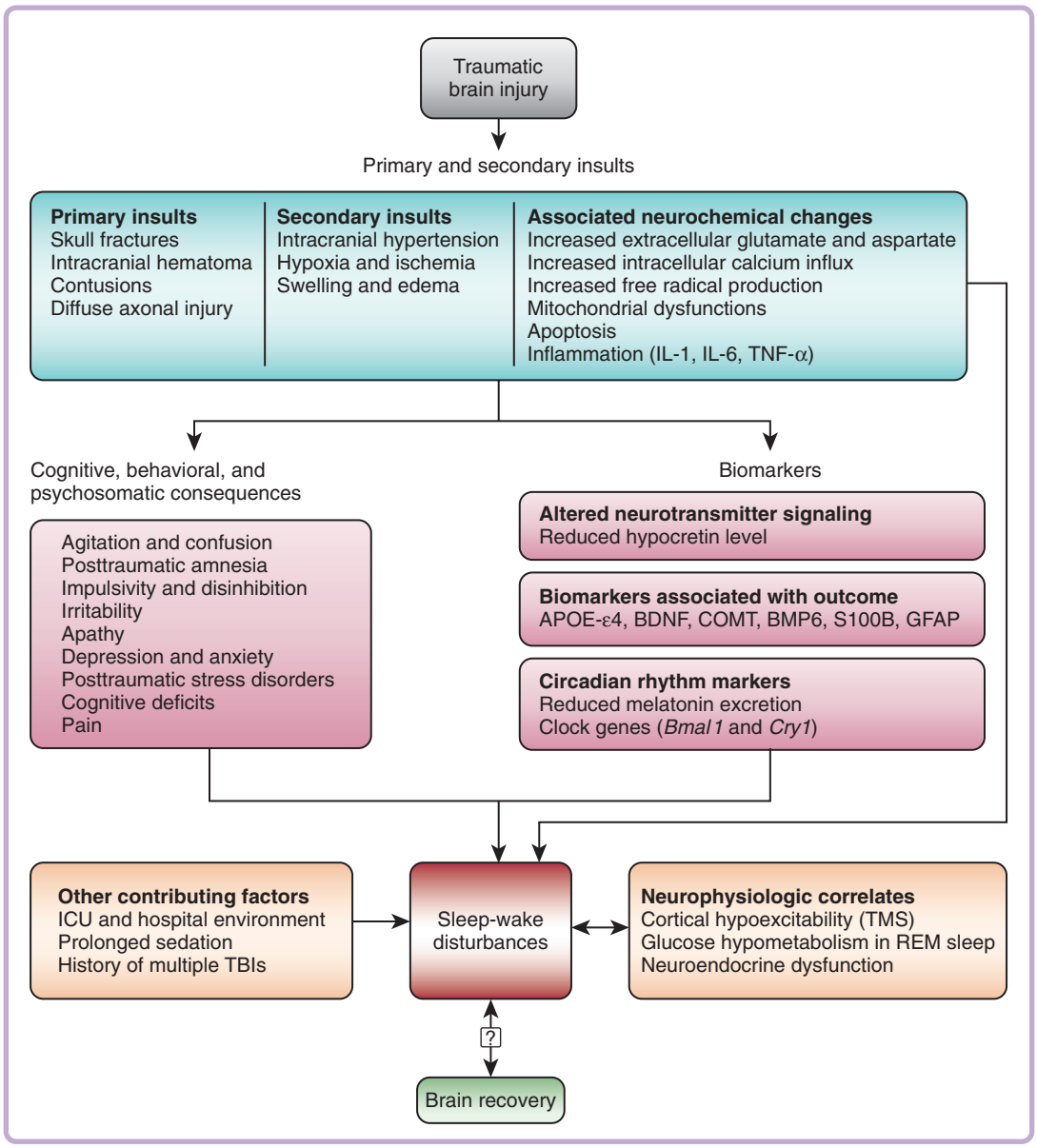


Figure 25-1 Integrative model of potential contributors to sleep-wake disturbance (SWD) after traumatic brain injury (TBI). All potential contributors to SWD after mild, moderate, or severe TBI are included. TBI causes primary and secondary insults that may directly influence sleep-wake patterns, but these injuries also have cognitive, behavioral, and/or psychosomatic consequences known to increase the risk of SWD. Several biomarkers have been associated with outcome, and an increased risk of SWD has been identified for certain of these markers. Other significant contributing factors are environment, medication, and history of multiple TBIs. SWDs have been associated with physiologic changes in patients with TBI, more specifically, cortical hypoexcitability, reduced glucose metabolism in rapid eye movement (REM) sleep, and neuroendocrine dysfunction. Posttraumatic SWDs possibly influence brain recovery after TBI. APOE, Apolipoprotein; BDNF, brain-derived neurotrophic factor; BMP6, bone morphogenetic protein 6; COMT, catechol-*O*-methyl transferase; GFAP, glial fibrillary acidic protein; ICU, intensive care unit; IL, interleukin; TMS, transcranial magnetic stimulation; TNF, tumor necrosis factor.

peduncules, and brainstem.[33] In the Noain et al. study, rats that were subjected to TBI exhibited an increase in NREM sleep at 1 month after injury, and in contrast with the latter study, sleep was less fragmented.

In conclusion, different models of rodent TBI have been used to examine posttraumatic SWD, both behaviorally and by means of EEG/EMG recordings. This heterogeneity of applied methods probably is responsible for the somewhat inconclusive results so far. Nevertheless, most studies found an increased amount of sleep and a reduction of the capability to maintain wakefulness in the first few days, but also 2 to 4 weeks after TBI. Issues to be investigated in the future include whether this observed increase in sleep after TBI is a mere reaction to the trauma and consecutive disturbed signaling within sleep-wake modulating pathways, or whether sleep is critical for recovery.

Table 25-1 Animal Models of Posttraumatic Sleep-Wake Disturbances

Study	TBI Model	Measure of Sleep and Wakefulness	Delay after TBI	Observations for Sleep-Wake Patterns	Other Observations
Row et al., 2014a[32] and 2014b[34]	Midline fluid percussion injury in mice	Piezoelectric cage	6 hours 2-5 weeks	Increase in sleep No change in sleep-wake patterns	Increased proinflammatory cytokine IL-1β
Lim et al., 2013[36]	Lateral fluid percussion injury in mice	EEG/EMG recording	8-12 days	Decreased wakefulness and more sleep-wake transitions	Decreased hypocretin neuronal activation during wakefulness but normal hypocretin neuron numbers
Boone et al., 2012[62]	Lateral fluid percussion injury in rats	Locomotor activity	2-50 hours	Disruption of circadian locomotor activity rhythms	Dysregulated expression of key circadian clock genes (*Bmal1, cry1*)
Skopin et al., 2014[37]	Lateral fluid percussion injury in rats	EEG/EMG recording	6, 19, and 29 days	Decreased ability to maintain consolidated wake bouts during the active phase	Reduced number of orexin-A positive neurons in the lateral hypothalamus
Willie et al., 2012[38]	Controlled cortical impact in mice	EEG/EMG recording	3 days	Reduction of wakefulness and shorter wake bouts	Depressed hypothalamic hypocretin levels (microdialysis)
Harza et al. 2014[31]	Controlled cortical impact in mice	Automated behavioral analysis system	1 month	Increased number of awakenings from sleep	Increase in reactive microglial cells in thalamic region and delayed reactive astrocytosis in the thalamic reticular nucleus
Sabir et al., 2015[39]	Weight drop in mice	EEG/EMG recording	24 hours	Decreased long bouts of wakefulness	Sleep deprivation decreased the expression of *Arc*, *EfnA3,* and *Homer1a* only in mice with mild TBI
Noain et al., unpublished	Weight drop in rats	EEG/EMG recording	1 month	Increased NREM sleep and less fragmented sleep	Fewer histaminergic neurons in the tuberomammillary nucleus
Petraglia et al., 2014[74]	Single and repetitive TBI, weight drop in mice	EEG/EMG recording	1 month	Reduction in NREM sleep, increase in NREM sleep fragmentation, and an increase in wake time over 24 hours in both groups	

EEG, Electroencephalogram; EMG, electromyogram; IL, interleukin; NREM, non-rapid eye movement; TBI, traumatic brain injury.

Inflammation Mediators from Animal Studies

Mediators of inflammation that are secreted as a result of TBI are important potential contributors to SWD. Acute inflammatory response occurring after TBI is characterized by an increase in interleukin 1 (IL-1), interleukin 6 (IL-6), and tumor necrosis factor alpha (TNF-α). Proinflammatory cytokines IL-1β and TNF-α are known to increase sleepiness, prolong duration of slow wave sleep, and increase EEG delta power.[40] TNF-α also is known to inhibit melatonin synthesis and hypocretin (orexin) activity in rats,[41,42] which can have a significant impact on sleep and wakefulness. To date, only one mouse study has suggested an association between proinflammatory cytokine IL-1β level and increasing sleep duration in the hours following a mild TBI.[32] The investigators argued that increased sleep in their TBI mice may result from the inflammatory response associated with the secondary injury, but this association needs to be confirmed in future studies.

Impaired Neurotransmitter Signaling in Animal Models of Traumatic Brain Injury

Brain structures, especially the hypothalamic, midbrain, and brainstem networks that are involved in sleep, wake, and circadian rhythmicity, can be impaired or damaged in a

proportion of patients with TBI, potentially explaining some forms of SWD observed in this population. In this line of thought, consecutive impairment of neurotransmitter signaling would contribute to altered regulation of behavioral states.

Some studies performed intracerebral microdialysis after controlled cortical impact in mice (as described earlier) and found that hypothalamic hypocretin levels were depressed.[38] Furthermore, the associations between hypocretin levels and wakefulness were diminished. Abnormal hypocretin dynamics were not associated with acute loss of hypocretin neurons but with hypothalamic astrogliosis. Lim and colleagues, who performed fluid percussion injury in mice (see earlier), measured c-Fos protein in the lateral hypothalamus, where the hypocretin neurons reside.[36] Mice with TBI exhibited significantly decreased hypocretin neuronal activation during wakefulness but, again, normal hypocretin neuron numbers, suggesting that nonfatal injury in these animals primarily affects hypocretin physiology rather than marked cell loss.

In addition, one study investigated the immunoreactivity of hypocretin-1 receptors after murine TBI, and the reporting authors found transiently increased immunoreactivity in the surrounding penumbra of the injury.[43] Again, all of these results suggest that that hypocretin and its receptors may play a role in behavioral abnormalities after TBI, but the exact nature of dysfunctional hypocretin signaling and its pathophysiologic impact remain elusive thus far.

As suggested by the experience of the Noain group of investigators, mentioned previously (Noain et al.; see earlier), additional observations in post-TBI rats point toward a critical role of the histamine system in pathologic sleep-wake behavior after trauma. Stereologic cell counts in these animals revealed fewer histaminergic neurons in the tuberomammillary nucleus, and the number of histamine-producing neurons was inversely correlated with the total amount of sleep per 24-hours period. These findings might be among the first initiatives for delineating the mechanisms behind sleep disturbances after TBI.

Impaired Neurotransmitter Signaling in Human Traumatic Brain Injury

For obvious reasons, human data on neurotransmitter signaling after TBI are sparse. The first hint toward dysfunctional hypocretin signaling after TBI appeared in the supplementary data of a publication on hypocretin levels in neurologic disorders, showing pathologically decreased hypocretin levels in the cerebrospinal fluid (CSF) of some brain-injured patients.[44] A systematic approach in prospectively assessed CSF from patients with TBI within the first 4 days after trauma confirmed a massive decrease in hypocretin levels, similar to that in narcolepsy patients.[45] Several months later, however, hypocretin levels mostly had recovered to normal (Figure 25-2).[46] Altogether, these CSF-based results suggested that impaired signaling of wake-promoting hypocretin neurons may at least contribute to decreased arousal after TBI.

A pilot study in four deceased patients with fatal TBI reported a significant 27% loss of hypocretin neurons in the hypothalamus.[47] Similarly, in 12 brains from patients who died of fatal TBI, a significant 21% loss of hypothalamic hypocretin neurons was documented.[48] The most marked cell loss (41%), however, involved the wake-maintaining histamine-producing neurons in the tuberomammillary nucleus. Again, these data suggest that damage to this susceptible and exposed posterior hypothalamic area might be involved in the SWD we often see after TBI.

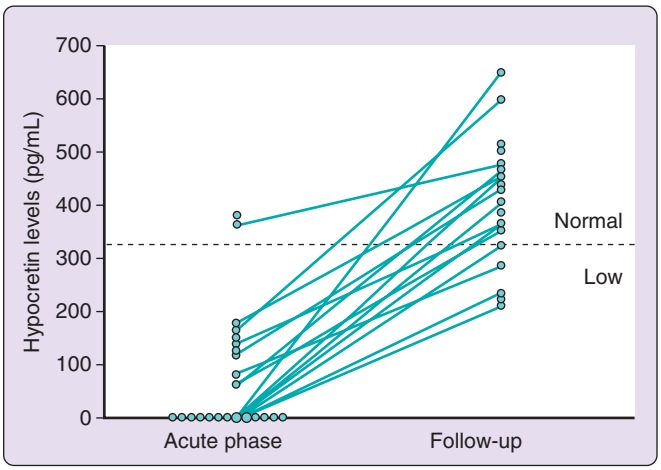


Figure 25-2 Cerebrospinal fluid hypocretin (orexin) levels in subjects with traumatic brain injury (TBI). Levels were measured 1 to 4 days after trauma (acute phase; $n = 27$) and 6 months later (follow-up; $n = 21$). *Lines* connect hypocretin values from the same patient ($n = 15$). Many patients had undetectable levels in the acute phase (plotted as 0). In all patients, levels recovered toward normal, but in some patients, they remained low. (From Baumann CR, Werth E, Stocker R, et al. Sleep-wake disturbances 6 months after traumatic brain injury: a prospective study. *Brain* 2007;130:1873-83.)

Cortical Hypoexcitability and Transcranial Magnetic Stimulation

As suggested earlier, lesions to midbrain or brainstem structures presumably could lead to posttraumatic SWD. One group of investigators used transcranial magnetic stimulation (TMS) in humans to determine whether SWDs after trauma are linked to pathologic excitability of the cerebral cortex.[49] The study included patients with TBI demonstrating excessive daytime sleepiness, fatigue, or increased sleep need and appropriate control subjects. Only in those patients with objective excessive daytime sleepiness (as confirmed by multiple sleep latency tests) were correlates of cortical hypoexcitability found. The investigators concluded that this finding on a cortical level might reflect the deficiency of the excitatory hypocretin system. In fact, earlier functional imaging studies showed that brain perfusion abnormalities, as assessed by single photon emission computed tomography (SPECT), are present predominantly in the midbrain structures and less so in cortical areas.[50]

Neuroimaging of Sleep

Functional neuroimaging comprises noninvasive techniques that allow exploring patterns of brain activation during wakefulness and sleep in humans. Although some studies using functional neuroimaging during sleep that included normal control subjects were published in past years, only one study was recently performed in persons with mild TBI.[51] That study included five consecutive in-laboratory polysomnographic recordings in 14 military veterans with a history of blast exposure and/or mild TBI (42.6 ± 26.9 months after injury) using [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET) during wakefulness, REM sleep, and NREM sleep. After controlling for effects of posttraumatic stress disorder, it was found that these patients exhibited

decreased glucose metabolism during wakefulness and REM sleep in the amygdala, hippocampus, parahippocampal gyrus, thalamus, insula uncus, culmen, visual association cortices, and midline medial frontal cortices when compared with control subjects (other veterans). Hypometabolism was found despite normal sleep architecture. These findings suggest that blast exposure and/or mild TBI may produce long-lasting neural effects that can be observed during both wakefulness and REM sleep, but further studies are needed to understand how these brain metabolic changes may explain persistent SWD.

Circadian Rhythms

One factor potentially contributing to SWD is the presence of alteration in the circadian timing system. Specifically, irregular circadian rhythms are known to be translated into increased daytime sleep, reduced nocturnal sleep, and fragmented sleep.[52] Circadian disruption may occurs when the main biologic clock of the hypothalamus is not synchronized to the 24-hour day and/or when it produces a circadian signal too weak to entrain the peripheral clocks of the brain and body. Short (acute phase) and long-term studies were conducted after TBI, and it is obvious that the paucity of literature necessitates caution in interpreting the following data.

Circadian Studies in Acute Traumatic Brain Injury

Thus far, only a few studies have investigated melatonin, cortisol, and body temperature rhythms, three well-known markers of circadian rhythms, in analgosedated patients with TBI while they were hospitalized in intensive care units. In a study performed in three patients with TBI and eight patients with other acute neurologic injuries, an absence of circadian rhythms was found for plasmatic melatonin, plasmatic cortisol, and body temperature.[53] When results were compared with those for critically ill patients without neurologic injury, alteration in circadian rhythms was greater in patients with brain lesions. These results suggest that brain injuries are in part responsible for altered circadian rhythms in intensive care units. In another study,[54] reduction in serum melatonin concentration was observed in addition to irregular circadian cycles in eight sedated patients with TBI, and these alterations were correlated with brain injury severity. Using microdialysis, an absence of cortisol circadian rhythm also was found in 10 patients with severe TBI under analgosedation.[55] Among mechanisms that may explain the absence of circadian rhythms in critically ill brain-injured patients is a cerebral lesion in the suprachiasmatic nucleus, because this region is known as the master clock that synchronizes most endogenous circadian rhythms.[56] Other possible causes include an altered pattern of light exposure (such as with too-little contrast between day and nighttime light exposure),[57] hypoxia,[58] and anesthesia/sedation.[53]

Circadian Studies in Post–Acute Traumatic Brain Injury

In a preliminary study performed in an intermediate care unit, actigraphic sleep-wake cycle and urinary melatonin circadian rhythm (measured hourly for 26 consecutive hours) were evaluated in nine nonsedated patients with severe TBI.[59] Nocturnal 6-sulfatoxymelatonin excretion showed an increased nighttime concentration (13.0 ± 17.9 ng/mL) beyond daytime concentration peaks (3.0 ± 5.2 ng/mL). Despite the presence of melatonin circadian rhythm, only two of the nine patients exhibited a consolidated sleep-wake cycle, which suggests that a nonconsolidated sleep-wake cycle is not due to an absence of melatonin circadian rhythm. Such study design is an example of a method that could be used for future investigations of circadian rhythm alterations in acute and postacute TBI.

Circadian Studies in Chronic Traumatic Brain Injury

Several case studies reported circadian rhythm disturbances associated with chronic TBI (see Chapter 99 for more details), and two studies specifically investigated salivary dim-light melatonin onset in patients with chronic TBI. The first study reported no group difference in the timing of nighttime melatonin onset, but a high variability among the subjects was noted.[60] A second study conducted by the same investigators found lower levels of evening melatonin production in patients with mild to severe TBI than in control subjects, but again, no changes were found for dim-light melatonin onset[61] (Figure 25-3). Because saliva melatonin concentration was measured only for 6.5 consecutive hours during the evening (i.e., from 6:00 PM to 12:30 AM) in this latter study, it was not possible to determine whether a phase delay of melatonin secretion could explain the reduced melatonin production in the TBI group.

Clock Genes

More recently, it has been hypothesized that circadian rhythm disruption after TBI is mediated by changes in expression of clock genes in the suprachiasmatic nucleus and hippocampus. It was found that in rats with fluid percussion–induced TBI, in comparison with sham-operated ones, the circadian expression of two key clock genes (i.e., *Bmal1* and *Cry1*) was altered in both the suprachiasmatic nucleus and the hippocampus.[62] In this study, altered daily rhythm of locomotor activity coincided with the deregulation of clock genes. The investigators suggested that a disturbance in the transcriptional-translation

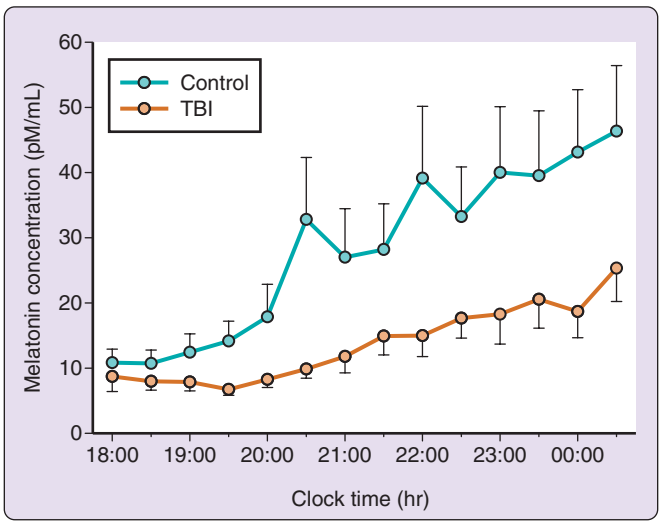


Figure 25-3 Nighttime salivary melatonin levels in patients with traumatic brain injury (TBI) and control subjects. Salivary melatonin levels of TBI (*orange*) and control group (*green*) from 6:00 PM to 12:30 AM. Data are presented as mean ± standard error. Salivary samples were collected every half-hour. The TBI group had lower melatonin concentrations across the sampling period compared with the control group. (From Shekleton JA, Parcell DL, Redman JR, et al. Sleep disturbance and melatonin levels following traumatic brain injury. *Neurology* 2010;74:1732-8.)

feedback loops that modulate circadian timing could be induced by TBI. Obviously, the role of these genes needs to be confirmed in humans (see Chapters 26 to 31 for more information on genes in relation to sleep and circadian rhythms).

Neuroendocrine Dysfunctions

Neuroendocrine dysfunctions, mostly hypopituitarism, are common in patients after TBI. The prevalence is estimated at 15% to 68% in the acute phase of TBI, with the highest prevalence among persons with moderate and severe TBI, but tends to decrease over the first year after the injury.[63] Chronic hypopituitarism may lead to growth hormone and gonadotropin deficiency and to hypothyroidism. These alterations in neuroendocrine functioning are important to consider in the pathophysiology of SWD after a TBI, because bidirectional interactions between hormonal secretion and sleep are well documented.[64] Among the most important hormonal disturbances in this context is the growth hormone deficiency that was found to be closely associated with increased slow wave sleep and EEG delta power but with worsened subjective sleep quality and sleepiness. To date, few studies have investigated neuroendocrine changes in relation to sleep alterations and fatigue after TBI, and no clear evidence exists for a role of hypopituitarism and SWD or fatigue in the TBI population. In one study, patients with TBI reporting fatigue exhibited an overall higher prevalence of growth hormone deficiency than that for patients not reporting fatigue.[65] In another study, however, no correlation was observed between abnormal endocrine function and fatigue in a sample of 119 patients with TBI tested at least 1 year after injury.[66] In these two studies, sleep was assessed with questionnaires, and no objective measures of sleep were performed. Further studies are indicated to investigate whether neuroendocrine dysfunctions among patients with TBI explain not only subjective evaluation of sleep quality but also abnormal polysomnographic findings in this population.

Effects of Prolonged Sedation

Sedative and analgesic agents are commonly administered to patients with severe TBI admitted to the intensive care unit for prolonged periods regardless of the patient's specific circadian rhythm. These agents are used to prevent agitation, facilitate mechanical ventilation, and reduce pain and also may improve intracranial pressure and cerebral perfusion pressure. Still not clear, however, is whether sedation and analgesia share the restorative properties of sleep. Some studies have shown that prolonged sedation with propofol does not result in sleep deprivation: After a 12-hour infusion of propofol followed by withdrawal, rats did not show the typical increase in slow wave activity that occurs in response to sleep deprivation.[67] Controversial results, however, were reported for isoflurane anesthesia and its recovery effect.[68,69] In any case, sedation and analgesia have adverse effects on circadian rhythms, particularly when these agents are not administered during biologic night—that is, when they are given during the day, they impair the sleep-wake cycle mechanism more potently than when they are administered during the night.[70] How sedative and analgesic agents interact with an injured brain to promote the development of sleep and circadian rhythm disorders is still not well understood, but one study reported greater alteration in circadian rhythms among patients with neurologic injury (traumatic and nontraumatic injuries such as stroke) under analgosedation than among those without neurologic injury.[53] For more information on mechanism and effects of opioid analgesia on sleep, see Chapter 23.

Repetitive Traumatic Brain Injury

In many people with TBI, a presumed underlying pathologic process involves multiple lifetime traumatic injuries, particularly among military personnel and contact sports athletes. Repetitive acceleration and deceleration forces directed to the brain may cause chronic traumatic encephalopathy, but the underlying pathophysiology is unknown.[71] Increasing evidence indicates that repetitive TBI accentuates residual symptoms; for instance, multiple TBIs have been found to be associated with a higher risk for suicide.[72] Using patient self-report measures and clinical interviews, a recent study showed that repetitive TBI increases the severity of posttraumatic SWD, namely, insomnia.[73]

To characterize more fully the clinical entity of chronic traumatic encephalopathy and its consequences, a novel mouse model was introduced whereby a series of 6 concussive impacts were delivered daily, for 7 days. This model allowed the comparison of groups of single-injury and sham-injury control animals.[74] EEG/EMG was recorded at 1 month after injury for a period of 24 hours; animas in the single- and repetitive-TBI groups exhibited a reduction in NREM sleep, more NREM sleep fragmentation, and an increase in wake time in comparison with control animals. Moreover, EEG spectral analysis during NREM sleep demonstrated reduced power for very low frequencies (1 to 2 Hz) and increased power in the theta band in single- and repetitive-TBI groups compared with the control group. Still needed are larger studies to support these human and animal observations that repeated trauma leads to a higher SWD burden, not only in relation to insomnia but also in association with other SWDs.

Pain

A case–control study showed that both pain and SWD are prevalent after TBI, and that pain was strongly associated with insomnia.[75] Anxiety, depression, and pain influence sleep quality; however, only anxiety and pain seem to explain 32% of the variance in the sleep quality assessments.[76] Conversely, after multiple stepwise analyses controlling for many factors, another study reported that depression mainly accounts for the poor sleep quality complaints as assessed by the Pittsburgh Sleep Quality Index.[77] Such discrepancy in findings is not surprising, because chronic pain is multifactorial; it is a frequent comorbid accompaniment of TBI and is reported to be independent of posttraumatic stress disorder and depression in brain-injured patients.[78] The interaction between pain and mood on the one hand and long-term outcome on the other deserves studies with larger sample sizes, taking into account the heterogeneity of the population under investigation. It is likely that many factors in addition to pain contribute to the different patterns of SWD within the first weeks after trauma, with more insomnia at the acute phase and a higher prevalence of disorders of arousal such as sleepiness or increased sleep need in the chronic post-TBI phase.[79]

In view of the probable impact of pain on sleep quality, one group of researchers aimed to identify how pain affects sleep quality and sleep microstructure using quantitative and high-temporal-density EEG analysis.[77] Compared with patients with TBI not afflicted by pain, those with pain

showed increased rapid EEG frequency bands (9 to 50 Hz), mostly during REM sleep, and beta bands (16 to 30 Hz) in NREM sleep. The specificity of this phenomenon remains to be demonstrated, because intrusion of fast EEG activity in sleep of TBI patients with pain is not an exclusive phenomenon. Such differences also might reflect altered nociceptive processing at the central nervous system level, exacerbation of hypervigilance, and presleep cognitive arousal, as seen in the sleep of chronic pain patients.

Psychiatric Comorbid Illness

Approximately 65% of patients with TBI suffer from at least one psychiatric disorder, with depression, anxiety, and posttraumatic stress disorders being among the most common.[80] The associations between these psychiatric disorders and sleep disturbances, mostly insomnia, have been well described in the non-TBI population.[81] In agreement with this strong association, sleep-wake disturbances are included among the diagnostic features of major depressive disorder, generalized anxiety disorder, and posttraumatic stress disorder.[82]

For instance, most studies that investigated the prevalence of comorbid SWD and psychiatric syndromes among patients with TBI consistently found a higher burden of psychiatric disorders among those with sleep complaints than among those without sleep complaints. For example, in two different cohort studies of chronic TBI, depression and anxiety symptoms were found to be among the variables that significantly contributed to the prediction of the presence of insomnia syndrome.[83,84]

Still poorly understood, however, is how psychiatric disorders interact with TBI in the emergence and/or the persistence of SWD. Recent studies point to the predictive role of SWDs in the emergence of mental health problems. One retrospective study in 443 patients with mild TBI showed that reporting sleep complaints at 10 days after injury is associated with a 9.9- and 6.3-fold increased risk of feeling depressed at 10 days and 6 weeks, respectively.[85] Similarly, it was observed that onset of sleep disturbances within 3 months of injury predicted emergence of symptoms of depression and anxiety 12 months after TBI.[86] The role of sleep disturbances in the development of posttraumatic stress disorder and depression was recently confirmed by a large cohort study of 29,640 U.S. Navy and Marine Corps men, in whom sleep problems mediated 26% of TBI's effect on development of posttraumatic stress disorder and 41% of its effect on development of depression.[87] The specific mechanism underlying the link between sleep disturbances and the development of depression, anxiety, or posttraumatic stress disorder remains to be investigated.

SLEEP AND BRAIN RECOVERY

The role of sleep in memory and neural plasticity is now increasingly recognized.[88] In fact, sleep is known to optimize the consolidation of memory by strengthening new synapses. More recently, sleep has been shown to have a crucial role in neurogenesis.[89] Indeed, chronic sleep deprivation or fragmentation (>3 days) in rodents was associated with a 30% to 80% reduction in hippocampal cell proliferation and with a decrease in the percentage of cells that mature and develop into adult neurons.[90,91]

Extrapolation of the significance of these findings to human TBI suggest that chronically restricted or disrupted sleep in persons with acute TBI should reduce learning abilities, cerebral plasticity, and generation of new neurons in their brain, particularly in the hippocampus. This association is particularly important in the context of brain injury, where recovery is dependent specifically on new learning, cerebral plasticity, and neurogenesis. In humans, SWDs have been associated with worse functional outcome in the acute stage of TBI, whereby consolidated sleep-wake cycle predicted resolution of posttraumatic amnesia,[92,93] and in the chronic stage of moderate and severe TBI.[94]

The question remains whether sleep abnormalities in the acute phase after TBI influence recovery, on both a clinical and a neuropathologic level. In a recent animal study, the impact of acute 6-hour sleep disruption after TBI was investigated in adult mice; no effect on recovery of neurologic and cognitive functions was found.[95] This study suggests that short-duration sleep disruption after relatively mild TBI does not affect functional outcome. Other studies, however, found that 24 hours of sleep deprivation in rats before TBI results in faster recovery.[96] Also, 24 hours of sleep deprivation after TBI was reported to reduce morphologic damage and to improve behavioral recovery in rats.[97] Furthermore, unpublished evidence suggests that both sleep induction and sleep deprivation immediately after TBI in rats improve functional recovery and histologic outcome (Morawska et al., unpublished results).

In an animal study, the reverse association between sleep disruption and recovery was found.[39] Genes associated with sleep regulation and neuronal plasticity were investigated in adult mice with mild TBI before and after two consecutive 6-hour periods of sleep deprivation during the light period of the day-night cycle. Quantification of messenger RNA levels was performed in the cerebral cortex and the hippocampus, and in a region incorporating the thalamus and the hypothalamus. In the hippocampus, sleep deprivation decreased the expression of *Arc* and *EfnA3* only in mice with mild TBI. In the thalamus-hypothalamus, sleep deprivation decreased *Homer1a* only in mice with mild TBI. These results point to an accentuated effect of sleep deprivation on sleep- or plasticity-related genes after mild TBI.

In addition, an EEG and magnetoencephalographic study in humans showed that sleep spindles may reflect recovery of consciousness and cognitive function after diffuse axonal injury, suggesting that some aspects of sleep—that is, pharmacologically induced sleep or sleep rebound—might underlie neuronal plasticity and clinical recovery after TBI.[98]

Altogether, and despite contradictory results in the previous literature, the possibility remains that some aspects of sleep may potentially influence recovery after TBI, and the challenge is to define the mechanism and how such an improvement would be mediated.

CLINICAL PEARLS

- Both TBI and comorbid conditions, such as pain and psychiatric disorders, are potential causes of posttraumatic SWD in the acute and chronic phases of injury.
- Acute SWD may adversely influence brain recovery after TBI, as well as contributing to comorbid posttraumatic symptoms.
- Identifying mechanisms associated with the pathogenesis of posttraumatic SWD and factors that contribute to the persistence of TBI can be expected to lead to improved management in the chronic phase.

SUMMARY

Fatigue, excessive daytime sleepiness, and pleiosomnia are common SWDs after TBI. Recently developed animal models have conferred new understanding of posttraumatic SWD; the translation of animal findings to human TBI suggests that specific factors are associated with the pathophysiology of TBI and contribute to emergence of SWDs. Consistent with animal models, an important decrease in hypocretin occurs in acute human TBI, which is reversed in the months after the injury. Prolonged use of sedative and analgesic in the intensive care unit, altered circadian timing, comorbid pain, anxiety, depression, and posttraumatic stress disorder are among the factors associated with the emergence and/or persistence of SWD after a TBI. How acute sleep disturbances and sleep disruption improve or impede brain recovery after TBI needs further investigation, but interesting preliminary observations suggest a role for an accentuated effect of sleep deprivation on sleep and/or the plasticity related to some specific genes after TBI. The differences observed between acute and chronic phases suggest that studies should be conducted within a longer time frame to elucidate TBI and SWD pathophysiology and derive best management strategies.

Selected Readings

Baumann CR, Bassetti CL, Valko PO, et al. Loss of hypocretin (orexin) neurons with traumatic brain injury. *Ann Neurol* 2009;**66**(4):555–9.

Baumann CR, Werth E, Stocker R, et al. Sleep-wake disturbances 6 months after traumatic brain injury: a prospective study. *Brain* 2007;**130**(Pt 7):1873–83.

Duclos C, Dumont M, Blais H, et al. Rest-activity cycle disturbances in the acute phase of moderate to severe traumatic brain injury. *Neurorehabil Neural Repair* 2013;**28**(5):472–82.

Gilbert KS, Kark SM, Gehrman P, Bogdanova Y. Sleep disturbances, TBI and PTSD: implications for treatment and recovery. *Clin Psychol Rev* 2015;**40**:195–212.

Khoury S, Chouchou F, Amzica F, et al. Rapid EEG activity during sleep dominates in mild traumatic brain injury patients with acute pain. *J Neurotrauma* 2013;**30**(8):633–41.

Nardone R, Bergmann J, Kunz A, et al. Cortical excitability changes in patients with sleep-wake disturbances after traumatic brain injury. *J Neurotrauma* 2011;**28**:1165–71.

Paul T, Lemmer B. Disturbance of circadian rhythms in analgosedated intensive care unit patients with and without craniocerebral injury. *Chronobiol Int* 2007;**24**(1):45–61.

Rowe RK, Striz M, Bachstetter AD, et al. Diffuse brain injury induces acute post-traumatic sleep. *PLoS ONE* 2014;**9**:e82507.

Sabir M, Gaudreault PO, Freyburger M, et al. Impact of traumatic brain injury on sleep structure, electrocorticographic activity and transcriptome in mice. *Brain Behav and Immun* 2015;**47**:118–30.

Shekleton JA, Parcell DL, Redman JR, et al. Sleep disturbance and melatonin levels following traumatic brain injury. *Neurology* 2010;**74**(21):1732–8.

Stocker RPJ, Cieply MA, Paul B, et al. Combat-related blast exposure and traumatic brain injury influence brain glucose metabolism during REM sleep in military veterans. *Neuroimage* 2014;**99**:207–14.

Vermaelen J, Greiffenstein P, deBoisblanc BP. Sleep in traumatic brain injury. *Crit Care Clin* 2015;**31**:551–61.

A complete reference list can be found online at ExpertConsult.com.

Section 4

Genetics and Genomic Basis of Sleep

Chapter 26

Introduction: Genetics and Genomics of Sleep

Fred W. Turek; Ravi Allada

A section on the "Genetics of Sleep" was an important addition to the preceding edition of the *Principles and Practices of Sleep Medicine,* and inclusion of this new topic was heralded as signifying that the genetic control of the sleep-wake cycle had become an active area for sleep research, both for characterizing regulatory mechanisms underlying the control of the sleep-wake cycle and for elucidating the function(s) of sleep. The inclusion of the term "genomics" in the title of this section in the sixth edition of the book is a heralding event as well, in confirming that sleep researchers are beginning to take a systemwide genomic approach for investigating this interplay of genes and gene networks regulating the multiple phenotypes or traits that constitute the sleep-wake cycle. Indeed, as genetic and genomic approaches are being used for the study of sleep in diverse species from flies to mice to humans (see Chapters 28 to 30), the evolutionary significance of the many functions of sleep that have evolved over time is becoming a tractable subject for research as many investigators are bringing the tools of genetics and genomics to the sleep field. The complexity of the sleep-wake phenotypes and the difficulty in collecting phenotypic data on a large enough number of animals or humans to begin to unravel genetic mechanisms have been partly responsible for why few comprehensive attempts have been made to identify "sleep genes" beyond the circadian clock genes regulating the timing of sleep (see Chapter 27).

The focus on genetic and genomic approaches to sleep is predicated in large part on the notion that the mechanisms of sleep regulation are conserved from simple organisms highly amenable to genetic analysis such as the fruit fly and zebrafish to conventional rodent models and even humans (see Chapters 28 to 30). The best example of such evolutionary conservation comes from the discovery of the core circadian clock genes that regulate the diurnal sleep-wake cycle, as well as most, if not all, 24-hour behavioral, physiologic, and cellular rhythms. These core clock genes were first identified in fruit flies, employing mutagenesis and the screening of thousands of flies for mutant phenotypes, eventually leading to the finding of the same genes in mice and humans. Indeed, similar screens in rodents leveraging the simplicity and ease of monitoring a representative "output rhythm" of the central circadian clock from literally thousands of rodents in a single laboratory, such as the precise rhythm of locomotor activity or wheel running in rodents, also was a major factor in uncovering conserved molecular transcriptional-translational feedback loops that give rise to 24-hour output signals (see Chapter 27). The discovery of the core clock genes controlling the timing of sleep-wake cycles represents one of the great triumphs of behavioral genetics and serves as a road map to uncover the mechanisms of sleep homeostasis. The relatively recent search for sleep genes in flies has been an active area of study for only the past decade or so, and this approach is uncovering new sleep-related genes with clear mammalian orthologs.

In addition, substantial evidence has now accumulated demonstrating that deletion or mutations in many canonical circadian clock genes can lead to fundamental changes in other sleep-wake traits including the amount of sleep and the response to sleep deprivation.[1] The recent finding that different alleles of a core circadian gene, *per,* first identified in flies, can affect the homeostatic response to sleep deprivation and/or the amount of slow wave sleep (Chapter 30) argues that

uncovering sleep genes in flies will lead directly to changes in sleep-wake traits in mammals. As discussed in Chapter 29, on sleep genetics in rodents, one can argue that many circadian clock genes are also "sleep genes." As noted by Dijk and Landolt (Chapter 30), sleep is a rich phenotype that can be broken down into a wide variety of sleep-wake traits based on the electroencephalogram and electromyogram. Furthermore, the "genetic landscape" for regulating multiple sleep-wake traits clearly is going to involve hundreds of genes and integrated molecular neurobiologic networks.[2,3] Although early work by Valtex and colleagues using inbred strains of mice[4] and early human twin studies[5] provided considerable evidence of a strong genetic basis for some sleep-wake traits, little has been done to unravel the complex network of genetic interactions that must underlie this universal behavior in mammals. The fact that the environment and the subject's own volition also can have major effects on sleep-wake traits, particularly sleep duration in humans, also makes it difficult to uncover the underlying genetic control mechanisms. Indeed, although a large number of genome-wide association studies in humans have identified multiple genetic loci and genes involved in the regulation of a wide variety of physiologic systems and disease states, only recently has this approach been used to uncover sleep-wake genes (Chapters 30 and 31). Franken and colleagues pioneered the use of quantitative trait loci in recombinant inbred mouse strains, which has led to the identification of a small number of genes that are associated with specific sleep-wake properties[6] (see Chapter 27). More recently, the first attempt to record sleep in a large genetically segregating population of mice revealed considerable complexity to the genetic landscape for multiple sleep-wake traits.[2,3] Uncovering these loci and elucidating how gene-environment interactions contribute to different sleep-wake states are expected to not only reveal the molecular events underlying the sleep-wake cycle but also identify new targets for drug discovery. New therapies based on the genetic control of sleep may be particularly important for treating both genetically and environmentally based disorders of sleep (Chapter 31).[7]

A complete reference list can be found online at ExpertConsult.com.

Chapter
27

Genetics and Genomics of Circadian Clocks

Martha Hotz Vitaterna; Fred W. Turek; Peng Jiang

Chapter Highlights

- Circadian (near-24-hour) rhythms can be produced by individual mammalian cells in a self-sustaining manner. These rhythmic patterns result from coordinated daily oscillations in the transcription and translation of key clock component genes. The positive elements CLOCK and BMAL1 regulate transcription of the *Per* and *Cry* genes, forming the core feedback loop of the timekeeping mechanism.
- Interactions between these genes and their products, as well as other proteins, tune the clock and link it to other cellular pathways. The genetic clock also regulates a tissue-specific circadian program of gene expression.
- Rhythms in histone acetylation/deacetylation, chromatin modifications, and posttranslational changes are beginning to be explored, adding epigenetic elements to the circadian clock network.
- Signals from the suprachiasmatic nucleus and from behavioral states (sleep/wake or feed/fast) can also influence clocks in peripheral tissues.

Over the past three decades, remarkable progress has been made in elucidating the molecular substrates that underlie the generation of 24-hour rhythms in mammals. A major finding that has emerged over the past decade is that most cells and tissues of the body contain and express the core 24-hour molecular clock mechanism. Although normally coordinated, individual tissues and even cells are capable of producing sustained rhythms in isolation. These rhythms result from oscillatory expression of a core set of interrelated circadian genes. This chapter describes the genes expressed in cells of the hypothalamic suprachiasmatic nucleus and other oscillators and reviews the current understanding of the roles they play in this daily rhythmicity. In addition, in view of the homology of mammalian clock genes with those in the fruit fly, where appropriate, a discussion of the discovery of fly and mammalian genes is provided.

THE MAMMALIAN CELLULAR CIRCADIAN CLOCK

Several lines of evidence pointed to the suprachiasmatic nucleus (SCN) as the site of the master circadian "pacemaker," beginning in the 1970s. (For a discussion of central and peripheral circadian clocks see Chapter 38.) Destruction of the SCN abolishes circadian oscillations in the plasma concentration of cortisol[1] and in locomotion and drinking.[2] These oscillations are independent of inputs from the eye,[1] although an autonomous circadian clock had been demonstrated to exist within the eye that controls, among other functions, the shedding of rod outer segment disks.[3] Normal circadian rhythms can be restored to an SCN-lesioned animal by transplantation of fetal SCN tissue, but not by transplantation of fetal tissue from other regions of the brain.[4] Transplant into an SCN-lesioned animal of fetal tissue from the SCN of a circadian-mutant animal confers the short period of the donor,[5] indicating that the properties of the rhythm are determined by the SCN, rather than by other tissues or brain regions. Thus several lines of evidence point to the SCN as the site driving or controlling circadian behavior for mammals. More recent studies of rhythms in gene expression indicated that persistent rhythms can be observed in tissues throughout the organism, even in tissue explants kept in culture for extended periods.[6,7] The phase of these peripheral tissue rhythms differs from that of the SCN but nonetheless appears to be coordinated by the SCN. In SCN-lesioned animals, these peripheral rhythms persist but no longer exhibit the consistent phase seen in intact animals.[7] Some environmental manipulations, such as temperature cycles or restricted feeding, can alter the phase of peripheral rhythms.[8,9] In addition, studies confirm the presence of oscillations in gene expression throughout the body, with different phases in different tissues.[10-13] Loss of many of these rhythms is reported with SCN lesions. Of note, however, these studies cannot discriminate between a loss of rhythmicity by individual cells and a loss of synchronicity among the cells. Thus the roles of the SCN and of the peripheral oscillators in the mammalian circadian system continue to be defined.

CIRCADIAN CLOCK PROPERTIES AND CLOCK GENES

More than half a century ago, the formal properties of the circadian "clock" function had been well characterized. Among the 16 "empirical generalizations about circadian rhythms" defined by Colin Pittendrigh[14] were the following: Circadian rhythms are ubiquitous in living systems; they are endogenous; they are innate; they are not learned from or impressed by the environment; they occur autonomously at both cellular and whole-organism levels of organization; the free-running period of circadian rhythms is so slightly temperature-dependent that it is proper to emphasize its near-independence

of temperature; and the rhythms are surprisingly intractable to chemical perturbation. These six observations already suggest what is now definitely known to be the case: A cell-autonomous program of gene expression makes up the mammalian circadian "clock" that produces these rhythms.

How do individual cells generate rhythmic activity with a period of approximately 1 day? Many pacemaker neurons generate oscillatory activity, such as rhythmic patterns of action potentials, and these relatively rapid oscillations can be explained by the concerted action of a small number of ion channels. The much slower oscillations of the individual SCN neurons, however, are not likely to involve the same mechanisms. Pittendrigh's observations that circadian rhythms are not sped up or slowed down by changes in temperature and that they are relatively impervious to chemical perturbations similarly argue against such a neuronal mechanism underlying the generation of 24-hour oscillations. In fact, the finding that nonneuronal tissues (Pittendrigh's cell autonomy) can produce sustained circadian rhythms, as well as the prevalence of circadian rhythms in plants and unicellular organisms (Pittendrigh's ubiquity), argue against a neural process underlying circadian rhythm generation.

Indeed, it appears that the synthesis of proteins by each oscillatory cell is central to the mechanism for the generation of 24-hour rhythms. The initial evidence for this role is that application of protein synthesis inhibitors in the region of the SCN shifts the circadian phase of activity of animals by an amount and in a direction that depends on the time at which the inhibition is imposed.[15,16] A similar shift in the phase of vasopressin release from explanted SCN also results from inhibition of protein synthesis.[17] Thus gene expression is central to the generation of circadian oscillations.

As discussed in detail further on, gene expression profile studies, in which expression levels are sampled at regular time points in constant darkness (free-running conditions) focusing on the SCN and on peripheral tissues, reveal that approximately one third of the transcriptome is rhythmically expressed, even in peripheral tissues.[18-21] Reporter gene constructs combining genes known to be rhythmically expressed with luciferase (thereby allowing visualization of expression in culture by means of the luciferase-induced glow) demonstrated in rats and in mice that peripheral tissues are capable of self-sustained rhythms.[6,7] With so many genes exhibiting circadian expression, and with competent oscillators present in such a variety of tissues, one cannot assume that either rhythmic expression or expression in the SCN constitutes a valid criterion for a "clock gene." Identification of which genes are central to the generation and maintenance of circadian cycles thus represents a challenge.

A potential solution to the challenge of identification of clock genes (versus clock-controlled genes) is to refocus attention on the formal properties of the circadian system. Arnold Eskin and others promoted this conceptual framework in the late 1970s by characterizing rhythm properties as arising from the input pathway, the clock itself, or the output pathway[22] (Figure 27-1). As articulated for pharmacologic approaches (but equally valid for genetic ones), manipulations that produce phase-dependent shifts in the rhythm (a phase response curve), or changes in the phase response curve or the free-running period, are likely to be affecting the clock, or at least not affecting an output process.[23] Applying this logic, a genetic perturbation that alters the free-running period in constant darkness, or the phase response curve to light pulses, or the persistence of rhythmicity in constant conditions may be likely to be a perturbation in a clock gene, although no one alteration alone is necessarily a clock change (as opposed to input or output).

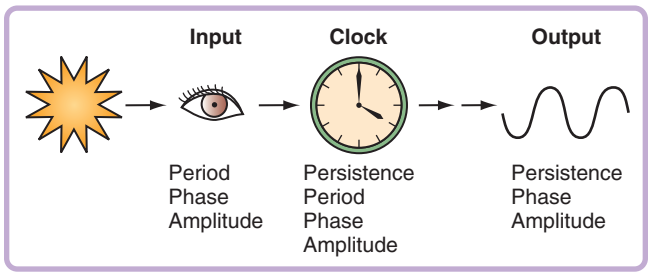


Figure 27-1 Classical View of Circadian System and Circadian Clock Properties. At minimum, the circadian clock system would have an input pathway by which entraining signals are received (light in this example), a clock mechanism, and output pathways. No single property of observed circadian rhythms is necessarily determined by the clock mechanism; these properties may be affected by changes in the input or output. However, when a mutation is observed to affect multiple properties of circadian rhythms, then that genetic change may most parsimoniously be attributed to a change in the clock mechanism itself.

Zatz[24] and Aronson[25] and their colleagues proposed more restrictive criteria: Null mutations should abolish rhythmicity; the gene's protein product level or activity should oscillate and be reset by light pulses; changes in amount or activity should result in phase shifts; and prevention of oscillation of protein levels/activity should result in loss of rhythmicity.[24] In the parlance of the field, these criteria would define a "state variable"—a rate-limiting element that itself defines the phase of the core oscillation. A self-sustaining clock would require at least two state variables,[26] although more clearly are possible. To date, no single gene in the mammalian system has satisfied all of the criteria for a state variable. Indeed, a hallmark of the mammalian circadian clock seems to be the multiple homologs of many genes that appear to play distinct but partially redundant roles. It may thus be that so-called state variable status is actually shared by related groups of genes in the mammalian system.

POSITIVE ELEMENTS

Clock

In the early 1990s, no genes in mammals had been identified as even possible candidate circadian clock genes, leading us to undertake a mutagenesis and screening in mice in an effort to identify mammalian circadian clock genes. For this, we used the C57BL/6J mouse strain, in which wild-type mice show robust entrainment to a light-dark cycle and exhibit a circadian period between 23.6 and 23.8 hours under free-running conditions in constant darkness (DD). In a screen for mutations of more than 300 progeny of mutagen-treated mice, one animal was found to have a free-running period of approximately 24.8 hours—more than 6 standard deviations longer than the mean.[27] In the homozygous condition, this mutation results in a dramatic lengthening of the period to approximately 28 hours, which usually is followed by the eventual loss of circadian rhythmicity (i.e., arrhythmicity) after 1 to 3 weeks in DD. The affected gene was mapped to mouse chromosome 5 and named *Clock*.[27,28] The *Clock* gene was cloned using a

combination of genetic rescue and positional cloning techniques. *Clock/Clock* mutant mice were phenotypically rescued by a bacterial artificial chromosome (BAC) transgene that contained the *Clock* gene, allowing for functional identification of the gene.[29] The *Clock* gene encodes a transcriptional regulatory protein comprising a basic helix-loop-helix DNA-binding domain, a PAS dimerization domain, and a Q-rich transactivation domain. The mutant form of the CLOCK protein (CLOCK Δ19) lacks a portion of the activation domain found in wild-type protein; thus, although it is capable of protein dimerization, transcriptional activation is diminished or lost. The PAS domain is named for the genes originally identified with this protein dimerization domain: ***p****er* (a fly clock gene), ***ARNT,*** and ***sim.*** *Clock* mRNA is expressed in the SCN as well as other tissues, but it has not been found to oscillate in a circadian fashion.[30]

Bmal1

The presence of the PAS dimerization domain in CLOCK protein suggested that it may form a heterodimer similar to that of PER and the protein product of another *Drosophila* clock gene, TIM.[31] A screen for potential partners for the CLOCK protein using the yeast two-hybrid system revealed that a protein of unknown function, BMAL1 (Brain and Muscle ARNT-Like 1), was able to dimerize with the CLOCK protein.[32] Creation of mice harboring a null allele of *Bmal1* (also referred to as *MOP3*) demonstrated the critical role of this gene in circadian rhythm generation. These mutant mice, while displaying light-dark–responsive differences in activity level, lose circadian rhythmicity immediately upon release in constant darkness.

Recently, additional actions of the CLOCK:BMAL1 heterodimer have become clear. Whereas *Clock* mRNA does not oscillate, its protein's nuclear versus cytoplasmic localization does so.[33] By studying the intracellular localization of CLOCK and BMAL1 in cultured fibroblasts of mouse embryos with mutations in different clock genes, and ectopically expressing the proteins, it was found that nuclear accumulation of CLOCK was dependent on formation of the CLOCK:BMAL1 dimer, as was phosphorylation of the complex and its degradation.[33] Other PAS domain–containing proteins failed to affect the localization of CLOCK, indicating that these posttranslational events are specific to the CLOCK:BMAL1 dimer.

NEGATIVE ELEMENTS

The *Period* Genes

The first identified gene (defined as a *Mendelian gene* as opposed to a sequenced, cloned gene) that encodes a clock component, *period,* denoted by the symbol *per*, was discovered in 1971 in *Drosophila* using a forward genetics approach consisting of chemically inducing random mutations in the genome, and detecting those mutations that affect circadian rhythms by screening the progeny of the mutagenized individuals for altered rhythmicity.[34] This approach has the advantage that no assumptions are made about the nature of the genes or gene products involved but is based on the presumed existence of genes that, when mutated, will alter rhythms in a detectable manner. At the time, this presumption of the existence of genes that regulate a complex behavior was considered radical but has proved to have been a field-defining moment.

Initially, three alleles of the *per* gene were identified by the process of mutagenesis and screening. Flies carrying these alleles had no apparent rhythm in either eclosion (emergence from the pupal case) or locomotion, or had either long (e.g., 29 hours) or short periods (e.g., 19 hours) for the rhythms of eclosion and locomotor activity.[34] It is important to note that the finding of three alleles with three different phenotypes made it possible to have confidence in the conclusion that the *per* gene encodes a protein that is a clock component. Had only an arrhythmic mutant been found, then the alternative explanation could be proposed that the lack of circadian behavior was secondary to another primary defect that did not lie in a clock component, such as in a clock output process. This approach of mutagenesis and screening also has been successful in identifying circadian clock genes in other organisms such as *Neurospora crassa*,[35] plants,[36] and cyanobacteria,[37] as well as mice. A discussion of these important findings, however, is outside the scope of this chapter.

Confirmation of the importance of the *per* gene as a central circadian clock component was made by the rescue of the mutant phenotype after introduction of the wild-type allele of the *per* gene into mutant flies.[38,39] The level of the mRNA transcript encoded by the *per* gene was shown to oscillate in a circadian fashion[40] as a result of transcriptional regulation,[41] and the levels of the PER protein were shown to lag the *per* mRNA levels.[42] In fact, shifts in the circadian phase can be evoked by the induction of PER protein under the control of a noncircadian promoter.[43] Thus many lines of evidence indicate that the *per* gene encodes a protein that is a clock component. Three orthologs of the *per* gene, *Per1, Per2*, and *Per3*, have now been identified in the mouse, and the levels of their mRNA also have been shown to oscillate with a circadian period.[44-48]

After the identification of CLOCK:BMAL1 dimerization, the ability of this heterodimer to regulate transcription was tested using a reporter construct based on the upstream regulatory elements of the *per* gene. The *per* gene of *Drosophila* contains an upstream regulatory element, the "clock control region," within which is contained a sequence needed for positive regulation of transcription, the E-box element (CACGTG).[49] CLOCK-BMAL1 heterodimers were found to activate transcription of the *Per1* gene in a process that requires binding to the E-box element.[32] The CLOCK-Δ19 mutant protein, however, was not able to activate transcription, consistent with the finding that exon 19, which is skipped in *Clock* mutant animals,[30] is necessary for transactivation. Thus CLOCK protein interacts with the regulatory regions of the *per* gene to allow transcription of the *per* mRNA and eventual translation of PER protein. A similar activation of transcription of the *tim* gene by the CLOCK-BMAL1 heterodimer also occurs in flies.[50] This positive regulation alone, however, will not produce an oscillation in *per* mRNA levels, which is known to be responsible for the oscillation in PER protein levels.[41] Findings that the *Clock* mutation dramatically decreases *per* gene expression also confirm the positive regulation of CLOCK:BMAL1 on *per* transcription in situ.[51,52] Mice with null mutations of *Per1, Per2,* or *Per3* alone display altered circadian periods,[53,54] whereas mice with both *Per1* and *Per2* null mutations lose circadian rhythmicity. *Per3* null mutant mice exhibit only a subtle alteration in rhythmicity, and *Per1/Per3* or *Per2/Per3* double mutants are not substantially distinct from the *Per1* or *Per2* single mutants. These

findings suggest there may be some compensation of function among the different mammalian *per* genes and raise the question of the significance of *Per3* for the generation of mammalian circadian rhythms.

Cryptochromes

Cryptochromes are blue light–responsive flavoprotein photopigments related to photolyases, so named because their function was cryptic when they were first identified. In mammals, two cryptochrome genes, *Cry1* and *Cry2,* have been identified and were found to be highly expressed in the ganglion cells and inner nuclear layer of the retina as well as the SCN,[55] and their messenger RNA (mRNA) expression levels oscillate in these tissues. Targeted mutant mice lacking *Cry2* exhibit a lengthened circadian period, whereas mice lacking *Cry1* have a shortened circadian period; mice with both mutations experience immediate loss of rhythmicity on transfer to constant darkness.[56-58] Thus, like the mammalian *period* genes, the cryptochrome genes appear to have both distinct (given their opposite effects on circadian period) and compensatory (given that either gene can sustain rhythmicity in the absence of the other) functions.

Because of their expression pattern, the cryptochromes were thought to be the long-unidentified mammalian circadian photoreceptors (see later); accordingly, light responses were examined in characterizing the null mutants. *Cry2* mutant mice exhibit altered phase-shifting responses to light pulses.[56] *Cry1/Cry2* double mutants exhibit impaired light induction of *Per1* in the SCN, whereas light induction of *Per2* in double mutants remains intact.[57,59] Neither *Per1* nor *Per2* exhibits persistent oscillations in expression in the SCN in constant conditions in *Cry1/Cry2* double mutants.[57,59] Thus, although the cryptochromes are not the mammalian circadian photoreceptor, they do appear to play a central role in the generation of circadian signals.

Further evidence for a central clock function is the finding that the cryptochromes appear to share a number of regulatory features with the *period* genes. In *Clock* mutant mice, the mRNA levels of *Cry1* and *Cry2* are reduced in the SCN and in skeletal muscle,[60] suggesting that the *cryptochrome* genes also are induced by CLOCK:BMAL1 transactivation. In experiments using mammalian (NIH 3T3 or COS7) cell lines, CRY1 and CRY2 were found by coimmunoprecipitation techniques to interact with PER1, PER2, and PER3, leading to nuclear localization of the CRY:PER dimer, as indicated by cotransfection assays with epitope-tagged proteins.[60] Luciferase assays indicate that CRY:CRY or CRY:PER complexes are capable of inhibiting CLOCK:BMAL1 transactivation of *Per1* or *vasopressin* transcription.[60] Thus the CRYs as well as the PERs are capable of a negative feedback function, inhibiting CLOCK:BMAL1-induced transcription.

MODULATORS AND OTHER COMPONENTS OF THE CLOCK

Timeless

How is the level of the PER protein regulated by the circadian clock? The first hint came from the identification of the *timeless* gene *tim*, which when mutated produces abnormal circadian rhythms in *Drosophila*.[61] The levels of the mRNA encoded by the *tim* gene oscillate with a time course that is indistinguishable from those of *per* mRNA.[62] The levels of the TIM protein lag behind those of *tim* mRNA by several hours,[63] similar to the finding with *per* mRNA and PER protein. The PER and TIM proteins form heterodimers[64] that are transported to the nucleus.[65] The finding that the heterodimer is transported to the nucleus suggested that it might be involved in the regulation of transcription of the *per* or *tim* genes. Indeed, recent experiments have shown that the transcription of the *per* and *tim* genes is repressed by the PER-TIM protein heterodimer.[50] This finding is very important because it demonstrates that the production of mRNA encoded by a clock component gene, the delayed accumulation of the encoded protein, and later feedback to the clock gene's promoter in the nucleus are able to explain the basic features of the circadian clock in *Drosophila*.

PER-TIM interactions are not sufficient, however, and the basic mechanism does not become clear until interactions with other clock genes are considered. In experiments using a luciferase reporter assay, the luminescent luciferase protein was expressed under the control of the promoter regions of the *Drosophila per* and *tim* genes. It was found that the fly homolog of *Clock*[66] was capable of driving expression of luciferase[50] in cells that have high endogenous levels of the *Drosophila* homolog of BMAL1, CYC (encoded by the *cycle* gene). The effect of the PER-TIM heterodimer on the ability of the CLOCK-CYC heterodimer to drive the transcription of the *per* and *tim* genes was tested by cotransfecting the encoding genes into the cells that expressed the luciferase reporter gene. Indeed, it was found that the expression of both the *per* and *tim* genes was reduced by their own protein products. This negative feedback has recently been found for a mammalian heterodimer consisting of homologs of the TIMELESS and PER1 proteins.[67]

Whether the mammalian *tim* homolog identified[67,68] actually represents an orthologous gene has been called into question.[69] This issue has been difficult to resolve, because gene targeting to create a null mutant resulted in early embryonic lethality. Differences in results obtained by different groups of researchers examining the oscillation of *Tim* expression could result from existence of both a full-length and a truncated protein, with only the full-length form oscillating.[70] Using antisense oligodeoxynucleotides directed against *Timeless* in rat SCN slice preparations results in a disruption of neuronal oscillations in vitro, suggesting a role in rhythmicity.[70] True functional homology of *Timeless* in mammals, however, remains to be demonstrated in vivo.

The *CASEin Kinase 1* Gene

The *tau* mutation of the hamster arose spontaneously in a laboratory stock.[71] The mutation is semidominant and shortens the period from 24 to 22 hours in heterozygotes and 20 hours in homozygotes. This mutation has been of great importance for several reasons. First, the *tau* mutation predated the *Clock* mutation and demonstrated that single-gene mutations could profoundly alter the circadian clock in mammals, just as in flies and *Neurospora*. Second, *tau* mutants display several other physiologic phenotypes such as alteration of the responses of males to photoperiod length[72] and effects of the estrous cycle in females,[73] which gave further insights into the importance of the circadian clock for other biologic cycles. Finally, the evidence that the SCN is indeed the site of the master circadian oscillator (see earlier) was demonstrated unequivocally using transplantation of the SCN that employed

the *tau* mutation. These manipulations also gave rise to the evidence necessary to conclude that the *tau* mutation encodes a protein that is a clock component. Unfortunately, the genetic tools needed for cloning this important and interesting gene were not available for the hamster, so its molecular identity could not be determined by conventional genetic mapping/positional cloning approaches.

Lowrey and colleagues were able to identify a genomic region of conserved synteny (a grouping of genes together on a chromosome), in hamsters, mice, and humans, that encompassed the *tau* mutation.[74] *Tau* was thus identified as being a mutation in the *Casein Kinase 1, epsilon* gene (*CK1e*), the mammalian ortholog of the *Drosophila doubletime* gene. Sequencing of the gene identified a point mutation that leads to altered enzyme dynamics and autophosphorylation state. In vitro assays demonstrated that CK1e can phosphorylate PER proteins, and that the *tau* mutant enzyme is deficient in this ability. Thus CK1e may lead to degradation of PERs, slowing the accumulation of PER in the nucleus and thus repression of CLOCK:BMAL1. The *Casein Kinase 1, delta* gene (*CK1d*) also has been implicated in mammalian circadian rhythmicity.[75-78]

Rev-erb alpha and *ROR*

Whereas the negative feedback of PER and CRY proteins on their own CLOCK:BMAL1-induced may be sufficient to explain the oscillations in expression of *Per* and *Cry* genes, the rhythmic expression of *Bmal1* with an opposite phase is not explained by this feedback. What regulatory elements produce the rhythmic transcription of *Bmal1*, with an antiphase relationship to the *Per*s? *Rev-erb alpha,* an orphan nuclear receptor, may act as the missing link. Its promoter region contains three E-boxes, and transcription is therefore positively regulated by CLOCK and BMAL1.[79] Its transcription is negatively regulated by PER and CRYs; *Rev-erb alpha* mRNA is at a minimum when PER2 is at a maximum, and it is constitutively expressed at intermediate levels in *Cry1/Cry2* or *Per1/Per2* double knockouts. REV-ERBa protein appears to drive the circadian oscillation in *Bmal1* transcription: The *Bmal1* promoter includes two RORE sequences (enhancer sequences that recognize members of the REV-ERB and ROR orphan nuclear receptor families), and *Bmal1* expression is drastically reduced in *Rev-erb alpha* null mutants.[79] Thus *Rev-erb alpha* may act to link the positive and negative regulatory signals of other clock genes to the transcription of *Bmal1*. In view of the importance of orphan nuclear receptors in regulating cellular metabolic properties,[80] interaction with circadian clock genes may, at the molecular level, form the links between circadian clocks and metabolic regulation, with important implications for health and disease. Such molecular links between circadian clocks and cellular metabolism also include a clock-metabolism feedback loop. CLOCK:BMAL1 activity is suppressed by a NAD^+-dependent deacetylase SIRT1, which is in turn regulated by the circadian control of NAD^+ biosynthesis.[81-84] See Bass (2012) in Selected Readings for more detail.

Rev-erb alpha also contributes to the differences between the phase of *Cry1* mRNA rhythms and that of other clock genes whose transcription is enhanced by CLOCK:BMAL binding to E-boxes. The *Cry1* gene has three candidate REV-ERB/ROR binding sites[85]; in vitro assays indicate that REV-ERBa binds to two of these sites. Luciferase reporter assays indicate that REV-ERBa protein can inhibit transcription of *Cry1* through binding at these two sites. REV-ERBb also appears to share some functional redundancy with REV-ERBa.[86]

Fbxl3 and *Fbxl21*

In addition to *Rev-erb* modulation of *Cry* genes (described earlier), other genes appear to modulate the activity of the Cryptochromes. The *Overtime* mutation in mice was identified in a mutagenesis screen based on a lengthened free-running circadian period.[87] The responsible mutation was ultimately localized to a known gene encoding the F-box protein *Fbxl3* but previously unknown to be involved in circadian rhythmicity. $Fbxl3^{OVTM}$ mutants appear to be functionally comparable to null mutants. FBXL3 protein leads to degradation of CRY1; the OVTM mutant protein is less effective in this capacity. Thus the period lengthening may be a direct result of a delay in degradation of CRY, effectively preventing the core cycle from restarting. Recently, the closely related protein FBXL21[88] has been found to function in a similar manner.[89,90]

NPAS2

NPAS2 (neuronal PAS family member 2) shares the closest homology with CLOCK of all identified bHLH-PAS family members. Null mutants for this gene exhibit altered circadian activity patterns, notably the absence of a "siesta" in later subjective night, but no dramatic alterations in circadian free-running period or persistence.[91] However, when *Clock* null mutants showed less dramatic phenotypes than for the Δ19 mutant,[92] the role of NPAS2 was reexamined. In the absence of functioning CLOCK, NPAS2 appears to be able to partially compensate.[93]

Dec1 and *Dec2*

Like other clock genes, *Dec1* and *Dec2* encode basic helix-loop-helix transcription factors that bind to E-boxes. DEC1 and DEC2 have been found to inhibit transactivation of Per by CLOCK and BMAL1.[94] DEC1 and DEC2 form dimers.[95] The inhibition of CLOCK and BMAL1 transactivation may be related to interactions with BMAL1 but also can be attributed to binding to (and thus possibly competition for) E-boxes.[96]

Recently, a human allelic variant in DEC2 has been linked with total sleep time.[97] This functional relationship has been confirmed in transgenic mice expressing the human *Dec2* allele.

Other Regulators of the Clock

The discovery of the core molecular machinery of the circadian clock has allowed identification of additional regulators of the circadian clock by characterizing proteins that interact with known clock components. For example, an RNA- and DNA-binding protein NONO and a subunit of histone methyltransferase complexes WDR5 were identified as interacting proteins and modulators of PER1. Knockdown of *NONO* using RNA interference (RNAi) in mammalian cells and *NONO* mutant in flies both disrupt circadian rhythms.[98] Another study later showed that NONO and PER1 are involved in the regulation of circadian gating of cell cycles via controlling the circadian expression of a cell cycle checkpoint gene *p16-Ink4A*.[99] Similarly, RACK1 (receptor for activated

C kinase 1) and PKCα (protein kinase C alpha) were found interacting with BMAL1 and modulating CLOCK-BMAL1 transcriptional activity. Knockdown of either PKCα or RACK1 by RNAi shortens circadian period.[100] Clock modulators identified by studying proteins interacting with known clock components also include E3 ligases Arf-bp1 and Pam, which are involved in the regulation of the degradation of REV-ERBa.[101]

Furthermore, high-throughput screening studies using the combination of luciferase reporter assays in cell-based systems with large-scale RNAi libraries also have identified many other genes that modulate the circadian oscillation of core clock genes. One study focused on known and predicted kinases as well as phosphatases and F-box proteins and identified 22 kinases, 7 phosphates (or regulatory subunits), and 6 F-box proteins as potential components of the clock.[102] Among those, casein kinase 2 (CK2) was shown to phosphorylate PER2 and regulate its nuclear accumulation and stability. Using a similar RNAi screening approach, Zhang and colleagues searched the entire genome and identified more than 200 candidates that strongly affected the circadian period or amplitude when knocked down.[103] As indicated by protein interaction network analysis, most of these identified candidates are directly or indirectly connected with the core clock components, suggesting that large-scale and complex interactions among the molecular clock mechanisms are likely to be involved in the regulation of circadian timing.

Finally, along with the accumulation of "-omic" (i.e., genomic, proteomic, interactomic, metabolomic) data sets, it recently became possible to predict clock components using in silico modeling approaches. The first computer-assisted study to identify clock components was performed by Anafi and colleagues.[104] These authors collected a number of genomewide data sets and established metrics that describe key features of core clock components, including (1) 24-hour oscillatory expression, (2) disruption of circadian rhythms when mutated or knocked down, (3) interaction with other core clock proteins, (4) ubiquity expressed in multiple tissues, and (5) conservation between vertebrates and flies. Although these clock features are not absolute, the core clock metrics allowed the investigators to use a machine learning algorithm to identify genes that share similar metrics with known core clock genes. The top identified genes include many that have already been implicated in the circadian clock machinery. In addition, a previously uncharacterized gene, *Gene Model 129* (*Gm129*), also was among the top candidate clock genes, and the investigators renamed it *Chrono*, for "*c*omputationally *h*ighlighted *r*epressor *o*f the *n*etwork *o*scillator." Validation experiments show that *Gm129* expression exhibits strong circadian rhythms in the liver, heart, and adipose tissue. GM129 protein interacts with the carboxyl terminus (C-terminus) of BMAL1 and suppresses CLOCK/BMAL1 transcriptional activity. Finally, free-running periods of activity are lengthened in *Gm129*-knockout mice. Of interest, the role of *Gm129* in the circadian clock network also was independently discovered using conventional biochemical approaches,[105] at roughly the same time of this in silico–assisted study. With the rapid growth of bioinformatics techniques and databases, it can be expected that computational approaches will be highly complementary to traditional biochemical approaches, contributing to a full understanding of the circadian clock network.

An important point is that many regulators of the clock also are key regulators in other cellular processes, including metabolism and redox homeostasis (such as *Rev-erb alpha* and *Sirt1*), cell cycle (such as *NONO*), cell signaling (such as *PKCα*), and many others. The interactions between clock and other cellular pathways are further supported by results from large-scale RNAi screen studies because pathway analysis suggests that the identified clock modulators are overrepresented for genes involved in insulin and Hedgehog signaling, the cell cycle, and folate metabolism.[103] Many of these pathways and key regulators that influence the circadian clock are themselves regulated by the clock. Taken together, these findings suggest that the molecular clock machinery is hardwired to diverse cellular processes and pathways, providing a basis for functional coordination of multiple pathways by the circadian clock.

OUTPUT REGULATION

Transcriptional Output of the Clock

As described earlier, the core clock components form a transcriptional and translational feedback loop, which produces a program of oscillatory gene expression involving roughly one third of the transcriptome. Much effort has been devoted to clarifying the details of how circadian transcriptome is regulated by the clock. Bioinformatics analyses of promoter regions of oscillatory genes in multiple tissues suggest enrichment of classical enhancer elements such as E-box and D-box.[106,107] Many of the clock-controlled, E-box–containing genes are themselves transcription factors, which are thought to be important for expanding the repertoire of clock-regulated transcription to a wide range of genes. Of interest, bioinformatics studies also have suggested that rather than a single E-box, an arrangement of two closely spaced (six to seven base pairs apart) E-box–like elements—namely, E1 and E2 elements—are critical for robust oscillations in many of the clock-regulated genes in both flies and mice.[108,109] Using chromatin immunoprecipitation (CHIP) combined with deep sequencing (CHIP-seq), one study characterized direct DNA-binding targets of BMAL1 in mouse liver.[110] A total of 2049 BMAL1 target–binding sites were identified, 60% of which exhibited rhythmic binding. Sequence analysis suggests that 13% of the BMAL1–binding sites consist of a pair of E1-E2 elements, and these E1-E2 sites are associated with more robust rhythmic BMAL1 binding, whereas a single E-box alone is sufficient for rhythmic CLOCK:BMAL1 binding.[110] Another study using CHIP-seq compared the binding targets of seven clock components, CLOCK, BMAL1, NAPS2, PER1, PER2, CRY1, and CRY2, in mouse liver.[111] Remarkably, although more than 1400 target sites were found to be common to CLOCK, BMAL1, PER1, PER2, CRY1, and CRY2, distinct target profiles also were apparent for each of these components. Targets of all clock components commonly are enriched with E-boxes; sites that are specifically bound by PER2, CRY1, or CRY2, however, show a reduction in E-boxes, as well as an enrichment for nuclear hormone receptors—consistent with known partnerships between the negative elements of the clock and nuclear receptors.[112,113] Finally, an interesting finding is that E-box elements can be occupied by other transcription factors. For example, Usf1, a suppressor of the $Clock^{\Delta 19}$ mutant phenotype, competes with CLOCK:BMAL1 for E1 sites, modulating the circadian transcriptome.[114]

Many other mechanisms are likely to be involved in the regulation of the circadian transcriptome. For example, multiple lines of evidence have supported a role for rhythmic histone modification and chromatin remodeling in the regulation of circadian transcription. P300, a histone acetyltransferase, immunoprecipitates together with CLOCK in liver nuclear preparations, with a peak at CT 6 and minimum at CT 18.[85] P300 enhances CLOCK:BMAL1-mediated transcription of reporter gene, and this increase in expression is inhibited by CRY1 and CRY2, suggesting a potential mechanism by which CRY proteins can preclude CLOCK:BMAL1 transactivation of target genes. It also has been demonstrated that CLOCK protein itself also can function as a histone acetyltransferase.[115] Indeed, genome-wide circadian chromatin modifications have been extensively described in a number of studies.[110,116,117] Remarkably, the binding of CLOCK:BMAL1 to target DNA promotes removal of nucleosomes, which leads to chromatin opening at the binding site, allowing for rhythmic binding of other transcription factors.[118] Finally, a recent study has reported clock-controlled long-range chromosomal interactions, suggesting a genomic configuration that is coordinated with the rhythmic gene expression.[119]

In addition to direct transcriptional activation and chromatin remodeling, posttranscriptional mechanisms are likely to be involved in shaping output programs of the clock (as described by Lim and Allada[120]). By separating intronic and exonic signals in whole-genome RNA-seq data, Koike and coworkers were able to infer the oscillatory patterns of pre-mRNA from total mRNA.[111] Of interest, genes that exhibited both exonic and intronic cycling signals account for only 22% of the exonic cycling genes and 30% of the intronic cycling genes, suggesting that de novo transcription contributes to only a relatively small portion of circadian program of the transcriptome. Similar results were observed on comparing rhythmic expression pattern of nascent and total mRNA using RNA-seq in both mice and flies,[121,122] indicating significant contributions of posttranscriptional mechanisms in regulating circadian transcriptome.

A complex profile of rhythmic gene expression arises from clock-dependent transcriptional, epigenetic, and posttranscriptional regulations. Early microarray studies have suggested that up to a third of the expressed genes exhibit circadian oscillations.[18-21] More recent studies with higher temporal resolutions have added a great amount of detail to the overall picture of the circadian transcriptome, including transcripts that exhibited clock-controlled harmonic oscillations (i.e., 12-hour and 4-hour rhythms)[123] as well as rhythmically expressed noncoding regulatory RNAs.[117,124] Of interest, only a small fraction of rhythmically expressed genes commonly are found in SCN and liver,[20] in liver and heart,[21] as well as in liver and skeletal muscle[125] (Figure 27-2). Adding together all oscillatory genes from all the tissues that have been examined thus far, a total of 43% of protein-coding genes oscillate in the body.[124,126] At least in flies, tissue-specific programming of circadian gene expression is correlated with a tissue-specific preference in *cis*-regulatory motifs of genes that are bound by CLOCK and other partner transcription factors.[127] Gene ontology suggests that rhythmically expressed genes are involved in diverse pathways and cellular functions. In addition, the tissue specificity of circadian gene expression is tightly related to the specific functions of the tissue. For example, genes cycling in the SCN include those involved in protein or neuropeptide synthesis, processing, and degradation, as well as genes known to be important for circadian locomotor activity, whereas genes cycling in the liver are involved in nutrient metabolisms and regulation of metabolic intermediates.[20] Nonetheless, expression of genes involved in essential cell functions, such as redox homeostasis, appear to be under circadian regulation across different tissues. Remarkably, a recent study suggests that the targets of best-selling drugs and World Health Organization (WHO)-designated essential medicines are highly enriched with oscillating genes.[124] This finding has significant clinical implications, in that the treatment outcome drugs with short half-lives could potentially be improved by circadian-timed administration.

Figure 27-2 Rhythmic Gene Expression Varies among Cells and Tissues. Gene expression profiling studies[124,126] have revealed that up to a third of genes expressed in any given tissue may do so rhythmically. Although the core circadian clock genes that constitute the transcription-translation feedback loop are rhythmically expressed in all tissues, and some common regulatory factors may be expressed in multiple tissues, the vast majority of these rhythmic transcripts are specific to the cell type and functions of the particular tissue. Still to be delineated are the contributions of various regulatory mechanisms in determining the tissue specificity. (Data from Miller BH, McDearmon EL, Panda S, et al. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc Natl Acad Sci U S A* 2007;104:3342–7; McCarthy JJ, Andrews JL, McDearmon EL, et al. Identification of the circadian transcriptome in adult mouse skeletal muscle. *Physiol Genomics* 2007;31: 86–95; and Hogenesch JB, Panda S, Kay S, Takahashi JS. Circadian transcriptional output in the SCN and liver of the mouse. *Novartis Found Symp* 2003;253:171–80.)

Of note, the oscillatory gene expression pattern observed outside of the SCN is driven by both local cell-automatous clock machineries as well as system cues originating from the SCN. Using an inducible hepatocyte-specific *Rev-erb alpha* transgene, Kornmann and associates were able to arrest the liver clock by suppressing *Bmal1* expression on activation of the transgene.[128] Of interest, 10% of the liver circadian transcriptome, including the transcript of core clock gene *Per2*, remained cycling when the hepatocyte clock was arrested. Circadian expression of these genes are thus attributed to the output signals from the master clock, including direct neural and humoral signals from the SCN as well as indirect cues such as those associated with clock-controlled sleep-wake, feed-fast, and body temperature rhythms. In particular, sleep-wake affects the expression of at least 5% of the genes,[129] and large-scale interactions between circadian rhythms and sleep-wake have been recently documented[130,131] (for detailed discussion, see Chapter 29). Recent studies also suggest that

the circadian transcriptome can be reprogrammed by environmental conditions. For example, compared with normal-chow feeding, feeding with a high-fat diet suppressed cycling of more than 1000 genes (including those involved in insulin signaling) while eliciting rhythmic expression of hundreds of other genes.[132] In addition, of the genes cycling under both conditions, 66% exhibited significant shifts in circadian phase. These findings suggest that the circadian program of gene expression is highly plastic in response to environmental conditions and associated functional challenges.

Output from the Suprachiasmatic Nucleus

Both neural and humoral mechanisms are likely to be involved in the output of the master clock. The molecular clock controls a firing rhythm in SCN neurons, and the SCN neurons project to many other brain areas.[133] On the other hand, encapsulated SCN grafts are capable of rescuing the circadian locomotor activity in SCN lesion animals,[134] suggesting that secreted factors are also involved in SCN output. To date, a few proteins, such as PK2, TGF-α, CLC, and VIP, have been implicated in the SCN humoral outputs.

Prokineticin 2 (PK2) is rhythmically expressed in the SCN,[135] and infusion of PK2 into the cerebral ventricles inhibits locomotor activity. Mice with a null mutation in the PK2-encoding gene exhibit dramatically reduced levels of activity[136] with reduced circadian amplitude. These mice also exhibit attenuated rebound in non–rapid eye movement, rapid eye movement, and delta wave power on electroencephalogram after sleep deprivation.[137]

The peptide transforming growth factor alpha (TGF-α) was identified in a screen for SCN factors that might inhibit locomotor activity; when infused into the third ventricle, this peptide inhibits locomotor activity. Mice with targeted mutations of the epidermal growth factor receptor (the receptor likely to bind TGF-α) also display disruption of activity rhythms.[138] These effects are attributable to actions on the epidermal growth factor receptors.

Mice that lack the peptide receptor VPAC2 show abnormal entrainment and disrupted rhythms, indicating that VIP signaling in the SCN may be necessary for normal expression and coordination of rhythms.[139]

Cardiomyotrophin-like cytokine (CLC) also is expressed in the SCN in a rhythmic manner in vasopressin neurons. Infusion of CLC into the third ventricle (near the SCN) dramatically inhibits locomotor activity, whereas infusion of antibodies to the CLC receptor increases activity.[140]

INPUT REGULATION

Melanopsin

The circadian rhythms of many humans who are blind, with no conscious perception of light, are nevertheless able to be entrained by light.[141] This intriguing observation led to studies of the circadian light input pathway and the photoreceptors and photopigments in mammals.

In experiments using mouse mutations that result in degeneration of rods[142] or both rods and cones,[143] light entrainment of the circadian rhythm was preserved.[144] However, the eye must be the site of the light-entraining pathways in mammals because enucleated mammals are not capable of light entrainment.[142] Indeed a morphologically distinct set of retinal ganglion cells projects to the SCN via the retinohypothalamic tract.[145] Ablation of the SCN abolishes circadian rhythmicity and ablation of the retinohypotalamic tract abolishes light entrainment.[146] Thus the light signal responsible for light entrainment enters the SCN via a unique axonal pathway from the eye.

Melanopsin, a member of the opsin family of photopigments, was first identified in the inner retina[147] and later found to be expressed in the somata and dendrites of retinal ganglion cells of the retinohypothamamic tract.[148] Neurons that contribute axons to the retinohypothalamic tract were found to express the marker pituitary adenylate cyclase–activating polypeptide (PACAP).[148] PACAP was used as a marker for rat retinohypothalamic tract neurons; every PACAP-positive neuron was found to express melanopsin, and every melanopsin-positive neuron was PACAP-positive.[148]

Further evidence confirming the role of melanopsin as the phase-shifting pigment has come from genetically engineered mice in which the gene encoding melanopsin was disrupted.[147,149,150] Two behavioral measures of circadian rhythm responses to light were altered in these mice. The *phase-shifting response* to a discrete light pulse was of lesser amplitude in the knockout mice than in wild-type mice, and the *free-running periods* of the knockout mice were lengthened less by exposure to constant light than the periods of wild-type mice. Hence it appears that melanopsin represents a primary photopigment with other photopigments also having input to the circadian system.

Rab3a

The *Rab3a* gene was identified in a mutagenesis screen (*earlybird*) based on an advanced phase angle of entrainment and shortened circadian period. Null mutant mice display a similar phenotype.[151] Furthermore, both the *Rab3a* null and *earlybird* mutants exhibit alterations in the homeostatic response to sleep deprivation,[151] as well as alterations in emotional behavior.[152]

CONCLUSIONS AND FUTURE DIRECTIONS

The core circadian oscillator is autonomous to individual neurons of the SCN and is the result of the daily oscillation in the levels of several clock component proteins. The basis for this oscillation in mammals, as in other organisms, lies in rhythmic feedback regulation of transcription of the genes encoding these proteins. The levels of the PER and CRY proteins alter the rate of transcription of their own genes. This alteration is achieved by inhibition of the enhancement of transcription that results from binding of the CLOCK-BMAL1 heterodimer to the E-box element of the promoter region of the *Per* and *Cry* genes. Additional interactions between circadian clock proteins may slow the time course of this feedback, achieving the near-24-hour interval. The phosphorylation of PER by CKIε may lead to its degradation, and the association with BMAL1 apparently is needed for CLOCK to be present in the nucleus. Rhythmic transcription of *Bmal1* appears to result from regulation mediated by REV-ERBα, itself regulated by E-box elements. Finally, rhythms in histone acetylation evidently contribute to the circadian expression pattern of some core circadian genes. Additional genes have been identified on the basis of altered circadian rhythms in mutants, although the roles of these genes in the circadian system remain to be determined.

It is of interest that a majority of the core genes have been identified in mice or in flies by forward genetics, in which mutations were induced in the genome randomly, and those mutations that specifically affect the circadian oscillator were identified with carefully crafted circadian phenotypic screens. Now that these clock component proteins have been identified, it will be easier to find the proteins that serve the input and output pathways of the circadian oscillator, and to identify the components that are out of order in disease states affecting circadian rhythms. Furthermore, with the great advances achieved thus far in genomic, proteomic, and metabolomic techniques, future studies are expected to integrate and depict the circadian program at a multiscale level. Such studies will eventually facilitate a comprehensive understanding of how the clock is coupled with other cellular and physiologic functions. It is fortuitous that the unraveling of the molecular basis for circadian rhythmicity is occurring at a time when the general public is becoming aware of the importance of normal circadian timekeeping for human health, safety, performance, and productivity.

CLINICAL PEARL: How Many Levels of Genetic Regulation of Clocks Exist?

The developing picture of the molecular genetic clock is becoming many-layered and complex. At its core is a transcriptional-translational feedback loop of CLOCK:BMAL1 and PER:CRY. This loop interacts intimately with a cellular metabolism loop, by way of NAD and SIRT. The genetic circadian clock regulates rhythms of gene expression through direct transcriptional regulation, mediated by histone acetylation/deacetylation, chromatin modifications, and possibly posttranscriptional modifications as well. These modifications may be tissue-specific but can be influenced by the SCN.[153] Rhythms or behaviors (driven by the SCN) such as melatonin levels,[154] body temperature,[155] feeding,[156] or sleep[131,157] also can influence gene expression rhythms in the periphery. It thus seems that the so-called cell-autonomous genetic circadian clock is subjected to a wide array of modulators ranging from the intracellular to nutritional, endocrine, and behavioral. How cell-autonomous is the human circadian clock? Whether such disruptions in physiology and behavior are due to altered circadian information from the SCN or to local tissue-specific changes in the molecular circadian clock is not known. Regardless of the mechanisms, such results point to a very central role of circadian clock genes in regulating biochemical, metabolic, and physiologic processes at many different levels of organization.

SUMMARY

The cell-autonomous circadian clock found in all mammalian cells and tissues has, at its core, a feedback loop of transcription and translation of a key set of "clock genes." Identification of clock genes has been guided by examining the effects of gene alterations on clock properties. A substantial proportion of expressed genes exhibit circadian variation; however, distinct cell types appear to have their own unique sets of "clock-controlled genes," while the core clock genes remain the same. The transcription-translation feedback loop is subject to regulation at multiple levels, including histone, chromatin, and posttranscriptional modifications. Signals from the suprachiasmatic nucleus of the hypothalamus and from behavioral states such as sleep/wake or feed/fast also can influence clocks in peripheral tissues.

ACKNOWLEDGMENTS

Preparation for this chapter was supported in part by National Institutes of Health (NIH) grant PO1 AG 11412.

Selected Readings

Baggs JE, Hogenesch JB. Genomics and systems approaches in the mammalian circadian clock. *Curr Opin Genet Dev* 2010;**20**(6):581–7.

Bass J. Circadian topology of metabolism. *Nature* 2012;**491**:348–56.

Bedont JL, Blackshaw S. Constructing the suprachiasmatic nucleus: a watchmaker's perspective on the central clockworks. *Front Syst Neurosci* 2015;**9**:74.

Buhr ED, Takahashi JS. Molecular components of the mammalian circadian clock. *Handb Exp Pharmacol* 2013;**217**:3–27.

Laing EE, et al. Exploiting human and mouse transcriptomic data: identification of circadian genes and pathways influencing health. *BioEssays* 2015;**37**:544–56.

Mauvoisin D, Dayon L, Gachon F, Kussmann M. Proteomics and circadian rhythms: it's all about signaling! *Proteomics* 2015;**15**:310–17.

A complete reference list can be found online at ExpertConsult.com.

Genetics and Genomic Basis of Sleep in Simple Model Organisms

Chapter 28

Ravi Allada; Mark Wu

Chapter Highlights

- The availability of simpler animal model systems, such as nematodes, flies, and fish, has greatly facilitated the genetic analysis of sleep.
- These simpler animal models exhibit the defining features of sleep, including behavioral quiescence, reduced responsiveness to sensory stimuli, and homeostatic regulation.
- Genetic studies in these simpler models have revealed both circuit and molecular pathways involved in sleep regulation, including many that are conserved with rodents and even humans.

The complexity of sleep in mammals has led to a movement toward examining related phenomena in simpler organisms, such as nematodes, flies, and fish, that exhibit sleep (or sleep-like states), with development of model systems that harbor technical advantages not present in more conventional rodent, feline, or primate models. One such system is that starring the fruit fly *Drosophila melanogaster,* best known for its use in genetic studies. Remarkably, the mammalian homologues of many *Drosophila* genes have been found to function in a manner similar to that observed in flies. Indeed, most human disease genes have clear fly homologues. In view of the notable genetic similarity between *Drosophila* and mammals, it is not surprising that fruit flies exhibit many of the defining features of sleep, including circadian-dependent quiescence, reduced responsiveness to sensory stimuli, and homeostatic regulation. Furthermore, preliminary indications are that even the genetic and pharmacologic underpinnings of sleep are conserved between flies and mammals. Primary topics addressed in this chapter are (1) use of the *Drosophila* model for sleep studies, (2) the available evidence on the circadian regulation of behavior, and (3) recent insights into sleep regulation. Also discussed are the unique features of both the zebrafish and nematode *Caenorhabditis elegans* models of sleep, along with relevant technological advances. Both experimental observations in reported studies and the insights they engender could prove to be important in elucidating the genetic basis of human sleep and, ultimately, answering the question of why organisms sleep.

In contrast with studies of naturally occurring genetic variation, it is possible to induce mutations in animal models to test whether a given gene is important for sleep. One strategy to investigate the molecular basis of complex behaviors such as sleep is classical or *forward genetics*.[1] Here a population of animals is randomly mutagenized using DNA-alkylating agents, such as ethyl methane sulfonate (EMS), or mobile DNA transposable elements. This mutagenized population is screened for a mutant phenotype of interest, such as altered sleep. Molecular genetics techniques can then be applied to identify the mutant gene responsible for the mutant phenotype. Thus forward genetics can be used to establish causal relationships between the function of individual genes and otherwise complex phenotypes. Forward genetics is unbiased and does not require any previous knowledge about the genetic basis of the phenotype of interest and is therefore an ideal approach for studying sleep. By contrast, *reverse genetics* starts with a disrupted gene in search of a phenotype. The finding of a gene can provide insight into biochemical and cellular pathways that are important for sleep, perhaps even providing novel diagnostic tests or targets for drug development for sleep disorders.

DROSOPHILA AS A MODEL SYSTEM FOR GENETICS

Many of the "genetic" model organisms, such as zebrafish (*Danio rerio*) and the nematode (*C. elegans*), recently have been adopted for the study of sleep because they are highly suited to the forward genetics approach.[2] One of the premier model systems for genetics is that of the fruit fly, *D. melanogaster* (Figure 28-1). The fruit fly has been a workhorse for genetic studies since the pioneering work of Thomas Hunt Morgan in the early 20th century.[3] A major advantage of *Drosophila* over mammalian model systems is the ability to grow and handle large numbers relatively easily and cheaply.[1] A single female can produce hundreds of offspring. In addition, it has a short generation time, approximately 10 to 12 days from fertilized egg to fertile adult at room temperature. Because of these traits, *Drosophila* has been a model par excellence for high-throughput screening of mutants with altered phenotypes. The facility of genetic mapping, gene disruption using transposable elements (mobile DNA), and the full genome sequence allows rapid identification of mutant genes responsible for mutant phenotypes.[4] Remarkably (to some researchers), the mammalian versions (i.e., homologues) of *Drosophila* genes have been found to function in a manner similar to that for their *Drosophila* counterparts. Indeed, entire signaling pathways are shared between *Drosophila* and their mammalian counterparts. For example, most human disease genes have

Figure 28-1 *Drosophila*, a Genetic Model Organism. Shown is a fruit fly attached to a tether with recording electrodes implanted for measurement of electrical correlates for behavioral states. (Courtesy B. Van Swinderen.)

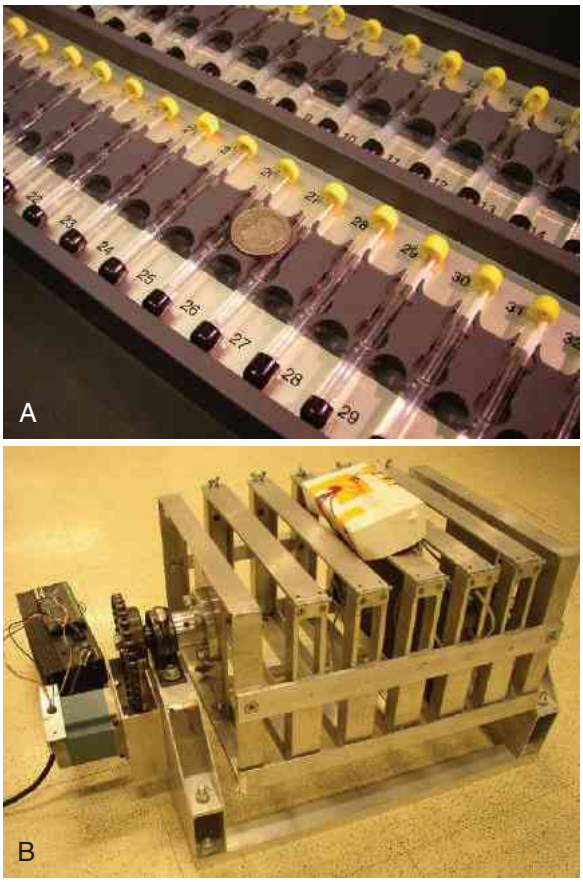


Figure 28-2 The *Drosophila* Activity Monitoring (DAM) System and Rotating Sleep-Depriving Box. **A,** The DAM system. A U.S. dime (diameter = ~1.5 cm) is shown for scale, placed over the location of the infrared emitter/detectors. **B,** *Drosophila* sleep deprivation apparatus. A DAM monitor can be placed into a slot, and then the box is rotated randomly to disrupt fly sleep. (Courtesy B. Chung.)

clear fly homologues.[5] Thus genes identified in flies are likely to serve comparable functions in more complex mammalian systems, including that of humans.

The conservation between flies and mammals extends to the nervous system. Although flies have approximately 1/1,000,000th the number of neurons in humans (10^5 versus 10^{11}), the fly and human genomes are surprisingly similar in gene number (14,000 and approximately 22,000, respectively), with differences largely due to gene duplication.[6] In fact, the fly brain uses comparable neuronal machinery, including neurotransmitters, ion channels, receptors, and signal transduction pathways. Consequently, flies have been used as valuable nervous system models for olfaction,[7] vision,[8] hearing,[9] sexual behavior,[10] synaptic transmission, axon guidance,[11] and learning and memory.[12] Flies also have been exploited as models for numerous human diseases and conditions including diabetes,[13] aging,[14] pain,[15] Alzheimer and Parkinson diseases,[11] epilepsy,[16] and fragile X mental retardation.[17] Finally, flies have been successfully used to study the response to clinically important drugs such as ethanol,[18] cocaine,[18] and general anesthetics.[19] In many of these cases, genes identified in *Drosophila* serve similar functions in mammalian systems.

DROSOPHILA AS A MODEL FOR STUDIES OF SLEEP

Studies of *Drosophila* sleep have been predicated on a small but noteworthy literature examining sleep-like states in other invertebrate models such as mollusks[20] and insects including cockroaches[21] and honey bees.[22,23] These classical descriptions of sleep behavior formed the basis for pursuing similar studies in *Drosophila*.

Sleep in the fruit fly typically is measured behaviorally using the *Drosophila* Activity Monitoring (DAM) System (developed by Trikinetics Inc., Waltham, Massachusetts) (Figure 28-2), which allows for high-throughput analyses. Single flies are placed into a small transparent glass tube plugged on one end by agar food and the other end by a porous cap, allowing air passage. Each tube is placed into a monitor that contains a series of 32 infrared emitter-detector pairs, one for each tube. An awake fly will move back and forth in the tube, periodically breaking the infrared beam. Independent methods indicate a close correlation between infrared beam breaks and overall activity. A 5-minute period of inactivity (i.e., no beam breaks) has been found to be a reliable indicator of sleep. Video-based monitoring also has been coupled to measurements of beam breaks to provide higher spatial resolution analysis of fly sleep behavior.[24,25] The use of consolidated inactivity to measure sleep in flies is similar to the use of actigraphy to measure sleep in humans.[26]

In view of the remarkable similarity between *Drosophila* and mammals, it is not surprising that fruit flies exhibit many of the defining features of sleep. Flies exhibit extended periods of behavioral quiescence that can last for hours, and a majority of the sleep-like episodes typically occur in bouts

longer than 30 minutes.[24] In addition, sleeping flies exhibit reduced responsiveness to sensory stimuli.[2,24,27-29] Indeed, *Drosophila* sleep studies do not rely solely on measures of spontaneous movement but also assess responsiveness to sensory stimuli. Arousal threshold is assayed by application of a stimulus and measuring a behavioral response, typically induction of locomotor activity. During periods of extended immobility, flies are less likely to respond to a range of sensory stimuli, including social, mechanical, vibratory, thermal, and visual.[24,27,30,31] Although this responsiveness typically is measured behaviorally, it also can be uncoupled from movement using electrophysiologic measures.[32] The typical fly demonstrates an increase in arousal threshold, reaching a plateau after 5 minutes.[27,31] The 5-minute criterion for fly sleep is in part based on this observation. More recent analysis suggests that like mammals, flies exhibit deeper stages of sleep, characterized by more elevated arousal thresholds after more than 10 minutes of inactivity.[33] Nonetheless, quiet wakefulness can be distinguished from sleep by assessing arousal threshold.

Of importance, fly sleep is under homeostatic regulation—that is, flies deprived of sleep will exhibit increases in sleep duration and intensity (the latter as measured by sleep bout length or arousal threshold) the following day. Flies typically are deprived of sleep mechanically using automated devices or by tapping the flies by hand[24,27,31,34] (see Figure 28-2). Sleep rebound is not observed, or is much less evident, if similar deprivation protocols are applied to flies that are already awake, arguing strongly against nonspecific stress effects of mechanical disruption.[24,27,31] Continuous sleep deprivation ultimately results in premature death,[34] as it does in some mammals.[35] Thus sleep is essential for life in the fly.

Although flies do not display the precise electroencephalographic signatures of mammalian sleep (that is, the synchronous changes in neural activity seen as slow waves on the electroencephalogram [EEG] that are seen in mammalian sleep have not been observed in *Drosophila*),[30] they do exhibit electrical correlates of sleep behavior, providing behavior-independent state markers. In vivo electrical correlates in behaving flies have relied on the development of novel approaches to study the fly brain. In this approach, recording electrodes are inserted into the center of the *Drosophila* brain and into the optic lobes of a tethered fly[30] (see Figure 28-1). Local field potentials (LFPs) are measured, reflecting neural activity near the electrode. Leg movement in the tethered fly is simultaneously monitored. There is a general-but-not-perfect correlation between spike-like potentials recorded from the central brain and waking movement. If flies are exposed to a rotating stripe, LFPs in the 20- to 30-Hz frequency are observed, reflecting attention to the stimulus, but these LFPs are reduced when the fly is asleep.[32] In addition, periods of poor correlation between LFPs and movement are associated with increased arousal threshold and precede behavioral quiescence.[32] These approaches clearly demonstrate that differences in arousal states can be characterized using electrophysiologic correlates as they are in more complex organisms. The differences between fly and mammalian neuroanatomy are likely to account for the differing electrical manifestations of sleep even if the underlying molecular and cellular mechanisms are similar. Studies of eye development provide some precedent for the idea that common genetic mechanisms can underlie anatomically distinct but functionally analogous structures.[36]

In addition to the core features of sleep, flies also display age-related changes in sleep architecture similar to those in aging mammals. Directly after emergence from the pupal case, young flies exhibit increased amounts and depth of sleep, as in their mammalian counterparts.[27,37] With increasing age, sleep becomes more fragmented and less consolidated.[38] In addition, drugs that increase oxidative stress can induce changes that mimic these effects.[38] Thus the fruit fly has the potential to become a valuable model for the analysis of aging effects on sleep.

Mammalian sleep enhances various aspects of memory consolidation.[39] Similarly, flies also display sleep-loss related deficits in learning and memory, using a number of different learning paradigms. For example, flies normally exhibit phototaxis but can be trained to avoid light using aversive stimuli. However, flies that have experienced reduced sleep perform more poorly on this task.[40] In courtship conditioning, male flies learn to stop courting females that have already mated. Under the appropriate conditions, males can remember this experience for longer than 24 hours; however, if flies are subjected to sleep deprivation after training, that is, during the period of presumed memory consolidation, they fail to retain this memory.[41] Finally, waking experience, in particular, social experience, can increase subsequent sleep amount, a process that may utilize dopamine and cyclic adenosine monophosphate (cAMP) pathways.[41] Taken together, these data implicate a reciprocal relationship between sleep-wake regulation and plasticity and memory in *Drosophila,* as is proposed for mammals.

DROSOPHILA CIRCADIAN BEHAVIOR REVEALS CONSERVED MECHANISMS BETWEEN FLIES AND HUMANS

The best case for the argument that *Drosophila* genetics will illuminate the genetics of human sleep has emerged from studies of circadian behavior. As in most (but not all) organisms, sleep is under temporal control of a circadian clock in *Drosophila.*[24,27] The first-identified fly circadian mutants displayed short- and long-period rhythms in constant conditions and phase-advanced and delayed activity in light-dark conditions, analogous to human advanced and delayed sleep phase syndromes.[42,43] Cloning of the genes responsible for these fly phenotypes led to breakthroughs in the understanding of the core biochemical mechanisms of circadian timing. These studies also provide an experimental roadmap for elucidating basic mechanisms of sleep homeostasis. Although circadian clocks often have been viewed solely as timekeepers, both circadian genes and their accompanying neural circuits extensively regulate sleep-wake, perhaps independent of their timing functions (see later). Thus a deeper molecular understanding of the circadian system should provide insights into the control mechanisms for sleep.

Most aspects of the fly molecular clockwork are conserved with mammals, including humans (Table 28-1).[44,45] People affected by familial advanced sleep phase syndrome exhibit an advanced phase of sleep-wake rhythms and shortened circadian period that is inherited in a mendelian dominant manner.[43] Mutations in the human *PER2* and *CK1delta* genes, orthologues of fly circadian genes *period* and *double-time,* respectively, are responsible for this advanced sleep phase.[46,47] These data argue that the basic architecture and

Table 28-1 *Drosophila* Clock Genes and Their Highly Conserved Mammalian Homologues

Drosophila	Mammals
Period	*Period 1, 2, 3*
Timeless	*Timeless*
Clock	*Clock, NPAS2*
Cycle	*Bmal1*
Doubletime	*CK1δ/ε*
CK2	*CK2*
Cryptochrome	*Cryptochrome 1, 2*
Clockwork orange	*Dec1, 2*
Slimb	*β-TRCP*

core components of circadian clocks can be traced back to the shared ancestor of flies and humans hundreds of millions of years ago. In view of the close association of circadian with sleep behavior, the basic mechanisms of sleep homeostasis presumably also may be conserved with flies.

CELLULAR AND MOLECULAR BASIS OF *DROSOPHILA* SLEEP

As described earlier, considerable evidence indicates that fruit flies display the core characteristics of sleep. Substantial progress has been made in identifying discrete molecular and neural circuits that convey signals to time sleep and wake behavior. Presented next is an overview of the neural circuits that contribute to sleep-wake behavior, along with the genes that regulate sleep and that are regulated by sleep. The emerging picture (and "take-home message") is that the molecular mechanisms governing *Drosophila* sleep may be shared with animals with more complex nervous systems.

Specific Neural Circuits Are Important for Sleep-Wake Regulation

A theme of mammalian sleep studies is the notion that discrete neural circuits are important for initiating and maintaining sleep and wake states. Using genetics-based tools to study such circuitry in live animals, distinct neural circuits have been identified that regulate sleep in *Drosophila*. Thus far, several anatomically defined loci have been implicated in sleep-wake regulation: the mushroom bodies, the dorsal fan-shaped body (dFB), the pars intercerebralis (PI), and the circadian pacemaker neurons—the small and large ventral lateral neurons (sLNv and lLNv neurons) (Figure 28-3). In addition to these loci, other circuits have been defined according to their specific neurotransmitter (e.g., dopamine). These circuits are discussed further later on. To discover novel circuits involved in sleep regulation, an approach akin to forward genetics has been employed with the modification that, instead of screening for genes, circuits are screened in an unbiased manner for behavioral functions.

A cornerstone of this approach is the binary GAL4/UAS system.[48,49] In one parental strain, the yeast transcription factor GAL4 is placed under the control of a tissue-specific promoter. In the second strain, the upstream activating sequence (UAS) bearing GAL4-binding sites is fused to an effector gene of interest. In the progeny of these two strains, the effector gene is expressed in the distribution specified by the tissue- or circuit-specific promoter driving GAL4. A plethora of GAL4 lines are available that provide a nearly limitless display of temporal and spatial expression patterns.

In addition to the multitude of GAL4 lines, numerous UAS effector lines have been generated, including those

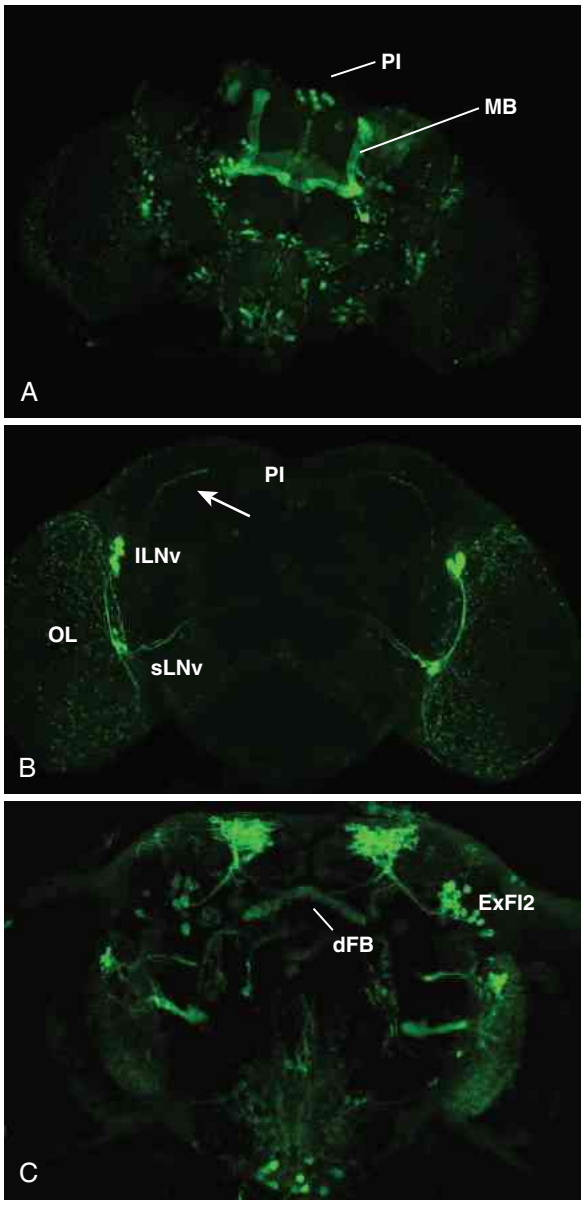


Figure 28-3 Neuroanatomy of *Drosophila* Sleep-Wake Circuits. **A,** The sleep regulatory mushroom bodies (MB) and pars intercerebralis (PI) neurons are labeled with green fluorescent protein (GFP). **B,** Large and small ventral lateral neurons (lLNv and sLNv, respectively) labeled with GFP. The arousal-promoting large subset of neurons send projections to the ipsilateral and contralateral sLNv and optic lobes (OL). sLNv sends projections to the PI (*arrow*). **C,** The sleep-promoting ExFl2 neurons are labeled with GFP. The ExFl2 neurons send projections to the terminal dorsal fan-shaped body structure (dFB) in the central complex.

engineered to alter specific cellular properties such as membrane excitability or synaptic transmission. Although these approaches are well developed in *Drosophila,* they have inspired even more sophisticated strategies in mammalian models.[50] A number of cellular effectors have been successfully utilized for *Drosophila* sleep studies. One tool that has been used to conditionally block synaptic transmission is the UAS-driven *shibire*^{ts1} (*shi*^{ts1}) transgene.[51] This transgene encodes a multimeric guanosine triphosphatase (GTPase) required for vesicle scission, a process that is in turn required for synaptic vesicle recycling and hence the maintenance of fast synaptic transmission. UAS-driven expression of the *shi*^{ts1} allele in a wild-type neuron can block synaptic transmission at an elevated restrictive temperature (e.g., 29° C) but not at the permissive temperature (e.g., 21° C). Using the GAL4/UAS system, and given the fact that flies are not homeotherms, it is possible to specifically manipulate synaptic transmission in discrete neural circuits in a live-behaving animal, with temperature acting as a remote control, and then assay the behavioral consequences of circuit modulation. Tools to manipulate cellular excitability also have been developed and applied using ectopically expressed constitutive and conditionally active ion channels that can activate or silence neuronal activity. For example, a bacterial depolarization-activated sodium channel, NaChBac, and the thermosensitive TrpA1 channel have been used to increase cellular excitability, the latter in a temperature-dependent manner.[52,53] Conversely, a variety of engineered potassium channels have been employed to silence neurons.[54,55]

The use of these transgenic tools in combination with various GAL4 drivers led to the discovery of a sleep regulatory role for the mushroom bodies (MBs), paired (bilateral) neuropil wells known for their role in learning and memory.[12,56,57] The MBs are analogous to the mammalian cerebral cortex and hippocampus.[58,59] Conditional inhibition of the MBs using *shi*^{ts1} as well as chemical ablation reduces overall sleep levels,[56,57] and in both cases the sleep decrease was largely due to reduced sleep bout length, that is, an inability to maintain sleep.[56] In addition, flies with impaired or absent MBs exhibited a reduced lifespan consistent with a loss of restorative sleep.[56] Thus these flies are analogous to insomniac patients who are unable to maintain sleep and who may suffer adverse consequences as a result. Other data using a different MB-GAL4 line that can be conditionally activated in adult flies[60,61] in combination with activating and silencing tools[57] suggest that the MBs also may promote wakefulness. One possibility is that different subsets of MB neurons play opposing roles in sleep regulation. Sleep and memory functions may overlap in the MBs, an important site for the function of adenylate cyclase in short-term memory[62] and cAMP-dependent protein kinase A (PKA) activity in sleep.[57] The connection between sleep and memory is particularly intriguing in view of the proposed function of sleep in regulating synaptic plasticity.

Flies that lack MBs demonstrate both spontaneous sleep (albeit reduced) and a robust homeostatic response, indicating that other brain loci promote sleep. Decades of research in mammals has led to the identification of discrete sleep and arousal centers in the brain.[63] For example, the gamma-aminobutyric acid (GABA)-ergic ventrolateral preoptic nucleus (VLPO) located in the anterior hypothalamus promotes sleep, whereas monoaminergic and acetylcholinergic nuclei in the brainstem and posterior hypothalamus enhance wakefulness.[63] The dFB in *Drosophila* appears to be an analogue of VLPO in mammals because activation of the dFB using TrpA1 induces sleep (see Figure 28-3).[64] In addition, in a manner similar to how VLPO activity is increased with sleep deprivation in mammals,[65] sleep deprivation in flies enhances the excitability of dFB neurons.[66] These data suggest that VLPO in mammals and the dFB circuit in flies act downstream of sleep-homeostatic mechanisms to regulate sleep. Of interest, inducing sleep by activation of the dFB circuit facilitates memory consolidation, further emphasizing the intimate relationship between sleep and memory.[64]

Another sleep regulatory locus is the PI, a neuroendocrine cluster considered genetically analogous to the mammalian hypothalamus, an important player in sleep regulation.[67] Targeted decreases in epidermal growth factor (EGF) function in the PI result in reduced sleep.[68] EGF and its related set of ligands (e.g., transforming growth factor alpha) appear to serve similar functions in promoting sleep in *C. elegans*[69] and mammals,[70,71] suggesting an ancient sleep function for EGF.

As discussed further on, specific PI neurons act downstream of monoaminergic circuits to promote arousal.[72] In addition, the PI may be a direct target of circadian pacemaker neurons. The dorsal projections of a subset of circadian pacemaker neurons, the sLNv neurons, terminate in close proximity to the PI neuron soma (see Figure 28-3), and loss of the key peptide transmitter of these neurons, pigment-dispersing factor (PDF), affects molecular circadian rhythms in PI neurons.[73]

In addition to sleep-wake circuits in the MBs, dFB, and PI, the circadian pacemaker LNv neurons also regulate wakefulness and sleep (see Figure 28-3). The *Drosophila* clock circuit comprises approximately 150 neurons,[74] which together are analogous to the mammalian circadian pacemaker, the suprachiasmatic nucleus. Within this network are the two subgroups of LNv neurons mentioned earlier, the large (lLNv) and small (sLNv) clusters, which have opposite roles in regulating sleep. The lLNv neurons promote arousal, whereas the sLNv neurons induce sleep. Excitation of the lLNv using NaChBac or TrpA1 reduces sleep, especially at night.[75-77] Selective ablation of the lLNv results in increased sleep.[76] The vigilance effect is similar to that observed in animals in which the mammalian circadian pacemaker, the suprachiasmatic nucleus, is ablated.[78] The activity of these arousal-promoting neurons appears to be inhibited by GABA,[77] a relationship that is reminiscent of a similar organization of mammalian sleep circuits.[79] By contrast, the sLNv neurons appear to enhance sleep by inhibiting the lLNv cells through the action of short neuropeptide F (sNPF) (which is related to mammalian NPY).[80] These studies demonstrate that distinct clock subcircuits can promote wake or sleep, implying that similar principles may apply to the mammalian suprachiasmatic nucleus.

Arousal Neurotransmitters: Monoaminergic Arousal Pathways

A number of different neurotransmitters and neuromodulators have been shown to play important roles in sleep regulation. Whereas the anatomic organization of the *Drosophila* brain is distinct from that in mammals, flies use a similar system of neurotransmitters and receptors. Indeed, it appears that the transmitters used by flies to regulate sleep are similar to those used by mammals. In mammals, monoamine transmitters,

such as dopamine, histamine, and norepinephrine, generally promote wake-arousal.[81] Similarly, in flies, these monoamines, or their fly counterparts, also promote wakefulness, suggesting that the nervous system of the common ancestor of flies and mammals used similar arousal transmitters.

The monoamine most strongly linked to arousal in *Drosophila* is dopamine. Flies with mutations of the dopamine transporter gene *fumin* exhibit dramatically reduced sleep duration (approximately 50%).[82,83] Of note, mutations in this gene also were identified in a large-scale unbiased genetic screen for sleep mutants.[83] The psychostimulant methamphetamine, which is thought to increase dopaminergic neurotransmission, also reduces sleep.[84] When awake, flies with enhanced dopamine neurotransmission display increased spontaneous locomotor activity as well as hyperresponsiveness to mechanosensory stimuli, suggesting that these flies were hyperaroused.[82-84] In addition, although its mechanism of action remains unclear, the clinically prescribed wake-promoting drug modafinil may operate by enhancing dopaminergic neurotransmission.[85] Of importance, modafinil has similar wake-promoting properties in *Drosophila.*[86]

In mammals, the predominant model for how sleep and wake circuits intersect involves mutually inhibitory interactions, promoting fast switching between bistable states.[63] Of interest, in flies, arousal-promoting dopaminergic neurons project to and inhibit the sleep-promoting dFB circuit.[87,88] This dopaminergic (DA-dFB) circuit also plays a critical role in age-dependent effects on sleep. This circuit is hypoactive when flies are very young (0 to 1 days old) compared with that in adults (approximately 8 days old).[37] Finally, the dopamine-dFB circuit also is involved in mediating the effects of a volatile anesthetic on arousal in flies,[89] which parallels the finding in mice of increased activation of the VLPO circuit by volatile anesthetics.[90] Dopaminergic neurons densely innervate the MBs,[91] and type D_1 dopamine receptor (dDA1) activation in the MB may mediate the wake-promoting effects of caffeine[92] and can mitigate the effects of sleep loss in an MB-dependent learning paradigm.[40] Taken together, these data indicate that dopamine is an important transmitter for arousal and cognitive function in *Drosophila*.

Two other monoaminergic transmitters implicated in fly arousal are octopamine and histamine. Octopamine is thought to serve as a functional homologue of mammalian norepinephrine. Genetically reduced octopamine synthesis or octopamine neuron activity results in increased sleep which can be rescued in the former case by pharmacologically restoring octopamine.[93] These effects require PKA but do not operate through the MB.[93] Instead, specific wake-promoting octopamine cells send projections to the PI circuit discussed earlier.[72] Octopamine acts to inhibit the function of a Ca^{2+}-dependent potassium channel and also increases cAMP signaling in these PI neurons to promote wakefulness.[72] Histamine has been implicated principally by pharmacology-based studies. The histamine H_1 receptor antagonist hydroxyzine induces sleep and reduces sleep latency in flies,[27] suggesting conserved functions for histamine. A large-scale small-molecule screen was performed in adult flies that identified reserpine, which is an inhibitor of vesicular monoamine transporter (VMAT), as causing an increase in sleep.[94] In light of the fact that the function of VMAT is to load monoaminergic neurotransmitters in vesicles, these data further reinforce the importance of monoamines in promoting arousal in flies. It should be mentioned that not all monoamines are arousal-promoting in *Drosophila* (e.g., serotonin).[95] Nonetheless, the role for monoamines in promoting arousal in mammals is largely preserved in *Drosophila*.

Sleep Neurotransmitters: Gamma-aminobutyric Acid and Adenosine Sleep Pathways

A crucial neurotransmitter for sleep promotion in both flies and mammals is the inhibitory neurotransmitter GABA. Many of the most commonly prescribed hypnotics act at ionotropic GABA receptors, promoting GABAergic neurotransmission.[96] Silencing of GABAergic neurons reduces sleep, a finding consistent with a sleep-promoting role.[97] In *Drosophila*, a mutation in one $GABA_A$ receptor subunit gene is responsible for resistance to the dieldrin insecticide; hence its name, *Resistant to DieLdrin* (*Rdl*). These receptors rapidly desensitize on GABA activation, and this process is reduced in insecticide-resistant *Rdl* mutants, prolonging GABA-activated currents.[97] In these *Rdl* mutants, the latency to sleep after lights-off is increased, an observation consistent with an important role for GABA in promoting sleep in *Drosophila.*[97] *Rdl* may promote sleep in part by reducing the activity of PDF arousal-promoting neurons (see earlier; Figure 28-3).[77,98,99] In view of the conserved role for ionotropic GABA receptors in sleep regulation, it will be of interest to determine whether flies respond to many clinically used hypnotics that target this receptor class.

Adenosine is thought to play a key role in sleep homeostasis in mammals.[100] Adenosine also has been implicated in the promotion of sleep in *Drosophila.* Adenosine, a metabolic product of adenosine triphosphate, acts through specific G protein–coupled receptors.[100] In mammals, the stimulant effects of caffeine are thought to operate by antagonizing adenosine receptors.[101] Flies fed caffeine exhibit reduced sleep, whereas flies administered cyclohexyladenosine, an adenosine agonist, exhibit increased sleep.[24,27] However, caffeine may function differently in flies versus mammals. Besides antagonizing adenosine signaling, caffeine, like other methylxanthines, has been shown to inhibit cAMP phosphodiesterase activity.[101] Deletion of the single adenosine receptor in flies did not block the wake-promoting effects of caffeine.[102] Instead, reduction of neuronal PKA activity largely suppressed the effects of caffeine on sleep.[102]

GENETICS AND PHARMACOLOGY OF SLEEP: WHICH MOLECULES REGULATE SLEEP?

The power of the *Drosophila* system lies in the ability to identify genes whose function is important for sleep. One strategy is to manipulate the function of genes and assay the consequences on sleep. In addition, traditional pharmacologic approaches have complemented genetics to identify pathways whose function is important for sleep. Finally, genomic approaches have been applied to identify genes whose expression is regulated by sleep-wake state. The discovery of novel sleep genes in *Drosophila* using genetics has relied on a combination of candidate gene approaches and classical forward genetics. In addition, the molecular identification of quantitative trait loci contributing to sleep using inbred strains also

has identified numerous candidate sleep genes.[103] Of note, unbiased large-scale screens are especially powerful because they tend to identify the strongest contributors to a process among thousands of mutagenized candidates. Not surprisingly, identified genes play a role in various aspects of neural function, including genes involved in the circadian system, stress and immune responses, signal transduction, neurotransmitter/neuromodulator signaling, and cellular excitability. These findings suggest that many sleep pathways are conserved between flies and mammals and support the notion that studies in *Drosophila* should yield insights into the molecular basis of sleep in more complex systems.

Circadian Clock Pathway

As in many mammalian species, sleep is under the control of a circadian clock in *Drosophila.*[24,27] Mutations in the core clock transcription factors Clock (Clk) and cycle (cyc) result in reduced sleep.[99,104] However, in certain arrhythmic mutants, such as *per*^{01}, sleep homeostasis is intact, as is consistent with the two-process model.[24,27] Some of these clock gene effects are likely to be mediated by the PDF neuropeptide through its function in the arousal promoting lLNv[76,77] (see Figure 28-3).

Although a detailed molecular understanding of the nature of the core circadian oscillator has emerged, the mechanisms by which this core clock regulates sleep is poorly understood. A novel molecule named Wide Awake ("WAKE") is a likely candidate for a key clock output molecule that specifically regulates the timing of sleep onset.[99] WAKE exhibits cycling expression in the lLNv clock neurons, peaking in the evening, when it acts to silence these arousal-promoting cells by upregulating GABA transmission. Strikingly, a single homologue of WAKE has been identified in mice, which is enriched in the suprachiasmatic nucleus, the master circadian pacemaker in mammals, suggesting that its function in mediating circadian regulation of sleep is conserved.[99]

Unbiased genetic screens often uncover unexpected molecular pathways in sleep regulation. Using in vivo RNA interference, the essential cell cycle gene, encoding cyclin A, functions in nondividing neurons to regulate sleep duration and homeostasis.[105] Of note, this cyclin is expressed in a small number of adult brain neurons that are in proximity to clock neurons.[105] As supported by the extensive understanding of both molecular and circuit basis of circadian behavior that has been achieved, elucidating the clock links to sleep should continue to be a fruitful area of research.

Dopamine Arousal Pathways

Dopamine is a major arousal-promoting neuromodulator in flies (see earlier). A key player in this process is protein ubiquitination, mediated by the E3 ubiquitin ligase cullin-3 (Cul3) and the BTB Cul3 substrate adaptor insomniac (inc). Post-translational covalent modification of proteins by ubiquitin polypeptides typically targets these proteins for degradation but also influences diverse processes like receptor trafficking. Loss of either *Cul3* or *inc* dramatically reduces sleep levels and sleep bout lengths and suppresses homeostatic responses to sleep loss.[106,107] Inhibition of DA synthesis can reverse these sleep phenotypes.[106]

BTB protein-protein interaction motifs also have been implicated as a genetic contributor to restless legs syndrome (RLS). Loss of the *Drosophila* homologue of the RLS gene *BTBD9* (*dBTBD9*) results in fragmented sleep and increased locomotor activity, which parallels the clinical disorder.[108] The dBTBD9 protein appears to function in dopaminergic neurons and is important to maintain dopamine levels, consistent with the established role for dopamine in RLS.[108] These data provide important validation of using fly models for human sleep disorders.

Stress and Immune Pathways

Studies of the role of the circadian clock gene *cyc* (*Bmal1* in mammals) led to the discovery of a role for heat shock stress response genes in sleep homeostasis. Female, but not male, *cyc* mutants display an exaggerated sleep rebound.[34,104] In addition, *cyc* mutants, both male and female, but not other clock mutants, are hypersensitive to the lethal effects of sleep deprivation.[34] These effects appear to be due to inadequate expression of heat shock proteins, protein chaperones important in the cellular response to stresses, such as elevated temperature.[34] This study suggests that the heat shock stress response is an important pathway for defending against the adverse consequences of sleep loss (e.g., death).

In addition to the heat shock stress pathway, endoplasmic reticulum (ER) stress responses also may be important for sleep homeostasis. The ER stress pathway is both regulated by and regulates sleep. The ER chaperone BiP is upregulated during wake and sleep deprivation.[27,109] In addition, the amount of rebound sleep after sleep deprivation is dependent on BiP levels.[109] BiP plays an important role in the stabilization and translocation of newly synthesized secretory proteins from the cytosol to the ER. BiP also is upregulated as part of the unfolded protein response (UPR), which is activated in response to an abundance of unfolded proteins in the ER. The UPR is itself upregulated in mammals subjected to sleep deprivation.[110] These data suggest that the UPR response, and subsequent BiP activation, may occur as a consequence of extended wakefulness and support a significant role for ER stress pathways in sleep homeostasis.

Another important regulator of *Drosophila* sleep is the immune response. Infection resistance and immune response genes, including the immune system master regulator nuclear factor-kappaB, Relish, are upregulated in response to sleep deprivation.[111] Decreased Relish function in the fat bodies, a key immune response tissue in *Drosophila,* results in reduced sleep.[111] Of note, the immune-related cytokines are important regulators of sleep in mammals.[112]

Membrane Excitability

A role for membrane excitability in sleep regulation is evident from studies using engineered heterologous ion channels that regulate membrane excitability (see earlier); however, these studies leave open the question of which specific channels normally underlie sleep function. Studies in *Drosophila* have highlighted the function of the voltage-gated potassium channel *Shaker* (*Sh*). The *Sh* mutant was discovered several decades ago as a mutant whose legs shake under ether anesthesia.[113] Positional cloning of this mutant led to the identification of the first voltage-gated potassium channel and subsequently several similar channels in mammals, highlighting the similarity in fly and mammalian nervous system components.[114-116]

Independent unbiased mutagenesis screens identified mutants in the Sh potassium channel and a novel SH regulator, called Sleepless (SSS), that exhibit dramatically reduced sleep amounts, with loss of as much as 80% of total sleep in a *sss* mutant.[117,118] The gene *sss* encodes a glycosylphosphatidyl-linked membrane protein,[118] and strikingly it was found that SSS directly regulates the levels, localization, and function of Sh channels.[119] In addition, mutants of an Sh regulatory subunit, *Hyperkinetic* (*Hk*), also exhibit a reduced sleep phenotype.[120] Short-sleeping *Sh* and *Hk* mutants display reduced memory in a short-term memory paradigm, providing a genetic link between reduced sleep and cognitive function.[120] The function of Sh in sleep is highly conserved as genetic inactivation of mammalian Sh orthologues also results in reduced sleep.[121,122] Of importance, severe insomnia is a prominent complaint in persons with Morvan syndrome and other autoimmune disorders characterized by antibodies to voltage-gated potassium channels,[123,124] which further highlights the conserved role of Sh and membrane excitability in sleep regulation. Another molecule recently shown to regulate sleep through changes in excitability is the rho-GTPase–activating protein Crossveinless-C (CV-C).[66] CV-C acts specifically in the dFB circuit to regulate its excitability and appears to influence sleep amount in response to homeostatic inputs, although precisely how CV-C modulates excitability remains unknown.[66]

Signal Transduction

Components of the cAMP signaling pathway also play conserved roles in sleep regulation. Various neurotransmitters act at cell surface G protein–coupled receptors to activate intracellular signal transduction cascades via metabotropic receptors (e.g., dopamine). Activation of G protein–coupled receptors such as dopamine receptors leads to an increase or decrease in adenylate cyclase activity, which modulates cAMP levels. Cyclic AMP in turn activates PKA, which phosphorylates a number of targets, including the transcription factor CREB (cAMP response element–binding protein). Mutants affecting this pathway that increase activity (e.g., increase cAMP or PKA activity) result in increased wake, whereas mutants that decrease cAMP levels generally reduce wake.[57,125] Furthermore, CREB activity is linked to sleep homeostasis. A CRE reporter gene is upregulated in response to sleep deprivation, and reduced CREB activity results in an elevated sleep rebound.[125] In addition to a wake-promoting role for this pathway in *Drosophila,* the cAMP pathway serves similar functions in both nematodes and mice.[126,127] Many of the mutations that affect the cAMP pathway were originally isolated in unbiased genetic screens for mutations that disrupt learning and memory. Thus signaling important for sleep and memory may intersect at cAMP pathways.

In mammals, a substantial body of evidence has demonstrated that acetylcholinergic signaling, originating from the lateral dorsal and pedunculopontine tegmental nuclei, promotes wakefulness and also REM sleep. Recent work in flies suggests that nicotinic acetylycholine receptors play a role in the sleep homeostatic output pathway. The *redeye* (*rye*) gene, which encodes a nicotinic acetylcholine receptor alpha subunit,[128] recently has been shown to be required for normal amounts of sleep in flies. What was unusual about Rye, however, is that levels of this protein cycle in response not to circadian time but rather to sleep need, suggesting that it acts as a sleep homeostatic output molecule. At present, however, the circuits that Rye regulate remain unclear, and it is unknown whether Rye itself acts in dFB neurons or elsewhere to modulate sleep.

WHICH GENES ARE REGULATED BY SLEEP-WAKE?

Although the focus of classical genetics is to identify those genes whose function is important for a process of interest, it also is of interest to determine how sleep-wake may in turn regulate gene function or expression. A number of sleep-wake–sensitive changes in gene expression have been described, including the activity of a CRE reporter,[125] upregulation of the ER chaperone BiP,[109] and immune system genes.[111] The most extensive attempts at identifying sleep-wake–regulated genes have used DNA microarrays to assess genome-wide gene expression under conditions of sleep, wake, and sleep deprivation.[129,130] To control for circadian regulation of gene expression, sleep-deprived and spontaneously asleep animals are analyzed at the same circadian time. By identifying the genes that correlate with a behavioral state, one is presumably identifying both factors that sustain and/or reflect that state. In addition, one can monitor gene expression linked to homeostatic drive by screening for genes whose expression increases with the duration of wakefulness. These changes in gene expression may provide clues to the molecular processes occurring during sleep and thus may reveal its underlying function.

One of the most compelling theories for the function of sleep relates to its role in homeostatic control of synaptic strength, the so-called synaptic homeostasis hypothesis. Sleep deprivation and also waking experience in *Drosophila* are accompanied by elevated expression of synaptic proteins throughout the brain,[131] as well as in larger and more numerous synapses.[132] In addition, social experience increases sleep levels and synaptic proteins in a subset of clock pacemaker neurons, the PDF^+ large LNv, which in turn mediate social experience–induced effects on sleep.[133] A number of plasticity-related genes have been implicated in mediating the sleep-plasticity link.[132,133] These observations indicate that the fly model will be a valuable platform to reveal the molecular links between sleep and plasticity.

SUMMARY: *DROSOPHILA*

The fruit fly *Drosophila melanogaster* is now well established as a genetic model organism for studying sleep. The track record of *Drosophila* genetics in unraveling the molecular mechanisms underlying the circadian clock supports the idea that this approach will yield key insights into sleep as well. The neural circuitry regulating sleep in flies remains less well characterized than in mammals, where these circuits have been studied for more than 50 years. In view of the ever-expanding technological resources available in fly model systems, however, rapid progress in delineating the circuits regulating sleep-wake is expected. Reflecting the intersection of sleep with multiple biologic processes, a number of pathways and genes have been identified that affect sleep. A current challenge is to integrate these findings to determine how these molecules fit into a hierarchical framework within the context of specific circuits in the regulation and function of sleep.

NEWER GENETIC MODEL SYSTEMS FOR STUDYING SLEEP

The success of the fruit fly as a model for sleep and circadian research has in part motivated the development of other genetically tractable animal models for studying sleep. Discussed next are the advantages and disadvantages of two such models, the zebrafish *Danio rerio* (Figure 28-4) and the nematode worm *C. elegans* (Figure 28-5), and how these newer models are providing insights into the molecular and cellular basis of sleep.

ZEBRAFISH AS A MODEL SYSTEM FOR GENETICS

In the 1970s, seminal work by George Streisinger and colleagues laid the groundwork for the use of the zebrafish *Danio rerio* as a model system for the genetic analysis of vertebrate development.[134] Zebrafish are particularly well suited for these studies owing to their small size (approximately 0.5 cm for larvae and 2.5 cm for adults), ease of maintaining large collections, rapid development, and transparent body (during embryonic and larval stages), which facilitates visualization of internal structures. After hatching, zebrafish spend approximately 1 month in a larval stage and can live up to 2 to 3 years as adults. Zebrafish mating pairs can yield hundreds of progeny, but one relative disadvantage is that their generation time approaches 3 months, which is similar to that for mice and much longer than for flies (approximately 10 days).

Large-scale chemical mutagenesis screens for developmental mutants have been carried out in zebrafish,[135-137] illustrating their suitability for powerful forward genetics approaches. In addition, generating transgenic zebrafish is routine. However, unlike with flies and worms, other genome-wide genetic technologies remain less developed, such as large libraries of transgenic lines for Gal4/UAS binary expression systems, mutation-causing transposon insertions, or RNA interference (RNAi) knockdown (to study the function of genes in specific neural circuits). Another disadvantage for genetic analysis is that the genome of zebrafish is relatively large. Members of a given gene family often are more numerous than in mammals, which can obscure analysis of mutant phenotypes.[138,139]

In contrast with other commonly used genetic model systems such as flies and worms, zebrafish are vertebrates and accordingly have a nervous system whose organization and neuropeptides and hormones are more similar to those in humans.[140-142] Moreover, the nervous systems of zebrafish are much simpler than in mammals (approximately 100,000 neurons in larvae), which facilitates fine-grained dissection of neural circuits. For example, mice have approximately 5000 hypocretin/orexin (Hcrt/Orx) neurons, whereas larval and adult zebrafish have approximately 10 and 50 such neurons, respectively.[143-145] Another advantage of zebrafish models is the ease of performing large-scale drug screens in these animals, which can then be used to inform mechanisms of action of drugs or dissect molecular pathways underlying a behavior of interest, as well as to discover novel potential pharmaceuticals.[146,147] Finally, the transparency of larval zebrafish (and in adults of certain strains) makes possible powerful in vivo imaging approaches to study the activity of neural circuits in real time.

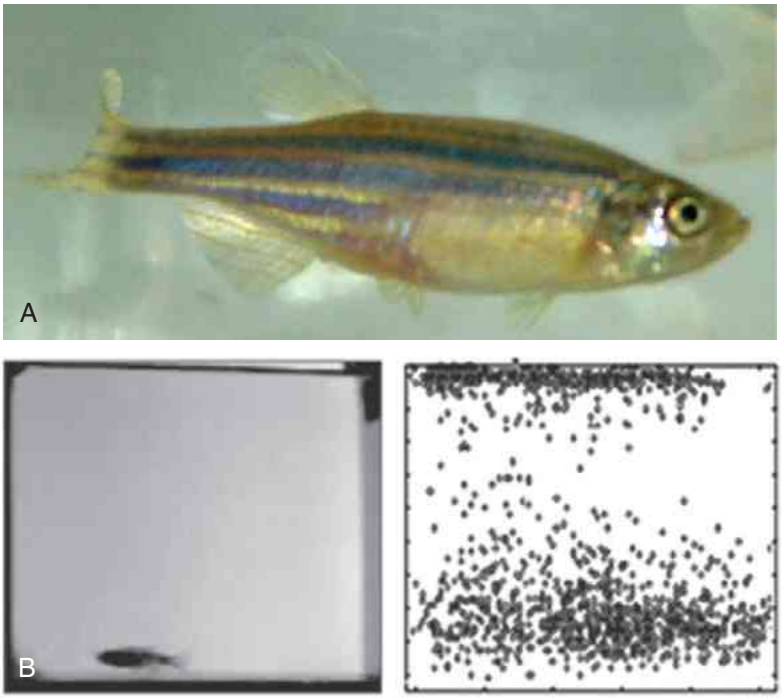


Figure 28-4 Zebrafish, an Emerging Genetic Model Organism to Study Sleep. The zebrafish, *Danio rerio,* is a relatively new genetic model organism for studying sleep in a vertebrate. **A,** An adult zebrafish, which typically grow to approximately 1 inch long. **B,** *Left,* An adult zebrafish sleeping with a drooping caudal fin. *Right,* Preference for the top or bottom of the tank during sleep; *dots* represent location of sleep bouts. (**A** Courtesy S. Liu. **B,** From Yokogawa T, Marin W, Faraco J, et al. Characterization of sleep in zebrafish and insomnia in hypocretin receptor mutants. *PLoS Biol* 2007;5:e277.)

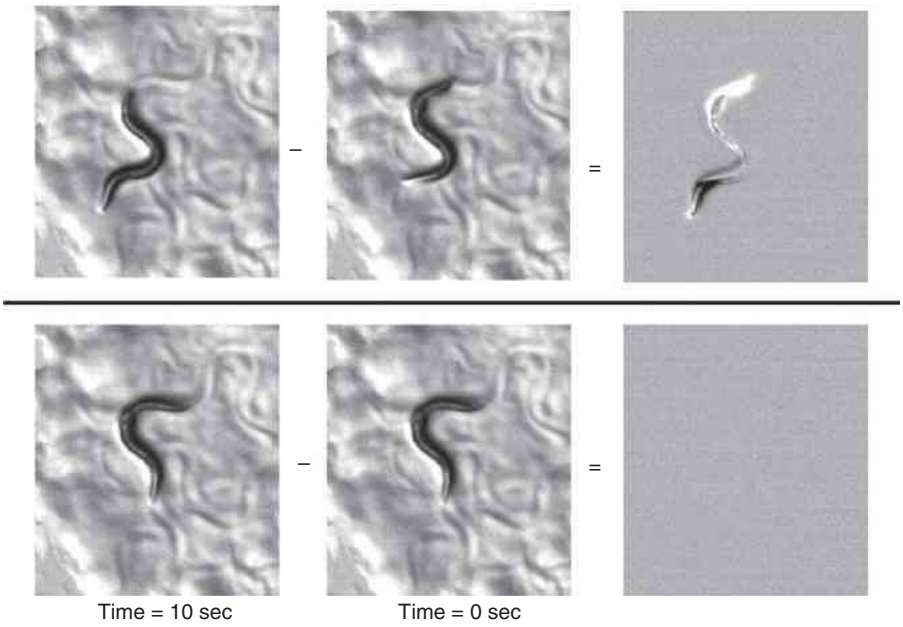


Figure 28-5 *Caenorhabditis elegans* Lethargus as a Model for Studying Sleep. Lethargus is a quiescent developmental stage in the roundworm *C. elegans,* which shares behavioral, molecular, and genetic characteristics with sleep in other animals. Behavioral quiescence in worms is measured using a frame subtraction method, as shown. *Dark pixels* indicate where the animal has moved to, and *white pixels* show where the animal has moved from. (From Macmillan Publishers Ltd: Raizen DM, Zimmerman JE, Maycock MH, et al. Lethargus is a Caenorhabditis elegans sleep-like state. *Nature* 2008;451:569–72; with permission, copyright 2008.)

ZEBRAFISH AS A MODEL SYSTEM FOR STUDYING SLEEP

Like fruit flies and humans, zebrafish are diurnal, with elevated locomotor activity during the day as compared with the night. Sleep in zebrafish has been characterized during two developmental stages: larval and adult stages.[140,141,144,148,149] Zebrafish sleep was first characterized in midstage larvae[148] and later in early larvae[144] and adults.[149] As in fruit flies and other nonmammalian models, sleep is defined in larval and adult zebrafish using behavioral criteria.[140,141] In both stages, zebrafish exhibit periods of quiescence under circadian control, during which they are less responsive to external stimuli (manifested as increased arousal threshold). Furthermore, in general, this behavior is under homeostatic control and associated with postural changes. High-resolution video tracking is employed to measure sleep in zebrafish, and utilizing changes in arousal threshold as criteria, immobility for 1 minute (for larvae) or 6 seconds (for adults) is used to identify a minimum bout of sleep. Also, as in most other organisms, sleep amount in zebrafish changes throughout the lifespan—sleep amounts are higher in larval fish than in adults.[150]

Sleep behavior in zebrafish is under circadian control, as indicated by persistence of its variation throughout the day in the absence of environmental cues. Molecular components of the core circadian oscillator are well conserved in zebrafish,[151] although only one mutation has been identified in a core clock gene (*Clock1*^{dg3}).[152] It is unclear whether zebrafish have a suprachiasmatic nucleus, and even if present, it may not have the central importance in organizing rhythms as in mammals. Instead, because zebrafish are largely transparent and have many light-entrainable cells throughout their body,[153,154] there may not be a need for a single "master clock."

Zebrafish sleep also is under homeostatic control. To examine these phenotypes, sleep deprivation typically is performed using mechanical perturbation or mild electric shock.[148,149] In larval zebrafish, sleep deprivation induced using mechanical vibration led to an increase in "rebound sleep" (i.e., reduced locomotor behavior associated with an increase in arousal threshold).[148] However, homeostatic regulation may not be as pronounced in adult zebrafish. Although shorter periods of deprivation for several hours led to an increase in sleep behavior in adults,[149] the data have been described to be less consistent with more prolonged sleep deprivation.[140] In addition, sleep deprivation induced by prolonged exposure to light does not result in an increase in sleep rebound,[149] although the absence of rebound may reflect the effect of light on other processes, such as circadian clock function.

Signaling Mechanisms Regulating Sleep in Zebrafish

As discussed earlier, a number of monoaminergic and cholinergic cell groups in the brainstem and hypothalamus promote wakefulness in mammals. A majority of these groups—namely, the tuberomamillary nucleus (histamine), the raphe nucleus (serotonin), and locus ceruleus (norepinephrine)—are well conserved in zebrafish.[155] In terms of sleep-promoting centers, zebrafish have a cluster of GABAergic neurons in the anterior hypothalamus,[149] which may be analogous to VLPO in mammals, although the function of this circuit has not yet been tested. In addition to these neuromodulatory transmitters, zebrafish sleep also is regulated by the hormone melatonin and the neuropeptide hcrt/orx, discussed in more detail further on.

Melatonin Promotes Sleep

In mammals, melatonin is produced by the pineal gland and is released at night in both diurnal and nocturnal animals.

Similarly, melatonin is secreted by the pineal gland of zebrafish under circadian control, with higher circulating levels at night. Also as in mammals, direct projections from Hcrt neurons to the pineal gland have been identified in zebrafish, suggesting that the melatonin/pineal gland system probably can be modulated by noncircadian neuromodulatory circuits.[156] However, unlike in mammals, where circadian release of melatonin is driven by the suprachiasmatic nucleus, in zebrafish, the pineal gland is thought to function as an independent circadian oscillator able to generate rhythmic melatonin release.[157,158] This distinction may reflect anatomic differences, whereby the pineal gland can be directly entrained by light in zebrafish, whereas it is not subject to direct light exposure in mammals. In zebrafish, melatonin has significant sleep-promoting effects. Administration of melatonin to larval zebrafish during the daytime leads to a marked increase in sleep time.[148] These effects appear to be more pronounced than those seen in humans, in whom sleep time is only mildly increased with administration of melatonin.[159-161] However, this difference may reflect administration of melatonin during the night done in a majority of studies, when sufficient levels of melatonin are already present.[159] Indeed, if melatonin is administered to humans during the daytime, significant effects of sleepiness can be observed.[162] These studies in zebrafish highlight the role of melatonin as a sleep-promoting factor, which is regulated by circadian timing mechanisms. Of interest, melatonin does not promote sleep in nocturnal rodents,[163,164] and commonly used laboratory mouse strains (e.g., C57BL/6 and 129/Sv) do not synthesize melatonin.[165] Thus zebrafish may be a better model organism than mice for studying the effects of melatonin on sleep.

Hypocretin Stabilizes Sleep-Wake States

Impairment of Hcrt signaling is the key pathogenic mechanism underlying narcolepsy. To gain further insights into Hcrt function and circuitry, investigators have turned to the zebrafish system. As in mammals, a single pre-pro-Hcrt protein is differentially cleaved to yield two distinct peptides, Hcrt-1 (i.e., orexin-A) and Hcrt-2 (orexin-B). However, only a single Hcrt receptor exists in the zebrafish proteome (HcrtR2), in contrast with mammals with their two receptors. Because of the simplicity of the zebrafish nervous system, the number of Hcrt neurons is substantially smaller (approximately 10 per brain hemisphere in larvae and 50 per hemisphere in adults). In larvae, these cells are located in the lateral hypothalamus[143] and have been shown to send extensive projections to multiple arousal centers, including noradrenergic cells of the LC.[144,145] Consistent with this observation, Hcrt2 is broadly expressed and found in monoaminergic arousal centers.[144] Of interest, however, another team of investigators reported that in 2-day-old larvae and adults, Hcrt2 expression was detected not in these monoaminergic cells but rather in GABAergic cells.[149] The difference in these findings may reflect the use of fluorescence versus chromogenic in situ hybridization assays, but as discussed later, the balance of evidence currently favors the model that Hcrt neurons in zebrafish promote wakefulness by signaling to arousal-promoting monoaminergic neurons.

To address the function of Hcrt in zebrafish, one group of workers overexpressed the Hcrt pre-propeptide in larval zebrafish using an inducible heat-shock promoter and observed an increase in wakefulness and arousal.[144] Inducible ablation of Hcrt neurons in larval zebrafish increases daytime sleep and the number of sleep-wake transitions, supporting the notion that Hcrt promotes wakefulness and sleep-wake stability.[166] Moreover, as discussed later, monitoring Hcrt neuronal activity in vivo reveals that these neurons are active during periods of robust locomotor activity, when larvae presumably are highly aroused.[167] By contrast, another group of investigators found that a null mutation in the Hcrt2 receptor led to a reduction in sleep time during the night, with increased fragmentation, in adult zebrafish,[149] which on the surface would suggest that Hcrt signaling promotes sleep. Differences may be due to inducible versus constitutive manipulations. Although it is well recognized that mouse models of narcolepsy as well as human subjects so affected exhibit sleep fragmentation during the night,[168,169] the primary clinical feature in patients with narcolepsy is their profound sleepiness during the day.[170] In summary, these studies suggest significant conservation of the neuronal circuitry and function of Hcrt between zebrafish and mammals and point toward the promise of the zebrafish system in unraveling molecular and cellular mechanisms that regulate this pathway.

Neuropharmacology Regulating Sleep

As is the case for other genetic model organisms for sleep, including mice and fruit flies, zebrafish sleep is responsive to sleep- and wake-promoting drugs. GABA receptor agonists such as diazepam and pentobarbital promote sleep,[148] whereas modafinil promotes wakefulness.[171] However, a clear strength of the zebrafish system is the ease of performing drug-feeding assays in a high-throughput manner. Drugs can be added to the water, and because zebrafish larvae are approximately 0.4 mm long, individual animals can be monitored in 96-well plates.[172] Furthermore, zebrafish larvae readily take up small molecules and lack a functional blood-brain barrier.[147] Recently, a large-scale screen of approximately 4000 unique compounds was carried out using zebrafish larvae to assay the effects of these drugs on sleep-wake behavior.[147] By profiling and comparing the behavioral responses to each compound, this study found conservation of known molecular pathways regulating sleep in other animals, such as monoamines, GABA, adenosine, and Shaker-type potassium channels. Moreover, novel roles for ether-a-go-go–related gene (ERG) potassium channels and L-type calcium channels in sleep-wake regulation were identified. Finally, such drug profiling screens have the potential to reveal mechanisms of action for compounds that are poorly characterized.[147]

In Vivo Analysis of Circuit Function in Zebrafish

As mentioned previously, the transparency of zebrafish embryos and larvae facilitates live imaging approaches in intact animals. Coupled with their useful genetic traits and their relatively simple neural networks, zebrafish hold tremendous potential for the analysis of long-term in vivo imaging as well for systems neuroscience. A growing body of evidence suggests that synaptic structure is regulated by circadian- and sleep-dependent processes.[132,133,173-175] For example, the *Drosophila* PDF^+ sLNv neurons exhibit circadian-dependent changes in their terminal projections.[176,177] To address this issue in larval zebrafish, a recent study utilized repeated two-photon imaging of the same Hcrt neurons over a 24-hour period. This study reported an increase in synaptic terminal

number of Hcrt neurons projecting to the pineal gland during the daytime, suggesting that Hcrt synaptic terminal structure is under circadian control.[178]

Besides neuronal structure, investigations into the function of neural circuits have been revolutionized by the use of the genetically encoded neuronal activity reporters including GCaMP, which measures Ca^{2+} levels.[179] For example, zebrafish researchers have used GCaMP to monitor real-time dynamics of brainwide circuit function underlying optomotor behavior.[180] One drawback for the use of GCaMP for studying sleep-wake circuits is that the light used for excitation will affect behavior, so this technique may not be suitable until the development of Ca^{2+} reporters with excitation spectra in the infrared range. In addition, the identification of which specific cells are active will require the generation of more cell-specific transgenic driver lines. Until then, bioluminescence approaches, which do not require excitation and can measure neuronal activity over longer time scales (minutes to hours), can be used to measure neuronal activity in vivo in freely behaving zebrafish. To examine Hcrt neuron activity in freely behaving zebrafish larvae, the Ca^{2+}-activated bioluminescent reporter green fluorescent protein (GFP)-apoAequorin was expressed in these neurons. The aggregate activity of these neurons, as assessed by total neuroluminescence, was continuously measured and found to be increased during periods of arousal,[167] similar to results obtained in rodents.[181,182] At present, these bioluminescence techniques lack the spatial resolution to distinguish changes in activity in distinct cells. Ultimately, however, the use of genetically encoded reporters should allow researchers to probe brainwide dynamics of sleep-wake circuits in zebrafish, yielding critical information about how these networks function and interact under different circadian and sleep homeostatic conditions.

SUMMARY: ZEBRAFISH

The zebrafish *Danio rerio* has a well-established history as a powerful genetic model system for the study of vertebrate development and has more recently been utilized to study sleep. Advantages of this system include (1) the ability to study the molecular and cellular mechanisms underlying sleep in a vertebrate animal with a relatively simple nervous system; (2) a higher degree of conservation (compared with invertebrates) of sleep-related neuropeptides and hormones; (3) the ability to perform high-throughput behavioral assays; (4) transparency of early stage zebrafish which facilitates in vivo neuronal imaging; and (5) the ease of performing large-scale drug screens. Disadvantages include (1) a long generation time, which can hinder outcrossing to control for genetic background; (2) a relative dearth (in comparison to *Drosophila* and *C. elegans*) of genome-wide transgenic reagents; and (3) a relatively large genome, with a significant number of paralogs, which can inhibit genetic analysis. Until now, sleep studies in zebrafish have largely focused on determining basic characteristics of sleep behavior as well as confirming the function and circuits of previously identified molecules underlying sleep, such as melatonin and Hcrt. With respect to future investigations, the two main strengths of this system relate to the power of large-scale drug screens and the ability to perform in vivo imaging activity of neural networks regulating sleep at a systems level.

CAENORHABDITIS ELEGANS AS A MODEL SYSTEM FOR GENETICS

At first glance, the nematode worm *C. elegans* would seem to be an unlikely candidate for a model organism to study sleep behavior. Adult animals are small (approximately 1 mm long) and have a very simple nervous system, comprising only 302 neurons.[183] Recent work, however, has indicated that *C. elegans* is the simplest animal that exhibits a sleep-like state, termed *lethargus*. As described next, studying this quiescent behavior in worms can reveal shared mechanisms underlying sleep and potentially identify novel genetic mechanisms.

Since the 1970s, when Sydney Brenner first established *C. elegans* as a model system for the genetic analysis of developmental biology and neurobiology,[184] the nematode worm has been a workhorse for the identification of important molecular mechanisms underlying processes such as apoptosis, cell fate decisions, RNA interference, and the neural circuitry underlying simple behaviors. *C. elegans* shares some advantages with zebrafish in that the worms also can be easily cultivated in large numbers (approximately 10,000 in a single petri dish) and are transparent. In contrast with zebrafish, however, adult *C. elegans* nematodes live only 2 to 3 weeks and have a short generation time (approximately 3 days), during which they develop through four larval stages before becoming an adult. Furthermore, the *C. elegans* genome is compact, with approximately 100 million base pairs, and a number of genomewide transgenic reagents are available in worms, including RNA interference libraries and knockout strains. However, one of the most appealing features of the worm system for studying neuroscience is the presence of a complete "connectome." The entire neural network of *C. elegans* has been mapped, with every connection between each of the 302 neurons in their nervous system identified at the ultrastructural level.[183,185] This comprehensive circuit atlas is a great boon for researchers studying how neural circuits regulate behavior in worms.

CAENORHABDITIS ELEGANS AS A MODEL SYSTEM FOR STUDYING SLEEP

As discussed previously, behavioral criteria have been used to define sleep in small nonmammalian animal models.[186] On the surface, worms do not meet the traditional criterion of "behavioral quiescence under circadian control." That is, these worms do not undergo daily cycling of rhythmic behavioral quiescence. It has long been appreciated, however, that the *C. elegans* nematode exhibits consolidated episodes of behavioral quiescence during development.[187] During larval development, these worms undergo a stage termed "lethargus" that precedes each larval molt, in which the animals generally do not move or feed. An important clue linking lethargus to sleep was the finding that the Period homologue in worms, *lin-42,* did not exhibit a daily cycling rhythm as seen in other animals but rather showed episodic expression, coordinated with the timing of lethargus.[188] In further support of *lin-42* regulating rhythmic timing of this behavior, loss of *lin-42* causes arrhythmic molting.[189] Lethargus is thus a behaviorally quiescent state under control of a molecule best known for its role in regulating the circadian clock. Moreover, quiescence during lethargus meets all other criteria for sleep and was

thus suggested to represent a sleep-like state.[126] First, during lethargus, worms assume a specific posture with reduced body curvature.[190] Second, quiescent animals during lethargus exhibit an increased arousal threshold. Third, depriving *C. elegans* of lethargus behavior leads to an increased drive for this behavior, indicating that it is under homeostatic control. Furthermore, bouts of quiescence are longer in duration in the early part of a lethargus period.[190] Strikingly, as is the case for rats and flies,[34,35] complete deprivation of this sleep-like behavior is associated with lethality in worms.[191] Finally, as discussed further on, a number of molecular pathways that have been shown to modulate sleep in other organisms also do so in worms. Because lethargus is associated with molting behavior, an important point is that this relative immobility is not simply due to physical constraints related to the molting process.[126,190]

As is the case for larval zebrafish, measurement of lethargus quiescence in worms uses high-resolution video analysis. These studies typically focus on the lethargus state occurring during the fourth larval stage and have mainly used a frame-subtraction algorithm for image analysis.[25,126,192] Lethargus is a relatively brief state, lasting 2 to 3 hours, and within this period, hundreds of bouts of quiescence, each lasting approximately 30 seconds and seen predominantly during the early part of this stage, are interspersed with bouts of activity. As discussed previously, measurements of sleep in fruit flies and zebrafish use a specified duration of inactivity, associated with a change in arousal threshold, as the minimum bout length for sleep. By contrast, worm researchers measure the fraction of time a worm is quiescent in a moving 10-minute window during the lethargus state, which is easily recognized owing to both developmental time and the absence of feeding activity (pharyngeal pumping). Because these quiescence bouts during lethargus states are relatively brief, and because adult worms do not exhibit this behavior, certain aspects of sleep, including the effects of aging on sleep, probably cannot be addressed using this model system. Nonetheless, the power to perform genetic and neural circuit analyses is nearly unrivaled in worms and makes this system a compelling one to address specific aspects of conserved quiescent behaviors and potentially ancient functions of sleep.

Shared Molecular Mechanisms between Lethargus and Sleep in Other Animals

The worm system is a relative newcomer to the study of sleep, but already molecular pathways are being identified in *C. elegans* that demonstrate conservation of these sleep-regulating pathways in other animals. These studies support the notion that the developmental lethargus stage is analogous to a sleep state in other animals.

Protein Kinase Signaling Can Promote or Inhibit Sleep-Like States

Two kinases have been implicated in regulating lethargus behavior in worms. The cyclic guanine monophosphate (cGMP)-dependent protein kinase (PKG) has been shown to promote behavioral quiescence, whereas the cAMP protein kinase (i.e., PKA) inhibits quiescence during lethargus. For example, loss-of-function mutation in *egl-4,* a worm homologue of PKG, leads to a decrease in lethargus behavior, suggesting that PKG signaling promotes sleep; similar observations were subsequently made in flies.[126,193] In mice, reduced PKG activity is associated with decreased power in the delta band of the electroencephalogram, suggesting a reduced drive to sleep.[194] By contrast, PKA signaling has been shown to promote wakefulness in flies and mice.[57,125,127] In *Drosophila*, overexpression of PKA reduces sleep time, whereas increasing cAMP levels (in the cAMP phosphodiesterase mutant *dunce*) and decreasing cAMP levels (in the adenylate cyclase mutant *rutabaga*) cause a decrease and an increase in sleep time, respectively.[125] In worms, increasing PKA activity through a loss-of-function mutation in the PKA regulatory subunit *kin-2* or a gain-of-function mutation in the adenylyl cyclase *acy-1* reduces the amount of quiescent behavior during lethargus.[190,195]

Growth and Differentiation Pathways Modulate Sleep

Two signaling pathways with broad roles in development and cellular differentiation—EGF and Notch signaling[196,197]—also have been shown to regulate lethargus in worms. The first signaling pathway identified to regulate lethargus was the EGF pathway.[69] In mammals, EGF signaling had previously been implicated in circadian regulation of locomotion and sleep.[71] Further work in flies revealed that EGF signaling enhanced sleep amount.[68] Similarly, ectopic overexpression of the EGF homologue, Lin-3, induced behavioral quiescence, as seen in lethargus.[69] Finally, Notch signaling recently has been shown to regulate sleep-like behavior in both worms and flies. In worms, overexpression of the Notch coligand OSM-11 induces lethargus behavior,[198] and this phenotype depends on Egl-4/PKG as well as the ALA interneuron.[198] Furthermore, double mutants bearing mutations in OSM-11 and another related protein, OSM-7, display a reduction in quiescence during lethargus. In flies, Notch signaling has been implicated in regulation of "rebound" sleep, the increased sleep seen after sleep deprivation. Overexpression of the Notch ligand Delta in the MBs reduces sleep rebound, which also is seen in a Notch gain-of-function mutation.[199] These data thus demonstrate that classical growth and differentiation signaling pathways also can play a role in modulating sleep-like states.

Neuromodulatory Signaling and Lethargus

The activity of neural circuits often is subject to regulation by neuromodulators, which tunes the function of a circuit in response to internal states of the organism or the environment. These neuromodulators include monoaminergic neurotransmitters, GABA, and neuropeptides and modulate a wide range of behaviors and physiologic processes, including sleep.[200-202] As discussed earlier, dopamine signaling promotes wakefulness in flies.[82,87,88,203] A similar arousal-promoting function for dopamine also has been found in *C. elegans* lethargus. A loss-of-function mutation in the dopamine transporter *dat-1* (which leads to excess dopamine in the synaptic cleft) reduces lethargus quiescence, whereas reduced activity of the dopamine receptor *dop-1* results in increased lethargus quiescence.[204] Neuropeptides also play a key role in regulating sleep behavior.[202,205] These neuropeptides typically activate G protein–coupled receptors and, unlike classical neurotransmitters, can act on multiple neurons at once.[201] In *Drosophila*, the neuropeptide PDF plays a key role as a neuromodulator that synchronizes circadian clock circuits and promotes wakefulness.[77,98,99,206,207] Two homologues of PDF exist in *C. elegans*, PDF-1 and PDF-2, and PDF-1 has been shown

to promote arousal.[208] The release of PDF is itself regulated by neuropeptidergic signaling. The FMRF-like peptides FLP-18 and FLP-21 are ligands for the neuropeptide receptor NPR-1, and this signaling pathway inhibits PDF release during lethargus.[208]

Another neuropeptide recently shown to regulate quiescence behavior during lethargus is the neuropeptide-like protein NLP-22.[209] The expression of NLP-22 mRNA cycles, reaching a peak a few hours before the onset of a lethargus period, and can be regulated by Lin-42/PERIOD. Overexpression of NLP-22 during the normally active adult stage induces behavioral quiescence, while loss of NLP-22 function decreases quiescence during lethargus. Of interest, this neuropeptide bears some similarity to neuromedin-S, which is produced in the mammalian suprachiasmatic nucleus and has been suggested to play a role in regulating circadian phase.[210] Thus the finding that NLP-22 is a rhythmic molecule that mediates behavioral quiescence in worms raises the possibility that neuromedin-S may similarly regulate clock-dependent regulation of sleep. Collectively, these findings from kinase, growth/differentiation, and neuromodulatory signaling pathways suggest that lethargus behavior is regulated by similar mechanisms that modulate sleep in other organisms and support the notion that lethargus in *C. elegans* is a sleep-like state.

Simple Neural Circuits Regulating Lethargus

As mentioned previously, every neuron in the *C. elegans* nervous system has been identified and characterized in terms of its connections to other neurons. Thus studying the circuits that mediate lethargus holds significant potential for precisely dissecting the minimal circuits required for a sleep-like state, as well as how these circuits are modulated according to changes in internal or external states.

Individual Neurons that Regulate Behavioral Quiescence in Caenorhabditis elegans

Recent studies have led to the identification of individual neurons required for lethargus behavior in worms. The first neuron shown to regulate lethargus was the ALA interneuron, which receives the Lin-3/EGF sleep-promoting signal.[69] Elimination of the ALA neuron by laser-induced cell ablation suppressed the lethargus-promoting effect of Lin-3 overexpression. More recent data suggest that one of the key functions of the ALA interneuron is to mediate stress-induced quiescence behavior.[211,212] Exposure of *C. elegans* to treatments that induce cellular stress (e.g., heat) lead to behavioral quiescence resembling that seen during lethargus. This stress-induced quiescence behavior has protective effects, suggesting the possibility that one of the ancient functions of sleep may be for recovery from cellular stress.[212] Of interest, stress-induced quiescence behavior depends on a neuropeptide, FLP-13, which shares some homology with the sleep-promoting sNPF neuropeptide in flies.[211]

Another neuron regulating quiescence during lethargus was identified from a forward genetic screen. From a screen of approximately 4,000 mutant strains, a mutation in an AP2 transcription factor homologue, *aptf-1,* was identified that exhibited strongly reduced quiescence during lethargus.[213] Characterization of this gene led to the identification of a single neuron—the RIS neuron—that promotes quiescence during lethargus. Using GCaMP imaging, the RIS neuron was found to exhibit a sharp increase in activity near the onset of quiescent behavior, and laser-mediated ablation of this cell markedly reduced quiescence behavior during lethargus. Moreover, optogenetic activation of the RIS neuron using Channelrhodopsin-2 (a light-activated ion channel) induced behavioral quiescence, which was reversible, and appeared to do so by suppressing the activity of downstream locomotor command neurons.[213] These properties of the RIS neuron resemble that of the sleep-promoting center VLPO in mammals. Thus the identification of these neurons that promote quiescence in worms may provide an entry point to elucidate how various mechanisms such as homeostatic regulation act to influence sleep.

Gating of Sensory Stimuli Is an Important Mechanism Regulating Behavioral Quiescence

An important mechanism involved in regulating sleep is "sensory gating." In mammals, the thalamus is thought to gate sensory input, such that sensory information is processed at a subcortical level during sleep and wakefulness.[214] Such mechanisms would protect sleep continuity as stimuli from minor noises would be filtered out and not reach the cortex to disturb sleep. Emerging studies using genetic and electrophysiologic approaches are addressing these circuit mechanisms in mammals,[215] but studying lethargus in the nematode *C. elegans* also should provide insights into the mechanisms underlying sensory gating. For instance, the effects of Egl-4/PKG and Notch signaling on lethargus depend on their function in sensory neurons.[126,198] Ca^{2+} imaging studies using GCaMP have supported the notion that the activity of sensory neurons is inhibited during lethargus. For example, the ALM mechanosensory neuron exhibits calcium transients in response to gentle mechanical stimulation, and these calcium transients are substantially reduced in amplitude during lethargus.[216] Another sensory neuron that exhibits state-dependent changes during lethargus is the ASH neuron, which responds to aversive mechanical and olfactory stimuli.[217] Functional imaging of the ASH sensory neuron demonstrates reduced activation after exposure to aversive chemical stimuli during the lethargus state. Also during lethargus, the activity of interneurons downstream of ASH also becomes asynchronous, and synchronizing the activity of these neurons promotes arousal.[217]

An additional example of a sensory gating mechanism regulating lethargus is the PDF-1 neuropeptide discussed previously. During lethargus, the release of PDF is inhibited by NPR-1, a neuropeptide receptor that has sequence similarity to the neuropeptide Y (NPY) receptor in mammals. Outside of lethargus, PDF-1 acts on its receptor PDFR-1 in mechanosensory neurons to enhance their sensitivity to touch stimuli, thus promoting arousal. PDFR-1 is required for the increased touch-evoked calcium transients in these neurons in *npr-1* mutants.[208] Collectively, these data suggest that NPR-1 inhibits PDF-1/PDFR-1 signaling during lethargus, which dampens the response of sensory neurons to external stimuli. Such sensory gating mechanisms would promote sleep by inhibiting its interruption by external stimuli. Despite the many advantages of studying a small and simple neural network for sleep, such as in worms, there are also potential disadvantages. Owing to their small size and small numbers, the firing patterns of neurons in *C. elegans* are distinct from those in mammals and even flies. For example, the current

evidence suggest that neurons in *C. elegans* fire classical action potentials; rather, they appear to produce "plateau potentials," with prolonged depolarized or hyperpolarized states.[218,219] Thus how neural circuits in worm encode information may differ significantly from that in other organisms.

SUMMARY: CAENORHABDITIS ELEGANS

The advantages of using *C. elegans* to study sleep include the powerful molecular genetic techniques available for this system, as well as the simple nervous system with a fully defined wiring diagram for its 302 neurons. In support of the hypothesis that lethargus represents a sleep-like state, previous work has established that a number of signaling mechanisms known to regulate sleep in other animals also regulate lethargus in *C. elegans*. Disadvantages of this system include the question of how closely related lethargus is to sleep behavior in other animals. Just as these worms do not have typical 24-hour behavioral and physiologic rhythms, it may be that an animal with so few neurons and whose lifespan is only 2 weeks long may not have the same requirements for specific functions of sleep. In view of current knowledge on this topic, however, it is likely that many genetic and molecular pathways underlying sleep will be conserved in worms, and in future investigations, the major strengths of this system will be the identification of novel molecules and how they act within a precisely defined neural circuit to regulate sleep.

CLINICAL PEARL

Chronic sleep loss can lead to adverse health consequences. Indeed, sleep deprivation in the fruit fly leads to premature death.

SUMMARY

Over the past 15 years, the fruit fly model of sleep has been validated as an important model for the study of sleep. The fly has many of the core features of sleep in common with mammalian systems. Many features of *Drosophila* sleep are independent of changes in spontaneous movement and include elevated arousal threshold, homeostatic regulation, electrophysiologic correlates, and conserved responses to sleep-wake regulatory drugs. In addition, genetic screens have identified shared genes/pathways of sleep control with their mammalian brethren. Zebrafish and *C. elegans* have more recently joined the fruit fly as established nonmammalian genetic model organisms for dissecting the molecular and cellular mechanisms underlying sleep, and each system has intriguing advantages for such study. Future work exploiting the power of genetics in these model organisms promises to reveal the underlying molecular and cellular mechanisms of sleep regulation and ultimately provide some clues to the function of sleep.

ACKNOWLEDGMENTS

We thank David Raizen and David Prober for comments on this chapter.

Selected Readings

Allada R, Chung BY. Circadian organization of behavior and physiology in drosophila. *Annu Rev Physiol* 2010;**72**:605–24.
Axelrod S, Saez L, Young MW. Studying circadian rhythm and sleep using genetic screens in Drosophila. *Methods Enzymol* 2015;**551**:3–27.
Chiu CN, Prober DA. Regulation of zebrafish sleep and arousal states: current and prospective approaches. *Front Neural Circuits* 2013;**7**:1–14.
Cirelli C. Genetic and molecular regulation of sleep: from fruit flies to humans. *Nat Rev Neurosci* 2009;**10**:549–60.
Gandhi AV, Mosser EA, Oikonomou G, Prober DA. Melatonin is required for the circadian regulation of sleep. *Neuron* 2015;**85**(6):1193–9.
Nelson MD, Raizen DM. A sleep state during C. elegans development. *Curr Opin Neurobiol* 2013;**23**:824–30.
Sehgal A, Mignot E. Genetics of sleep and sleep disorders. *Cell* 2011;**147**:207.

A complete reference list can be found online at ExpertConsult.com.

Chapter
29

Genetics and Genomic Basis of Sleep in Rodents

Bruce F. O'Hara; Peng Jiang; Fred W. Turek; Paul Franken

Chapter Highlights

- Sleep, the genome, and the brain are all well conserved across mammalian orders. Rodents, and especially the house mouse, *Mus musculus*, offer the ability to dissect the role of genes in multiple pathways including those that regulate, influence, or underlie sleep and wake.
- A number of these genes, or potential candidate genes, have been identified by a wide variety of genetic and genomic approaches, including studies of gene expression that differ in sleep versus wake, alterations or "knock out" of genes by mutagenesis or directed approaches, and the use of many well-characterized mouse strains that differ in multiple sleep traits and can be crossed to map the causative gene alleles that contribute to this variation.
- Knowledge of these genes and gene alleles has suggested both known and new pathways that influence sleep, including the so-called clock genes that may underlie both circadian rhythms and sleep homeostasis, which occurs most likely in forebrain structures such as the cerebral cortex.
- Genetic and genomic approaches have also been useful as markers of neural activity or used to identify key brain regions involved in arousal state control.

Genetic approaches and gene manipulation have recently added to our understanding of genes involved in sleep architecture and function, identified new brain regions involved in regulation of sleep, and characterized genes associated with sleep disorders and disease-related sleep aberrations. New sleep-related discoveries made by integrating classical and newer genetic methods have validated the utility of genetic approaches for identifying and understanding genes involved in sleep in both health and disease. At this time, the mouse represents the best animal model for such studies, which can take advantage of the rich genetic resources that have already been established in this species, including many genetically modified lines and genetic reference populations that can be used for a wide range of biomedical research. Given the similarities in physiology and genetics across eutherian mammals, it is likely that genes influencing sleep traits in mice also influence sleep traits in humans, and genetic studies of sleep using mouse models will continue to help identify critical molecular pathways and networks underlying sleep functions and regulation.

Genes and their allelic variants are important for an understanding of sleep for a number of reasons. For example, genetic approaches have identified critical functions and the core molecular pathways involved in normal sleep and in sleep disorders. Although the biologic need for sleep itself has been more difficult to study, identification of key regulatory genes involved in sleep homeostasis provides a mechanism to study the fundamental question of the function of sleep. This includes exploring the possible substrates of the restorative function of sleep, such as energy state, protein levels, and synaptic optimization, as well as examining the importance of sleep for mental and physical health. Genes and gene regulation may or may not be directly critical to the regulation of sleep-wake transitions; nevertheless, genes still code for the proteins that are critical to all biologic functions, and genetic analyses may still provide the keys for finding these functions. For example, although changes in gene expression are probably too slow to directly underlie sleep-wake transitions, one could imagine that general levels of gene expression could influence the probability of transitions between non–rapid eye movement (NREM) sleep, rapid eye movement (REM) sleep, and wake. Such genes could encode proteins involved in post-translational changes in ion channels or other proteins that alter membrane potential, and their allelic variants could thus contribute to variations in sleep-wake behavior in mouse and human populations.

This chapter addresses various approaches to find genes and allelic variants that influence sleep behavior. As the functions of sleep are still not completely understood, genetic approaches have been most valuable in identifying genes that affect measurable sleep traits, such as the time spent asleep and its distribution over the day. The genes that regulate sleep may or may not be different from those that are involved in the functional restorations that sleep may provide. Similarly, some genes that alter traits related to specific electroencephalography oscillations may or may not be central to the understanding of the function of sleep. In addition, a large group of genes vary their expression across sleep and wake, some of which are probably important to downstream functions relevant to sleep and wake even if they themselves are not "core" sleep-regulating genes. To help understand the current state of the art in these various approaches toward understanding

sleep, this chapter is divided into sections based on the most important general methodologies used to identify genetic mechanisms:

- Changes in gene expression, or more precisely, changes in steady-state–specific messenger RNA (mRNA) levels, during sleep and wake. The majority of these genes may simply be responding to arousal state, but some of these genes could be critical components in the regulation of sleep-wake traits.
- Identification of "naturally occurring" allelic variants that underlie individual differences or strain differences in sleep-related traits. Some of these genes are likely to be important in sleep regulation or functional aspects of sleep, although many may represent pleiotropic genes that only indirectly alter sleep. Nevertheless, such genes and associated genetic networks may still be important to understanding of the underlying physiologic and pathophysiologic mechanisms that underlie normal sleep as well as sleep disorders. We highlight the quantitative trait loci (QTL) approach in uncovering the natural allelic variants influencing sleep-wake phenotypes.
- Identification of genes that regulate sleep or influence sleep-related traits by mutagenesis and transgenic strategies. These approaches artificially alter or knock out genes to find those that influence sleep.

Although these three methodologies differ in their approach, they complement each other, and to ensure progress in identifying the genes and gene networks important to sleep, they are often used in parallel. For example, transgenic approaches are used as follow-up studies for candidate genes located within QTL. Moreover, the QTL approach is now more and more integrated with genome-wide expression profiling to model gene networks and key regulatory pathways that underlie multiple interrelated sleep phenotypes. Such multiscale systems approaches have revealed intriguing gene functional organizations not only important for sleep but also linking sleep to other major functions of the brain.

GENE EXPRESSION, mRNAs, AND MICROARRAY STUDIES

The investigation of variation in gene expression patterns between sleep and wake began with targeted investigation of single genes of interest and later used microarrays and now RNA sequencing to compare global expression changes between sleep and wake in brain and other tissues to identify genes with no prior assumptions about their functional involvement in sleep. The first class of genes that was clearly shown to vary across sleep and wake were rapid response genes, often called immediate early genes (IEGs), such as c-*fos* (FBJ osteosarcoma oncogene).[1-3] Changes in IEG expression are of interest for at least two reasons. Because most IEG mRNAs and proteins increase with neuronal activity, they can be used to identify brain regions activated by changes in arousal state. Second, because most IEGs are transcription factors, they may represent "master-switch" genes that initiate a complex of molecular signaling cascades.[4] These pathways may be important for longer term homeostatic control or restorative functions critical to sleep and wake.

Most brain regions are more active during wake than during sleep and thus show higher levels of expression of IEGs during wake.[1-3] Exceptions to this expression pattern point to brain regions that are active during sleep and may play regulatory roles in sleep. In the cerebral cortex, a small subset of GABAergic neurons that express neuronal nitric oxide synthase (nNOS) increase c-*fos* expression during NREM sleep.[5] This finding led to subsequent studies that further characterized cortical sleep-active nNOS neurons and their role in NREM sleep homeostasis. Specifically, these cortical nNOS neurons coexpress the NK1 receptor (or *Tacr1*, tachykinin receptor 1), and the activities of these neurons regulate the amount of NREM sleep, NREM sleep bout duration, and electroencephalogram (EEG) delta power during NREM sleep.[6,7] The last measure quantifies the amplitude and prevalence of delta or slow waves (1 to 4 Hz) characteristic of the NREM sleep EEG and is widely used to estimate homeostatic sleep need. Another interesting region where c-*fos* is higher in sleep than in wake is the ventrolateral preoptic area (VLPO) of the anterior hypothalamus, consistent with the VLPO's being a "sleep-active" region.[8] Neuroanatomic tracings linked this sleep-active region containing inhibitory GABAergic neurons to wake-active structures, including histaminergic neurons in the posterior hypothalamus, suggesting a reciprocal interaction between these two structures.[9] Later, other hypothalamic nuclei were integrated into a model indicating that several hypothalamic regions combine to play a critical role in sleep-wake regulation.[10] Among these other regions are the hypocretin/orexin neurons that are central to our growing understanding of narcolepsy. The exciting breakthrough that led to the elucidation of a critical role for the hypocretin/orexin neurons in narcolepsy came from a combination of dog genetics,[11] mouse knockouts,[12] human pathology and pathophysiology,[13] and, earlier, subtractive hybridization and rat neuroanatomy.[14] Research on the role of hypocretin/orexin in narcolepsy is described in Chapter 89. Expression of IEGs has also been used to support and to extend our understanding of the dorsolateral pontine region in the control of REM sleep[15] and the locus ceruleus in wake.[16] Mapping of sleep-active neurons by increased c-*fos* expression was first used to identify potential sleep-active neurons in the brainstem that respond to REM sleep perturbations.[17] More recently, c-*fos* expression was used to identify a sleep-active region of the brainstem, called the parafacial zone; loss of the sleep-active neurons in this region causes increased daily wakefulness and reductions in both NREM and REM sleep.[18] A recent follow-up study demonstrates that projections from sleep-active parafacial neurons inhibit wake-promoting neurons in the medial parabrachial nucleus, which in turn send glutaminergic innervations to cortical-projecting magnocellular regions of the basal forebrain, suggesting a neural circuit that may be important in triggering NREM sleep and promoting cortical EEG synchronization.[19]

Other studies have shown interesting regional differences throughout the brain for the expression of specific IEGs during sleep deprivation and recovery sleep[20] and similarly for some of the so-called heat shock or stress response genes.[21] The results from studies in rats and mice are consistent with each other,[22] suggesting an important generalizability across rodents and probably all mammals. In addition, diurnal squirrels have peaks of IEG expression and heat shock protein 70 during the day (instead of the night, as in nocturnal rodents), consistent with high levels of mRNA correlating with wake.[23,24] Although it is still unclear what role these IEGs and

heat shock mRNAs and proteins play, their consistent increase during periods of wake and neuronal activity suggests the possibility of some restorative role or, more specifically with IEGs, the need for transcriptional activation during wake. It is possible that transcription is preferentially activated during wake while translation into proteins is increased during sleep. This is supported by some earlier protein work using ^{14}C-Leu autoradiography in both rat and monkey, which showed that the rate at which labeled leucine was incorporated into the brain was positively correlated with the occurrence of slow wave sleep, suggesting the exciting possibility that one function of sleep might be the restoration of proteins through increased synthesis.[25,26] In general, protein data across sleep and wake periods have been limited in part because of the greater difficulty of identifying and quantifying proteins (see[27] for review) compared with measuring mRNA levels. For example, microarray technology allows hundreds or thousands of mRNAs to be compared across conditions and has generally supported these results.[22,28]

The first large-scale microarray study investigating sleep versus wake was done in rats.[28] Several interesting observations were made in this extensive survey that compared transcripts derived from sleeping (undisturbed) versus sleep-deprived rats killed at 6 PM (sleep deprivation was for 8 hours) as well as from rats sacrificed at 6 AM that had spent the majority of the night awake. More than 15,000 transcribed sequences were found to be present in the cerebral cortex. Of these 15,000, about 10% differed between day and night, and about half of these, 5%, varied between sleep and wake regardless of time of day. The cerebellum, a structure not generally associated with sleep, had similar changes. Again, about 5% of the detectable transcripts were differentially expressed between sleep and wake. Although the cerebellum does not display the electrographic signs of sleep comparable to EEG recordings of the cerebral cortex, this result is consistent with the idea that all or most neurons in the brain may require similar cellular restoration during sleep. The few areas of the brain that have higher activity during sleep, such as the VLPO, could potentially perform these tasks during wake. Different brain regions are likely to express a different constellation of genes. As discussed before, the hypothalamus alone has many different nuclei that appear to play central and counterbalancing roles in sleep and sleep regulation during sleep and wake. In keeping with this diversity, the hypothalamic gene expression changes across sleep and wake appear to be greatly different relative to the cerebral cortex and basal forebrain[22] and to have a smaller number of significant mRNA changes,[29] perhaps owing to its functional diversity.

The nature of the restorative or other biologically important processes occurring during sleep is not well understood. During wake, there is a notable increase in certain categories of mRNAs, including those involved in oxidative phosphorylation (e.g., mitochondrial genes) and other energy-related processes (e.g., *Glut1*). Other genes showing increased activity during wake involve transcription factors such as the IEGs noted before and some of the clock-related genes (e.g., *Per2*, discussed in detail later), stress response factors (e.g., heat shock proteins), glutamatergic neurotransmission–related genes (e.g., *Narp*, *Homer1*), and mRNAs related to activity-dependent neural plasticity and long-term potentiation (e.g., *Arc*, *Bdnf*).[28] In contrast, during sleep, different categories of mRNAs appear to be upregulated, including those involved with translational machinery (e.g., initiating factor 4A-II), membrane trafficking (e.g., *Rab*s, *Arf*s), promoting hyperpolarization (e.g., *Trek-1*, *Task-1*), and synaptic plasticity related to memory consolidation (e.g., *Camk4*, *Calcineurin*). An issue in this and many other studies is how to define the sleep-wake state for each group of animals from which tissues are collected for microarray analysis. For example, whereas mice and rats spend the majority of the day sleeping and the majority of the night awake, they are often awake one third of the day and sleeping one third of the night (with sleep and wake periods typically alternating over periods of minutes or even seconds). In rats and mice, one cannot assume that the animals are "sleeping" during the light or daytime or are "awake" during the dark or nighttime at the time of sample collections. Also, some mRNAs may change rapidly, in which case the prior hour or two before collection of expression data may be of paramount importance, whereas other mRNAs may be influenced over many more hours. Last, it is not clear in sleep deprivation versus control studies whether a gene is sleep induced or goes down during prolonged waking. Incorporating recovery sleep periods after sleep deprivation can partially address this issue, especially if multiple recovery time points are examined (e.g., 1 hour, 2 hours, and 4 hours after sleep deprivation). This has been done for several clock-related genes, such as *Per1*, *Per2*, and *Dbp*,[30,31] which generally return to baseline levels after recovery sleep. *Per1*, *Per2*, and other clock genes are central to the circadian pacemaker located in the suprachiasmatic nuclei (SCN) of the hypothalamus, as discussed in Chapter 27. However, in other brain regions, these genes appear to be responsive more to sleep, wake, and activity state rather than to their internal clock-time.[30-32] Mutations and knockouts of these genes in mice produce not only the expected circadian disturbances but also, unexpectedly, fundamental changes to sleep homeostasis. This is discussed in detail later.

Subsequent sleep microarray studies generally confirmed gene functional groups (particularly genes involved in metabolic homeostasis and synaptic functions) that are associated with sleep and wake status. These more recent studies have also discovered additional groups of sleep-wake–dependent genes. The coordinated changes in particular categories of genes associated with sleep or wake provided indications or support for various hypothesized sleep functions. For example, one study found genes that exhibited decreased expression in mouse hippocampus after 5-hour sleep were overrepresented by those involved in translation, including *Mtor* (mechanistic target of rapamycin), a key regulator of protein synthesis.[33] This finding is consistent with the hypothesis that sleep allows protein synthesis, which is also required for long-term memory formation. Another microarray study examined sleeping versus sleep-deprived mice at multiple time points across the entire day, a study design allowing identification of a large number of differentially expressed genes between sleep and wake. In particular, many mRNAs that were higher in the sleeping state are involved in macromolecule biosynthesis and transport.[29] This was most apparent for genes involved with cholesterol synthesis and lipid transport. In general, this supports the hypothesis that one function of sleep may be the restoration of certain molecular components that are perhaps depleted during wake. Furthermore, because microarrays have been used to investigate gene expression changes during sleep and wake in multiple studies and organisms, it is now possible

to characterize consistent patterns of sleep-wake–dependent gene expression across species and experimental conditions. As one of these attempts, a computational study performed a meta-analysis using microarray data from multiple sleep deprivation studies and found that IEG transcription factors, such as the *Egr* and *Nr4a* gene families, are sleep-wake–regulated genes conserved in mouse, rat, fruit fly, and white-crowned sparrow.[34] In addition, gene regulatory network analysis revealed that cyclic adenosine monophosphate–responsive element (CRE) is likely to be one of the key *cis*-regulatory elements responsible for sleep deprivation–induced transcriptional changes. Because glutamate receptors regulate the phosphorylation and activation of CRE-binding protein (CREB), this observation may highlight the role of glutamate neurotransmission during wakefulness, although it is also possible that it merely reflects the fact that CRE is one of the most well characterized transcriptional regulatory elements. Nonetheless, this study demonstrated the utility of bioinformatic tools in the efforts to eventually convert this kind of data to our increased understanding of sleep regulation, sleep functions, and sleep disorders.

In addition to protein-coding RNAs, studies have also reported changes in the expression of noncoding transcripts, such as microRNAs (miRNAs), during sleep and wake.[35-37] The miRNAs play a role in the post-transcriptional regulation of gene expression, and increasing evidence reveals their importance in central nervous system functioning. By silencing and destabilization of mRNAs, miRNAs thus add an additional level of gene regulation through which sleep and wake could alter the transcriptome. Interestingly, intracerebroventricular injection of a specific inhibitor to one of the identified miRNAs, *miR-138*, reduced total sleep time and EEG delta power during NREM sleep,[36] suggesting that some miRNAs can alter sleep regulation.

Gene expression changes during sleep and wake are likely to be influenced by the genetic background of the animals tested. One study addressed this issue by comparing sleep-deprived and control mice of three different inbred strains at four different time points over the 24-hour day.[35] Of more than 2000 brain transcripts found to vary as a function of time of day under control conditions, only 391 remained rhythmic when mice were sleep deprived, suggesting that most diurnal changes in gene expression are, in fact, sleep-wake dependent instead of being under direct circadian control. This study also demonstrated that many of the changes in gene expression were strain specific, and only a relatively small number of transcripts changed consistently in all three strains. Of these, *Homer1a* showed the most consistent and dramatic changes. *Homer1a* is a truncated form of *Homer1* that is involved in glutamate neurotransmission and probably in intercellular calcium homeostasis. Thus, it may be important for neuronal recovery or optimization after periods of wakefulness. The *Homer1* gene is also in the middle of a chromosome region shown to influence recovery from sleep deprivation[36] (see later), a finding that may lead to more definitive evidence of a role for this gene in sleep.

Another issue for studying gene expression changes after sleep deprivation is that sleep deprivation is generally stressful, leading to an increase in glucocorticoids and other stress-related changes. To address this issue, Mongrain et al[38] examined changes in gene expression with sleep deprivation in adrenalectomized versus control mice from three different strains that differ in their homeostatic response to sleep deprivation. Many genes involved in the pathways that were previously linked to sleep and wake were no longer affected by sleep deprivation after adrenalectomy, whereas others, such as *Homer1a*, remained affected by sleep and wake. *Per2*, the most consistent clock gene responding to changes in sleep and wake, continued to show a significant increased expression after sleep deprivation in adrenalectomized mice. Finally, similar to protein-coding mRNAs, miRNAs were also affected by both sleep deprivation and adrenalectomy.

In summary, gene expression studies have suggested that at least 5% of mRNAs vary across sleep and wake, but many of these changes may not be consistent across strains and species and may be dependent on the specific conditions of each experiment, including the level and type of stress. It seems likely that an even higher percentage of transcripts vary slightly with arousal state, but the present methods are generally not sensitive enough to pick up small changes in steady-state mRNA levels. Recent studies have demonstrated that genes showing time-dependent changes across the day in various tissues added to a total of 43% of all protein-coding genes.[39] Many of these genes may be driven by sleep-wake, the timing of which is in synchrony of the circadian clock under undisturbed conditions. Nevertheless, the mRNAs with measurable sleep-wake–dependent changes code for many different proteins that are likely to reflect specific functions of sleep and wake, but more data from different levels of organization are needed to document these specific functions. One concern raised by this kind of approach is the time course issue. This is perhaps most relevant with the changes in transcription factor mRNAs, which are among the most consistent and reliable changes documented thus far. Presumably, to produce a functional change in the brain, the mRNAs must first be made into proteins, return to the nucleus, and activate or inactivate other genes that alter their mRNA levels, and then these mRNAs must be made into proteins and perform some function that ultimately alters a neuronal property (such as resting membrane potential) before there is any fundamental change in sleepiness or other sleep-related variables. In most cases, this will take several hours. Of course, other molecular changes, such as phosphorylation of proteins, can take place in seconds and alter many neuronal properties; however, this does not elucidate how changes in mRNA levels ultimately affect sleep and wake. As discussed before, homeostatic need for sleep can build up during many hours and days, and certainly these long-term processes could be both monitored and regulated by these relatively slow transcriptional or translational changes. However, it has been shown in mice that sleepiness begins to accumulate after as little as 1 hour of wake,[31,36] raising the question of whether these slow mechanisms are fast enough to underlie the physiology of sleepiness. It is of course possible that there are both slower and faster homeostatic responses to different arousal states, as has been suggested for REM sleep,[31] and changes in gene expression and mRNA levels may underlie only the longer ones. It is also important to consider that most changes in steady-state mRNA levels are probably driven by sleep-wake changes and not the other way around.

Finally, an intriguing and rather surprising finding is that sleep deprivation results in a greater number of transcripts being significantly affected in the liver than in the brain.[35] Such a finding adds to the cautions of extrapolating from

changes in gene expression across arousal state to a functional role of these genes in sleep-wake regulation. If sleep is "of the brain, by the brain, and for the brain,"[40] how does one account for this observation? Perhaps the brain is protected from changes in transcription during sleep deprivation, or perhaps the increased number of significantly changed transcripts in the liver simply reflects its greater homogeneity compared with the brain with its numerous cell types, layers, nuclei, and structures that are functionally diverse and that can respond with equal diversity to sleep deprivation; or perhaps the liver in fact needs these greater changes to respond to increased wake. Is it even possible that the liver or other tissues are in fact sleeping or awake in any meaningful way? For example, sleep might define a neuroendocrine state, allowing peripheral organs to engage in activities that are not possible during wake. Until we understand the functions of sleep with more certainty, it is difficult to define what sleep may mean for peripheral tissues. In any case, a number of studies have reported sleep-wake–dependent gene expression in peripheral tissues. For example, one study in mice has demonstrated that the expression of 3% of the transcripts in the lung and 6% of the transcripts in the heart is influenced by sleep and wake state, highlighting tissue-specific metabolic processes.[37] Interestingly, results from this study also suggest that sleep modulates the circadian program of gene expression in the lung. Similar observations have also been made in the circadian transcriptome and metabolome in human peripheral blood.[41-43] Future studies are needed to further characterize the relationship between sleep-wake–dependent modulation of the transcriptome and the activities of peripheral organs, which may provide indications for the functional significance of sleep and wake to peripheral tissues. Nevertheless, accumulating evidence now links sleep deprivation to profound changes in metabolism[44] and the expression of genes in many peripheral tissues, indicating that sleep is more than "just for the brain" and that sleep-wake states and durations have pronounced effects on gene expressions throughout the entire body.[45]

IDENTIFICATION OF NATURALLY OCCURRING ALLELES THAT INFLUENCE SLEEP OR SLEEP-RELATED TRAITS

Although gene expression studies are likely to lead to a better understanding of certain aspects of sleep, there are many limitations to this approach. Therefore, it is important to use other complementary genetic approaches. There is abundant evidence to indicate that allelic differences in sleep-related genes exist, and the identification of these alleles and the genes themselves is likely to lead to important sleep-related functions. Many aspects of "normal" sleep as well as several sleep disorders have strong genetic components (reviewed by[46-49]). Results from twin studies make this especially clear (Figure 29-1). First, brain architecture and regional activity are much more similar in monozygotic versus dizygotic twins[50] (Figure 29-1, *A*). Second, EEG patterns of monozygotic twins have a much higher concordance than those of dizygotic twins, with the patterns in monozygotic twins being nearly as similar as the same individual recorded on two different occasions[51-53] (Figure 29-1, *B* and *C*). These results support the hypothesis that complex EEG traits are largely controlled by genes and that environmental factors play a lesser role. In fact, for many EEG traits, well above 80% of the variance appears to be accounted for by genetic factors, whereas for other sleep traits, such as the amount or timing of sleep, the relative importance of genes and environment is probably closer to 50% (again, see reviews cited before). Like twin studies in humans, genetic studies of sleep in the mouse, pioneered by Valatx, yielded substantial support for the genetic control of sleep. In the early 1970s, Valatx's group initiated a series of crossing experiments and recorded sleep in hundreds of inbred, recombinant inbred, and hybrid mice mainly to follow the segregation of REM sleep.[54-56] However, until very recently, none of the genes underlying these or any other sleep traits had been identified. The first significant breakthrough in the field occurred in 1999, with the discovery that a mutation of the hypocretin-2 receptor gene underlies canine narcolepsy[11] (see Chapter 89). This gene was certainly not one that would have been predicted to have a role in narcolepsy, highlighting the strength of the genetic approach.

In a genome-wide search for genes affecting a particular phenotype, no a priori assumptions are made on the genes involved. Although this approach may lead to already known physiologic mechanisms, its strength lies in the capability of uncovering previously unknown systems or pathways involved in sleep. Going from variability in a reliable sleep-related phenotype to the underlying genotypic variability of differing alleles (sometimes called forward genetics or traditional genetics) often involves the following approach.[57-60] In the first step, the mode of inheritance for a trait of interest is determined in segregating offspring, although for most complex traits this will not appear as simple mendelian inheritance patterns. Next, the localization of the underlying genes is mapped by examining the entire genome at regular intervals using polymorphic markers (e.g., restriction fragment length polymorphisms, simple sequence length polymorphisms, single nucleotide polymorphisms). Traits generally cosegregate with the markers most closely linked to the underlying genes. For simple mendelian traits, initial mapping in a few hundred offspring can yield sub-centimorgan resolution (typically on the order of 1 million base pairs, with perhaps 10 genes in the defined genetically linked region). However, for nonmendelian complex traits, in which several genes and environment contribute to a single trait, this initial step will usually narrow down the region to only about 10 to 30 cM. Subsequent fine mapping, if feasible, may further reduce these regions. Positional cloning techniques can then be applied to find and to characterize candidate gene sequences, with potential follow-up studies using gain- or loss-of-function knockout and transgenic mouse models (see later). This approach has been most successful with mendelian monogenic traits, such as for canine narcolepsy as discussed before, and for studies involving mutagenesis as discussed later, in which a single mutated gene usually accounts for the phenotype. However, positional cloning is becoming increasingly successful in even the more common and difficult cases in which multiple genes and environmental influences interact to produce a wide phenotypic range of a quantitative trait, such as differences in the spectral composition of the EEG signals among inbred strains of mice.

The standard approach to mapping of chromosomal regions that underlie quantitative traits is referred to as QTL analysis. QTL analysis is a good method to genetically dissect complex traits, like sleep, because naturally occurring allelic variations

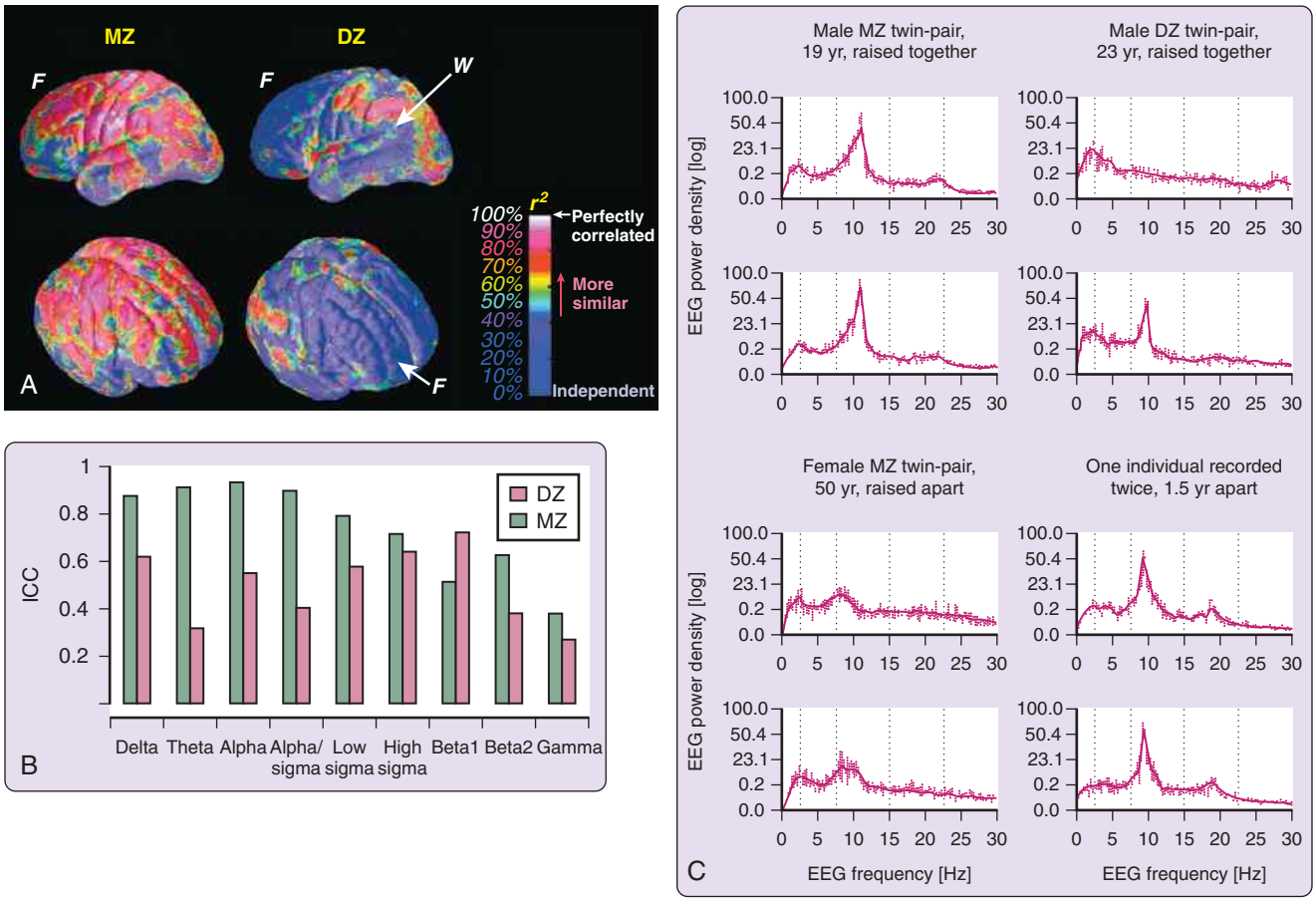


Figure 29-1 Basic brain structure and electroencephalogram (EEG) patterns are among the most highly heritable complex traits. **A,** Brain architecture is highly genetically determined. Genetically identical (monozygotic; MZ) twins are almost perfectly correlated in their gray matter distribution. Fraternal (dizygotic; DZ) twins are significantly less alike in frontal (F) cortices but are 90% to 100% correlated for gray matter in the perisylvian language-related cortex, including supramarginal and angular territories and Wernicke's language area (W). The significance of these increased similarities, visualized in color, is related to the local intraclass correlation coefficients (*r*). **B,** The spectral composition of the EEG during non–rapid eye movement (NREM) sleep is highly genetically determined. Intraclass correlation coefficients (ICCs) of EEG power in various frequency bands indicate a much higher concordance in MZ compared with DZ twin pairs, especially for lower frequencies (delta, theta, and alpha). **C,** The spectral composition of the waking EEG is equally under strong genetic control, and twin studies identified heritabilities of up to 90%, indicating that 90% in the variance of the phenotype can be accounted for by additive genetic factors. Note the higher similarity in MZ versus DZ twins even when raised apart. MZ twins are nearly as similar as the same subject recorded on two different occasions. (**A,** From Thompson PM, Cannon TD, Narr KL, et al. Genetic influences on brain structure. *Nat Neurosci* 2001;4:1253-8; **B,** From Ambrosius U, Lietzenmaier S, Wehrle R, et al. Heritability of sleep electroencephalogram. *Biol Psychiatry* 2008;64:344-8; **C,** From Stassen HH, Lykken DT, Propping P, Bomben G. Genetic determination of the human EEG. Survey of recent results on twins reared together and apart. *Hum Genet* 1988;80:165-76.)

or gene mutations with smaller effects can be mapped.[57-60] QTL analysis can be performed in segregating mouse populations that involve intercrosses, backcrosses, recombinant inbred (RI) strains, or heterogeneous stocks (outbred mice derived from breeding multiple inbred strains together). Often, two inbred mouse strains differing in a trait of interest are crossed, and their F1 offspring are then intercrossed to generate F2 offspring. To generate RI sets, F2 mice are inbred by brother-sister matings for more than 20 generations until essentially full homozygosity is achieved, thereby "fixing" a unique set of recombinations of the parental genomes in each RI strain.

Many aspects of sleep and the EEG parameters measured during sleep differ dramatically among different inbred strains of mice.[36,55,61-65] Such differences are likely due to genetic factors. As mentioned before, QTL analysis can be simplified in sets of RI strains that are derived from two parental inbred strains. In its most simple form, this mapping entails point-correlations between the strain-distribution pattern of the phenotype and the strain-distribution pattern of the genotype at each marker (see Figure 29-2 as an example of this approach). Polymorphic markers at a specific chromosomal locus that correlate significantly with the quantitative trait are presumably linked and cosegregate with the actual gene (or genes) whose differing alleles, derived from the two parental strains, contribute to the phenotypic variance. QTL analysis of traditional crosses, such as intercross and backcross panels, also follows these same principles, but with two primary

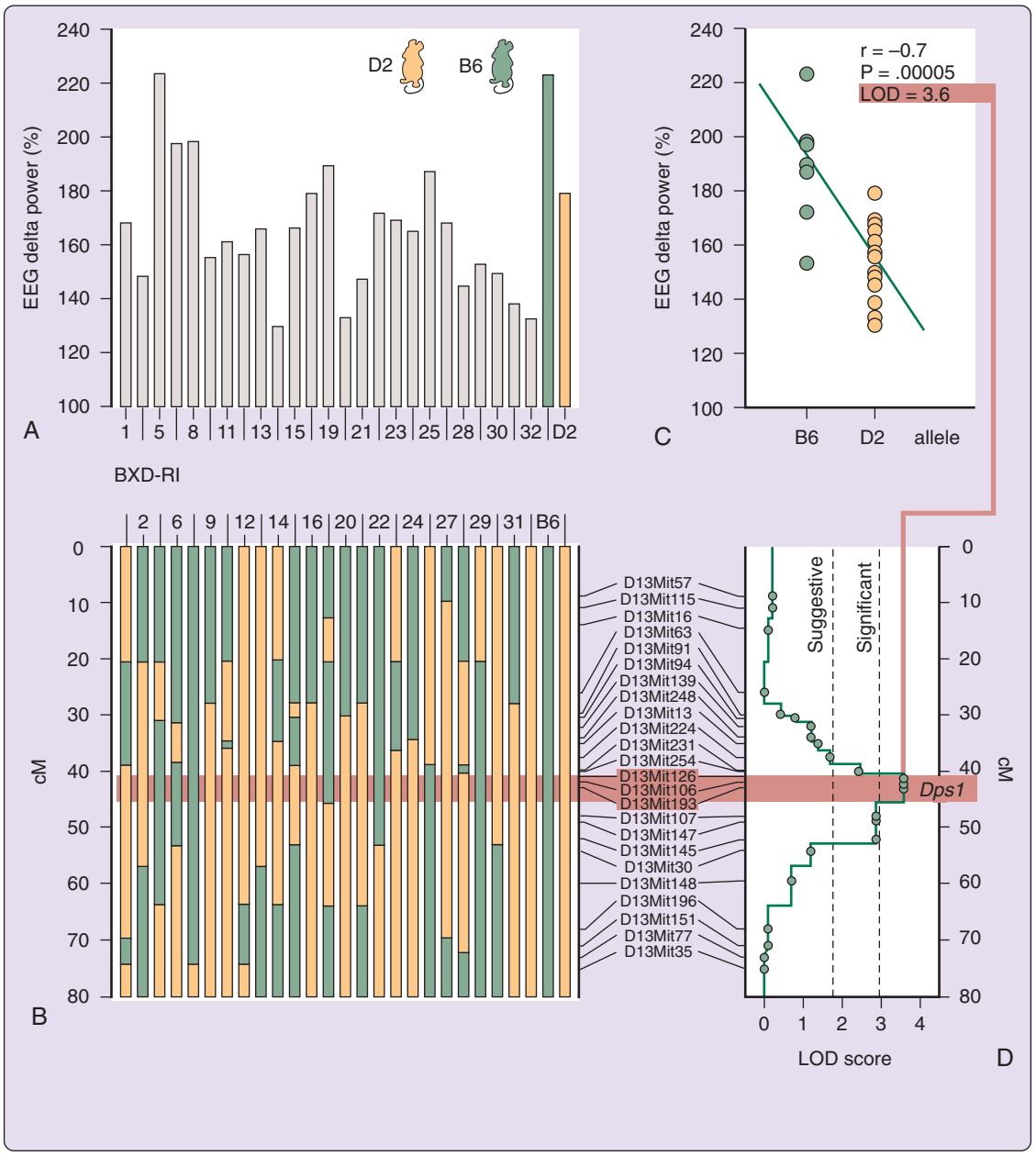


Figure 29-2 Quantitative trait locus (QTL) analysis for a sleep recovery trait in BXD recombinant inbred (RI) mice. QTL analysis is illustrated for the chromosome 13 *Dps1* QTL (see text and reference 35 for additional details). **A,** The strain-distribution pattern (SDP) of the sleep phenotype. In 25 BXD RI strains (BXD-1 to BXD-32) and their parentals C57Bl/6J (B6; *green*) and DBA/2J (D2; *orange*), the rebound in electroencephalogram (EEG) delta power was measured after 6 hours of sleep deprivation (*bars* indicate mean strain values; $n = 128$; 4-7/strain). **B,** BXD RI recombination pattern for chromosome 13 (B6 alleles in *green*; D2 alleles in *orange*). This pattern is based on the 24 Mit markers polymorphic between B6 and D2 that were genotyped in the BXD RIs. Relative map positions in centimorgans from the centromere. For QTL mapping for each marker, the genotype SDP is correlated with the SDP of the phenotype. **C,** The SDP for markers D13Mit126, 106, and 193 (along the *red horizontal bar* in **B**) yielded the best correlation coefficient (r), which was highly significant (P) and translated into a LOD score of 3.6. **D,** Among all 788 Mit markers used, only for these three markers was a genome-wide significant level ($P < .05$) obtained. This QTL was named *Dps1* (*delta power in sleep 1*). The underlying assumption of the QTL approach is that a gene (or genes) within the *Dsp1* region segregated with the three D13Mit markers in this BXD RI panel and that the B6 and D2 alleles for this gene are functionally different and modify the rebound in delta power after sleep deprivation. *Dsp1* has been confirmed for EEG delta power at sleep onset under baseline conditions, and refined mapping is in progress.

disadvantages. First, each offspring from these crosses is genetically unique and must be genetically mapped to determine the pattern of recombination of the parental genomes, unlike RI strains, in which the recombination pattern is fixed by inbreeding and has already been mapped. Also, because each individual offspring is unique in traditional crosses (and cannot be made again), it is not possible to average over multiple mice. By contrast, in RI strains, multiple identical mice can be averaged, thereby reducing the phenotypic variability. Moreover, using RI strains, one can study various phenotypes over development and aging, sex differences, and effects of environmental challenges in individual mice of the same genotype. A major advantage of traditional crosses is that a large number can be generated, unlike the limited number of RI strains. The limitation in numbers of RI strains is partially compensated by the increased number of recombinations in each strain during the inbreeding process, which produces improved mapping power per animal by approximately fourfold. The limitation in the number of RI strains is being addressed by expanding existing genetic reference populations, such as the BXD derived from the two inbred lines C57BL/6J (B6) and DBA2/J (D2),[66,67] and by creating new ones, such as the Collaborative Cross (CC), a very large effort that could revolutionize complex genetics in mice for the study of almost any trait of interest including sleep.[68] The CC line includes eight parental strains chosen for both their high genetic diversity and high diversity in almost every tested phenotype from cancer susceptibility to behavior. With plans for more than 1000 strains, and each strain having a very large number of historical recombination events fixed by inbreeding, it should be possible to map many QTL at subcentimorgan resolution, sometimes down to individual genes.[69] Because genetic reference populations have to be mapped only once, in theory, one can look at any trait in these mice, analyze the correlations with allelic distributions, and obtain a very good idea of the gene alleles contributing to this trait. In a few years, when the CC mice are available, combined with improved sequences for all eight founder strains, it may even be possible to immediately predict which nucleotide difference among strains is responsible for a given QTL (this is sometimes referred to as a QTN for quantitative trait nucleotide).

Until recently, although only small sets of RI strains were available, these strains nonetheless contributed significant advances to the understanding of genetic influences on sleep homeostasis. For example, the original BXD RI strains were examined for sleep-related traits by EEG analysis to identify QTL for the homeostatic regulation of NREM sleep.[36,47] The trait of interest was defined as the level of EEG delta power in NREM sleep reached after a 6-hour sleep deprivation (see Figure 29-2 for variance in this trait among the two parental strains D and B and the 25 BXD RI strains). Large interstrain differences were observed in this trait, and approximately 37% of the total variance could be attributed to additive genetic factors (i.e., heritability). The contribution of the chromosome 13 QTL to the total genetic variance amounted to 49%, suggesting the presence of a "major" gene. Confirmation of the chromosome 13 QTL was obtained in baseline recordings of the same animals. This QTL was designated *Dps1* for delta power in sleep 1. The basic assumption underlying the QTL analysis is that the identified chromosomal regions contain genes with functionally different alleles that somehow alter the sleep homeostatic process. As discussed in the previous section, *Homer1*, particularly its short splicing variant, *Homer1a*, is a favorable candidate gene for *Dps1* by subsequent molecular and bioinformatic studies.[35,70] *Homer1* encodes postsynaptic density scaffolding proteins for group 1 metabotropic glutamate receptors and is thought to be important for sleep control through regulating downstream effects of glutamate neurotransmission and intracellular calcium homeostasis. HOMER1A protein functions as a dominant negative regulator of the longer HOMER1 isoform. The expression of *Homer1a* is elevated after sleep deprivation and reduced to baseline after recovery sleep, consistent with its proposed role in the regulation of sleep homeostasis. However, a recent study challenges this view by showing that the specific knock out of *Homer1a* (with no disruption of the full-length transcript) has no effect on homeostatic EEG responses after sleep deprivation.[71] It is possible that the allelic difference between B6 and D2 in *Homer1a* indeed influences homeostatic EEG activity after sleep deprivation, whereas its role is somehow compensated by other functionally related genes in the *Homer1a* knockout mice. It is also possible that genetic variations of other genes in the *Dps1* region are responsible for the observed QTL effect. Ongoing efforts are under way to elucidate how the *Dps1* locus influences homeostatic EEG activity after sleep deprivation.

The same BXD RI lines also led to another QTL on chromosome 14 that contributes to the delta oscillations (1 to 4 Hz) that mark NREM slow wave sleep.[72] Although the study began as a QTL analysis, the data suggested that a single locus accounted for the majority of the variance in this trait. Assisted by additional polymorphic markers in this region, haplotype analysis of additional mouse inbred strains that are phenotypically alike to either B6 or D2 defined a 350-kilobase region as the smallest genomic region associated with this trait. Within this region was retinoic acid receptor beta (*Rarb*) that contained a restriction fragment length polymorphism that cosegregated 100% with the trait. *Rarb* has two different promoters that produce four different transcripts. Targeted deletion of these different transcripts, coupled with gene sequencing and real-time reverse transcription polymerase chain reaction, clearly documented that alleles of *Rarb* did in fact underlie this EEG phenotype. Retinoic acid receptors and retinoid X receptors are nuclear receptors that form heterodimers, are highly expressed in the brain, and are implicated in neuronal functions such as brain development, control of locomotion, long-term potentiation, and effects on dopaminergic and cholinergic neurotransmission.[73] These last effects may underlie the role of *Rarb* in regulating the cortical synchrony that determines this EEG phenotype. This same research group was able to use similar methods to identify another major gene, acyl-coenzyme A dehydrogenase (*Acads*), that, when mutated, dramatically alters the frequency of the theta oscillations characterizing the REM sleep EEG.[74] Microarray analysis of gene expression in mice with mutations in *Acads* also implicated *Glo1* (glyoxylase 1) as a key factor. This work suggested the surprising involvement of a metabolic pathway involving fatty acid beta-oxidation in regulating theta oscillations during REM sleep. Because this work began as a QTL analysis, it would be fair to say that it represented the first successful identification of a gene that underlies a sleep-related QTL and one of the first for any brain or behavior phenotype.

Another recent example of successful QTL dissection of sleep traits is the case of "early running" mice that show activity traits similar to humans with advanced sleep phase syndrome (ASPS).[75] Virtually all "normal" mice begin their peak activity very close to dark onset. In contrast, a wild-derived inbred strain of mice, CAST/EiJ (CAST), and a significant portion of the B6 × (B6 × CAST) backcross progeny begin their activity 2 to 6 hours before dark onset (i.e., early runner), similar to ASPS humans who wake up several hours earlier than "normal." Early runner mice also have shorter free-running periods than B6 mice, but this does not fully account for the several-hour phase advance in activity onset observed in these mice. Some cases of human ASPS are highly familial and are called familial ASPS (see Chapter 31); they can be virtually mendelian with mutations in specific genes, such as *Per2*[76] or *CKI-delta*.[77] Alternatively, more subtle allelic effects, such as those occurring in human *Per3*, may influence "morningness" and "eveningness"[78,79] along with environmental factors. At some point, extreme morning or evening preference can become ASPS or delayed sleep phase syndrome. The dramatic cases with known mutations are rare (and therefore are referred to as mutations rather than alleles). In contrast, the *Per3* alleles mentioned before are common. Early runner mice are more like this example in humans in that the primary QTL on chromosome 18 accounts for only about 10% of the total variance in this trait,[75] although it contributes a higher percentage of the genetic variance and the total variance in general daytime activity. Congenic mice with CAST alleles at the QTL region with a more than 99% B6 background in the rest of the genome do not start their activity until lights-off because of direct inhibition of activities by light.[80] However, their circadian phase angle of entrainment (defined as the time of activity onset during constant darkness extrapolated to the last day of the prior light-dark cycle) was approximately 20 to 30 minutes earlier than that of the B6 control mice, indicating the CAST allele at the QTL region indeed contributes to an earlier timing of the activity-rest cycle. Interestingly, such a change is not accompanied by a shortening of circadian period, as is common in *Per2* or *CKI* mutant models of familial ASPS. Detailed study in the parental CAST mice suggests that the earlier activity might result from the output rather than from the central components of the circadian clock.[80,81] Such mice may be useful in testing new pharmaceuticals or nonpharmaceutical approaches to the treatment of ASPS.[82]

Apart from these studies, surprisingly few other QTL studies have investigated "natural" sleep and EEG traits, especially sleep architecture, duration of vigilance states, sleep fragmentation, and many other sleep phenotypes in large segregating populations. However, a study of natural sleep and EEG and electromyogram activity in a large B6 × (BALB/c × B6) backcross mouse population (n = 269) identified 52 significant QTL, representing at least 20 genomic loci, as being involved in the regulation of 20 diverse sleep-wake traits during the 12-hour light period, the 12-hour dark period, or the entire 24-hr day.[83] Because multiple sleep traits were investigated at the same time, it was possible to demonstrate the linkage relationships of their underlying QTL (the genetic landscape of sleep phenotypes). Sleep traits were first grouped into five distinct categories, including sleep fragmentation, state length, latency, REM sleep, and EEG power bands, based on exploratory factor analysis. Not surprisingly, QTL for the same category of traits tend to map to the same genomic location. However, QTL underlying different groups of traits can also colocalize with each other. For example, the number of brief arousals (a fragmentation trait), latency of REM sleep (a latency trait), and theta power during REM sleep (an EEG power band trait) all mapped to chromosome 1 at approximately 64 cM, suggesting perhaps shared genetic regulatory mechanisms for these seemingly different traits. Conversely, a number of QTL underlying traits thought to be closely related, such as the average duration and number of REM sleep bouts, were mapped to distinct loci, indicating diverse genetic control of related aspects of sleep. Interestingly, complex interactions between time of day and QTL were also observed: as many as 24 QTL influence sleep traits at a specific time of day (i.e., light versus dark period); some even exert the opposite effects during the light period compared with the dark period. These findings speak further to the complexity of the genetic landscape of the sleep.

Controversy exists concerning the efficacy of the QTL approach in identifying genes, and in the past, forward genetics by genome-wide mutagenesis has been favored because of greater ease of identifying the underlying gene once a genomic region of interest has been identified.[84] Nonetheless, 2000 or so QTL have now been mapped in mice with a high level of confidence for a variety of phenotypes, and although only about 100 of these have been identified at the gene level, improved methods are making positional cloning of QTL tractable, even for genes that only mildly affect the quantitative phenotype. QTL approaches may find sets of genes different from those uncovered in mutagenesis experiments and are thus complementary.[85] In addition, although it has been difficult to directly localize the actual gene (either coding or regulatory region) underlying a QTL because of the large candidate regions identified from the initial QTL analysis, recent developments combining QTL analysis for specific sleep traits with transcriptome profiling have provided an alternative method to identify candidate genes whose expressions are correlated with the QTL. For example, in the B6 × (B6 × BALB/c) study that revealed 52 QTL, one QTL located on chromosome 17 was found to influence the amount of REM sleep, number of REM sleep bouts, and wake amount.[83] Incorporating the expression of 28,053 transcripts by microarrays in three brain regions (frontal cortex, hypothalamus, and thalamus/midbrain) subsequently found that this QTL is tightly linked to an expression QTL that regulates the expression of *Ntsr1*, which encodes a receptor for neurotensin, a neuropeptide implicated in several psychiatric disorders.[86] Knock out of *Ntsr1* in mice confirmed its role in regulating REM sleep, particularly during the night.

Furthermore, with more sophisticated statistical methods, such as conditional independence models and Bayesian approaches, it is possible to mathematically disentangle the correlations and to establish causal relationships among phenotypic, genotypic, and gene expression data. This approach was also applied to the B6 × (B6 × BALB/c) sleep data set described before to identify "causal" candidates underlying two QTL regulating REM sleep and another QTL regulating amount of wake. By use of a causal inference test based on conditional independence models,[87] 65 genes in three brain regions were identified as causal candidates, mediating the effect of QTL and causal to observed variations in the amount

of REM sleep or wake.[88] Highlighted by gene network analysis, two candidate genes for REM sleep, *Acad10* and *Ncor2*, are particularly interesting. *Acad10* encodes acyl-coenzyme A dehydrogenase 10. As mentioned before, another acyl-coenzyme A dehydrogenase, encoded by *Acads*, influences theta oscillations during REM sleep.[74] *Ncor2* encodes a nuclear receptor corepressor that interacts with peroxisome proliferator-activated receptors, and the ligands of peroxisome proliferator-activated receptors have been implicated in the homeostatic regulation of sleep.[89] Both *Ncor2* and *Acad10* are involved in lipid and fatty acid metabolism, and they interact with each other in a large transcriptional regulatory network in the thalamus, suggesting a role of basic cellular metabolism in the thalamus in regulating REM sleep.[87]

The application of causal inference tests was further extended to identify candidate genes underlying all 52 QTL identified in the B6 × (B6 × BALB/c) sleep data set. To test the validity of the identified candidate genes, six of the candidates for which selective pharmacologic tools were readily available were tested for their sleep regulatory role in rats.[90] Remarkably, significant sleep regulatory effects that were consistent with the prediction of the causal inference test were observed for all six genes on perturbation by selective pharmacologic agents. Interestingly, this study also incorporated gene expression in the liver, an organ not traditionally thought to be involved in sleep regulation. *Glp1r* (glucagon-like peptide 1 receptor), whose expression in the liver was tested causal to amount of total and NREM sleep as well as the latency to NREM sleep, is among the six pharmacologically tested candidates. Administration of GLP1R agonists, exendin-4 and liraglutide, reduced wake in rats during the dark phase of the day and increased light sleep at the expense of deep slow wave sleep during the light period of the day.[90] Although current data do not indicate whether the site of sleep-regulating actions of GLP1R agonists is indeed in the liver, they raise an intriguing possibility that peripheral pathways could potentially affect sleep, which traditionally was thought to be under exclusive control of the central nervous system. As discussed before, sleep and wake states are associated with profound differences in gene expression in the brain as well as in peripheral tissues. Whereas sleep (or wake) might be associated with specific neuroendocrine environments that are permissive for certain cellular metabolic functions in peripheral organs, it may also be possible that the activities of peripheral tissues could set a global metabolic environment permissive for sleep (or wake), providing a feedback signal. Hence, as already stated earlier, sleep may not be just "of the brain, by the brain, and for the brain" as traditionally thought.[40] Future studies may further challenge this view by studying sleep and wake in animals with tissue-specific manipulations of target sleep genes.

Finally, the integration of multiscale data (phenotypic, genomic, transcriptomic, and other *-omic* data) allows modeling of multilevel interactive networks that link multiple sleep-wake phenotypes.[83,88,91] In addition, as comorbidities are frequently observed between sleep disruptions and neuropsychiatric and neurodegenerative disorders, ongoing work is using this "systems" approach to characterize interactive gene networks underlying the interaction between sleep and affective-like behaviors, stress responses, and neurodegeneration in large genetically segregating populations (Figure 29-3). In the future, such work can benefit from more advanced genetic reference panels, such as the CC lines mentioned before, which capture as much as 89% of genetic variance seen in mice and have the capacity of detecting small-effect QTL and mapping QTL to very refined regions.[92,93] Indeed, even before the CC lines have reached fully inbred, they have already demonstrated great potential in QTL analysis. A study screened incomplete-inbred CC mice using a high-throughput piezoelectric sleep recording system and has identified two sleep-wake QTL to small genomic regions.[69] One QTL influencing the time of peak activity after sleep deprivation even achieved sub-megabase resolution; it was mapped to a 530-kilobase region (29.70 to 30.03 Mb on chromosome 9), containing only three genes, including *Ntm* (neurotrimin), *Snx19* (sorting nexin 19), and a miRNA gene. The advantages of large RI panels are not limited to fine QTL mapping resolution. More importantly, they allow community efforts that accumulate multilayer *-omic* data (transcriptomic, metabolomic, proteomic, microbiome data) as well as a large collection of phenotypes (i.e., the "phenome").[94] Mining such databases will enable the identification of the interaction between sleep and other biologic functions, which will ultimately translate to what sleep means to the organism and what the functions of sleep are. In addition, incorporation of multilayer *-omic* data will allow establishment of models that describe how the functions of sleep are achieved or regulated by genetic information propagating through complex and interactive molecular networks.

MUTAGENESIS AND KNOCKOUT APPROACHES

The QTL analysis aims to identify "naturally" occurring allelic variants or gene mutations that modify sleep; in mutagenesis studies, however, gene function is assessed by randomly inducing mutations. Because mutagenesis studies involve a progression from phenotype to genotype (finding the mutation that alters the screened phenotype), it is also considered a forward genetic approach, similar to attempts to identify QTL associated with a particular phenotype or trait. In contrast, beginning with a gene of interest and using targeted deletion in mouse embryonic stem cells to knock it out is a reverse genetic approach because the progression is from gene to phenotype (see later). As with other approaches, these distinctions can blur. For instance, if a study is initiated on a large collection of knockout mice, without identified genotypic effects, one might begin with phenotypic assessment, an approach that would have similarities to mutagenesis and forward genetic approaches. This is now becoming feasible with the increasing number of knockout and other transgenic lines of mice and with high-throughput sleep analysis systems (see later). Whether it is forward or reverse genetics depends on whether one is screening more or less randomly or selecting only genes with suspected roles in the traits under study.

Mutagenesis has been a successful technique for the identification of genes that regulate circadian rhythms (as discussed in Chapters 27 and 28). A mutagen like *N*-ethyl-*N*-nitrosurea mutates spermatogonia at an average rate of 0.001 mutation/locus/gamete.[60] With high-throughput screening of several hundreds or thousands of offspring for dominant, semidominant, or recessive mutations, a major effect on a given trait can be expected to be identified.[95,96] The individual mouse, fruit fly, or other organism for which an

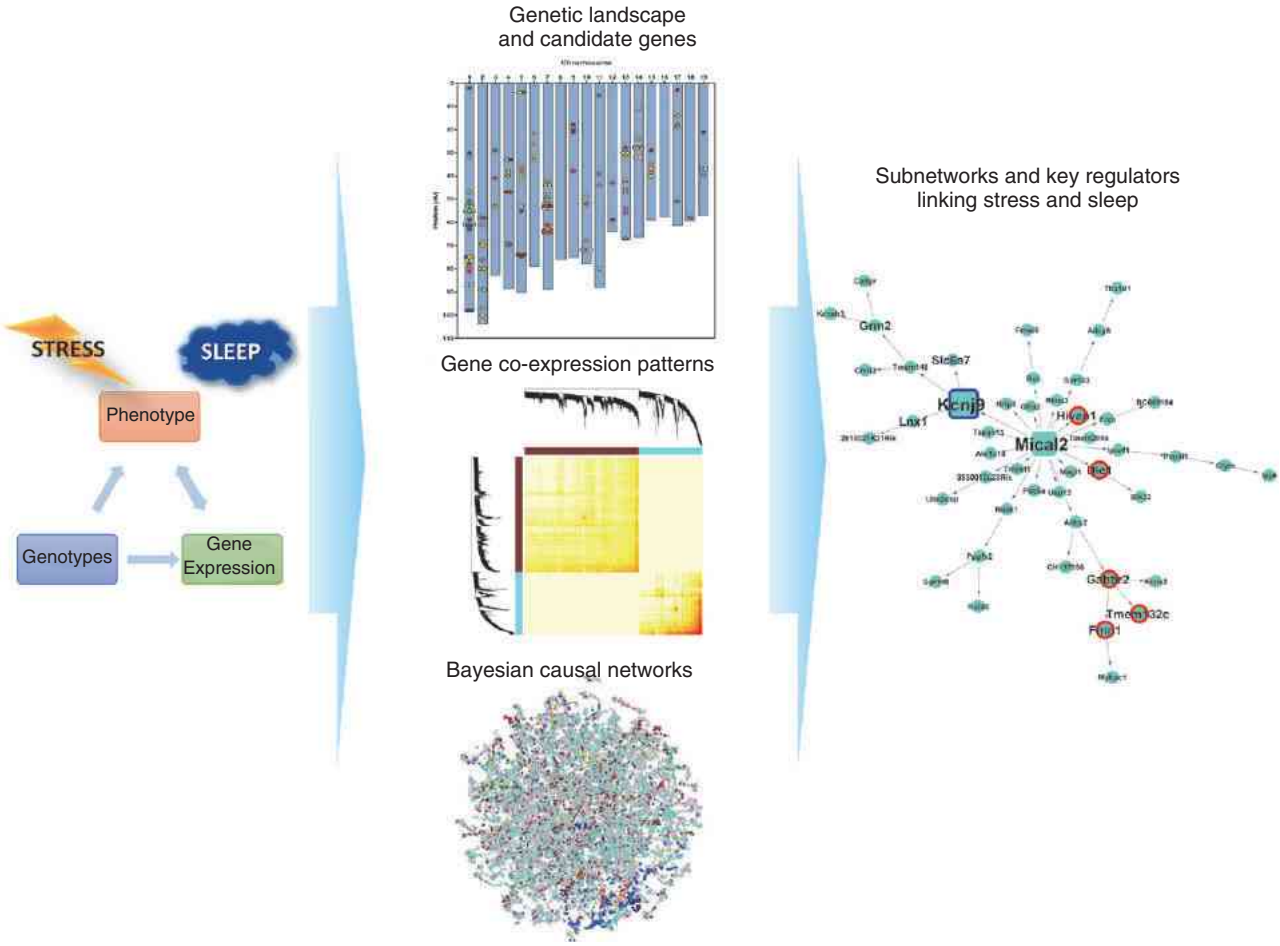


Figure 29-3 Systems approaches integrating genotypic and phenotypic data with gene expression can be used to identify gene regulatory networks underlying the interactions between sleep and other complex physiologic processes. In a chronically stressed (B6 × A/J) F2 population, Jiang and Scapa et al[162] characterized a total of 328 behavioral, physiologic, and electroencephalographic sleep phenotypes and identified more than 100 quantitative trait loci (QTL). By incorporating gene expression data in the striatum, it is possible to identify genes whose expression is underlying these QTL and causal to phenotypic variations. In addition, groups (or modules) of genes whose expressions are coregulated can be identified and associated with sleep and stress-related phenotypes. Furthermore, assisted by genotypic data, it is possible to decipher gene regulatory relationships from correlated gene expression and graphically represent these regulatory relationships in a Bayesian network. Combining all this information, one can identify subnetworks that are highly relevant to stress and sleep, such as a subnetwork driven by *Mical2*. Expression of genes in this network is highly correlated with a number of sleep and stress phenotypes. Genes downstream of *Mical2* include *Kcnj9*, whose expression was causal to stress-related behavioral phenotypes (*highlighted by blue rim*), as well as genes that were identified in genome-wide association studies as candidate genes for neuropsychiatric disorders (*highlighted by red rims*), suggesting potential molecular underpinnings linking sleep, stress, and neuropsychiatric disorders.

aberrant phenotype has been recorded then has to be crossed (usually with wild-type animals) to establish the mode of inheritance of this trait. The feasibility of this approach in the mouse was demonstrated by the isolation of the canonical circadian gene *Clock*.[97] Although some mutations may produce dramatic phenotypic changes, as was the case for the mutant *Clock* gene, others may produce only subtle effects that in addition can be confounded by epistatic interactions and genetic background (i.e., modifier genes).[98] In general, mutagenesis will probably be more successful for fully penetrant dominant or recessive mutations, whereas the QTL approach is more powerful in detecting natural allelic variations controlling complex traits.

The sleep field has been fortunate that gene alleles with very large effects on sleep-wake traits are present in the common inbred strains of mice. Whether QTL analysis or mutagenesis is used, both techniques require large numbers of mice to be screened to cover a majority of the genome (that is, to produce functional alterations in a majority of the estimated 30,000 or so genes).[95] Although screening a thousand mice is currently cumbersome with traditional EEG techniques, high-throughput methods to monitor sleep and wake in mice and other rodents may alleviate this problem.[99,100] One high-throughput, noninvasive technology uses piezoelectric films across the cage floor that act as an extremely sensitive motion detector. During sleep, the primary movement is

breathing, and thus the system records a consistent periodicity of about 3 Hz, a rate representative of the respiratory rate in mice. During wake, a variety of movements produce a more erratic signal because even during "quiet" wake mice are grooming or making many more postural adjustments than during sleep. It may eventually be possible to distinguish REM versus NREM sleep because respiratory variability increases during REM sleep, but this has not yet been achieved.[100,101] Not only could this make mutagenesis screening for sleep traits more feasible and take advantage of the large CC panel,[68] it could also be used as an initial screen for sleep- or wake-promoting drugs or to screen the increasing number of available transgenic mice (see later). Video methods for scoring sleep-wake traits are also showing increasing promise as a high-throughput method,[102,103] either in isolation or in combination with the piezo method.

The use of targeted gene deletion and other transgenic approaches has been an exciting area for sleep research. As discussed before, the use of these mouse models generally begins with a gene of interest that may influence some sleep phenotype and so is generally considered a reverse genetic approach (from gene to phenotype) as opposed to forward genetic approaches that assay the phenotype first. Reverse genetic approaches in mice have been made possible by the development of a range of techniques over the past several decades.[104,105] The field of sleep and circadian biology has benefited from knockout technology in which the insertion of a DNA construct into an exon results in a nonfunctional protein in mice that is then bred to homozygosity. Advances have also come from transgenic methodology or nonhomologous (illegitimate) recombination, in which one or more DNA construct copies are inserted into the genome at undefined locations, typically after injection of naked DNA into one of the two pronuclei at the one-cell stage.[104] These models generally produce loss- or gain-of-function mutations, respectively. The recently developed RNA-guided CRISPR-Cas9 nuclease system has dramatically transformed our ability to edit the genome in many species and is likely to replace the now classical transgenic techniques mentioned earlier. Cas9-mediated genome editing through nonhomologous end joining or homology-directed repair can be used to excise or to introduce DNA fragments, including single nucleotide polymorphisms of interest.[106,107]

These models are also useful in confirming the role of genes that were identified by forward genetic approaches. Advantages and problems of these techniques have been addressed in other reviews.[108] One important concern with respect to sleep regulation is developmental compensation, whereby other molecules, perhaps from the same gene family, could compensate for the lacking protein.[109] Other concerns involve nonspecificity (the relevant protein is absent in all the cells of the organism instead of the tissue of interest) and genetic background (genes that cosegregate with the introduced gene might differ between the background strain, often C57BL/6, and the strain in which the altered embryonic stem cells were introduced, usually a 129 strain) that might affect the phenotype.[110] Some of these issues have been addressed by developing (tissue-specific) conditional or inducible knockout models in which the acute effects of loss of function can be studied in structures of interest,[111] which has recently been successfully applied to the study of sleep-associated genes.

Despite these limitations, considerable knowledge has been gained from the study of knockout and other genetically altered mice, implicating genes in sleep and sleep-related traits in often unexpected ways. The first sleep studies using transgenic mice appeared in 1996.[112,113] Early as well as recent studies have focused on pathways with previously described roles in sleep regulation, such as monoamine neurotransmitters,[114] their receptors, and their transporters (reviewed in[47,115]). Additional studies have supported a role for cytokine pathways in the regulation of sleep, including interleukin-1, interleukin-10, tumor necrosis factor, and their receptors.[116] Finally, considerable information on sleep effects has been discovered about genes that are regarded as canonical circadian genes, such as *Clock*, *Bmal1*, *Per1*, *Per2*, *Cry1*, *Cry2*, and *Npas2* (a homologue of *Clock* expressed in the forebrain but not in the SCN), and genes known to alter circadian rhythms, such as *Dbp* and *Rab3a*.

The identification of genes involved in the homeostatic regulation of sleep has been advanced by the comparisons between baseline sleep and the responses to sleep deprivation. For instance, mice overexpressing growth hormone spend more time in REM sleep under baseline conditions but show normal recovery patterns after sleep deprivation.[117] Disruption of c-*fos* increases wakefulness, whereas *FosB* deletion reduces REM sleep. In addition, c-*fos* knockout mice respond to sleep deprivation primarily with an increased latency to sleep onset.[118] In mice lacking the transcription factor albumin D-binding protein, the circadian distribution of sleep was flatter and sleep was more fragmented. The NREM sleep response and the delta power response to sleep deprivation, however, did not differ from wild type.[119] Also, in *Clock* mutant mice, the relative increase in NREM sleep after sleep deprivation was the same as in wild-type mice, although NREM sleep amount in baseline conditions was significantly reduced.[120] Whereas a change in baseline sleep may reflect a change in sleep homeostasis, the lack of change after sleep deprivation raises the fundamental issue of what qualifies as a true change in the homeostatic response. For instance, mice lacking functional genes for the serotonin 2C receptor[121] or *Rab3a*[122] were reported to have an altered NREM sleep rebound after sleep deprivation. In these cases, however, the difference was attributable largely to NREM sleep differences under baseline conditions because no differences in recovery sleep were observed. Ideally, claims regarding an altered homeostatic regulation should be substantiated by quantifying the relationship between wake duration and the subsequent response of the regulated variable. This can be achieved either by establishing a "dose-response" relationship, in which the duration of the sleep deprivation is varied, or by mathematical means, whereby the effects of spontaneous and enforced periods of wakefulness on a regulated variable are quantified (see[36] for both approaches, and see[123] for general review). Furthermore, especially where the regulation of NREM sleep is concerned, one cannot rely on only one aspect because changes in the duration and intensity or consolidation of sleep have to be taken into account.

Changes in REM sleep homeostasis have also been observed in several knockout models. In mice lacking serotonin 1A or 1B receptors,[124,125] *Dbp*,[119] or *Cry1* and *Cry2*[31] and in *Clock* mutant mice,[120] sleep deprivation–induced loss of REM sleep was followed by a compensatory increase in REM sleep that was smaller than in wild-type animals or lacking

altogether. Apart from the serotonin 1A and 1B receptor knockout mice that displayed increased REM sleep during baseline conditions,[124,125] these changes in the REM sleep response after sleep deprivation could not be attributed to genotype differences in REM sleep during baseline. Differences in REM sleep under baseline conditions were also recently observed between *CKI-epsilon* mutant and wild-type mice as well as between congenic mice that carry naturally occurring polymorphisms at the *CKI-epsilon* locus.[126] The *CKI-epsilon* mutant (*tau* mutant allele) has long been known to shorten the circadian period.[127] Thus, this recent finding adds another example of canonical clock genes influencing sleep architecture and homeostasis in addition to controlling the timing of sleep and wake.

One rationale for studying sleep in mice with modified or deleted genes critical for circadian rhythm generation, such as *Clock*, *Per1*, *Per2*, *Cry1*, or *Cry2*,[128] is that they provide a model in which sleep homeostasis can be studied in the presence of an altered or absent circadian modulation of sleep-wake time. Previously, the interactions between circadian and homeostatic influences on sleep were studied after circadian rhythms were eliminated by lesions of the SCN[129-132] or in studies in which subjects followed a forced desynchrony protocol.[133,134] SCN lesion studies in rats revealed that direct circadian effects on the sleep homeostatic process were small if present at all (also see[135]). However, SCN lesion studies in mice[136] and in squirrel monkeys have found that SCN lesions can influence not only the timing of sleep but also the amount of sleep. Deletion or mutation in circadian genes also influences sleep-wake traits beyond just the timing of sleep, including alterations in the homeostatic regulation of sleep. A clear demonstration of this was observed in *Cry1*/*Cry2* double knockout mice that under baseline conditions showed all the hallmarks of high NREM sleep pressure, including a higher amount of NREM sleep, increased NREM sleep consolidation, and higher NREM sleep delta power, compared with wild-type controls.[31] After 6-hour sleep deprivation, there was no further increase in NREM sleep time or consolidation or in the reduced rebound in EEG delta power. This suggests that apart from their role in regulating circadian rhythms, cryptochromes or genes regulated by cryptochromes (such as *Per1* and *Per2*) play a role in sleep homeostasis. As *Per1* and *Per2* mRNA levels are responsive to sleep deprivation,[28,31] knockouts of these genes might be expected to have effects on sleep homeostasis. In two independent sets of experiments examining mice with nonfunctional *Per1*, *Per2*, or *Per3* and double knockouts of *Per1* and *Per2*, the authors suggested that there were no substantial alterations in sleep homeostasis.[137,138] However, in our view, these data may actually support an important role for *Per* genes in sleep homeostasis as a number of sleep-related changes were found in these mice. Results from these studies showed that *Per* mutations affected the total sleep time, the timing of sleep, and the effects of light and dark on sleep patterns, REM sleep, delta power during sleep recovery after sleep deprivation, and certain other parameters. The authors noted that the most clear-cut differences from wild-type controls were in sleep distribution, consistent with circadian alterations, with *Per1* knockouts sleeping more at the dark/light transition and *Per2* knockouts sleeping more at the light/dark transition. However, other sleep perturbations were also observed that do not appear to be primarily circadian in nature. *Per1* knockouts and *Per1*$^{-}$/*2*$^{-}$ double knockouts spent a much longer time in elevated delta power after sleep deprivation, from 5 to 12 hours compared with 3 hours in the wild-type mice. This contrasts with results found in *Cry1*$^{-}$/*2*$^{-}$ double knockouts, which have a greater overall sleep drive under baseline conditions, including a higher delta power, but have elevated delta power for only 1 hour after sleep deprivation. One possibility for the limited increase in delta power in sleep-deprived *Cry* knockouts is that delta power is already so high it cannot be increased any further. Taken together, these results are consistent with the possibility that the *Per* genes are involved in the process that ultimately produces delta power. The absence of a more dramatic change in sleep in response to deletion of the *Per* genes is perhaps not surprising because there are three *Per* genes, and each may be able to compensate for a loss in one or two of these genes.

Other genes in the molecular circadian clock network may also compensate for the loss of sleep. This compensation may be even greater in regions of the brain outside the SCN. Further evidence that clock genes play a role in sleep homeostasis comes from work on *Npas2*-deficient mice. *Npas2* is a clock gene that like *Clock* can bind with *Bmal1* and, although it is difficult to detect in the SCN, can substitute for a lack of *Clock*.[139] *Npas2*-deficient mice have normal circadian rhythms but altered sleep patterns that are most likely not the result of a disruption in circadian rhythmicity. *Npas2* knockout mice lack the typical nap found in the latter half of the dark period that is invariably present in wild-type mice of this strain background.[32] After sleep deprivation, however, these mice were also incapable of initiating the appropriate compensatory sleep response during the circadian phase when mice are usually awake. These mice are unable to survive when food availability is restricted to the light portion of the diurnal cycle.[140] These results suggest that the "clock" genes in the forebrain integrate behaviors such as sleep, wake, and feeding with physiologic cues, and when disrupted, the SCN-dependent circadian drive may dominate, limiting the expression of adaptive behaviors that may differ from the normal circadian peak for that behavior. These *Npas2* knockout results may be of particular interest because NPAS2/BMAL1 and CLOCK/BMAL1 heterodimers are redox sensitive, which may tie this gene network to basic energy metabolism.[141,142] Because restoration of an optimal neural energy state has long been considered a possible function of sleep as discussed before,[143] this gene network might underlie a fundamental energy restorative aspect of sleep homeostasis.

The suggestion that clock genes are fundamental to sleep homeostasis (in addition to their role in the circadian pacemaker) is strengthened by observations in the fruit fly *Drosophila*; mutants carrying loss-of-function mutations for the canonical circadian genes *Per*, *Timeless*, *Clock*, or *Cycle* (the *Bmal1* orthologue) show a more pronounced sleep rebound after sleep deprivation than wild-type flies[144] (as discussed in Chapter 28). *Cycle* mutant flies were exceptional in this respect because they clearly overcompensated for the amount of sleep lost, and this increase in sleep seemed permanent. In addition, *Cycle* mutant fruit flies died after sleep deprivations of 10 hours or more, whereas wild-type flies typically survive for about 70 hours. It has become increasingly accepted that rest and activity in flies share many features with sleep-wake states in mammals.[145-152] Recently, mice homozygous for the *Bmal1* deletion showed an attenuated rhythm of sleep and wakefulness distribution across the

24-hour period. In addition, these mice showed increases in total sleep time, sleep fragmentation, and EEG delta power under baseline conditions and an attenuated compensatory response to acute sleep deprivation.[153] Therefore, lack of circadian genes does not only affect circadian rhythms. In humans, as discussed in Chapter 30, a simple repeat polymorphism in *Per3* has been shown to significantly affect performance and EEG delta power after sleep deprivation[78,154] in addition to the morningness-eveningness trait mentioned earlier. Mice expressing these human PER3 isoforms also exhibited differences in EEG features in response to 12-hour sleep deprivation, which is accompanied by differences in the expression of sleep homeostasis-related genes but not circadian clock genes.[155] In retrospect, this might not be too surprising because circadian genes are expressed throughout the brain and the body, not only in the SCN. In addition, many of the circadian clock genes are pleiotropic and affect a number of physiologic processes, not just their timing. Nevertheless, the involvement of circadian genes in sleep homeostasis remains an intriguing molecular pathway for continued research.[156]

One other area in which knockouts, knockins, and other transgenic mice have been valuable in addressing questions about sleep is with animal models of human disorders that have serious sleep alterations or in which sleep may be part of the pathogenesis of the disease. Even in diseases for which mice are not an ideal model of the human condition, such as Alzheimer disease (AD), genetic models in mice have shed light on the role of sleep. For example, there is increasing consensus that not only is sleep disrupted in AD but poor sleep may contribute to and exacerbate the condition (see[157] for review). Different mouse models of AD have advantages and disadvantages,[158] but several have been shown to have altered sleep and circadian rhythms, and normal mice also experience daily fluctuations of the A-beta peptides that form AD plaques.[159,160] Further support for the connection between sleep and AD is the exciting finding that a major function of sleep may be to flush the brain, through the hypothesized "glymphatic system," including the flushing of toxic A-beta peptides.[161]

CLINICAL PEARL

An understanding of the genes and gene alleles that influence sleep and wake quality and the susceptibility to sleep loss may suggest novel targets or approaches to improve sleep and wake as well as for the treatment of sleep disorders. As in other areas of medicine, allelic differences in these genes may suggest different treatments for different individuals with the same disorder because the pathophysiology may be distinct. Rodent studies using molecular, forward, and reverse genetic approaches, in combination with human studies, are likely to lead to novel insights into the regulation and function of sleep, which can then be translated into improved treatments for sleep-wake disorders as well as for other mental and physical disorders associated with disrupted sleep.

SUMMARY

This chapter explores the progress in and various approaches to identification of genes important for an understanding of sleep and sleep-related traits in rodents and, by extension, in all mammals. Genetic and genomic methods have also led to or solidified an understanding of the neuroanatomy and neural pathways involved in sleep and wake regulation. Similar to other areas of medicine, an understanding of the underlying genetics will become increasingly important in sleep medicine as well. At this time, mice represent the best animal models for such studies, especially the many genetically modified mouse lines that have been created and the increasing use of various mouse strains that form important genetic reference populations for a wide range of biomedical research. Given the similarities in physiology and genetics across eutherian mammals, it is likely that genes influencing sleep traits in mice will also influence sleep traits in humans or at least help us identify the most critical molecular pathways and networks.

Selected Readings

Franken P. A role for clock genes in sleep homeostasis. *Curr Opin Neurobiol* 2013;**23**:864–72.

Jiang P, et al. A systems approach identifies networks and genes linking sleep and stress: Implications for neuropsychiatric disorders. *Cell Rep* 2015;**11**(5):835–48.

Liblau RS, Vassalli A, Seifinejad A, Tafti M. Hypocretin (orexin) biology and the pathophysiology of narcolepsy with cataplexy. *Lancet Neurol* 2015;**14**(3):318–28.

Mang GM, Franken P. Genetic dissection of sleep homeostasis. *Curr Top Behav Neurosci* 2015;**25**:25–63.

Millstein J, et al. Identification of causal genes, networks, and transcriptional regulators of REM sleep and wake. *Sleep* 2011;**34**:1469–77.

Threadgill DW, Churchill GA. Ten years of the collaborative cross. *Genetics* 2012;**190**:291–4.

A complete reference list can be found online at ExpertConsult.com.

Chapter 30

Genetics and Genomic Basis of Sleep in Healthy Humans

Hans-Peter Landolt; Derk-Jan Dijk

Chapter Highlights

- Distinct characteristics of human sleep are regulated by different molecular and genetic mechanisms with different degrees of heritability.
- The genetics of sleep and sleep-wake regulation are still relatively unknown compared with the genetics of other complex traits.
- The sleep electroencephalogram is one of the most heritable traits in humans. Elucidating the underlying genes may reveal novel sleep functions.
- Independent replication and confirmation of most of the currently proposed genetic associations with distinct characteristics of sleep and sleep regulatory processes in high-quality protocols and data sets are required.

EVIDENCE FOR TRAITLIKE AND GENOTYPE-DEPENDENT DIFFERENCES IN DIURNAL PREFERENCE, SLEEP TIMING, SLEEP ELECTROENCEPHALOGRAPHY, SLEEP ARCHITECTURE, AND SLEEP DURATION

Many aspects of sleep and sleep-wake regulation are highly variable among individuals, yet highly stable within individuals. Uncovering genetic factors contributing to these traitlike individual differences in healthy humans constitutes one of the most promising avenues to foster our understanding of the neurobiology of sleep in health and disease. This chapter summarizes the current evidence for genotype-dependent differences in timing, duration, and structure of sleep, as well as the sleep electroencephalogram (EEG) in healthy individuals. Table 30-1 summarizes known variations in genes and their functional significance that were investigated to date and whether they contribute to genotype-dependent differences in diurnal preference or sleep timing, sleep EEG, sleep structure, and sleep duration. The chapter also reviews how these differences may relate to the homeostatic and circadian regulation of sleep. Several sleep characteristics differ between the sexes and ethnic groups, but these differences are not discussed here.

The manifestation and regulation of sleep and the sleep EEG reflect different aspects of complex behaviors. Each of these aspects is likely to be under the control of multiple genes, which may interact, and are also influenced by the environment and other factors such as age. In humans, little is currently known about the genes that contribute to the traitlike, individual "sleep phenotypes." Similarly, little is known about the genes that contribute to individual "circadian phenotypes," although a considerable number of genes that contribute to circadian rhythmicity have been discovered in animals.

Two main techniques for the genetic dissection of normal human sleep are currently available. The first is to examine the effect of candidate genes, for which evidence exists that they are implicated in sleep and sleep-wake regulation. With this method, individuals with distinct genotypes of known genetic polymorphisms are prospectively studied in the sleep laboratory. This approach precludes discovery of novel "sleep genes" but may help to explain the consequences of these polymorphisms for sleep physiology. By contrast, genome-wide association (GWA) studies may lead to the identification of novel "sleep genes," which may lead to the discovery of novel sleep regulatory pathways. These studies, however, require very large sample sizes and multiple replications. The weaknesses and strengths of these strategies have been discussed in detail.[1]

Large interindividual differences are observed in preferred time of day for completion of distinct cognitive tasks, sleep timing, sleep EEG, sleep structure, and sleep duration. Genes contribute to each of these phenotypes, and a high degree of heritability (i.e., the percentage of variance explained by overall genetic effects) has been demonstrated for these variables. For some of these variables the magnitude of interindividual differences exceeds by far the size of the effects of manipulations of sleep regulatory processes, such as sleep deprivation.[2]

GENES CONTRIBUTING TO HUMAN MORNINGNESS-EVENINGNESS AND TIMING OF SLEEP

Candidate Genes

The timing of the peaks and troughs of daytime alertness and the timing of nocturnal sleep (i.e., diurnal preference) are highly variable among healthy individuals.[3] Some of us go to sleep when others wake up. Self-rating scales such as the Horne-Östberg Morningness-Eveningness Questionnaire and the Diurnal Type Scale show normal distribution along an eveningness-morningness axis,[4,5] indicating the contribution of additive, small effects of multiple genes in combination with the environment. Recent studies in large numbers of monozygotic (MZ) and dizygotic (DZ) twin pairs and population- and family-based cohorts revealed roughly 50% heritability for diurnal preference[6] and 22% to 25% for habitual bedtime.[7,8]

Table 30-1 Genes Investigated to Contribute to Genotype-Dependent Differences in Diurnal Preference, Sleep Timing, Sleep Electroencephalogram, Sleep Structure, Sleep Duration, and Sleep Homeostasis

Family	Gene	NCBI SNP-ID	Base Change/Amino Acid Change	Diurnal Preference	Sleep Timing	Sleep EEG	Sleep Structure	Sleep Duration	Sleep Homeostasis	Reference(s)
Clock gene pathway	*CLOCK*	rs1801260	c.3111T>C	√	√	√				5, 14-16
	CLOCK	rs2070062	c.257T>G	√						16
	CLOCK	rs12649507						√		87-89
	PER1		c.2548G>A	√						18, 19
	PER1	rs2735611	c.2434T>C	√						19
	PER1	rs7221412	g.8137696A>G	√						20
	PER2		c.1984A>G / p.Ser662Gly		√					21
	PER2	rs2304672	c.111G>C	√	√					24
	PER3	rs57875989	VNTR / del(1011-1028 aa)	√	√	√	√	√	√	46, 47, 108
	PER3	rs228697	c.2590C>G / p.Pro864Ala	√						29
	PER3	rs10462020	c.1940T>G / p.Val647Gly	√						30
	AANAT		c.619G>A / p.Ala129Thr	√						31, 32
	AANAT		c.–263G>C		√	√				33
	BMAL (*ARNTL2*)	rs922270	g.24165C>T	√						30
	BHLHE41	MIM:612975	c.1151C>G / p.Pro384Arg					√		90
	BHLHE41		c.1151C>A / p.Pro384Gln					√	√	92
	BHLHE41		c.1086C>T / p.Tyr362His					√	√	92
Adenosine	*ADA*	rs73598374	c.22G>A / p.Asp8Asn			√	√	√	√	53-55
	ADORA2A	rs5751876	c.1976T>C			√	√	√	√	53, 117, 118
Neurotransmitters	*GRIA3*	rs687577	g.123445253A>C					√		93
	COMT	rs4680	c.544G>A / p.Val158Met	√	√	√	√	√	√	63, 64, 119
Transporters	*SLC6A3* (*DAT1*)	rs28363170	VNTR	√	√	√	√	√	√	59, 119
	SLC6A4 (*5-HTT*)	rs687577	5-HTTLPR					√		103
Potassium channel	*ABCC9*	rs11046205	g.102303C>T					√		34
	ABCC9	rs11046209	g.97663T>A					√		107
Signaling pathways	*BDNF*	rs6265	c.196G>A / p.Val66Met			√	√		√	66, 67
	PRNP	rs1799990	c.385A>G / p.Met129Val			√				72
	PAX8	rs1823125	g.114090412A>G					√		108
Immune response	*DQB1* *0602								√	135

Gene: National Center for Biotechnology Information (NCBI) gene symbol. NCBI SNP-ID number: NCBI single nucleotide polymorphism reference number. Base change: Nucleotide substitution at indicated position of coding DNA. Amino acid change: Amino acid substitution associated with base change. √: Possible contribution to phenotypic variation was investigated and reported.

Morningness-eveningness and timing of sleep are thought to be determined in part by characteristics of the central circadian oscillator, and associations between the intrinsic period or phase marker of this oscillator and diurnal preference have been reported.[9-12] These oscillators consist at the molecular level of a network of interlocked transcriptional and translational feedback loops, which involve several clock-related genes, including the transcription regulators *CLOCK*, *BMAL1*, *PER1-3, CRY1-2,* and other genes. This knowledge has provided an obvious rational basis for the search for associations between these genes and morningness-eveningness and altered sleep timing.

The effect of a single nucleotide polymorphism (SNP) in the 3′-untranslated region (UTR) of the human circadian locomotor output cycles kaput gene (*CLOCK*) located on chromosome 4 on diurnal preference was first studied in middle-aged adults. This SNP may affect stability and half-life of messenger RNA (mRNA)[13] and thus alter the protein level that is finally translated. Katzenberg and colleagues[14] reported that homozygous carriers of the 3111C allele have increased evening preference for mental activities and sleep, with delays ranging from 10 to 44 minutes when compared with individuals carrying the 3111T allele. A similar association with diurnal preference was found in a Japanese population, and Morningness-Eveningness Questionnaire scores were significantly correlated with sleep-onset time and wake time.[5] By contrast, studies in healthy European and Brazilian samples failed to confirm an association between genetic variation in *CLOCK* and diurnal preference.[15,16] Interestingly, an almost complete linkage disequilibrium was shown between the 3111T>C and the 257T>G polymorphisms located in the other extremity of this gene.[16] Full-length analysis of secondary mRNA structure revealed no interaction between the two polymorphisms.

Mouse *Per1* and *Per2* are importantly involved in maintaining circadian rhythmicity,[17] and possible associations between variation in these genes and diurnal preference were thus also investigated in humans. Screening for missense mutations and functional or synonymous polymorphisms in promoter, 5′- and 3′-UTR and coding regions of the period-1 gene (*PER1*) in volunteers with extreme diurnal preference and patients with delayed sleep phase syndrome remained initially unsuccessful.[18,19] By contrast, the distribution of the C and T alleles of a silent polymorphism in exon 18 was found to differ between extreme morning and evening types.[19] Thus the frequency of the 2434C allele was roughly double in subjects with extreme morning preference (24%) compared with subjects with extreme evening preference (12%). This polymorphism may be linked to another functional polymorphism or directly affect *PER1* expression at the translational level.[19] In a candidate gene association study with replication, a polymorphism in *PER1* (single-nucleotide polymorphism identification number: rs7221412) was found to be associated with sleep timing based on actigraphy.[20]

A missense mutation in the human period-2 gene (*PER2*) currently provides the most striking example of a direct link between genetic variation in a clock gene and changed circadian rhythms. Linkage analyses in families afflicted with familial advanced sleep phase syndrome (FASPS) revealed associations with functional polymorphisms of *PER2* that cause altered amino acid sequences in regions important for phosphorylation of this protein[21] and a mutation in caseine kinase delta (CKδ), which plays an important role in phosphorylation.[22] The subsequent finding in a transgenic mouse model expressing the human FASPS mutation that casein kinase I delta (CKIδ) can regulate circadian period through *PER2* provided further important evidence that this gene is importantly involved in the mechanisms of circadian rhythm regulation in humans.[23] In accordance with this notion, a C111G polymorphism located in the 5′-UTR of *PER2* modulates diurnal preference in healthy volunteers.[24] Thus the 111G allele is significantly more prevalent in subjects with extreme morning preference (14%) than in individuals with extreme evening preference (3%). Computer simulation predicted that the 111G allele has different secondary RNA structure than the 111C allele and that the two transcripts may be differently translated.[24]

Findings in mice suggest that *Per3* has primary functions outside the central circadian clock.[17,25] Nevertheless, a variable-number tandem-repeat (VNTR) polymorphism in the human period-3 gene (*PER3*) also appears to modulate morning and evening preference. A 54-nucleotide sequence located in a coding region of this gene on human chromosome 1 is repeated in either four or five units. This difference may alter the dynamics in PER3 protein phosphorylation. The longer five-repeat allele was associated in European and Brazilian populations with morning preference and the shorter four-repeat allele with evening preference, respectively.[26-28] More recently in a sample of 925 healthy Japanese controls, the *PER3* SNP rs228697, which is associated with a proline-to-alanine amino acid substitution, was shown to be associated with diurnal preference such that the major C allele was more prevalent in morning types and the minor G allele more common in evening types.[29] In addition, in a sample of 966 young adults in Britain, a significant association between SNP rs10462020 of *PER3* and diurnal preference was reported such that G/G individuals had an increased morning preference compared with T/G and T/T individuals.[30] In this study an association between a polymorphism (rs922270) in *BMAL* (*ARNTL2*) and diurnal preference was also reported.

The gene encoding arylalkylamine *N*-acetyltransferase (*AANAT*) is located on human chromosome 17q25. This enzyme plays a key role in melatonin synthesis and may, thus, be important for diurnal preference and circadian rhythm disturbances. Comparison in a Japanese population between 50 outpatients diagnosed with delayed sleep phase syndrome and 161 unrelated healthy controls suggested that the frequency of a seldom-occurring threonine allele at codon 129 is significantly higher in patients than in controls.[31] This association was not confirmed in a Brazilian population, in which virtually no allelic variation at this position was found.[32] In a small study conducted in Singapore, it was suggested that a commonly occurring, silent −263G>C polymorphism of *AANAT* modulates sleep timing and sleep duration (also see later) among healthy students.[33]

Genome-wide Association Studies

Only three GWA studies of sleep-related phenotypes are currently available in humans.[7,34,35] In the Framingham Heart Study 100K Project,[7] phenotypic and genetic analyses were conducted in 749 subjects and revealed a heritability estimate for habitual bedtime of 22%. This small study suggests that a nonsynonymous polymorphism in a coding region of the gene encoding neuropeptide S receptor 1 (*NPSR1*) is a possible

modulator of usual bedtime as obtained from a self-completion questionnaire. This polymorphism leads to a gain-of-function mutation in the receptor protein by increasing the sensitivity for neuropeptide S receptor 10-fold.[36] Although a possible association of *NPSR1* to weekday bedtime is interesting, it has to be kept in mind that the statistical power of this pilot study is limited and necessary replication of this finding in independent samples is lacking. A recent analysis of a larger sample of the Framingham Offspring Cohort did not parallel the prior result.[8]

THE SLEEP ELECTROENCEPHALOGRAM IS AMONG THE MOST HERITABLE TRAITS IN HUMANS

Visual sleep state scoring relies on arbitrarily defined criteria and can reveal only limited information about sleep physiology. To obtain more detailed insights, quantitative analyses of the EEG signal recorded during sleep have to be performed. A powerful approach to quantify amplitude and prevalence of EEG oscillations with distinct frequencies is power spectral analysis based on fast Fourier transforms.[37-39] Recent studies strongly suggest that especially the sleep EEG, but also the waking EEG, are highly heritable traits in humans. All-night sleep EEG spectra derived from multiple recordings in healthy individuals show large interindividual variation and high intraindividual stability.[39,40] Buckelmüller and colleagues[40] recorded in eight young men two pairs of baseline nights separated by 4 weeks. Although the spectra in non–rapid eye movement (NREM) sleep differed largely among the individuals, the absolute power values and the shape of each subject's spectra were impressively constant across all nights (Figure 30-1). The largest differences among the subjects were present in the theta, alpha, and sigma (approximately 5 to 15 Hz) range. Hierarchical cluster analysis of Euclidean distances based on spectral values as feature vectors demonstrated that all four nights of each individual segregated into the same single cluster.[40] Similar results were obtained in rapid eye movement (REM) sleep and, by other researchers, in men and women of older age.[39] These data strongly suggest that the sleep EEG contains systematic and stable interindividual differences, which are at least in part genetically determined.

This notion is further supported by two recent twin studies investigating the heritability of the sleep EEG. Ambrosius and colleagues[41] quantified the sleep EEG profiles in 35 pairs of MZ twins (17 male pairs, 18 female pairs; age range: 17 to 43 years) and 14 pairs of DZ twins (7 male pairs, 7 female pairs; age range: 18 to 26 years). Stable and robust interindividual differences in a broad range of the NREM sleep EEGs were observed. Furthermore, intraclass correlation coefficients (ICC) of spectral power were significantly higher in MZ twins than in DZ twins.[41] The ICC reflect within-pair similarity of twin pairs. In frequencies between 0.75 and 13.75 Hz, the ICC equaled roughly 0.8 in MZ twins and roughly 0.6 in DZ twins. The differences between MC and DC twin pairs appeared most pronounced in theta and alpha (4.75 to 11.75 Hz) frequencies (see also Landolt[42]).

De Gennaro and colleagues[43] also conducted a twin study to test the hypothesis that the EEG in NREM sleep reflects a genetically determined, individual "fingerprint." They recorded baseline and recovery sleep after sleep deprivation in 10 MZ and 10 DZ twin pairs (mean age, 24.6 ± 2.4 years; five male and five female pairs in each group) and observed highest variability in the 8- to 16-Hz range. In this frequency band, group similarity quantified by an ICC procedure was more than double in MZ pairs (ICC = 0.934; 95% confidence interval [CI] = 0.911 to 0.965) than in DZ pairs (ICC = 0.459; 95% CI = 0.371 to 0.546) (Figure 30-2). This difference

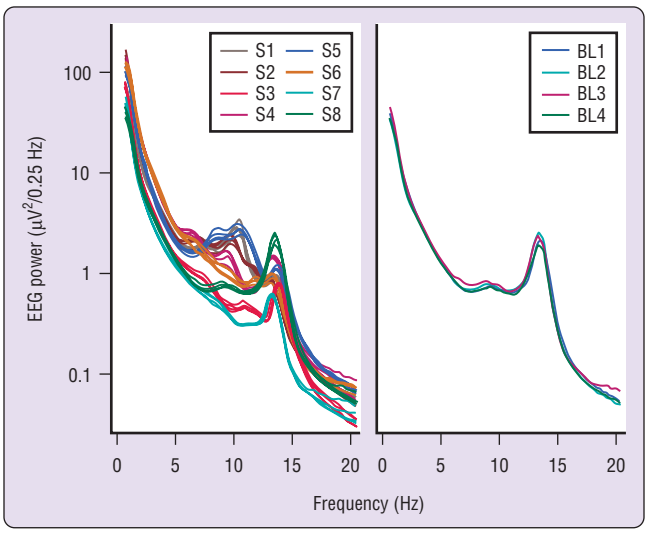


Figure 30-1 High *inter*individual variation (*left*) and high *intra*individual stability (*right*) in all-night electroencephalogram (EEG) power spectra in NREM sleep in 32 baseline nights of eight young men (S1 to S8). The largest interindividual variation is observed in theta, alpha, and sigma frequencies (~5 to 15 Hz). The spectra of all four baseline nights (BL1 to BL4) of one individual (S8) are virtually superimposable. (Modified from Buckelmüller J, Landolt HP, Stassen HH, Achermann P. Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience* 2006;138:351–6.)

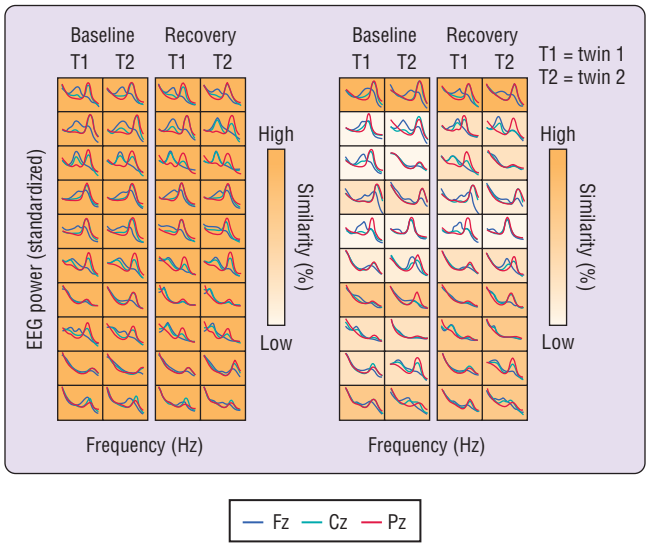


Figure 30-2 Heritability of NREM sleep electroencephalogram (EEG) is more than 90%. Panels show color-coded similarity indexes of 8 to 16 Hz activity in monozygotic (*left*) and dizygotic (*right*) twin pairs. The similarity index in each twin pair was scaled and color coded between minimal (0% similarity, *white*) and maximal (100% similarity, *dark orange*). *Black lines* indicate derivation Fz; *blue lines* indicate derivation Cz; *red lines* indicate derivation Pz (unipolar derivations referenced to averaged mastoid). (Modified from De Gennaro L, et al. The EEG fingerprint of sleep is genetically determined: a twin study. *Ann Neurol* 2008;64:455–60.)

suggested 95.9% heritability independently of sleep pressure.[43] As such, the sleep EEG qualifies as one of the most heritable traits known so far, only matched by heritability estimates for distinct brain characteristics like cortical gray matter distribution.[1] Thus it may be likely that trait characteristics of rhythmic brain oscillations during sleep and distinct neuroanatomic features are interrelated.

In conclusion, accumulating evidence suggests that the sleep EEG is a highly heritable trait, yet the underlying genetic determinants are largely unknown. Nevertheless, more and more studies investigate the effects of known allelic variants of candidate genes on the human sleep EEG (see Table 30-1). The findings demonstrate that genetic variation of various cells, molecules, and signaling pathways can profoundly modulate sleep EEG and other sleep phenotypes. Selected genes and pathways will be briefly discussed in the following paragraphs.

GENES CONTRIBUTING TO THE SLEEP ELECTROENCEPHALOGRAM

Circadian Clock Genes

A wealth of studies in genetically modified mice and flies demonstrates that circadian clock genes are strong determinants of major characteristics of the sleep EEG.[1,44] The only, yet intensively, studied "clock gene" variant in healthy humans is the previously mentioned VNTR polymorphism of *PER3* (rs57875989).[45] Apart from its impact on diurnal preference, this polymorphism also modulates the sleep EEG in NREM as well as in REM sleep. Compared with individuals with the $PER3^{4/4}$ genotype, young adult homozygous carriers of the long-repeat allele ($PER3^{5/5}$ genotype) exhibited higher EEG activity in the delta range (1 to 2 Hz) in NREM sleep and in the theta and alpha range (7 to 10 Hz) in REM sleep.[46] Partly similar observations were made in healthy older adults between 55 and 75 years of age.[47]

Adenosinergic Neuromodulation

The neuromodulator adenosine is released in an activity-dependent manner, and genes encoding adenosine-metabolizing enzymes and adenosine receptors are thought to play a major role in regulating the quality of sleep and wakefulness in animals and humans.[1,48] Adenosine kinase and adenosine deaminase (ADA) importantly contribute to the regulation of extracellular adenosine levels.[49] Genetic studies in mice suggest that both enzymes are involved in sleep-wake homeostasis.[50,51] In humans, the *ADA* gene is located on chromosome 20q13.11 and encodes two electrophoretic variants of ADA, referred to as ADA*1 and ADA*2 (rs73598374). The ADA*2 variant results from a guanine-to-adenine transition at nucleotide 22, which is translated into an asparagine-to-aspartic acid amino acid substitution at codon 8. The heterozygous ADA*1-2 (G/A) genotype shows reduced catalytic activity of ADA compared with homozygous individuals carrying the ADA*1 (G/G genotype) variant.[52] Rétey and colleagues[53] observed that this polymorphism affects the spectral composition of the sleep EEG. More specifically, EEG delta activity in NREM sleep (0.25 to 5.5 Hz) and REM sleep (2.0 to 2.25 Hz and 3.5 to 4.75 Hz) was higher in the G/A genotype than in the G/G genotype.[53] Inspired by studies in inbred mice showing that the genomic region encoding *Ada* modifies the rate at which sleep need accumulates during wakefulness,[51] it was then examined whether individuals with G/A and G/G genotypes respond differently to sleep deprivation. In accordance with the original study, delta (0.75 to 1.5 Hz) activity in NREM sleep was elevated in the G/A genotype compared with the G/G genotype in both baseline and recovery nights.[54] The *ADA* genotype-dependent EEG alterations, however, were not restricted to the low-delta range in NREM sleep but also included a pronounced increase in theta and alpha frequencies (~6 to 12 Hz) in NREM sleep, REM sleep, and wakefulness. Importantly, an independent study in a large epidemiologic sample confirmed that A-allele carriers have higher delta power in NREM sleep and increased theta power in NREM and REM sleep compared with homozygous G/G genotype carriers.[55]

The effects of adenosine on target cells are mediated through four different subtypes of G-protein-coupled adenosine receptors: A_1, A_{2A}, A_{2B}, and A_3 receptors. It is thought that adenosine modulates sleep primarily by binding to high-affinity A_1 and A_{2A} receptors.[48,56] No study yet investigated the possible effects of variants of the A_1 receptor gene on the human sleep EEG. By contrast, it was shown that the common variation rs5751876 of the adenosine A_{2A} receptor gene (*ADORA2A*) located on chromosome 22q11.2 affects the EEG in NREM and REM sleep.[53] This polymorphism is linked to a 2592C>T_{ins} polymorphism in the 3′–UTR of *ADORA2A* and may modulate receptor protein expression.[57] In a case–control study, Rétey and coworkers observed that EEG activity in the approximately 7 to 10 Hz range was invariably higher in all vigilance states in subjects with the C/C genotype of rs5751876 than in subjects with the T/T genotype.[53] Because the C allele is thought to facilitate A_{2A} receptor function compared with the T allele, these data may suggest that genetically increased A_{2A} receptor–mediated signal transduction enhances EEG theta and alpha activity independently of sleep state.

Neurotransmitters

Accumulating evidence suggests a contribution of dopamine to sleep-wake regulation in humans.[58,59] The enzyme catechol-*O*-methyltransferase (COMT) plays a major role in the metabolic degradation of brain catecholamines, including dopamine. The gene encoding COMT is located on human chromosome 22q11.2, in proximity to *ADORA2A*. Human *COMT* contains a common functional 544G>A variation that alters the amino acid sequence of COMT protein at codon 158 from valine (Val) to methionine (Met).[60] Individuals homozygous for the Val allele show higher COMT activity and lower dopaminergic signaling in prefrontal cortex than Met/Met homozygotes.[61,62] Sleep variables and the sleep EEG response did not differ between male carriers of Val/Val and Met/Met genotypes.[63] By contrast, the Val158Met polymorphism of *COMT* was associated with consistently lower EEG activity in the upper-alpha (11 to 13 Hz) range in NREM sleep, REM sleep, and wakefulness in Val/Val compared with Met/Met homozygotes.[64] The difference in NREM sleep was present before and after sleep deprivation and persisted after administration of the wake-promoting compound modafinil during prolonged wakefulness (Figure 30-3). These data demonstrate that a functional variation of the *COMT* gene predicts robust interindividual differences in the sleep EEG. In addition, this polymorphism profoundly affected the efficacy of modafinil to improve impaired well-being and

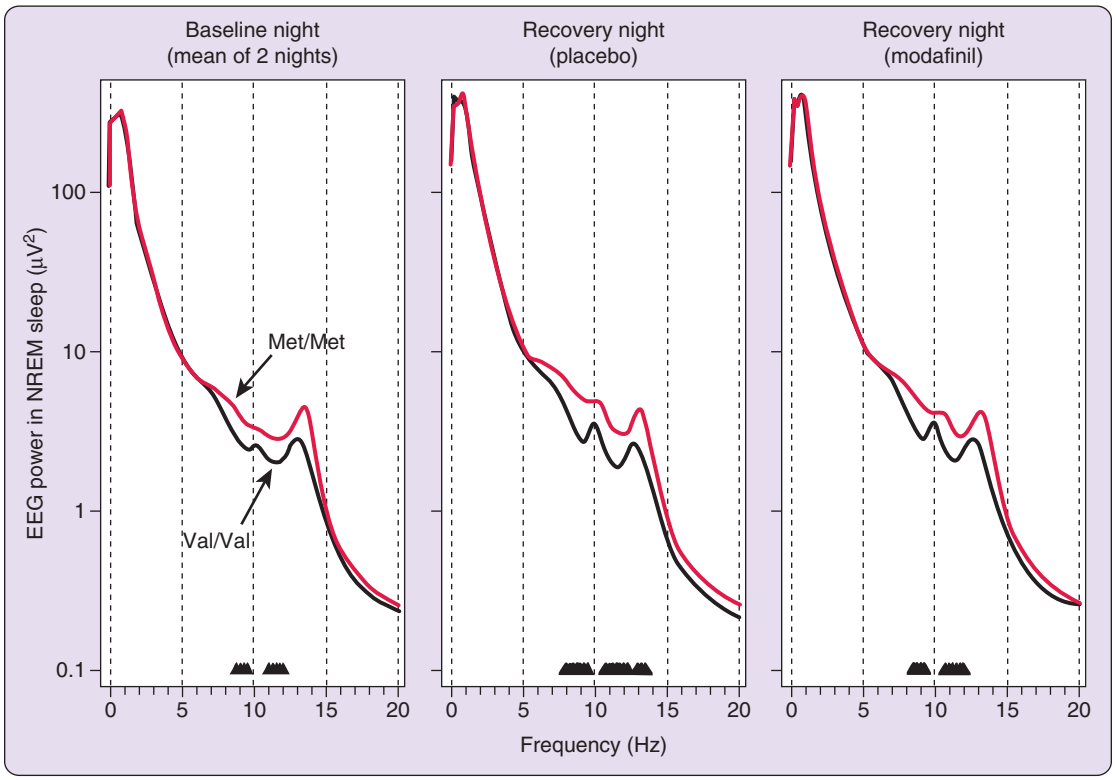


Figure 30-3 The Val158Met polymorphism (rs4680) of the gene encoding catechol-*O*-methyltransferase (COMT) modulates electroencephalogram (EEG) alpha activity in NREM sleep (all-night power spectra of stages 2 to 4). *Black triangles* at the bottom of the panels indicate frequency bins, which differ significantly between Val/Val ($n = 10$, *black lines*) and Met/Met ($n = 12$, *red lines*) genotypes ($P < .05$, unpaired, two-tailed t-tests). The frequency-specific effect of the genetic variation is robust against the effects of prolonged wakefulness and the stimulant modafinil. (Data from Bodenmann S, et al. The functional Val158Met polymorphism of COMT predicts interindividual differences in brain alpha oscillations in young men. *J Neurosci* 2009;29:10855–62.)

cognitive functions after sleep deprivation.[65] Thus two-time 100 mg modafinil potently improved vigor and well-being and maintained baseline performance of executive functioning and vigilant attention throughout 40 hours of prolonged wakefulness in 10 Val/Val homozygotes, yet the same dose was virtually ineffective in 12 Met/Met homozygotes. Interestingly, an opposite relationship between Val158Met genotype of *COMT* and measures of daytime sleepiness may be present in patients suffering from narcolepsy (see Clinical Pearl).

Signaling Pathways

Another functional polymorphism affecting the sleep EEG in theta and alpha frequencies is a guanine-to-adenine transition at nucleotide 196 of the gene encoding brain-derived neurotrophic factor (BDNF) (rs6265).[66,67] This polymorphism is located on human chromosome 11p13 and causes a valine-to-methionine amino acid substitution at codon 66 of the pro-BDNF sequence. In vitro studies suggest that the presence of a Met allele reduces intracellular trafficking and activity-dependent secretion of mature BDNF protein.[68] This polymorphism is typically associated with reduced performance on cognitive tasks that are also affected by sleep deprivation. Sleep and the sleep EEG were first investigated in case–control fashion in 11 carriers of the Val/Met genotype and 11 prospectively matched Val/Val homozygotes. It was found that the Val66Met polymorphism of *BDNF* not only reduced response accuracy on a verbal two-back working memory task but also modulated the spectral composition of the EEG in a frequency and vigilance state–specific manner.[66] More specifically, in baseline and recovery nights after sleep deprivation, delta, theta, and low-alpha activity in NREM sleep EEG was lower in Met allele carriers than in Val/Val homozygotes. Importantly, the genotype-dependent differences in the theta and low-alpha band (approximately 4 to 9 Hz) were recently confirmed in a large and ethnically diverse population-based epidemiologic sample.[67]

A point mutation at codon 178 (in rare cases also a mutation at codon 200) of the prion protein gene (*PRNP*) has been identified as the cause underlying the devastating disease, fatal familial insomnia.[69,70] Interestingly, although healthy relatives of fatal familial insomnia patients appear to have normal sleep EEG,[71] the polymorphic codon 129 of the *PRNP* gene may influence EEG activity during sleep.[72] Subjects with the Met/Val genotype showed lower slow-wave activity and higher spindle frequency activity than individuals with the Val/Val genotype, independent of codon 178.

GENES CONTRIBUTING TO SLEEP ARCHITECTURE

Not only the sleep EEG but also many variables characterizing sleep architecture demonstrate large variation among

individuals and high stability within individuals.[2,39,40,73] For example, the intraclass correlation coefficients, which estimate the intraindividual stability of a given variable across different conditions (i.e., baseline versus sleep deprivation), was reported to be 0.73 for slow wave sleep (SWS) and 0.48 for REM sleep.[2] This observation suggests the presence of traitlike, interindividual differences in sleep physiology, which have a genetic basis. Indeed, twin studies show striking similarity and concordance in visually defined sleep variables in MZ twins, yet not in DZ twins. The first polysomnographic sleep studies in MZ twins revealed almost complete concordance in the temporal sequence of sleep stages.[74] Subsequent work showed that in particular those variables, which most reliably reflect sleep need, are under tight genetic control. Apart from total sleep time, they include duration of NREM sleep stages, especially SWS, and density of rapid eye movements in REM sleep.[75-77] Linkowski[77] estimated that heritability of markers of sleep homeostasis is up to 90% (REM density).

Slow Wave Sleep

A few studies have conducted polysomnographic assessment in defined genotypes. The *CLOCK* genotypes that were associated with diurnal preference[14] did not significantly affect sleep variables derived from nocturnal polysomnography. By contrast, it was found that young homozygous carriers of the long-repeat genotype of *PER3* (*PER3*$^{5/5}$) fell asleep more rapidly and showed more SWS, particularly stage 4 sleep, compared with homozygous 4-repeat individuals.[46,78] A difference in SWS, yet on a lower level, was also observed in older people.[47]

Similarly, with respect to polymorphism rs73598374 of *ADA*, healthy carriers of the ADA*2 allele (G/A genotype) showed significantly more SWS than subjects with the G/G genotype.[53,54] All other sleep variables were similar in both *ADA* genotypes.

The impact of the Val66Met polymorphism of *BDNF* was also reflected in sleep architecture. In baseline and recovery nights, Val/Val allele carriers spent roughly 20 minutes more in deep stage 4 sleep than Val/Met allele carriers. By contrast, superficial stage 2 sleep was reduced.[66] Taken together, functional variation in the genes encoding *PER3*, *ADA,* and *BDNF* modulate not only the spectral characteristics of the sleep EEG but also sleep architecture.

GENES CONTRIBUTING TO HABITUAL SLEEP DURATION

Habitual sleep duration shows large variation among healthy individuals, and the physiologic sleep and circadian correlates of habitual short and long sleepers have been identified in small groups of subjects.[79-81] The temporal profiles of nocturnal melatonin and cortisol levels, body temperature, and sleepiness under constant environmental conditions and in the absence of sleep suggest that the circadian pacemaker programs a longer biologic night in long sleepers than in short sleepers.[81] Individual differences in this circadian program may contribute to the large variation in habitual sleep duration, which shows a perfect normal distribution in the general population.[82,83] Such a distribution is consistent with the influence of multiple, low-penetrance polymorphisms. Twin and GWA studies reported for sleep duration heritability estimates of 9% to 40%.[7,35,84-86]

Circadian Clock Genes

A candidate gene study of 194 SNPs in clock genes and self-reported sleep duration on the Munich Chronotype Questionnaire was recently conducted in a European population ($n = 283$).[87] The top two associations were both located in the gene *CLOCK* on chromosome 4. With one of these variants, rs12649507, sleep duration was significantly associated in the original discovery sample, in a replication sample ($n = 1011$), and in the meta-analysis of the two populations ($P < .009$).[87] Two recent studies aimed at replicating this initial finding; however, they revealed inconsistent results. Although Evans and colleagues[88] reported successful replication of the previously described association in 2527 male elderly participants, Lane and coworkers[89] found no evidence of an association. These authors collected objective polysomnographic data in three large independent cohorts of European ancestry. This analysis with more than 99% power to detect an effect of similar magnitude as previously reported did not support a significant association of *CLOCK* variants with sleep duration.

Evidence for a role of clock genes in modulating sleep duration also came from work in a small family who apparently needed just 6 hours of sleep per night.[90] This family-based candidate gene study revealed a point mutation in exon 5 of the gene encoding class E basic helix-loop-helix protein 41 (*BHLHE41*), also known as the transcriptional repressor gene *DEC2*. By this missense mutation (c.1151C>G), proline is replaced by arginine at amino acid position 384 (p.Pro384Arg) of BHLHE41 protein. This protein is part of the transcription factor family that is regulated by the mammalian circadian clock and influences the expression of *CLOCK/BMAL1*.[91,92] Interestingly, knocking-in the human mutation into mice and *Drosophila* species was reported to result in reduced sleep duration in transgenic animals.[90] Based on this study, other variants of the *BHLHE41* gene were searched for by DNA sequencing in two larger cohorts ($n = 417$) of healthy volunteers, and two other rare variants in the same exon of *BHLHE41* were found.[92] The phenotypic data reported in three carriers of the nonsynonomous variant c.1151C>A (p.Pro384Gln) and in one DZ twin pair discordant for the functional c.1086C>T (p.Tyr362His) polymorphism may suggest that variants that alter the suppression of *CLOCK/BMAL1* activation lead to short sleep, whereas a polymorphism that does not affect this suppression has no effect on sleep duration.[92]

Neurotransmitters

It is well established that the regulation of sleep and mood are closely related. A regression analysis of 23 risk variants of major depressive disorder covering 12 different genes with self-reported sleep duration was conducted in 3147 healthy individuals of two population-based Finnish cohorts. Polymorphism rs687577 (g.123445253A>C) of the gene *GRIA3* (ionotropic glutamate receptor, AMPA subunit 3) located on chromosome X was found to be significantly associated with sleep duration in healthy women.[93] More specifically, the frequency of the C/C genotype was highest in all age groups younger than 70 years in women reporting to sleep 8 hours or less. The frequency of this genotype decreased with longer sleep duration, and individuals with 9 to 10 hours of sleep showed higher frequencies of C/A and A/A

genotypes than midrange sleepers (7 to 8 hours). It was concluded that mood disorders and short sleep may share a common genetic and biologic background involving glutamatergic neurotransmission.[93]

Transporters

It has long been suggested that serotonin (5-hydroxytryptamine [5-HT]) is critically involved in sleep-wake mechanisms,[94] yet the specific roles for this neurotransmitter in sleep-wake regulation remain uncertain.[95] Current evidence supports the view that 5-HT contributes to the buildup of sleep need during wakefulness. Apart from its intracellular metabolism by monoamine oxidase, 5-HT is removed from the synapse by high-affinity serotonin transporters (5-HTT). In the brain, the 5-HTT is among the most important sites of action for many currently used antidepressant treatments.[96] A functional 44-base pair insertion and deletion polymorphism in the promoter region of the *5-HTT* gene (*5-HTTLPR*) located on chromosome 17q11.2 has been associated with neuropsychiatric diagnoses and individual responses to antidepressant treatments. Although this polymorphism can be subdivided further,[97] researchers commonly report it with two variations in humans: a long (L) or a short (S) variant allele. In vitro studies showed that basal transcriptional activity of the L allele is more than doubled when compared to the S variant allele.[98] Human individuals homozygous for the L/L variant show higher 5-HTT mRNA levels in postmortem brain tissue than subjects carrying the S allele (L/S + S/S).[99] Moreover, reduced transcription associated with the S allele may affect serotonergic tone and 5-HT receptor–mediated neurotransmission.[100] An association study in 157 patients suffering from primary insomnia suggested that the S variant is overrepresented in insomnia patients compared with healthy controls (n = 827).[101] Furthermore, this polymorphism may also mediate individual differences in the effects of chronic stress or stressful life events on impaired sleep quality and self-reported short sleep duration.[102,103] Nevertheless, other research indicated poorer sleep in L/L homozygotes than in carriers of at least one S allele, suggesting that the effects of this gene may be heterogeneous in different populations.[104]

Genome-wide Association Studies

The Framingham Heart Study 100K Project revealed a linkage peak to usual sleep duration on chromosome 3, including the gene encoding prokineticin 2 (*PROK2*).[7] This neuropeptide may be an important output molecule from the SCN, in particular in defining the onset and maintenance of the circadian night.[105,106] Because the danger of false-positive inferences from small GWA studies is high, the methodologic limitations of this work discussed previously also apply to this potential association. It was not corroborated in a larger sample of the Framingham Cohort.[8]

To identify novel genes associated with sleep duration, Allebrandt and colleagues performed GWA studies for self-reported average weekly sleep duration in seven discovery cohorts of a European consortium (n = 4251).[34] Meta-analysis revealed a genome-wide significant signal in the *ABCC9* (adenosine triphosphate [ATP]-binding cassette, subfamily C member 9) gene locus (rs11046205) that encodes one subunit of the ATP-sensitive potassium (K_{ATP}) channel.[34] The finding from the discovery cohorts was replicated when an *in silico* (GWA data) sample as well as a subgroup population of a large de novo (single genotyping) sample were additionally included in the meta-analysis. To confirm the role of *ABCC9* in modulating sleep duration, the homologue of this gene was knocked down in *Drosophila* species, which shortened nighttime sleep duration. Approximately 5% of the variance in sleep duration may be explained by this genetic variation in *ABCC9*.[34] In a candidate gene approach of another group attempting to replicate the proposed association, a significant association of the *ABCC9* gene with sleep duration was seen for a different polymorphism (rs11046209) and only in a rare homozygous genotype (n = 2).[107] By contrast, the previously suggested polymorphism of *ABCC9* (rs11046205) was associated with depressive symptoms.

A very recent study combining 18 community-based cohorts including more than 47,000 individuals of European ancestry revealed a genome-wide significant association with polymorphisms in a gene located on chromosome 2 encoding the thyroid-specific transcription factor *PAX8* (paired box gene 8).[108] The finding was replicated in an African American sample of about 4800 individuals. Although the finding is interesting, each copy of the minor allele only causes an estimated increase in usual sleep duration of approximately 3 minutes per copy and explains as little as 0.07% of variance in sleep duration.

In conclusion, no GWA studies of habitual sleep duration in humans have yet been convincingly reproduced or have explained a major portion of the variance in sleep length. Large sample sizes are needed for detecting genome-wide significant variants of genetically complex traits such as sleep duration. Thus the phenotypic data in the available studies typically rely on questionnaire-derived, self-reported sleep duration or time in bed. These measures differ when assessed with different questionnaires, as well as when compared with objectively verified sleep duration, which may challenge the reliability and reproducibility of the currently available studies.

GENETIC BASIS OF SLEEP-WAKE REGULATION: INTERACTION BETWEEN CIRCADIAN AND HOMEOSTATIC SYSTEMS

Many of the traits and genes described earlier concern sleep-wake characteristics as assessed under baseline conditions. How these alterations in sleep characteristics relate to sleep-wake regulation and how they may lead to functional consequences remain largely unexplored. The available data, however, already indicate that the effects cross boundaries between sleep and wakefulness and homeostatic and circadian aspects of sleep-wake regulation. For example, the polymorphisms in *PER3, ADORA2A,* and *COMT* affect the EEG in NREM sleep, REM sleep, and wakefulness. To investigate whether these changes reflect changes in EEG generating mechanisms with or without a relation to sleep regulatory processes requires these processes to be challenged by, for example, sleep deprivation.

Circadian Clock Genes

Comparing the effects of sleep deprivation with $PER3^{4/4}$ individuals revealed that the increase in theta activity in the EEG during wakefulness was more rapid in carriers of the $PER3^{5/5}$ genotype.[46] In addition, in recovery sleep following total sleep deprivation, REM sleep was reduced in $PER3^{5/5}$ individuals.

Finally, some data suggested that the increase in slow wave energy after sleep restriction was slightly higher in adults carrying the *PER3*$^{5/5}$ genotype than in *PER3*$^{4/5}$ and *PER3*$^{4/4}$ allele carriers[109] and also that the decline of cognitive performance during prolonged wakefulness and after sleep restriction differed as a function of the *PER3* genotype.[110-112] The differential susceptibility to the negative effects of sleep loss on waking performance was particularly pronounced in the second half of the circadian night and on tasks of executive functioning.[110] One interpretation of these data is that the VNTR polymorphism in *PER3* affects the dynamics of the homeostatic process, which then through its interaction with the circadian regulation of performance leads to differential sleep ability and vulnerability to the negative effects of sleep loss.[110,112] Indeed, it has previously been shown that individuals differ not only with respect to baseline characteristics of sleep but also in their response to sleep loss and that this vulnerability is a traitlike characteristic. The data suggest a contribution of *PER3* to individual tolerance to shift work and jetlag, which are highly prevalent in society.

A 6-hour sleep deprivation in mice carrying the Pro384Arg mutation of *BHLHE41* resulted in a smaller rebound in both NREM sleep and REM sleep and in a smaller relative increase in EEG delta power compared with control mice.[90] Furthermore, a functional variant (c.1086C>T) at another location in the same exon was studied in a DZ twin pair. The carrier of the variant was reported to have less recovery sleep following sleep deprivation and to produce fewer performance lapses during prolonged waking than the no-variant carrier. The variant reduced the ability of *BHLHE41* to suppress CLOCK/BMAL1 and NPAS2/BMAL1 transactivation in vitro, suggesting that genetic variants modifying the normal function of BHLHE41 may affect the homeostatic response to sleep deprivation.[92]

Adenosinergic Neuromodulation

Quantitative trait-locus analyses in inbred mouse strains revealed that a genomic region including *Ada* modifies the rate at which NREM sleep need accumulates during wakefulness.[51] Based on this observation, it was investigated whether human carriers of G/A and G/G genotypes of *ADA* respond differently to sleep deprivation.[54,113] Bachmann and colleagues first systematically studied attention, learning, memory, executive functioning, and self-reported sleep duration in 245 healthy adults.[54] They found that heterozygous carriers of the variant allele (G/A genotype, $n = 29$) performed significantly worse on the d2 attention task than G/G homozygotes ($n = 191$). To test whether this difference reflected elevated sleep pressure, sleep and sleep EEG before and after sleep deprivation were recorded in two prospectively matched groups of 11 G/A and 11 G/G genotypes. Corroborating two independent studies,[53,55] EEG delta activity and SWS were higher in the G/A than the G/G genotype. In addition, sustained attention (d2 and psychomotor vigilance tasks) and vigor were reduced, whereas EEG alpha oscillations in waking, as well as sleepiness, fatigue, and α-amylase activity in saliva (a proposed biomarker of sleep drive), were increased throughout prolonged wakefulness.[54] These convergent behavioral, neurophysiologic, subjective, and biochemical data demonstrated that genetically reduced ADA activity is associated with elevated sleep pressure. By contrast, the dynamics of the homeostatic response to sleep deprivation were not affected by *ADA* genotype.[54,113] Thus the data suggest an elevated level in overt NREM sleep propensity in the G/A genotype compared with G/G homozygotes, which may be due to elevated adenosinergic tone at the synapse because of genetically reduced ADA activity.

Convergent observations in candidate gene and GWA studies strongly suggest that genetic variation of *ADORA2A* is a determinant of individual sensitivity to subjective and objective effects of caffeine on sleep.[114,115] Interestingly, caffeine-sensitive and caffeine-insensitive individuals appeared to be differently affected by sleep loss.[116] These observations suggest that genetic variants of *ADORA2A* may alter the accumulation of homeostatically regulated sleep propensity during prolonged wakefulness. Convergent findings in mice[117] and humans[118] are consistent with this notion. They indicate that the sleep-deprivation–induced rebound of EEG delta activity in NREM sleep, the most reliable marker of sleep homeostasis, depends on the functional state of A_{2A} receptors.[119]

Neurotransmitters

Valomon and coworkers[120] recently investigated whether the Val158Met polymorphism of *COMT* (rs4680) affects actigraphy-derived rest-activity cycles and sleep estimates in 110 healthy adults. No genotype-dependent differences in actigraphy-derived circadian rest-activity patterns were found. Nevertheless, *COMT* genotype modulated the magnitude of sleep rebound on rest days compared with workdays. This difference is thought to reflect the compensation for a sleep debt accumulated during workdays ("social jetlag"). The Val/Val and Met/Met homozygotes significantly prolonged habitual sleep on rest days, whereas the Val/Met heterozygotes did not.[120] Similarly, neurophysiologic markers of sleep homeostasis did not differ between homozygous Val/Val and Met/Met allele carriers.[63,65] By contrast, one study suggested that the Val158Met polymorphism of *COMT* may be related to interindividual differences in sleep homeostasis and physiologic sleep responses to partial sleep deprivation.[121] To further tackle the question of whether COMT plays a role in sleep homeostasis, the effects of pharmacologic interference with COMT enzymatic activity on the consequences of sleep deprivation in different *COMT* genotypes may be studied.

Transporters

Genetically modified animals with reduced dopamine clearance exhibit an increased homeostatic response to prolonged wakefulness compared with wild-type animals. For example, mutant flies (*Dat*^{fo}) with reduced dopamine acetyltransferase activity show a greater sleep rebound after prolonged waking than wild-type controls.[122] Furthermore, *Drosophila* species and mouse mutants lacking functional dopamine transporter (DAT) exhibit prolonged wakefulness and shortened sleep.[123-125] In mammals, the DAT is highly expressed in basal ganglia where it is responsible for reuptake of dopamine and constitutes a rate-limiting mechanism of dopaminergic neurotransmission.[126] An important role for the basal ganglia in sleep-wake regulation has been recently suggested.[127,128] The response to sleep deprivation was studied in 57 adult volunteers genotyped for the 3′-UTR VNTR polymorphism (rs28363170) of the gene (*DAT1*, *SLC6A3*) encoding DAT. Ten (10R) or nine repeats (9R) of a 40-base pair sequence of this gene on chromosome 5p15.3 are most common, whereas the 10R-allele homozygotes have 15% to 20% reduced DAT

availability in the striatum compared with heterozygous and homozygous 9R-allele carriers.[129,130] Consistent with the evidence from transgenic animals, it was found that the sleep deprivation–induced increase in SWS, EEG delta activity, and number, amplitude, and slope of low-frequency (0.5 to 2.0 Hz) oscillations in NREM sleep was significantly larger in the 10R/10R genotype than in the 9R carrier genotype.[59] The data indicated an increased homeostatic response to sleep deprivation in 10R/10R allele carriers of *DAT1* compared with 9R allele carriers.

Signaling Pathways

Recent findings in rats suggested a causal relationship between BDNF and the regulation of EEG delta activity in NREM sleep.[131,132] Inspired by these studies, the possible effect of the Val66Met polymorphism on the regulation of neurophysiologic markers of sleep homeostasis was examined in humans.[66] Delta power in the first NREM sleep episode of a baseline, as well as of a recovery night after prolonged wakefulness, was specifically higher in Val/Val compared with Val/Met genotype subjects. By contrast, activity in high-alpha/low-sigma frequencies (approximately 10 to 13.5 Hz) was reduced. Thus *BDNF* genotype modulated established EEG markers of NREM sleep intensity, whereas the rebound in delta activity after sleep deprivation and its dissipation throughout the nights were only subtly affected. These findings suggest that Val/Val genotypes exhibit overall higher NREM sleep pressure than Val/Met genotypes, which may obscure subtle genotype-dependent differences in the dynamics of sleep homeostasis.

Immune Response

The human leukocyte antigen (HLA) *DQB1*0602* allele is the best HLA marker for narcolepsy, a neurologic disorder characterized by excessive daytime sleepiness, fragmented sleep, and shortened REM sleep latency. Although more than 90% of patients with narcolepsy-cataplexy carry HLA-*DQB1*0602*, 12% to 38% of allele-positive carriers are healthy sleepers.[133,134] A study in 129 healthy subjects suggested that *DQB1*0602*-positive individuals showed decreased sleep homeostatic pressure with steeper declines and greater sleepiness and fatigue in baseline.[135] During partial sleep deprivation, slow wave energy increased in positive and negative subjects, whereas *DQB1*0602*-positive individuals showed more fragmented sleep and altered REM and stage 2 sleep in baseline and during partial sleep loss. Although these preliminary findings are interesting, independent replication is critically required for their validation.

HUMAN SLEEP PHARMACOGENETICS

Individual responses to treatments with pharmacologic agents vary widely in healthy individuals and diseased patients. The differences may relate to weight, body composition, age, gender, and ethnic descent. Furthermore, genetic factors modifying pharmacokinetic or pharmacodynamic properties of molecules and constitutive pathways are becoming increasingly recognized as key determinants of individual responses to pharmacologic treatments. Apart from potentially important implications for the neurobiology of sleep-wake disorders and their pharmacologic management, sleep pharmacogenetics also offers a powerful novel approach to identifying molecular mechanisms contributing to sleep-wake regulation in humans. For example, pharmacogenetic studies of caffeine not only revealed insights into a distinct molecular contribution to individual caffeine sensitivity but also indicated that adenosine A_{2A} receptors and DAT are part of a biologic pathway that regulates sleep.

Adenosinergic Neuromodulation

Since people drink coffee, it is well known that some people are sensitive to its stimulant effects whereas some others are not. With respect to sleep disturbances, already the first scientific study 100 years ago showed that "a few individuals show complete resistance to the effects of small doses of caffeine."[136] Because subsequent work revealed no consistent pharmacokinetic differences between caffeine-sensitive and caffeine-insensitive subjects, endogenous diversity at its site of action was proposed to influence caffeine's effects on sleep.[137] Recent work in mice provided strong evidence that the stimulant promotes wakefulness primarily by blocking the A_{2A} subtype of adenosine receptors.[138] In humans, the variant rs5751876 in the coding region of the *ADORA2A* gene contributes to individual sensitivity to caffeine effects on sleep.[114] In 4329 responders to a brief Internet questionnaire, caffeine consumption was associated with subjectively reduced sleep quality in caffeine-sensitive respondents, but not in caffeine insensitive respondents, and the distribution of carriers of C/C and T/T alleles of *ADORA2A* differed between caffeine-sensitive and caffeine-insensitive individuals. Double-blind study of the effects of the stimulant on the sleep EEG confirmed the self-rated caffeine sensitivity, suggesting that genetic variation of *ADORA2A* is a determinant of individual sensitivity to the effects of caffeine on sleep.[114]

Indeed, Byrne and colleagues[115] provided a recent confirmation of a role for *ADORA2A* in caffeine-related sleep disturbances. They conducted a GWA study in 2402 twins and their families of the Australian Twin Registry. More than 2 million common polymorphisms were examined. Caffeine-associated sleep disturbance was based on the participants' report of whether or not they have ever experienced caffeine-induced insomnia, statistically corrected by a "general insomnia factor score" derived from a questionnaire. Importantly, the previously suggested association between genetic variation of *ADORA2A* and disturbed sleep after caffeine was successfully replicated. This finding is remarkable in the genetics of complex traits because only a small minority of candidate genes has typically been confirmed.[139] The original variant (rs5751876) was not typed in the GWA sample. Nevertheless, this variant forms a perfect linkage-disequilibrium with several other variants of *ADORA2A* that significantly affect caffeine-induced sleep disturbance.[115]

Rétey and associates[114] combined self-reports and polysomnography after double-blind caffeine administration to document individual differences in the effects of caffeine on sleep. By contrast, the replication study was restricted to self-classification of caffeine sensitivity. The successful replication with this less accurate and less reliable (i.e., subjective) phenotype indicates that questionnaires are useful in large-scale epidemiologic studies. Subsequent follow-up with objective measurements in animals and humans can provide novel insights into the molecular bases of healthy and disturbed sleep. Thus sleep pharmacogenetics of caffeine may have important implications for the pathophysiology and the

rational treatment of insomnia, as well as for recommendations for the critical use of caffeine, which is consumed on a daily basis by up to 90% of adults in Western societies.

Dopaminergic Neurotransmission

Apart from being an adenosine receptor antagonist, the stimulant actions of caffeine also depend on the dopaminergic system. Data in *Dat* knockout animals and human homozygous carriers of the 10R allele of *DAT1* (*SLC6A3*) suggest that reduced DAT expression is associated with elevated sensitivity to the stimulant.[59,125] Furthermore, Holst and colleagues[59] found that caffeine reduced distinct neurophysiologic markers of sleep homeostasis, such as number, amplitude, and slope of individual slow waves, in a *DAT1* genotype-dependent manner. This finding suggested that the interference of caffeine with neurophysiologic markers of sleep homeostasis not only relies on adenosinergic mechanisms but also involves dopaminergic processes.

Like caffeine, the potency of the wake-promoting compound modafinil shows pronounced interindividual variation. The neurochemical mechanisms and cerebral regions through which modafinil produces wakefulness are incompletely understood. However, modafinil reduces DAT-mediated reuptake of dopamine in animals[140] and humans.[141] Consistent with a dopaminergic mode of action of modafinil, the compound was ineffective in promoting wakefulness in *Dat* knockout mice[125] and attenuated elevated sleepiness after sleep deprivation reflected in EEG theta (5.5 to 7 Hz) power in sleep-deprived volunteers in a *DAT1* genotype-dependent manner.[59]

Functional variants in the gene encoding COMT also alter dopaminergic neurotransmission in the brain. They may, thus, also contribute to individual differences in the wake-promoting effects of modafinil. Support for this hypothesis was obtained in both sleepy patients (see Clinical Pearl) and healthy volunteers subjected to sleep deprivation.[142] In healthy young men, placebo-controlled, double-blind, randomized administration of modafinil (2 × 100 mg) during prolonged wakefulness similarly reduced subjective sleepiness and EEG 5- to 8-Hz activity in Val/Val and Met/Met allele carriers of *COMT*.[63] By contrast, modafinil differently affected the NREM sleep EEG in recovery sleep. Furthermore, it maintained sustained vigilant attention and executive functioning at baseline level throughout prolonged waking in Val/Val allele carriers, whereas the compound was virtually ineffective in the Met/Met genotype.[65] These data highlight a role for dopamine in impaired waking functions after sleep loss. The functional significance of the modafinil-induced, genotype-dependent effects on the NREM sleep EEG during recovery from sleep loss remains to be determined.

CONCLUDING REMARKS

Sleep is a complex behavior, and any functional genetic variation associated with changes in one of the many neurotransmitter and neuromodulator systems can be expected to affect sleep and the sleep EEG. Polymorphic variations in a number of genes have now been shown to affect several characteristics of sleep, and some of these genes may indeed be involved in sleep regulatory processes. However, many associations need to be replicated, and failure of replication is common. Nevertheless, after robust associations have been established, elucidating the signaling pathways that are affected will aid our understanding of individual differences in sleep-wake behavior.

CLINICAL PEARL

Distinct alleles and genotypes in the genes of monoamine oxidase type A (MAO-A)[143] (but see Dauvilliers and colleagues[144]) and COMT[144] are thought to be associated with the clinical manifestation of narcolepsy. The Val158Met polymorphism of COMT exerts a sexual dimorphism and a strong effect of genotype on disease severity.[144] More specifically, women narcoleptics with high COMT activity fall asleep twice as fast during the Multiple Sleep Latency Test than those with low COMT activity. An opposite relationship, although less pronounced, is observed in men. Also, the response to treatment with modafinil to control excessive daytime sleepiness differs between COMT genotypes. Patients (female and male) with the Val/Val genotype need an almost 100 mg higher daily dose than patients with the Met/Met genotype.[145] Intriguingly, in male healthy volunteers, the effect of the Val158Met polymorphism of COMT on the efficacy of modafinil to improve excessive sleepiness after sleep deprivation is opposite that in narcolepsy patients.[65]

SUMMARY

Sleep is a very rich phenotype, and many aspects of sleep differ considerably in the population of healthy individuals (even when only a very narrow age range is considered). Interindividual variation in sleep timing (diurnal preference), sleep duration, sleep structure, and the EEG in NREM sleep, REM sleep, and wakefulness have all been shown to have a genetic basis. The response to challenges of sleep regulatory processes such as sleep deprivation and circadian misalignment has also been shown to vary between individuals. Some of the polymorphic variations in genes contributing to variation in sleep characteristics have now been identified. They include variations in genes associated with the circadian system (e.g., *CLOCK, PER1, PER2, PER3, BHLHE41*), the adenosine system (*ADA, ADORA2A*), and the catecholaminergic system (e.g., *COMT, SLC6A3, SLC6A4*), as well as other signaling pathways (e.g., *ABCC9*, *BDNF*, *PRNP*). For some of these genes, so far only associations with one aspect of sleep have been reported (e.g., *PER2* and sleep timing). Variations in other genes have been shown to affect multiple aspects of sleep and wakefulness, as well as the response to sleep loss or pharmacologic interventions. For example, *PER3* and *ADA* affect the EEG and performance during prolonged waking, whereas *ADORA2A*, *COMT,* and *SLC6A3* modulate EEG and response to the stimulants caffeine and modafinil. All currently known polymorphic variations explain only a small part of the variation in healthy human sleep phenotypes, and many more genetic contributions remain to be discovered.

ACKNOWLEDGMENTS

The authors' research has been supported by the Swiss National Science Foundation, Zürich Center for Interdisciplinary Sleep Research, Clinical Research Priority Program "Sleep and Health" of the University of Zürich, Zürich Center for Integrative Human Physiology, Neuroscience Center Zürich, and Novartis Foundation for Medical-Biological Research (to HPL); and by the Biotechnology and Biological Sciences

Research Council, Wellcome Trust, Air Force Office of Scientific Research, Higher Education Funding Council for England, and a Wolfson-Royal Society Award (to DJD).

Selected Readings

Andretic R, Franken P, Tafti M. Genetics of sleep. *Annu Rev Genet* 2008;**42**:361–88.

Byrne EM, et al. A genome-wide association study of caffeine-related sleep disturbance: confirmation of a role for a common variant in the adenosine receptor. *Sleep* 2012;**35**:967–75.

Byrne EM, et al. A genome-wide association study of sleep habits and insomnia. *Am J Med Genet B Neuropsychiatr Genet* 2013;**162B**:439–51.

Dauvilliers Y, Tafti M, Landolt H-P. Catechol-O-methyltransferase, dopamine, and sleep-wake regulation. *Sleep Med Rev* 2015;**22**:47–53.

De Gennaro L, et al. The EEG fingerprint of sleep is genetically determined: a twin study. *Ann Neurol* 2008;**64**:455–60.

Dijk DJ, Archer SN. PERIOD3, circadian phenotypes, and sleep homeostasis. *Sleep Med Rev* 2010;**14**:151–60.

Hida A, et al. Screening of clock gene polymorphisms demonstrates association of a PER3 polymorphism with morningness-eveningness preference and circadian rhythm sleep disorder. *Sci Rep* 2012;**4**:6309.

Holst SC, Valomon A, Landolt HP. Sleep pharmacogenetics: personalized sleep-wake therapy. *Annu Rev Pharmacol Toxicol* 2016; in press.

Lazar AS, et al. Sleep, diurnal preference, health, and psychological well-being: a prospective single-allelic-variation study. *Chronobiol Int* 2012;**29**:131–46.

Mang GM, Franken P. Genetic dissection of sleep homeostasis. *Curr Top Behav Neurosci* 2015;**25**:25–63.

McCarthy MI, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;**9**:356–69.

A complete reference list can be found online at ExpertConsult.com.

Chapter
31

Genetics and Genomic Basis of Sleep Disorders in Humans

Allan I. Pack; Brendan T. Keenan; Enda M. Byrne; Philip R. Gehrman

Chapter Highlights

- This chapter provides an overview of the approach to genetic studies in humans.
- Known genetic risk factors for sleep disorders are described.
- Genetic determinants of normal variants of sleep duration, chronotype, and response to sleep deprivation are identified.
- Genetic studies in narcolepsy show that HLA variants confer increased risk for and protection from narcolepsy.
- Genetic studies in narcolepsy show not just HLA variants but also variants in T-cell alpha receptor, supporting the autoimmune basis of the disorder.
- Genetic studies of restless legs syndrome identify novel pathways whose role needs to be identified.
- Variations in clock-associated genes affect not only timing of sleep but also sleep duration and response to sleep deprivation.

APPROACH TO IDENTIFYING GENETIC VARIANTS IN HUMANS

Overview

The overwhelming majority of biologic traits and disorders in humans have a genetic component as part of their etiology. Sleep and disorders of sleep are no exception. The total proportion of variation in risk to a disease in the population that can be attributed to genetic variation is known as the heritability. There are now a large number of studies showing that sleep disorders are heritable, that is, genetics play a substantial role in their etiology. These are also reviewed in chapters that cover specific disorders (e.g., movement disorders, Section 13; sleep breathing disorders, Section 14; and narcolepsy, Chapter 89).

The role of genetic and genomic factors in human disease has been studied for decades, progressing from classical heritability and linkage studies to more focused candidate gene analyses, then to genome-wide analyses made possible by the sequencing of the human genome and more recently including whole exome and genome sequencing analyses as well as evaluation of epigenetic modifications. Since the sequencing of the human genome, biomedical research has made great progress in understanding the genetic architecture and molecular pathways underlying human disease.[1,2] Whereas a greater understanding of genetic factors underlying complex disease has been achieved, the large amount of so-called missing heritability, that is, the unexplained genetically inherited disease risk, suggests that there is still opportunity and need for important discoveries.[1,2] This is particularly true for sleep disorders, despite the established genetic heritability; to date, only a small number of validated genetic risk variants have been discovered for sleep-related traits. There are a number of reasons for this lack of discovery, including inadequate sample sizes, variable phenotypes that add noise, and numerous pathways to disease.

Heritability analyses are the first step in understanding the genetic underpinnings of disease. They establish whether there is a relationship between genetic risk factors and a disease phenotype by estimating the amount of disease variability that is explained by genetic variants. In the past, after it was established that a disorder is heritable, *linkage studies* were a likely next step to try to further our understanding of the existence of genetic etiology by trying to pinpoint specific chromosomal regions that harbor genetic variants influencing disease risk. *Candidate gene studies* can then be used to examine these identified regions in finer detail or, more recently, to replicate genes identified through genome-wide analyses. The Human Genome Project,[1,3-6] the International HapMap Project,[7,8] and The 1000 Genomes Project[9] have firmly established and characterized interindividual variability throughout the human genome.

The primary focus of current studies examining the association between genetic variants and disease has been on *single-nucleotide polymorphisms* (SNPs). This is because SNPs, which are a difference in the DNA sequence at one nucleotide among individuals, are the most frequent form of genetic variation. Owing to the block-like structure of the genome, where regions of the genome that are close together tend to be transmitted together, genotyping of one SNP can provide information on genetic variation at many nearby SNPs. Initial publications suggested that approximately 500,000 common polymorphisms provided power to capture 90% of the variability in the genome.[2,10] Analysis approaches distinguish between common polymorphisms (occurring with >5% frequency in the population), which are likely to confer small effect sizes for complex disease, and rare polymorphisms (<5% frequency), which may lead to larger effects but are more difficult to discover because of their less frequent occurrence (Figure 31-1). Whereas for the latter individuals variants may be rare, there may exist multiple different rare variants

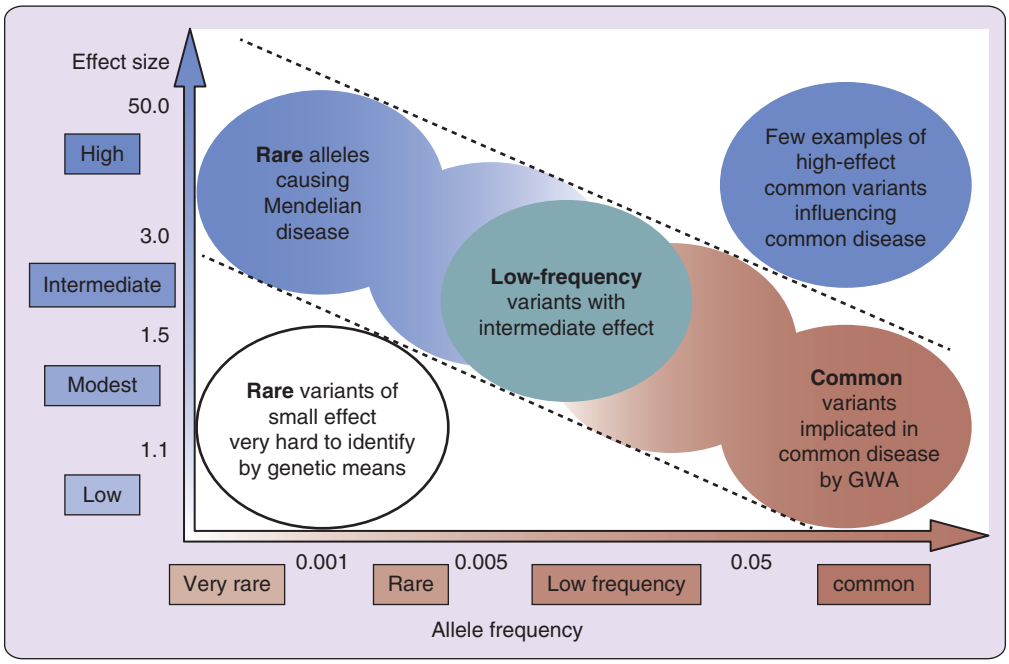


Figure 31-1 Relationship between allele frequency of variant (x-axis) and effect size of variant. Common variants at bottom right of figure have frequency in population of more than 5% but have smaller effect in controls. Variants to the left are rare but have large effects. GWA, Genome-wide association. (From Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747-53.)

in the same gene, and so one can evaluate the evidence for a role for rare genetic variants as a set rather than just individually.

In addition to SNPs, *copy number variation*,[11,12] including insertions and deletions, provides additional mechanisms for genetic underpinnings of disease.[11,13] Finally, there is an increasing focus on the role of epigenetics,[14,15] including DNA methylation, in elucidating the biological pathways to disease. The level of methylation in certain regions of genes or functional elements outside of genes affects the level of gene expression. Unlike the DNA sequence itself, methylation patterns are not fixed and can be influenced by a variety of environmental factors. It can occur in multiple sites in the genome. Because it can be permanent, it can be passed through the germline to offspring. Thus, environmental influences that individuals experience before they have children can influence which genes are expressed in particular cell types in their offspring. The following gives a brief overview of the potential approaches to genetic studies of sleep disorders. Later in the chapter, we explain what is known about the genetics of sleep and some common sleep disorders.

Heritability Estimation

Heritability can be broadly defined as the proportion of phenotypic variability that is attributable to genetic factors; higher estimates suggest that genetic variability has a large influence on the variability of a given trait in the population. Heritability analysis has been used for decades to estimate whether a given phenotype is influenced by genetic factors and how strong that influence is relative to nongenetic risk factors. Numerous techniques exist for estimating heritability; these range from use of phenotype information from twins[16] or family pedigree data[17,18] to more recently developed statistical techniques for estimating heritability based on genome-wide genotyped data on unrelated individuals.[19]

Estimating the heritability of a given trait with use of twin or family data does not require specific measurement of genetic variants. Rather, these methods take advantage of the known shared genetic variance among related individuals. The general principle behind heritability analysis is that people who are more genetically related to each other should be more similar to each other for the phenotype of interest. For binary traits like sleep disorders, one can measure the recurrence risk to relatives. That is, given that a family member has been diagnosed with a disorder, what is the risk in the family members of having the same disorder? This recurrence risk in families can be compared with the risk of disease in the general population to give an estimate of heritability. For heritable traits, the relative recurrence risk ratio should decrease as the family relationships examined become less similar genetically; for example, the recurrence risk in siblings of affected individuals should be greater than for first cousins of the affected.

Results from family studies are complicated by the fact that family members often share a similar environment. It can be difficult to parse out whether the higher risk in certain families compared with the general population is due to shared genetic risk factors, shared environmental risk factors, or a combination of the two. Twin studies help to separate out shared genetic variance from other sources of variance because twin pairs are assumed to share many common environmental factors—they are born at the same time, shared the same intrauterine environment, and often attend the same school. With this source of variance controlled for, the similarities and differences between twins can be separated into genetic and environmental sources. Estimates of heritability are derived by

comparisons between monozygotic and dizygotic twin pairs. Increased similarity in phenotype between monozygotic pairs (who are genetically identical) compared with dizygotic twin pairs (who share half of their genetic variants with each other) provides evidence for heritability. As presented in a recent publication establishing the heritability of performance deficit accumulation during sleep deprivation,[20] there are several complementary methods that may be used for evaluating heritability in twin samples. We briefly describe these methods.

As discussed with respect to performance deficits during sleep deprivation,[20] three methods for estimating heritability in twins are (1) classical heritability estimation, (2) analysis of variance (ANOVA) approach, and (3) likelihood-based estimation of variance components. Each of these methods can be useful in comparing the heritability to existing literature as well as in evaluating different assumptions. Classical heritability is derived using the differences in the intraclass correlation coefficient (ICC) statistics between monozygotic (ICC_{MZ}) and dizygotic (ICC_{DZ}) twin pairs.[21] Using these values, heritability (denoted h^2) is estimated as $h^2 = 2 \bullet (ICC_{MZ} - ICC_{DZ})$. In addition to estimating heritability, the classical approach can also provide an estimate of the shared common environmental variance, which is estimated as $2 \bullet ICC_{DZ} - ICC_{MZ}$. Next, the ANOVA approach uses combinations of the monozygotic and dizygotic within-twin and among-twin pair mean squares estimates in combination with specific assumptions about the variability (e.g., that total variability is equal in monozygotic and dizygotic twins).[22,23] Finally, the maximum likelihood variance components approach uses model-specific covariance matrices[23-25] and, importantly, allows the examination of specific patterns of genetic transmission and calculations of standard errors and *P* values associated with heritability estimates. These models of genetic transmission include components related to additive genetic effects (A), dominant genetic effects (D), common environmental effects (C), and unique individual effects (E).[25] For example, the ACE model assumes additive genetic effects, shared environments, and unique individual components of variability. By comparison of different models, specific questions about the mode of genetic inheritance can be assessed. Overall, we see that each method of heritability estimation has unique advantages. Whereas the classical method provides a more simplistic approach to the computation, the advantage of the ANOVA model is the ability to assess specific assumptions about the validity of the twin model. Although it is potentially more complex, the maximum likelihood variance components approach can provide information on specific genetic transmission models.

As discussed later in the chapter, more recently established techniques allow estimation of heritability in unrelated individuals by simultaneously examining the association between a given trait and all genotyped genetic polymorphisms.[19,26-28] These techniques have recently been used and extended to more accurately capture the amount of variability we can expect to explain through genome-wide association analyses.

Establishing that a given trait is heritable strongly implies that underlying genetic factors play a role in determining the phenotype. Many sleep-related disorders and intermediate phenotypes have been shown to be heritable in the last few decades, including sleep duration,[29-31] chronotype,[32-35] response to sleep loss,[20] restless legs syndrome (RLS),[36-38] insomnia,[29,39-41] parasomnia,[42] obstructive sleep apnea (OSA),[43-49] and key intermediate traits for OSA (such as craniofacial structures,[50] upper airway soft tissue volumes,[51] and ventilatory responses to hypoxia and hypercapnia[52]). Among behavioral traits, one of the most heritable is the spectral characteristics of the electroencephalogram (EEG) during sleep.[53]

Heritability is only an estimate for the specific population included in a study. There is not one true heritability for a given disorder or trait. Instead, the heritability can vary over time as environments change, and it can be different in specific ethnic groups or in particular age groups (see Visscher et al[19] for a review of heritability concepts). For instance, the heritability of sleep duration in adolescents is likely to be different from that in older adults. The relative importance of genes and environment in population variation may vary over the life span. Hence, estimates of heritability can vary substantially across studies.

Despite observing heritability estimates of more than 50% for some of these traits, the genetic variants discovered to date typically explain on the order of less than 5% of the known overall variability in any given phenotype. Finding the causes of this "missing heritability" is an ongoing area of research, and methods for determining heritability of a given phenotype continue to develop (for reviews, see[1,2,54-56]). Explanations for missing heritability are numerous, including a large number of common variants with small effects, multiple rare variants with large effects, insufficient tagging of causal variants in current genotyping platforms, gene-gene and gene-environment interactive effects, and other types of genetic variations (such as copy number variants and epigenetic modification). Ultimately, explaining the missing heritability is likely to require very large sample sizes and both rigorous and novel analytic techniques.

Linkage Analysis

Once a trait has been shown to be heritable, the next obvious question is, Which genes or regions are associated with the phenotype? One strategy for answering this question is to perform a genetic linkage analysis, which relies on the inheritance structure associated with family pedigrees. That is, genetic linkage analysis uses the observation that genes close to one another are more likely to be passed down to offspring than are genes that are farther away.[57-63] At their most basic level, linkage studies involve comparing chromosomal regions between affected and unaffected individuals to identify those DNA segments that are more commonly shared between affected relatives and not between affected and unaffected.[58,59] Genetic regions that are more likely to be shared among affected individuals in a family than between affected and unaffected individuals are expected to contain disease-related genetic factors. The approach does not involve initial assessment of genes but rather use of multiple polymorphic microsatellite markers that are equally spaced across the genome (typically 400 to 1000 markers). These markers in intergenic regions have high sequence variation between different individuals. The approach involves calculating the maximum likelihood log-odds (LOD scores) as a function of marker location. Because there are multiple such computations, false associations may occur by chance, and criteria to control for this have been developed (see later). Once a marker location with a high LOD score is identified, the genomic region can be narrowed and the linkage validated by determining LOD scores for an additional set of markers that are specific to this region—fine

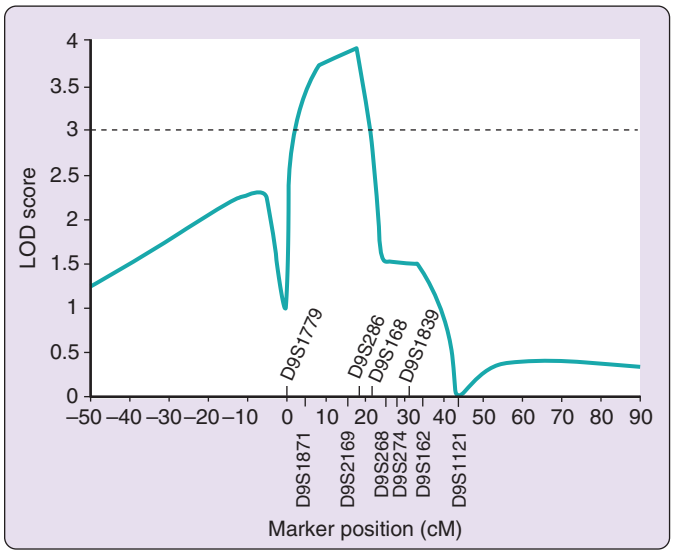


Figure 31-2 Example of identifying a region by linkage analysis. Marker location is on x-axis and LOD score on y-axis. When a significant linkage is found with equally spaced polymorphic markers across the genome (*dashed line* represents threshold for genome-wide significance linkage), LOD scores for additional markers in region (see annotations on x-axis) are calculated. As shown, this results in narrowing of the genomic region and leads to increased LOD scores. In this example, there is little doubt that there is a variant conferring risk for restless legs syndrome in this region. (From Chen S, Ondo WG, Rao S, et al. Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *Am J Hum Genet* 2004;74:876-85.)

mapping (see example, Figure 31-2). Linkage analyses should not be reported without this step. With the advent of extensive SNP data across the genome, an alternative strategy is to do case-control association analyses with all known SNPs in the linkage region. This limits the number of comparisons compared with that in genome-wide association analysis. This strategy has been successfully employed in studying the genetic risk factors for RLS[64] (see further later). Linkage analysis can also be extended to examine quantitative traits, rather than binary. Moreover, analyses can examine the linkage between single genetic loci or multiple loci.[60,63]

Multiple methods may be used for performing genetic linkage analyses, depending on the hypothesized model of inheritance, including parametric (or model based) and nonparametric (or model free).[57-63] Briefly, parametric linkage analysis involves testing whether the inheritance pattern follows a prespecified model, including disease and allele frequencies, and calculating the LOD score (equal to the log of the likelihood ratio) for each polymorphic genetic marker through statistical modeling.[57,58,63] Nonparametric linkage analysis, on the other hand, relies on a sib-pair design to test whether affected individuals are more likely to share genetic regions than would be expected under the assumptions of independence between trait and inheritance.[61-63] Nonparametric linkage is an important tool for complex traits, as parametric linkage has been shown to be quite sensitive to misspecification of the mode of inheritance.[58,63] Research has extended these methods to use linkage maps and a unified multipoint approach that includes all relevant pedigree information, rather than restricting to specific models or subsets of pedigree information.[60,63,65-67]

Importantly, previous research from Lander and Kruglyak[58] has provided LOD score thresholds for statistical significance in complex traits, depending on the mapping method and model of inheritance. These thresholds include a LOD = 3.3 for genome-wide significant linkage (corresponding to a $P = 4.9 \times 10^{-5}$) and a LOD = 1.9 for suggestive linkage ($P = 1.7 \times 10^{-3}$) for analysis in humans and slightly higher thresholds of LOD = 3.6 and LOD = 2.2 in sib-pair studies. Although suggestive signals are of interest, they should be interpreted with caution as they could be false positives, and many will likely not be replicated in independent samples. Subsequent to the identification of a genomic region with significant linkage, fine-mapping analysis, examining specific genetic markers within the identified region, is an important next step to narrow genomic regions and to identify specific variants associated with the phenotype of interest. Modern sequencing and genotyping technologies have facilitated these more detailed examinations.

Linkage studies have been around for several decades, and one of the most successful examples in sleep-related diseases is the identification of five linkage regions, that is, *RLS-1* through *RLS-5*, for RLS (Table 31-1).[68-72] This was based on studies of a number of large family pedigrees with RLS in different parts of the world. Subsequent fine-mapping analyses identified specific variants of interest in these genes,[64] leading to new insights into RLS genetics. In contrast to these positive findings in RLS, early linkage analyses for OSA in Europeans and African Americans have led to results that have not been replicated.[73-75] These studies reported linkage with most LOD scores not even in the suggestive range and did not do any fine mapping.[73-75] Thus, these early studies were underpowered for robust effects. It seems unlikely that linkage approaches will lead to identification of genomic regions linked to OSA.

Linkage analysis proved to be an important tool for mapping the variants underlying many single-gene disorders, such as Huntington disease and cystic fibrosis. Furthermore, linkage analysis aided in the identification of genes underlying polygenic disorders such as RLS, in which one variant or a very small number of variants explain a large proportion of the heritability in the population. However, linkage analysis has not proved a successful strategy for complex disorders for which hundreds of variants likely contribute to risk at the population level. This is primarily because linkage analysis requires family-based data. It is cumbersome to collect enough families to have enough power to detect variants that do not have a large effect on disease risk. Allied to this, linkage analysis primarily used microsatellite markers found throughout the genome, often at large distances from each other, and therefore a signal from a linkage analysis could implicate very large regions of the genome containing many genes. For these reasons among others, linkage analysis has now largely been replaced by association-based analysis using SNPs.

Candidate Gene Studies

A more directed approach to discovering important genes and genetic variants is by association studies using a candidate gene approach. Rather than having to focus on related individuals and family pedigree data, candidate gene studies can use unrelated individuals (e.g., cases and controls) as well as families and targeted genotyping to examine the association between variants within or near a priori hypothesized genes of interest and a given phenotype.[59,76] Hypothesized candidate

Table 31-1 Linkage Regions Identified in Studies of Restless Legs Syndrome Families

Locus (OMIM)	Reference	Chromosomal Location	Inheritance Mode	Parametric LOD Score
RLS-1	Desautels et al,[1] 2001	12q22-23.3	AR pseudodominant	3.42 (2P) 3.59 (MP)
RLS-2	Bonati et al,[2] 2003	14q13-22	AD	3.23 (2P)
RLS-3	Chen et al,[3] 2004	9p24-22	AD	3.77 (2P) 3.91 (MP)
RLS-4	Pichler et al,[4] 2006	2q33	AD	4.1 (2P)
RLS-5	Levchenko et al,[5] 2006	20p13	AD	3.34 (2P) 3.86 (MP)
—	Levchenko et al,[6] 2009	16p12.1	AD	3.5 (MP)
—	Winkelmann et al,[7] 2006*	4q25-26	AD	2.92 (MP)
—	Winkelmann et al,[7] 2006*	17p11-13	AD	2.83 (MP)
—	Kemlink et al,[8] 2008*	19p13	AD	2.61 (MP)

The linkage regions for restless legs syndrome are given with their chromosomal position by chromosome band, the proposed inheritance mode, and the logarithm of the odds (LOD) scores from parametric linkage analysis. Both 2-point and multipoint scores are reported. Numbering of loci has been indicated as listed in Online Mendelian Inheritance in Man (OMIM).
*Suggestive evidence only.
[1]Desautels A, Turecki G, Montplaisir J, et al. Identification of a major susceptibility locus for restless legs syndrome on chromosome 12q. *Am J Hum Genet* 2001;69:1266-70.
[2]Bonati MT, Ferini-Strambi L, Aridon P, et al. Autosomal dominant restless legs syndrome maps on chromosome 14q. *Brain* 2003;126:1485-92.
[3]Chen S, Ondo WG, Rao S, et al. Genome-side linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *Am J Hum Genet* 2004;74:876–85.
[4]Pichler I, Marroni F, Volpato CB, et al. Linkage analysis identifies a novel locus for restless legs syndrome on chromosome 2q in a South Tyrolean population isolate. *Am J Hum Genet* 2006;79:716-23.
[5]Levchenko A, Provost S, Montplaisir JY, et al. A novel autosomal dominant restless legs syndrome locus maps to chromosome 20p13. *Neurology* 2006;67:900-1.
[6]Levchenko A, Montplaisir JY, Asselin G, et al. Autosomal-dominant locus for restless legs syndrome in French-Canadians on chromosome 16p12.1. *Mov Disord* 2009;24:40-50.
[7]Winkelmann J, Lichtner P, Kemlink D, et al. New loci for restless legs syndrome map to chromosome 4q and 17p. *Mov Disord* 2006;21:S412.
[8]Kemlink D, Plazzi G, Vetrugno R, et al. Suggestive evidence for linkage for restless legs syndrome on chromosome 19p13. *Neurogenetics* 2008;9:75-82.
AR, Autosomal recessive; AD, autosomal dominant; 2P, two-point LOD score; MP, multipoint LOD score.
From Schormair B, Winkelmann J. Genetics of restless legs syndrome: mendelian, complex, and everything in between. *Sleep Med Clin* 2011;6:203-15.

genes may arise from several areas, including linkage analyses and genome-wide association results or from existing knowledge about the biological mechanisms involved in pathogenesis, and should be chosen by considering information relative to biologic pathways associated with the disease phenotype of interest. Using strong functional hypotheses is an important aspect for candidate gene studies as false-positive associations may arise because of linkage disequilibrium between candidate regions and true causal variants as well as because of population stratification (see further later).[59,76,77] This potential for confounding, coupled with the large number of gene associations that are not replicated in independent data sets, has led to considerable debate about the utility and interpretation of results based on these approaches.[59,76,78-81]

At their most basic level, association studies using candidate genes are no different from typical epidemiologic studies examining the relationship between nongenetic risk factors and a phenotype of interest. Candidate gene studies involve making a priori assumptions about which genetic regions are most likely to be associated with a phenotype and thus may be limited by the amount of information currently known about the functional effects of a given gene. Once candidate genes have been established, standard statistical modeling can be used to assess the correlation between the candidate alleles in these genes and the phenotype of interest. Compared with linkage analysis, candidate gene studies have been shown to have stronger statistical power for complex disease traits as they can identify genes or gene variants with smaller effects.[76,82] Like many types of association analyses, initial candidate gene studies are prone to overestimates of the true association (also referred to as winner's curse).[79-81] This initial overestimation, as well as the fact that many early studies use relatively small sample sizes, is a major reason that many novel genetic associations fail to replicate in later studies. A previous study by Ioannidis et al,[79] examining 36 genetic associations through a meta-analysis of 370 studies, showed that both a small sample size in initial publications and a large number of studies for a particular trait were independent predictors of discrepant results. With regard to sample size, they found statistically significant discrepancies in 5 of 7 cases in which the first publication had fewer than 150 subjects compared with only 3 of 29 when the initial sample size was more than 150.[79] This further emphasizes the importance of considering the size of the study population in interpreting genetic associations (for review, see[77]).

Detailed reevaluation of evidence for association for candidate genes in other fields with a larger genetics literature than for sleep disorders has indicated that very few if any associations with SNPs in candidate genes replicate in larger samples.[83,84] These findings mean that caution should be exercised in interpreting results from candidate gene studies for sleep disorders.

Another important consideration in interpreting the results of genetic association studies is population stratification,

which occurs when differences in disease prevalence and genotype frequency between populations of different ethnicities result in apparent associations between genotype and phenotype.[85-91] Because of evolution and genetic drift, SNP allele frequencies differ between populations. Moreover, different diseases are more or less prevalent in different parts of the world. The combination of a higher disease prevalence and higher or lower genotype frequency will result in inflated *P* values and significant associations in analyses that combine data from multiple ethnic populations, even if the two factors are unrelated. Approaches to correct for population stratification have been discussed in detail[85,88-91] and include correction by principal components estimated through ancestry informative genetic markers. Whenever possible, genetic analyses of individual variants should be performed separately within populations of differing ethnicity, in the case of both candidate gene studies and genome-wide associations. However, specific programs and genotyping platforms for capturing differences in associations between admixed populations have been developed.[92]

Candidate gene studies have been conducted for many sleep-related disorders, including chronotype[93-97] and OSA.[98-100] Candidate gene studies for chronotype have focused on the associations with clock genes (e.g., *CLOCK*, *PER1*, *PER2*, and *PER3*). Although not all studies have found significant associations, most likely because of small samples, the general pattern suggests significant roles of these functional genes on chronotype (see later). As extensively reviewed by Varvarigou et al,[100] candidate studies in OSA have focused on a wider variety of genes, including those related to obesity and craniofacial phenotypes, as well as specific genes with previously suggested associations (e.g., *APOE* ε2 and ε4); whereas some significant associations have been seen, only an SNP in the regulatory region of *TNFα* was found to be significant in the meta-analysis plan. This may not be related to risk for OSA but rather to the subjects with this SNP being more likely to be excessively sleepy and hence seeking evaluation.[101] As is the case with many candidate gene studies, those currently performed in OSA have had relatively small sample sizes and therefore have been greatly underpowered to find the type of effect sizes (e.g., odds ratio = 1.2) expected for common variants associated with complex diseases[100] (Figure 31-3). Much larger studies are needed.

By design, candidate gene studies focus primarily on variants in the coding regions of genes. However, it is becoming increasingly apparent that common genetic variation found

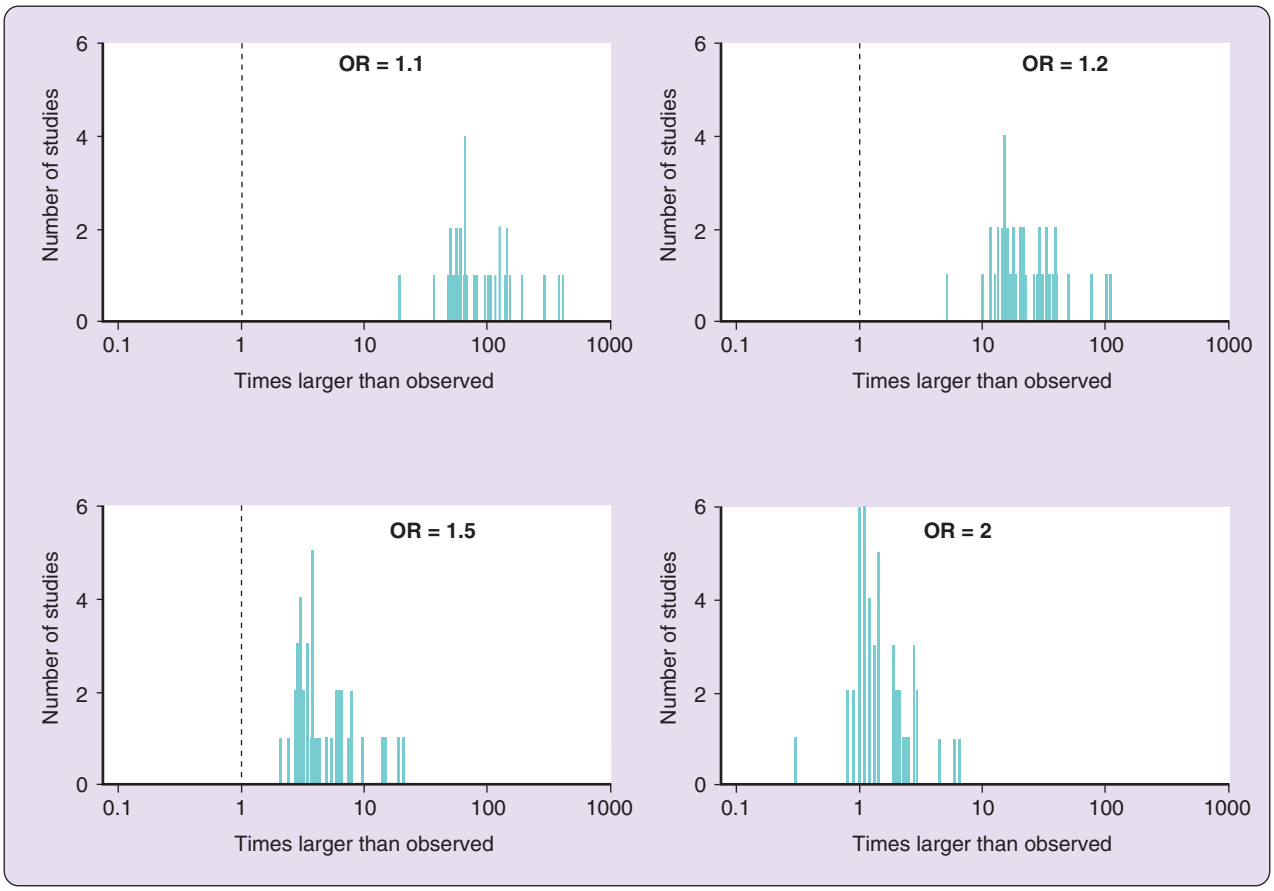


Figure 31-3 Calculation of how much larger candidate gene studies for obstructive sleep apnea would be needed to identify odds ratios (OR) of 1.1 (*top left panel*), 1.2 (*top right panel*), 1.5 (*bottom left panel*), and 2.0 (*bottom right panel*). Common variants typically result in low OR, such as 1.2. As can be seen for this OR, studies would need to be between 10 and 100 times larger than was used in the original studies. The majority of studies are underpowered to even detect an OR of 2.0, an effect typically not found with common variants. (From Varvarigou V, Dahabreh IJ, Malhotra A, et al. A review of genetic association studies of obstructive sleep apnea: field synopsis and meta-analysis. *Sleep* 2011;34:1461-8.)

outside of protein-coding genes contributes to the etiology of disease. Our understanding of noncoding functional areas of the genome has improved dramatically in the past few years, and it is no longer sufficient to focus solely on variants in the exome.

Genome-Wide Association Studies

The completion of the Human Genome Project and the subsequent mapping and sequencing projects that characterized genetic variation among individuals dramatically increased our ability to perform genetic association studies.[1,3-9] Rather than relying on the identification of broad regions through linkage or restricting our focus to hypothesized candidate genes, genome-wide genotyping allowed an unbiased examination of the relationship between a given trait and individual genetic variants throughout the genome. This type of "hypothesis-free" analysis is referred to as a genome-wide association study (GWAS). The first GWAS results were published with respect to age-related macular degeneration,[102,103] with the initial large-scale GWAS led by the Wellcome Trust Case Control Consortium, examining seven common diseases in 14,000 cases and 3000 controls.[104] Since these initial GWAS results, thousands of genetic loci have been associated with complex traits, as detailed by the National Human Genome Research Institute GWAS Catalog (available at http://www.genome.gov/gwastudies/).[105,106] These analyses are typically restricted to common genetic variants (i.e., with minor alleles occurring in more than 5% of the population), many of which have small but significant effects on complex diseases such as sleep-related disorders. GWAS analyses are proving to require very large samples of well-phenotyped individuals to obtain robust and reproducible results. Important considerations, discussed briefly, include interpretation of results, appropriate genome-wide significance thresholds, and the need for well-designed studies with replication samples and analyses.

GWAS rely on the haplotype or correlation structure among SNPs across the human genome, based on initial publications suggesting that approximately 500,000 common SNPs captured a substantial proportion of the variability in the genome.[2,10] Given this structure, rather than examine every SNP in the genome, designed genotyping platforms capture a majority of the information by genotyping several hundred thousand SNPs across the genome. Moreover, current statistical genetics approaches combine this individual-level genotype data with correlation structures inferred from established reference panels of fully sequenced individuals, such as the International HapMap Project[7,8] and The 1000 Genomes Project,[9] to impute nongenotyped SNPs for inclusion in GWAS analyses (for review, see[107]).

After genetic data have been collected, performing a GWAS involves conducting individual regression models assessing the relationship between each common variant and the phenotype of interest (e.g., a linear regression model for quantitative phenotypes or a logistic regression model for disease phenotypes). In these models, SNPs may be coded additively with respect to the number of copies of the allele of interest or assuming a dominant or recessive genotype effect. Because of the large number of tests resulting from all SNPs being examined individually, a multiple comparisons correction is necessary to determine statistical significance and to protect against false-positive associations. Whereas there are different methods for performing this correction, a P value less than 5×10^{-8} is typically required to claim genome-wide significance, reflecting a Bonferroni correction for 1 million independent tests of the hypothesis that SNPs are associated with a phenotype.[2,77,108,109] As the number of genetic variants included in analyses continues to increase, stricter criteria will be necessary to ensure robustness of associations.

As with candidate gene association studies (discussed earlier), although reaching statistical significance is important, determining a robust genetic association requires a critical review of initial studies and replication analyses. The National Cancer Institute–National Human Genome Research Institute (NCI-NHGRI) Working Group on Replication in Association Studies has published an excellent review on the interpretation, validity, replication, and publication of genotype-phenotype associations.[77] Criteria for establishing the validity of initial association reports include sufficient sample size and power to detect reasonable effects, appropriate correction for multiple comparisons across all tests, consistent results for any similar phenotypes and population subsets, standard and well-described quality control methods applied to genotypes, assessment of potential confounding by population stratification (as in candidate gene studies), similar associations between the most highly associated SNP and those in linkage disequilibrium, and reported results from replication studies (even if these are null).[77] Replication analyses are critical for genetic studies, given the likelihood of false positives; this is true for all approaches, not just GWAS. Replication analyses require reproducing observed SNP associations from initial discovery analyses within an independent population. As discussed in the NCI-NHGRI review,[77] many of the same criteria for establishing the soundness of initial results apply to replication studies, including adequate sample size, similarity in phenotypes, population and effect sizes, strong rationale for SNP selection, and, when possible, joint analysis combining discovery and replication analyses that results in lower P values. Meta-analyses and joint analyses combining observed results across multiple studies are typically a good way to ensure validity of genetic associations.[79] Ultimately, it is important to consider all of these issues in examining the robustness of genetic associations in complex diseases, including sleep-related disorders.

Despite the established heritability, GWAS for sleep-related traits have led to mixed results. These are discussed in detail in the sections that follow. Whereas there has been notable success in GWAS studies in RLS and narcolepsy, no GWAS has been reported for OSA. Some have argued that this is the next logical step in studying genetics of OSA.[110] Compared with other fields, understanding of the genetic architecture of sleep-related phenotypes through GWAS is in a relatively nascent stage. Because other fields have moved beyond GWAS, it is increasingly difficult to obtain funding for GWAS studies, making it even more challenging for studies of sleep and its disorders.

Rare Variant Analysis

Whereas GWAS analyses have increased our understanding of the genetic architecture underlying disease, analyses of genetic variants obtained through genotyping chips are typically restricted to common SNPs with minor allele frequencies of more than 5%. SNPs that occur with a frequency of less than 5% are typically referred to as rare variants and

require different analytical approaches. Rare variants may be analyzed in family or twin studies as well as in identified extreme phenotypes by examining the shared occurrence of rare variants in a particular region within individuals with the same phenotype.[111,112] As discussed in more detail later, rare variants may also be examined in unrelated individuals with more recently established techniques.[111,113-120] Information about SNPs of this frequency is typically obtained by deep-sequencing analyses, which can be performed either in a targeted fashion (capturing specific genes or regions) or, more recently, by whole exome or whole genome genotyping methodologies.[121] Exome sequencing looks for variants only in protein-coding regions of the genome. Whereas common variants tend to confer small effects for complex disease, rare variants are more likely to result in large effect sizes within subjects carrying the risk allele. The existence of rare variants with large effects is one explanation for the missing heritability in complex disease. Although early studies evaluating rare variants proposed that the large effects, combined with extreme phenotyping designs, would allow identification of important variants in even small samples, recent publications have indicated that sample sizes similar to those in common variant analyses are required to detect associations.[122]

As discussed with respect to GWAS analyses, common variant analyses in unrelated individuals involve examining the relationships between individual SNPs and a given phenotype. However, individual variant tests examining the associations between outcomes of interest and SNPs with a minor allele frequency of less than 5%, as is done with common variants, are typically underpowered. Instead, rare variants can be analyzed with a variety of established techniques, depending on certain assumptions, typically by collapsing all relevant variants across genes or regions of interest, including burden tests and kernel-based association methods.[113-120] The burden test is most powerful when the rare variants have effects going in the same direction (i.e., all harmful or all protective), whereas kernel-based association methods (e.g., SKAT) are more powerful when a region contains rare variants that are both harmful and protective. One recently established method, the SKAT-O test, has the burden and SKAT tests as special cases and is optimal across multiple scenarios in which rare variants are associated with complex disease; it can also be modified to increase power in small case-control samples.[114,115] In addition to basing these tests on specific chromosomal regions, tests can consider only functional rare variants, that is, those predicted to have a significant impact on the function of protein by altering the amino acid composition. Restricting only to these functional variants can lead to increased power for detecting rare variant association through elimination of neutral nonfunctional variants. Functional prediction scores can be incorporated into rare variant tests.[113] In addition to providing an optimal tool for analyzing rare variant associations within a given locus or gene, these methodologies can be extended to perform meta-analyses of the rare variant effects across regions[123] and to assess the combined effect of rare and common variants.[114,115]

Although rare variant analysis is an emerging field, developing along with the improvements in sequencing technology, rare variants in the *BHLHE41* gene (class E basic helix-loop-helix protein 41; also known as *DEC2*) have been shown to affect sleep duration in humans.[124,125] Using targeted sequencing of clock genes in one study[124] and focused sequencing of *DEC2* in another,[125] studies were able to identify specific mutations within a small number of subjects with short sleep duration (<6 hours) that showed no evidence of daytime impairment. Importantly, once rare variants with large effects were found, functional studies were performed to identify specific biological effects. Of three variants in *DEC2* found, this method of discovery followed by functional analysis showed that the variants with clear effects on sleep duration resulted in a reduction in the ability of *DEC2* to suppress CLOCK/BMAL1 transactivation.[125] Another variant with no obvious effect on sleep duration had no effect on this process.[125] Final proof of the functional significance of one of the variants was obtained by knocking the variants into both mice and *Drosophila* and replicating the phenotype.[124] This is one of the clear advantages of studies of rare variants with large effects, that is, using model systems to prove beyond any doubt their functional significance. Thus, we see the potential of rare variant analysis to inform about important biological disease pathways related to sleep and its disorders.

Current and Future Directions: Sequencing, Copy Number Variation, and Epigenetics

Whereas current sleep-related literature has focused on heritability, linkage, and candidate gene or genome-wide association analyses, emerging areas of focus in genetic research include whole exome and whole genome sequencing, study of copy number variation (CNV), and epigenetic effects.

As mentioned in the context of rare variant analysis, emerging sequencing technologies have resulted in the ability of researchers and clinicians to sequence large regions of the genome at lower cost. Thus, rather than relying on imputation or targeted sequencing of candidate genes for fine-mapping and rare variant analyses, more studies are performing whole exome (i.e., the protein-coding regions) or whole genome sequencing. Coupled with emerging bioinformatics techniques for variant calling, whole exome and whole genome sequencing provide accurate identification of all genetic variation within protein-coding genes or the entire genome, respectively.[121,126,127] Accurate variant identification through sequencing allows identification of important variants, including de novo mutations.[121,127]

Research has focused on relationships between single-nucleotide variation and phenotypic variability. However, two additional types of genetic variation may account for some of the missing heritability in current research. The first, CNV, is defined as structural variations in DNA that are greater than 1 kilobase in size, including deletions, insertions, and duplications.[11-13] CNVs have been shown to associate with complex disease, at times being tagged by genotyped SNPs, including body mass index.[128] In contrast to SNPs, where only a single base is changed in the DNA sequence, CNVs lead to the removal or duplication of large numbers of bases, including in some instances whole genes or groups of genes, and hence there is reason to believe that CNVs can have large effects on phenotypes. CNVs have been shown to confer risk to a wide range of disorders,[129] but particular focus has been on the role of CNVs in neuropsychiatric disorders, in which it has been established that although rare, they confer a large risk to disease in carriers.[130,131] Because they can have large effects on development and functioning, individual CNVs are often very rare, and hence association with individual CNVs can be difficult to detect. For this reason, many studies have

focused on evaluating whether there is an overall increase in the number of CNVs in cases compared with controls. To our knowledge, there has been only one study conducted on CNVs and sleep disorders. Yamasaki et al[132] found an enrichment of rare, large (<1% frequency and >100 kilobases long) CNVs in Japanese narcoleptic patients compared with controls. Their analysis also implicated the region near *PARK2* as harboring CNVs associated with narcolepsy. Given their impact on behavioral traits, analysis of the role of CNVs in sleep disorders is likely to prove fruitful in the future. In addition to CNVs, state-of-the-art genetic research has begun examining the impact of epigenetic modifications, including histone modifications and DNA methylation, on complex disease.[14,133,134] Epigenetic modification is expected to play an important role in disease, given its ability to regulate and to modify gene expression within specific cell types.[14] As the field continues to advance, recent progress includes the use of genome-wide characterizations of DNA methylation, resulting in the first epigenome-wide association studies, which have recently been published for a number of traits, including body mass index.[135,136]

As methodologies for analyzing these data continue to develop and prices for sequencing continue to decrease, research in these emerging genetic areas will further our current understanding of the genetic architecture that underlies complex diseases. Given the known heritability of many traits of sleep and circadian rhythm in humans and of sleep disorders, this will be a major opportunity for sleep research.

GENETICS OF SLEEP DURATION

The duration of sleep in different individuals varies substantially in the general population. In the Finnish population, 14.5% of the population has a sleep duration less than the average 7 to 8 hours, whereas 13.5% have more.[137] In the United States, there is a much larger percentage of the population with short sleep, with 35% sleeping less than 6 hours/night on average. Whereas much of this short sleep is behaviorally induced, being secondary to commitments of modern life, there is a genetic component. Classic twin studies analyzing the differences in sleep duration in monozygotic and dizygotic twins give estimates of heritability on the order of 0.31 to 0.44 for sleep duration.[29-31]

Both common variants associated with sleep duration and rare variants have been described. A number of GWAS have been reported. An initial GWAS, based on the Framingham cohort, used a relatively small sample ($n = 738$) and evaluated 70,987 SNPs. No genome-wide significant associations with sleep duration were identified. The most significant association was with sleepiness assessed by the Epworth Sleepiness Scale and an SNP in the intron of *PDE4D*, which encodes a cyclic adenosine monophosphate–specific phosphodiesterase, a plausible biological candidate.[138]

Another GWAS with a larger number of subjects ($n = 42,517$; all of European ancestry) found a genome-wide significant association in the discovery phase.[139] This was an intronic variant in *ABCC9*. This gene encodes one of the 17 transmembrane domains of the pore-forming subunit of an adenosine triphosphate–sensitive potassium channel. However, the association with this variant and sleep duration was not significant in the replication phase. Moreover, subsequent GWAS have failed to demonstrate this association.[140,141] A study specifically designed to confirm the association described by Allebrandt et al[139] failed to replicate this finding.[142] There was, however, a significant association with another variant in *ABCC9* and depression symptoms.[142]

Although it seems highly unlikely that this variant of *ABCC9* explains variation in sleep duration in human populations, Allebrandt et al[139] did examine further the role of this gene in sleep-wake control. Using a *Drosophila* model and expressing an RNA interference in neurons to knock down the expression of a *Drosophila* homologue of the gene, they found that it reduced sleep amounts at night, particularly in the early part of the night, but not during the day. Thus, this channel likely does play a role in sleep-wake control, and further investigation of this is warranted.

Ollila et al[140] also performed a GWAS based on a Finnish sample ($n = 1941$). They found no genome-wide significant associations. They proceeded, however, to test in a follow-up sample ($n = 6834$) the most suggestive associations. These showed some degree of association that was not significant after correcting for multiple comparisons. The most interesting was KLF6. This is a transcription factor expressed in the mouse cerebral cortex.[143] It can activate the promoter for inducible nitric oxide.[144] Interestingly, higher expression of KLF6 in circulating mononuclear lymphocytes was related to shorter sleep duration, and its expression in these cells was increased with experimental sleep restriction.[140]

The largest study reported to date comes from the CHARGE consortium.[141] The study used data from 18 community-based cohorts that had data on self-reported sleep duration and had genotyped their subjects. The sample size was 47,180. This large sample size led to the finding of not only genome-wide significant associations but also, for the first time, associations that were replicated in independent samples.

The most strongly associated locus that was found was located on chromosome 2 between two genes: PAX8 and cobalamin synthase W domain–containing strand (CBWD). PAX8 is a transcription factor involved in thyroid development but potentially more broadly. CBWD has unknown function but is widely expressed in brain. The initial association was found in individuals of European ancestry but replicated in an independent sample of African Americans ($n = 4771$). The causative variant remains to be identified.

That this large study involving cohorts from across the globe found genome-wide significant associations that were replicated, whereas other studies did not reach this threshold, is telling. Sleep duration, particularly by self-report, is a noisy quantitative phenotype. There are major behavioral influences that override the genetic contribution. Given this variance, it will take large sample sizes such as those used by Gottlieb et al[141] to reach definitive conclusions. Smaller studies in this area for common variants seem pointless. International consortia are needed. Moreover, it would be beneficial if phenotyping was by quantitative methods, such as actigraphy, rather than by self-report. The study by Gottlieb et al requires follow-up to identify likely causative mutations and to assess the role of the relevant genes in model systems and whether they affect sleep duration.

Rare variants affecting sleep duration have also been described.[124,125] Unlike the challenge with GWAS, which is about sample size, the challenge for identifying rare variants is selecting individuals with a clear extreme phenotype. The

seminal study of He et al[124] is based on only two subjects who slept from around 10:00 PM to 4:00 AM (i.e., 6 hours) without evidence of daytime impairment. Given that, in a sense, this is a timing difference, He et al sequenced all clock and clock-associated genes in these subjects. They found a mutation in exon 5 of *BHLHE41*. This mutation is in exon 5 of the gene at the amino acid position 384. The mutation results in a proline at this location being replaced by an arginine. What makes rare mutations easier to work with is that they lead to large effects. Thus, their role can be assessed functionally in both in vitro systems and model organisms such as *Drosophila* and mice. When He et al knocked this mutation into *Drosophila* or mice, it resulted in less sleep (i.e., shortened sleep duration). Moreover, in mice with this mutation, there was a marked reduction in recovery sleep after sleep deprivation compared with wild-type controls.

Given that genes that exhibit rare mutations with large effects may show multiple such mutations (i.e., they are "hot spots"), it is not unreasonable to suspect that other mutations of this gene may be found. Pellegrino et al[125] addressed this question and sequenced *DEC2* in two human samples, one from a previous twin study[20] and one from a study of chronic partial sleep deprivation.[145] Two new mutations in *DEC2* were identified. One was at a different location in the same exon. It also resulted in an amino acid change in the protein. This occurred in one member of a dizygotic twin pair. The twin with the mutation slept 2 hours less per day than the twin partner and had substantially less performance lapses during prolonged sleep deprivation. The other mutation was at the same site as that described by He et al[124] and was found in three unrelated individuals in the cohort who had chronic partial sleep duration. There was, however, no obvious effect of this variant on sleep duration.

To further investigate why some variants of *DEC2* had effects on sleep duration whereas others did not, Pellegrino et al knocked these different mutations into a cell-based system that used a PER2:luciferase reporter to assess rhythmic changes in expression of PER2. Both the variant described by He et al[124] and that in the twin (see before) resulted in reduced ability of *DEC2* to suppress CLOCK/BMAL1 transactivation (i.e., they had clear functional consequences). In contrast, the mutation found in the three unrelated individuals had no such functional effect. This is likely the reason that there is no effect on sleep duration in these individuals. Random mutagenesis of exon 5 in *DEC2* found a number of other mutations that reduced the ability of *DEC2* to suppress CLOCK/BMAL1 transactivation. Whether these mutations also occur in human populations remains to be determined. It seems likely, however, that other variants of *DEC2* will be identified that result in short sleep.

PER3 VARIABLE NUMBER OF TANDEM REPEATS

Perhaps the most studied gene variant with respect to its role in normal sleep-wake behavior is the polymorphism with variable number of tandem repeats in *PER3* component of a 54–base pair motif in exon 18 that is repeated either four or five times (for review, see[146]). This polymorphism is found only in primates and not in other species (e.g., mice and rats).[147,148] The number of repeats varies in different primate species, with a range of 2 to 11.[148] Humans can be homozygous 4/4, homozygous for the 5/5 repeats, or heterozygous 4/5. In populations of European ancestry, about 10% are $PER3^{5/5}$, with 50% homozygous for 4/4 and 40% heterozygous.[149] In New Guinea, the prevalences of the different genotypes are reversed.[149] Thus, it is a common polymorphism with likely a small effect. As such, as described previously, large samples with replications are needed to be sure of a real association. Studies in this area have, however, had very small sample sizes, on the order of 20. Power has been increased by selectively recruiting individuals on the basis of genotype, thereby enriching the sample studied for the less common $PER3^{5/5}$ genotype. Even with this, the studies remain underpowered. This issue of power likely contributes to the varying results in the literature. Moreover, this polymorphism has been associated with a large number of different phenotypes (Table 31-2), therefore further inflating the likelihood of finding spurious associations.

The initial claim for this polymorphism was related to diurnal preference. The $PER3^{5/5}$ genotype was found to be of higher prevalence in morning types and very low in individuals with delayed sleep phase syndrome.[94] The strength of the association with diurnal preferences attenuates with age.[150] The association with diurnal preferences was replicated in Brazil but not with delayed sleep phase.[151] Association with diurnal preference has also been found in South Africa[152] but not in Colombia[153] or in Norway.[154] The latter negative study cannot be attributed to age because it was conducted in Norwegian university students.

Following these initial observations, a more in-depth phenotyping study was done in a small sample of $PER3^{5/5}$ homozygotes ($n = 10$) and $PER3^{4/4}$ homozygotes ($n = 14$).[155] No difference was found in circadian behavior, including the timing of the onset of melatonin secretion. Instead, the major difference between genotypes was in sleep and wake behavior, particularly during sleep deprivation. Individuals with the $PER3^{5/5}$ genotype showed the following compared with the 4/4: (1) larger rise in theta power during constant wakefulness; (2) worse performance on a battery of tests during extended wakefulness; the $PER3^{5/5}$ homozygotes performed particularly poorly during the biologic night; and (3) stronger inhibition of rapid eye movement sleep during recovery sleep after sleep deprivation. A subsequent report highlighted that the *PER3* polymorphism affects the impact of sleep deprivation on performance in the early morning hours.[156] There are, however, again negative studies (see[20,157,158]). The best direct evidence that this polymorphism tandem repeat in *PER3* directly affects sleep homeostasis comes from recent elegant studies in mice.[159] Hasan et al created mice on a C57BL/6 background into which they knocked in the humanized 4/4 or 5/5 *PER3* tandem repeat. The phenotypes of these mice were assessed and compared with wild-type mice. There was no difference in baseline sleep-wake or circadian behavior. There was, however, a difference in the response to sleep deprivation. The increase in EEG delta power, a measure of sleep homeostasis, with sleep loss was greater in the $PER^{5/5}$ mice than in the other genotypes, and these mice more fully compensated for the effects of sleep deprivation. Changes in gene expression with sleep loss in cortex and hypothalamus were also different between genotypes.

These results in mice are compatible with the positive human studies. Ultimately, translation of these findings to humans will require study in a much larger sample of humans, a replication sample, and careful phenotyping. Currently, the

Table 31-2 Positive Associations Reported in Humans for the Tandem Repeat Polymorphism in *PER3*

Phenotype	Reference
Diurnal preference	Archer SN, Robilliard DL, Skene DJ, et al. A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* 2003;26:413-5.
Changes in electroencephalogram spectra during extended wakefulness	Viola AU, Archer SN, James LM, et al. PER3 polymorphism predicts sleep structure and waking performance. *Curr Biol* 2007;17:613-8.
Changes in cognitive performance during extended wakefulness	Groeger JA, Viola AU, Lo JC, et al. Early morning executive functioning during sleep deprivation is compromised by a PERIOD3 polymorphism. *Sleep* 2008;31:1159-67.
Sympathovagal heart rate balance during baseline and recovery sleep	Viola AU, James LM, Archer SN, et al. PER3 polymorphism and cardiac autonomic control: effects of sleep debt and circadian phase. *Am J Physiol Heart Circ Physiol* 2008;295:H2156-63.
Changes in functional magnetic resonance imaging–assessed brain response to an executive task in a period without sleep	Vandewalle G, Archer SN, Wuillaume C, et al. Functional magnetic resonance imaging–assessed brain responses during an executive task depend on interaction of sleep homeostasis, circadian phase, and PER3 genotype. *J Neurosci* 2009;29:7948-56.
Alerting response to light	Chellappa SL, Viola AU, Schmidt C, et al. Human melatonin and alerting response to blue-enriched light depend on a polymorphism in the clock gene PER3. *J Clin Endocrinol Metab* 2012;97:E433-7.
Suppression of melatonin secretion with blue light	Chellappa SL, Viola AU, Schmidt C, et al. Human melatonin and alerting response to blue-enriched light depend on a polymorphism in the clock gene PER3. *J Clin Endocrinol Metab* 2012;97:E433-7.
Insomnia severity in alcohol dependence	Brower KJ, Wojnar M, Sliwerska E, et al. PER3 polymorphism and insomnia severity in alcohol dependence. *Sleep* 2012;35:571-7.
Salivary cortisol secretion	Wirth M, Burch J, Violanti J, et al. Association of the Period3 clock gene length polymorphism with salivary cortisol secretion among police officers. *Neuro Endocrinol Lett* 2013;34:27-37.
Sleep ability	Maire M, Reichert CF, Gabel V, et al. Sleep ability mediates individual differences in the vulnerability to sleep loss: evidence from a PER3 polymorphism. *Cortex* 2014;52:47-59.

jury is still out as to the importance of this polymorphism in the human population.

CHRONOTYPE AND CIRCADIAN RHYTHM SLEEP DISORDERS

Chronotype and circadian rhythm sleep disorders (CRSDs) would be logical candidates for genetic studies, given the tremendous success in identifying the components of the molecular circadian clock intrinsic to all cells. The molecular circadian clock consists of an autoregulatory negative feedback loop involving the *Period* (*PER1*, *PER2*, and *PER3*) and *Cryptochrome* (*CRY1* and *CRY2*) genes.[160] Other genes involved in the molecular generation of circadian rhythms include *casein kinase I-delta* and *I-epsilon* (*CK1δ* and *CK1ε*), *CLOCK*, *NPAS1*, *NPAS2*, *DEC2*, and *BMAL1* and *BMAL2*.

With respect to chronotype, several twin and family studies have estimated the heritability of the broader trait of morningness and eveningness, rather than examining CRSD. Most of these studies assessed chronotype with the Horne-Östberg Morningness-Eveningness Questionnaire.[161] With use of twin data, the heritability of chronotype was estimated to be 54% in the United States[33] and 44% in the Netherlands.[34] Family-based studies of Hutterites[35] and in the Amazon[32] estimated lower heritability, with h^2 estimates of 14% and 23%, respectively. Taken together, results suggest that chronotype is a moderately heritable trait, with genetic factors potentially explaining as high as 50% of the variability within populations. The heritability of CRSD is currently unknown.

There have been a few reports of extended families with high rates of CRSDs, primarily advanced sleep phase syndrome (ASPS). In the first such report, three family members with very strong advanced sleep phase based on clinical history and objective markers were found to have a shorter intrinsic period of their endogenous rhythm.[162] In other families, it has been possible to identify specific genetic variants that segregate with ASPS. A serine to glycine mutation in the casein kinase I-epsilon binding region of *PER2* was reported in one study,[163] although there was no evidence of this mutation in two ASPS pedigrees in Japan.[164] In another pedigree, a missense mutation in the *CKI-delta* gene was identified as the causal variant.[164] Of note in this study, the gene was studied in transgenic *Drosophila* and mice, with the model systems showing opposing phenotypes of longer and shorter circadian period, respectively. This demonstrates the power of using model systems in combination with human studies to better isolate causal variants and their mechanisms of action.

Several studies have used a candidate gene approach and examined the association between circadian genes and either chronotype or CRSD. One study found an association between

the 3111C allele of the *CLOCK* gene and *PER3* and eveningness, with some[93,165] but not all[95,166,167] studies finding a significant association. Morningness has been found to be associated with polymorphisms in the clock genes *PER1* and *PER2* in other studies.[96,168] Associations have been found between delayed sleep phase syndrome and the *PER3* variable number of tandem repeats polymorphism[94,97,151] (for further discussion, see earlier) and 3111C allele of the *CLOCK* gene.[93] Polymorphisms in the *PER3* and *ARNTL2* genes were associated with chronotype in a sample of 966 British adults.[169] The results of these studies are therefore mixed, which is likely due to the relatively small sample sizes ($n < 450$), but the overall patterns suggest that not surprisingly, circadian genes play an important role in the determination of chronotype. Too few studies have been conducted in CRSD for any conclusions to be drawn.

GENETICS OF INSOMNIA

There are very few studies on the genetics of insomnia. In part, this may reflect uncertainty as to the appropriate phenotype. Insomnia as a clinical diagnosis is defined as difficulty in initiating or maintaining sleep that is associated with significant distress or daytime consequences. As such, a phenotype for genetic studies could rely on self-reported insomnia symptoms on a questionnaire such as the Insomnia Severity Index.[170] Alternatively, phenotypes could be based on quantitative parameters of questionnaires or sleep diaries, such as sleep latency or wakefulness after sleep onset. The former approach has the advantage of allowing case-control comparisons, which is the approach typically taken in genetic studies, although treating the parameters as quantitative traits is associated with greater statistical power. Another approach would be to use objective assessment of sleep by actigraphy or polysomnography to generate quantitative parameters. There is intuitive appeal to this approach, given that it avoids the cognitive biases associated with self-reports and considering that EEG-based parameters are some of the most heritable traits that have been measured.[53] A limitation of the objective assessment of insomnia is that there is often a discrepancy between these measures and self-reports of sleep in patients with insomnia.

A number of family and twin studies using self-reported measures of insomnia traits have demonstrated moderate heritability. In one of the only studies of childhood-onset insomnia, Hauri and Olmstead[39] found that patients whose insomnia began in childhood reported a positive family history of sleep complaints at a higher rate (55%) than did those with adult-onset insomnia (39%). In a larger cohort study,[40] there was no significant difference in positive family history of insomnia in those categorized as good sleepers, having symptoms of insomnia, and meeting criteria for a full insomnia syndrome (32.7%, 36.7%, and 38.1%, respectively). Significant differences were found only when the good sleepers were separated into those with and without a personal history of insomnia, and those without a personal history had a significantly lower rate of family history (29.0%) than those with a past history (48.9%). Several twin studies have examined the heritability of insomnia-related traits, such as sleep latency and sleep quality. A study conducted by the Australian Twin Registry of 1792 monozygotic and 2101 dizygotic twin pairs included several questions related to sleep quality, disturbance, and overall patterns.[29] Additive genetic influences were found for sleep quality ($h^2 = 0.32$), initial insomnia ($h^2 = 0.32$), sleep latency ($h^2 = 0.44$ for men and 0.32 for women), "anxious insomnia" ($h^2 = 0.36$), and "depressed insomnia" ($h^2 = 0.33$). In a study of twin pairs from the Vietnam Era Twin Registry,[41] heritability estimates were for trouble falling asleep ($h^2 = 0.28$), trouble staying asleep ($h^2 = 0.42$), waking up several times per night ($h^2 = 0.26$), waking up feeling tired and worn out ($h^2 = 0.21$), and a composite sleep score ($h^2 = 0.28$). Overall, self-reported insomnia traits are moderately heritable, with 30% to 40% of the variance attributable to additive genetic factors.

Only two GWAS analyses have been conducted in insomnia. One study, conducted in Korea, included 10,038 individuals who were classified as insomnia cases or controls on the basis of responses to a series of questions about their sleep patterns.[171] An association was found between case-control status and the *ROR1* gene, which modulates synapse formation, although this association did not reach genome-wide significance. In the Australian twin cohort mentioned before, a GWAS was conducted on 2323 individuals with more than 2,000,000 SNPs either tagged or imputed.[172] No SNPs reached genome-wide significance, but the strongest associations were between sleep latency and a group of SNPs in the third intron of the *CACNA1C* gene ($P = 1.3 \times 10^{-6}$), a gene repeatedly associated with bipolar disorder and schizophrenia in multiple studies. This finding did not replicate in a second cohort but was then subsequently replicated in a separate report.[142]

Clearly, the study of the genetics of insomnia is in its infancy. As has been the case in the psychiatric genetics field, significant genetic associations will likely require sample sizes, an order of magnitude higher than those currently reported.

NARCOLEPSY

It has been known since 1984, following a study in Japan, that narcolepsy is associated with specific HLA antigens.[173] HLA antigens are expressed in immune cells and are involved in presenting foreign peptides to receptors on T cells. In Japanese and those of European descent, the risk allele for narcolepsy is DQB1*0602, which occurs together with DQA1*0102 on a haplotype.[174] There are differences in African Americans.[175] In African Americans, the DQB1 alleles are found with distinct DRBI haplotypes—DRB1*1503, DRB1*1501, DRB1*1101, and DRB1*0806.[176,177] The association between DQB1*0602 and narcolepsy is particularly strong in cases of narcolepsy with cataplexy compared with cases without cataplexy.[178] Recent data from the European narcolepsy study showed a very high association in multiple European countries between narcolepsy with cataplexy and DQB1*0602, with a strong aggregate odds ratio of 251 (Table 31-3).[179] This may be the result of a clearer definition of cases.

There are also protective HLA class II haplotypes for narcolepsy.[180] These were demonstrated in a GWAS in European individuals with a case-control design with an independent replication sample. After identification of a protective variant near HLA-DQA2, analysis revealed that cases almost never carried a *trans* DRB1*1301-DQB1*0603 haplotype (odds ratio, 0.02; $P < 6 \times 10^{-14}$). The magnitude of these effects is such that studying HLA antigen profiles in suspected cases

Table 31-3 Association in European Countries with DQB1*0602 in Cases with Narcolepsy and Cataplexy

Country (Case, Control)	Case-DQB1+ No. (%)	Control-DQB1+ No. (%)	OR	*P*
DE (232, 296)	227 (97.84)	72 (24.3.2)	141.24	9.71E-26
CH (66, 473)	65 (98.48)	102 (21.56)	236.42	7.01E-8
NL (323, 469)	318 (98.45)	114 (24.31)	198.05	3.62E-30
PL (63, 197)	63 (100)	44 (22.33)	438.08	2.65E-09
SP (127, 1174)	126 (99.21)	170 (14.48)	744.14	5.25E-11
FR (341, 499)	335 (98.24)	94 (18.84)	240.56	1.18R-37
IT (66, 433)	64 (96.97)	30 (6.93)	429.87	3.21E-16
Mantel-Haenszel (meta-analysis)	1198 (98.36)	626 (17.68)	251.12	1.04-120

The frequency of DQB1*0602 in cases and controls with odds ratio (OR) for increased risk and the *P* value for the association are shown.
DE, Germany; CH, Switzerland; NL, Netherlands; PL, Poland; SP, Spain; FR, France; IT, Italy.
From Tafti M, Hor H, Dauvilliers Y, et al. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep* 2014;37:19-25.

of narcolepsy could have clinical utility. This remains to be assessed.

Although there is a very strong association between narcolepsy with cataplexy and DQB1*0602, this is a common variant in many individuals without narcolepsy. The frequency of this varies by ethnic group, with 12% in Japanese and 38% in African Americans.[174] Thus, having this variant, even if one is homozygous, is not sufficient to develop narcolepsy. There are two possible explanations for its role. First, narcolepsy could be a complex disorder with multiple gene variants other than DQB1*0602 being involved. Second, the presence of DQB1*0602 could make individuals susceptible to as yet largely unidentified environmental insults. These are not mutually exclusive possibilities.

To address the former, Hallmayer et al[181] did an innovative GWAS (Figure 31-4). They assembled multiple cohorts of cases with narcolepsy and cataplexy of different ethnicities. All individuals were positive for DQB1*0602. This ensured that this signal was removed so that other potential gene variants could be identified. Without this, a GWAS would have identified DQB1*0602 and overwhelmed other signals. With this strategy, they identified three SNPs within the T-cell receptor alpha locus that were significantly associated. The highest association had an odds ratio of 1.87 ($P = 1.9 \times 10^{-12}$). These associations were replicated in independent samples of Caucasian patients and Asians from Japan and Korea. There was a similar trend in the African American sample, but this was not significant. This may be an issue of sample size because this was the smallest sample studied. This association has already been replicated in other studies—in the European narcolepsy study[179] and in China.[182]

This association with the T-cell receptor alpha locus was also replicated in a study using an "immune chip" to evaluate variants relevant to the immune system.[183] This study also identified two other associations, one in cathepsin H and one in tumor necrosis factor (ligand) superfamily member 4 (TNFSF4). This led the authors to propose that antigen presentation to T-cell receptor is a key part of the pathogenesis of narcolepsy (Figure 31-5).

Another GWAS study in narcolepsy with cataplexy has identified an SNP in the 3′ untranslated region of the purinergic receptor subtype $P2Y_{11}$ gene (*P2RY11*).[184] The variant has an odds ratio of 1.28 (95% confidence interval, 1.19–1.39; $P = 6.1 \times 10^{-10}$). It seems that this variant may also play a role in the immune system. The variant is associated with substantial reduced expression of P2RY11 in $CD4^+$ T lymphocytes and natural killer cells. Thus, narcolepsy is a complex disorder, but all variants recognized to date are likely to affect the immune system. Thus, there is a strengthening basis to consider narcolepsy a specific autoimmune disorder.

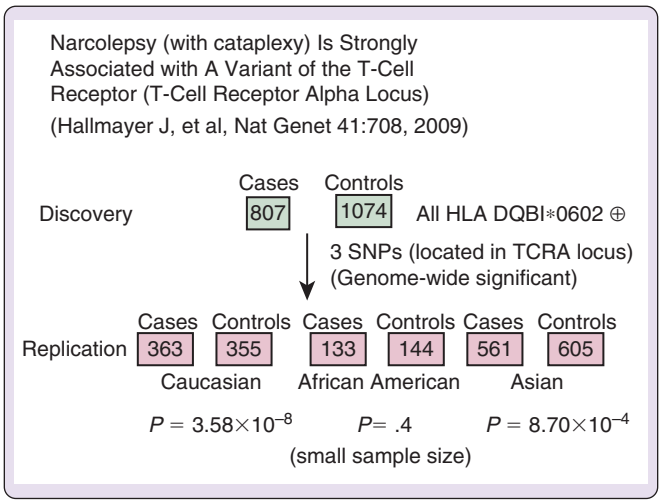


Figure 31-4 Design of the genome-wide association study in narcolepsy with cataplexy that identified the role of variants in the T-cell receptor alpha locus. All cases and controls in the study were positive for HLA DQB1*0602. There was an initial discovery phase followed by three replication samples in different ethnic groups. The variants found replicated in Caucasians and Asians but not in African Americans. There was, however, a much smaller sample of African Americans that might be the explanation for this failure to replicate. SNPs, Single-nucleotide polymorphisms; TCRA, T-cell receptor alpha. (From Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat Genet* 2009;41:708-11.)

As in other areas, there are also rare variants with large effects that can lead to narcolepsy. These are typically part of syndromes in which narcolepsy is part of the syndrome. One syndrome is autosomal dominant cerebellar ataxia, deafness, and narcolepsy. Symptoms typically develop at 30 to 40 years of age. Narcolepsy and deafness typically appear before ataxia. Exome sequencing in three individuals identified rare

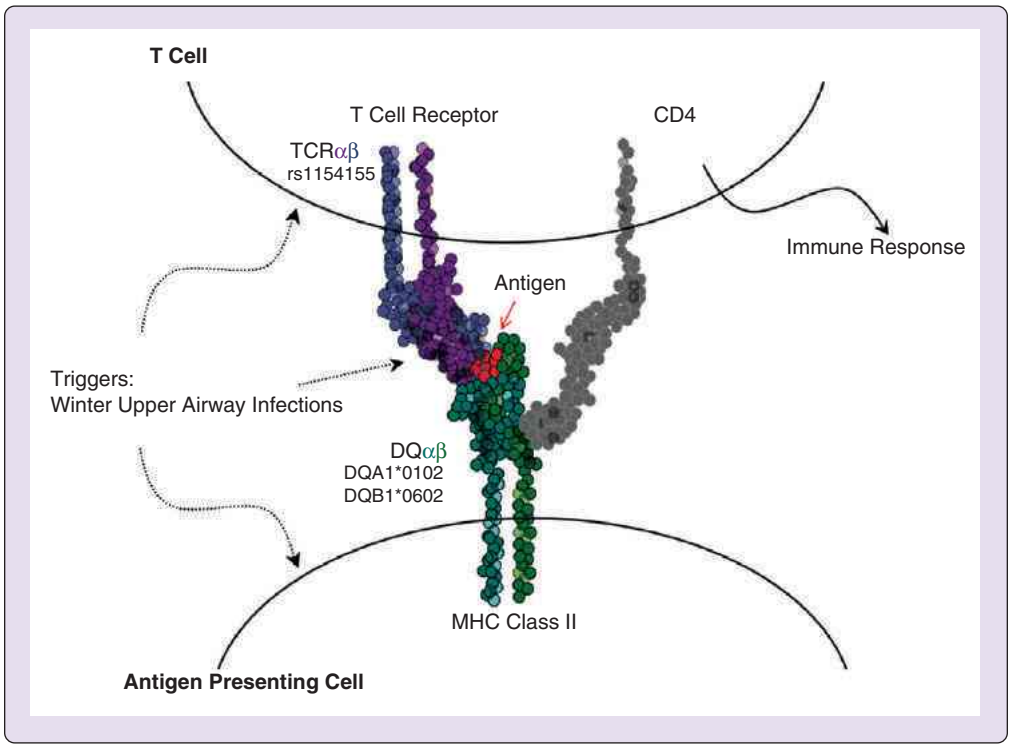


Figure 31-5 The pathogenesis of narcolepsy is likely autoimmune. Variants of HLA antigens confer risk or protection for narcolepsy. In addition, variants of the T-cell receptor alpha locus confer risk. These variants are thought to make individuals susceptible to an environmental challenge, such as upper airway infections. TCR, T-cell receptor; MHC, major histocompatibility complex. (From Faraco J, Mignot E. Genetics of narcolepsy. *Sleep Med Clin* 2011;6:217-28.)

mutations in DNA methyltransferase (*DNMT1*).[185] This enzyme is responsible for maintaining methylation patterns in development. It is known to be expressed in immune cells and is required for the differentiation of $CD4^+$ cells into T regulatory cells. Another variant of this gene was described in a single Brazilian patient.[186]

Another rare variant in myelin oligodendrocyte glycoprotein in a family with narcolepsy and cataplexy has been described.[187] This discovery was based on a linkage study of a single large family with 12 affected individuals. Exome sequencing identified a rare mutation in the second exon of the gene. This mutation was present in all affected individuals but absent in all unaffected family members and in 775 unrelated control subjects.

There are likely other rare variants conferring risk of narcolepsy, and studies with exome sequencing and whole genome sequencing are in progress.

Although it is clear that the presence of specific HLA-DQB1* alleles confers substantial risk or protection for narcolepsy with cataplexy,[179,180] a recent article from Ollila et al[188] presents the idea that HLA typing of HLA-DQA1*, in addition to HLA-DQB1*, to better characterize the specific haplotypes and possible heterodimers, may be an important next step in understanding this relationship. They propose an allele competition model, under which the risk conferred by specific DQB1* alleles is modified on the basis of the specific DQA1* alleles present because of the differences in binding affinity between specific DQA1* and DQB1* alleles. For example, whereas patients homozygous for the DQA1*01:02~DQB1*06:02 haplotype have the highest risk of narcolepsy, patients with only one DQA1*01:02~DQB1*06:02 haplotype may either have moderately increased risk (odds ratio, 1.0–1.5) if they have DQA1*01:02 and a DQB1*05/06 allele that is not DQB1*06:02 at the other chromosome or be protected from narcolepsy (odds ratio, 0.5) if they carry a DQA1*01 allele that is not DQA1*01:02 on the other chromosome. Although preliminary data support these assertions, the validity of the allele competition model hypothesis has been questioned.[189] Regardless, it remains clear that HLA-DQ alleles are strongly associated with narcolepsy.

RESTLESS LEGS SYNDROME

The discoveries related to RLS are one of the major accomplishments in the modern genetic era of elucidating gene variants for a sleep disorder. That this occurred where the primary phenotype is based on a questionnaire is particularly notable. That the key gene variants have now been replicated in multiple studies should put to rest the debate by critics as to whether this is a real disorder. This accomplishment was based on a solid base of defining diagnostic criteria.[190]

Genetic research in this area was initially stimulated by the findings of a large proportion of patients with RLS who had positive family histories.[191-193] Twin studies confirmed heritability.[36-38] Complex segregation analyses studied the mode of inheritance in German families with RLS.[194] For early age at onset of RLS (i.e., before 30 years of age), the pattern of inheritance was best described by an autosomal

dominant model with a single major gene. A similar study in the United States[195] came to the same conclusion.

Because family aggregation was so high in RLS, this led to use of linkage analyses within large family pedigrees. This was done in different parts of the world, including Canada in a French Canadian pedigree, northern Italy, North America, and southern Tyrol. These studies resulted in strong evidence of linkage that was genome-wide significant (see Table 31-1). A number of different linkage regions were found. However, until case-control association analyses of SNPs in these linkage regions were performed, linkage studies on their own did not identify specific gene variants.

The first linkage region to be investigated in this way was RLS-1 on chromosome 12.[64] Case-control association was conducted with 1536 SNPs in this region in the discovery phase, with 24 of the most significantly associated SNPs being investigated in an independent case-control replication sample. A significant association with an SNP in neuronal nitric oxide synthase was identified that was protective (odds ratio, 0.76; 95% confidence interval, 0.64–0.91).[64]

A similar strategy was used to evaluate the RLS-3 region on chromosome 9 with 3270 SNPs in the discovery phase and 8 in the independent replication sample.[196] Two different SNPs in introns of the protein tyrosine phosphatase receptor type delta (PTPRD) were shown to significantly associate with RLS. There are two independent association signals. The function of this gene has been studied in mouse knockout models.[197,198] PTPRD knockout mice show impairment in long-term potentiation in memory formation and abnormal axon targeting to motoneurons during development.[197,198] It is conceivable that this developmental role could underlie the role of this gene in RLS. An independent replication of this association has been reported.[199]

Whereas linkage studies have led to new insights with this combined approach, the seminal event in RLS genetics was the direct, simultaneous publication in 2007 of two independent successful GWAS.[200,201] The first study, conducted in Germany and in French Canadians,[200] used questionnaire-based case definition. The study employed a discovery phase with two independent case-control samples for replication. It identified three different loci: two intronic SNPs in *MEIS1* on chromosome 2 (odds ratio, 1.74); five intronic SNPs in *BTBD9* on chromosome 6p; and seven intronic or intergenic SNPs in chromosome 15q in the *MAP2K5* gene and the adjacent *SKOR1* gene. *MEIS1*[202] and *SKOR1*[203] play a role in development (for design of study, see Figure 31-6).

The other study, conducted primarily in Iceland,[201] used a discovery sample in Iceland and two independent replication case-control samples—one in Iceland, one in the United States. These investigators used a different phenotyping strategy as they measured, using actigraphy of the legs, the frequency of periodic limb movements over several nights of recording. They found an association with the same SNP for *BTBD9* (odds ratio, 1.7). This association was found only in cases with periodic limb movements and not in cases with only sensory symptoms. Thus, this specific gene variant likely contributes to the motor manifestations of RLS. Interestingly, this SNP was associated with ferritin levels in this study. It has long been known that alterations in iron metabolism can play a role in the pathogenesis of RLS (for review, see[204]). *BTBD9* has, however, not been found as a genetic variant determining iron metabolism in other studies.[205] Moreover,

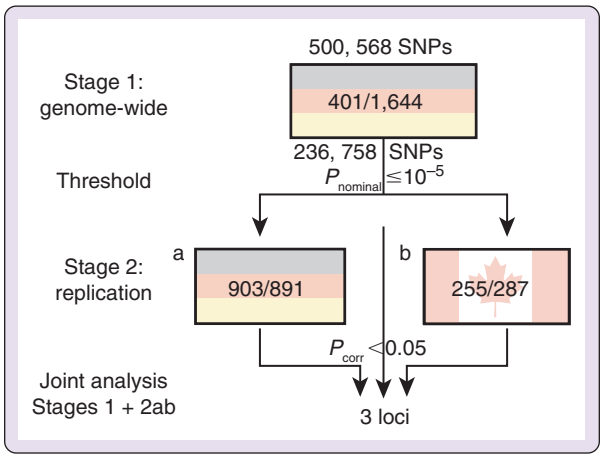


Figure 31-6 Design of the seminal genome-wide association study (GWAS) on restless legs syndrome. The stage 1 discovery phase involved assessing a large number of single-nucleotide polymorphisms (SNPs) in 401 cases and 1644 controls. The few SNPs that were nominally significant at $P < 10^{-5}$ (they were not genome-wide significant) were assessed in two replication samples: one had 903 cases and 891 controls, whereas the other had 255 cases and 287 controls. The SNPs that were significant in the replication samples were then assessed in a joint analysis across all three samples. (From Winkelmann J, Schormair B, Lichtner P, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* 2007;39:1000-6.)

studies looking for association of RLS with genes known to affect iron metabolism have been negative.[206]

The association of *MEIS1* and *BTBD9* has been confirmed in one study,[207] and all three loci were identified in other studies.[208-210] Interestingly, in a secondary form of RLS that occurs in patients with end-stage renal disease, the variant of *BTBD9* is again associated with RLS.[211]

Although these findings are exciting, the variants so far identified account for only around 3% of the heritability of RLS.[212] This, unfortunately, is not unusual for complex traits.[54] The reasons for missing heritability were described earlier, including the role of rare variants. Rare variants can be identified by deep resequencing of relevant genes and exome sequencing or whole genome sequencing. Efforts to use these approaches to investigate the genetics of RLS have begun but are in their infancy. Individuals with RLS have an excess of rare variants of *MEIS1* compared with controls.[213] In particular, there is an excess of loss-of-function alleles in cases with RLS.[213]

Exome sequencing has also started to be applied.[214] A variant of *PCDHA3* in a family with RLS that was absent in 500 controls suggests a functional role. This gene encodes protocadherin alpha 3 that is a member of the protocadherin gene family. It is expressed in neurons and is present at synaptic junctions, where it plays a role in neural cell-cell interaction.[215]

There is little doubt that a search for rare variants with new sequencing technology will continue in RLS, as will studies of CNV and altered methylation patterns. Given the large number of carefully assembled case-control and family-based cohorts, we can look forward to further exciting developments in this area.

For helpful recent reviews on genetics of RLS, see.[216,217]

GENETICS OF OBSTRUCTIVE SLEEP APNEA

OSA is a common condition.[218] The major risk factor in middle-aged adults is obesity.[218] With increasing obesity rates in the United Sates, the prevalence of this condition is increasing.[219] It has long been known that OSA aggregates in families. The initial finding that OSA has a major genetic component came from a study of a single family with a high prevalence of OSA.[49] Subsequently, it was shown that symptoms of OSA, such as habitual snoring, excessive sleepiness, snorting, gasping, and witnessed apneas, also aggregate in families.[48]

These observations led to studies including measurement of apneas and hypopneas during sleep. Such studies were conducted in the United States,[44,47] Israel,[46] Scotland,[45] and Iceland.[43] Because obesity, a major risk factor for OSA, is itself heritable,[220-223] whether family aggregation of OSA is simply due to obesity needs to be addressed. The important Cleveland Family Study addressed this issue and assessed increased relative risk of OSA in first-degree family members. This increased risk was not affected after controlling for body mass index.[47] Thus, family aggregation is unlikely to be simply explained by obesity.

Obesity as an explanation for family aggregation of OSA was more definitively addressed by Mathur and Douglas.[45] They examined the prevalence of OSA in first-degree relatives of less obese individuals with OSA (body mass index < 30 kg/m^2) and compared this with that in controls chosen at random from a list of patients in a primary care practice. Controls were matched for age, gender, height, and weight. The prevalence of OSA was significantly higher in first-degree relatives of OSA patients than in controls. First-degree relatives also had more retroposed mandibles and maxillae than controls did.[45] Thus, in these less obese cases, subtle differences in craniofacial structure likely played the key role, and hence genes for craniofacial structure could be involved. (For review of these early studies on genetics of OSA, see review by Redline and Tischler.[224])

Although family aggregation has been known for 2 decades, there has been little progress in identifying relevant gene variants. Studies in this area have been particularly problematic. Early family-based linkage studies were underpowered, did not find genome-wide significant linkages, and did not employ fine mapping to confirm the peak and to narrow the linkage region.[73-75] Candidate gene studies examining multiple candidates did not employ any replication samples.[98] Candidate gene studies have been massively underpowered, given that one is looking for small effects (typically, odds ratios of around 1.2) for common variants[100] (see Figure 31-2). Meta-analyses reveal that most claimed candidate gene variants, including *APOE* ε4, show no association.[100,225] The one exception is −308 tumor necrosis factor-alpha promoter polymorphism. This affects transcription of this gene.[226] Association with this SNP and OSA has been identified in European populations[227] and in Indians.[228] However, it seems unlikely that this gene variant actually contributes to the risk of OSA. In the pediatric population with OSA, excessive sleepiness is more pronounced in patients with OSA with this SNP than in those patients without the SNP.[101] Thus, this SNP may contribute to one of the key consequences of OSA that patients may recognize and have (e.g., it is more likely that they seek evaluation). This concept adds additional difficulty to determining the gene variants conferring risk for OSA.

The most convincing evidence of a relevant gene variant comes in a recent study that used subjects who were part of the Cleveland Family Study and the Sleep Heart Health Study.[99] They used a custom candidate gene array with 45,237 SNPs for more than 2000 candidate genes chosen because of their potential relevance to heart, lung, blood, and sleep disorders.[229] Replication samples came from the Western Australia Sleep Health Study for subjects of European ancestry and from the Cleveland Sleep Apnea Study and Case Transdisciplinary Research on Energetics and Cancer Colon Polyp Study for African Americans.

In African Americans, one SNP in the intronic region of the *LPAR1* gene showed genome-wide association with log apnea-hypopnea index (AHI). Using AHI as a quantitative phenotype is challenging, given the known night-to-night variability in the measure.[230] This association was greater in the nonobese than in the obese subjects. This association with OSA status was confirmed in African Americans in the replication sample. In European samples, this SNP also showed evidence of association to an apnea phenotype ($P = .01$) and a trend for association with log AHI ($P = .06$). This gene is lysophosphatidic acid receptor 1 (*LPAR1*). The gene is thought to play a proinflammatory role.[231,232] It is also expressed in developing cerebral cortex.[233] A mouse knockout of the gene results in changes in behavior as well as in craniofacial abnormalities.[234] Thus, the connection with OSA might be due to either neural differences in airway control or craniofacial alterations.

From the studies of cohorts of European ancestry, no SNP met criteria for a statistically significant association with log AHI. One SNP in the intronic region of the prostaglandin receptor 2 was significantly associated with OSA. There was some evidence of association in the replication sample.

That genetic studies of OSA have had such limited success is not only related to the poor study designs; it is also an extremely challenging area. There are multiple pathways contributing to OSA risk or protection (Figure 31-7), each with many likely variants.

As discussed, a major risk factor is obesity. We now know from multiple GWAS of body mass index that there are 97 loci that contribute to variations in body mass index.[235] The

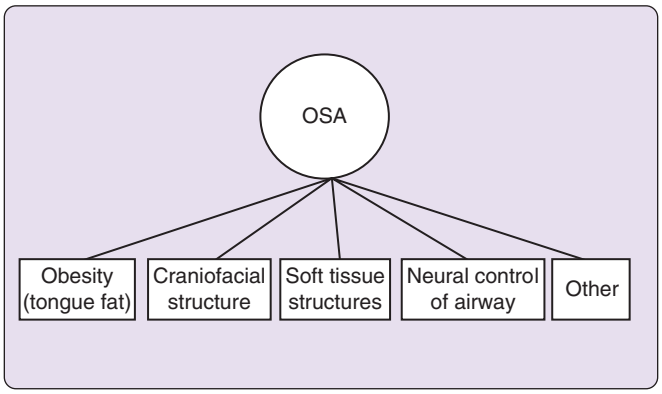


Figure 31-7 Multiple pathways to obstructive sleep apnea (OSA). Each is likely to have many genetic variants involved. These multiple pathways make finding gene variants conferring risk or protection for OSA challenging.

role of these in OSA has not been determined. It may, however, not simply be obesity but rather a particular distribution of fat. Novel imaging studies reveal that there is a difference in volume of fat in the tongue between cases with OSA and controls even after controlling for differences in overall body mass index.[236] There are gene variants associated with other specific distributions of fat (e.g., pericardial fat).[237] Determining whether tongue fat is a unique fat distribution is an area of opportunity.

The volume of soft tissue structures also likely contributes to the genetics of OSA. The volume of key upper airway soft tissues that confer risk for OSA is heritable.[51] The heritability of the volume of the lateral pharyngeal walls is 25.6%; for tongue volume, 37.8%; and for total volume of upper airway soft tissue structures, 41.3%. This is not simply related to obesity because after controlling for total neck fat volume, heritability estimates either stay the same or increase.

Soft tissue structures are not the only relevant structural risk factor; so too is craniofacial structure. Although several differences in craniofacial structures have been demonstrated between OSA cases and controls, the most robust finding is reduced mandibular length in OSA cases as revealed by meta-analyses.[238] Heritability of craniofacial dimensions by cephalometric analysis has been shown.[239,240] This heritability has been confirmed by three-dimensional analysis with magnetic resonance imaging.[50] Mandibular length and mandibular width are both heritable.

These different structural risk factors play a different role in different ethnic groups. An elegant study comparing whites with the same degree of OSA in Sydney, Australia, with Chinese OSA subjects in Hong Kong found that the Caucasians were more obese and had a larger tongue volume.[241] Craniofacial dimensions played a larger role in Chinese with more restriction of the craniofacial base.[241] Thus, the relative role of different genetically determined pathways to OSA varies by ethnic group.

There are, however, not only structural risk factors for OSA but also physiologic ones.[242] The key physiologic variables are overall loop gain, arousal threshold, and airway muscle responsiveness. Among these, overall loop gain is the most important.[242,243] There are individuals with OSA who do not have particularly collapsible airways but do have high loop gain.[242] The main determinant of loop gain is ventilatory response to hypoxia and hypercapnia. The hypoxic response is also a highly heritable trait as revealed by studies in twins.[52] Gene variants determining the hypoxic response have, however, not been identified to date. Thus, this is another potential genetic contribution to OSA.

These advances in understanding of the different phenotypic pathways to OSA underscore the inherent difficulty in determining genetic variants conferring risk for or protection from OSA. They also suggest that a productive approach, given the inherent complexity, is to address gene variants not for OSA itself but rather for each of the relevant intermediate traits.

GENETICS OF CONGENITAL HYPOVENTILATION

Congenital hypoventilation syndrome (also called Ondine's curse) is a rare disorder, but considerable progress has been made in understanding the genetic basis. It typically is manifested in early life. Individuals with this disorder ventilate normally during wakefulness. They can increase ventilation when asked to do so. The problem arises during sleep. When they fall asleep and loose the wakefulness drive to breathe, they hypoventilate. The problem is in their CO_2-dependent ventilatory system. They have very low or absent ventilatory responses to hypercapnia. This hypoventilation can be major with marked elevations in Pco_2. Hence, they require assisted ventilation during sleep. This is typically done by creating a tracheostomy early in life with positive pressure ventilation through the tracheostomy during sleep (for review, see[244]).

The relevant gene is *PHOX2B*. This is a transcription factor involved in the development of the autonomic nervous system. The role of this gene in this condition was identified by Amiel et al in 2003.[245] Mice that bear a frequent mutation of the disease in humans do not respond to hypercapnia and die in the early postnatal period.[246]

In humans, the most common mutation is expression of polyalanine repeats in exon 3 of the gene.[245] The upper limit of normal for the number of repeats is 20.[245] Most mutations occur de novo (i.e., there is no family history).[247] There are rare cases that are autosomal dominant, and up to 25% of cases show somatic mosaicism, that is, the mutation is found in some cells in the parents of cases[248] but not in all; 92% of cases with this condition have expansion of the polyalanine repeats (25 to 33 repeats).[249] There are case reports of individuals with 25 repeats who presented as adults.[250]

Patients with this disorder who do not have increased polyalanine repeats in exon 3 most commonly have other mutations in *PHOX2B*.[247,251] Patients with congenital hypoventilation syndrome have an increased rate of Hirschsprung disease and tumors of neural crest origin.[251] These associated conditions are more common in patients who have different mutations in *PHOX1B* compared with those with increased polyalanine repeats.[251]

CLINICAL PEARLS

- Clinicians should be aware of the relevance of sample size (power) and of the importance of independent replication in assessing literature on gene variants for sleep and its disorders.
- There are both many common gene variants with small effects and rare variants with large effects that contribute to risk for and protection from sleep disorders.
- There have been successful genetic studies in narcolepsy, RLS, and congenital hypoventilation syndrome.
- Variants of the clock-associated gene *DEC2* affect duration of sleep and response to sleep deprivation. Their functional role is confirmed by studies in model systems.

SUMMARY

In this chapter, we have provided an overview of the rapidly developing methodology being used to identify gene variants conferring risk for or protection from sleep disorders. We have explained the importance of study design, sample size, and replication. Unfortunately, much of the published literature on genetics of sleep disorders does not meet the criteria required for firm conclusions to be reached.

There have been important discoveries in narcolepsy, RLS, and congenital hypoventilation syndrome. For congenital hypoventilation syndrome, genotyping to identify the

causative mutation is now part of routine clinical care. For narcolepsy and RLS, the genetic variants identified have not at present altered current clinical practice. The findings in narcolepsy, in particular with respect to the role of different HLA antigens, do suggest that studying these variants in individual patients could have a clinical role. This remains to be determined. This is not the case for RLS because the variants identified to date explain so little of the heritability. The current key value of the results in this area is identification of novel genes and hence pathways. The question now is, What do these genes do? This opens up entirely new opportunities to understand the fundamental pathogenesis of sleep disorders and to develop novel targets for therapeutic intervention. One example of this is the gene *BTBD9*, which has been found to be associated with RLS in multiple studies, as described earlier. The approach to begin to identify its function is to study animal models in which expression of the gene is altered. For *BTBD9*, this has been done in both *Drosophila* and mice. A knockout of *BTBD9* in mice led to phenotypic differences.[252] Knockout mice had enhanced long-term potentiation in hippocampus and enhanced cued and contextual fear conditioning. Thus, *BTBD9* is involved in synaptic plasticity. Sleep was not studied to date in these mice. Studies in *Drosophila* tie *BTBD9* more directly to the function of dopamine neurons and to iron metabolism.[253] Flies with loss of function of *BTBD9* have fragmented sleep. Knockdown of *BTBD9* only in dopaminergic neurons using RNA interference leads to the same phenotype. Flies with loss of *BTBD9* had lower dopamine levels in brain, and agonists of dopamine that are used to treat RLS in humans reverse the sleep fragmentation phenotype in *Drosophila*. The connection to iron comes from studies in cell culture. Specifically, overexpression of *BTBD9* leads to reduction in iron-responsive element-binding protein with a resultant increase in ferritin levels. The connection between dopamine neuronal function and iron metabolism has not been investigated.

These functional studies are a start and in the future may lead to personalized approaches to managing patients.[254] Ultimately, we need to understand the role of these newly identified gene variants and whether they represent potential targets for future treatment development. Thus, we are at the early stages of this exciting journey, and much remains to be discovered.

Selected Readings

Andreani TS, Itoh TQ, Yildirim E, et al. Genetics of circadian rhythms. *Sleep Med Clin* 2015;**10**(4):413–21.

Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature* 2007;**447**:655–60.

Freeman AA, Rye DB. The molecular basis of restless legs syndrome. *Curr Opin Neurobiol* 2013;**23**:895–900.

Gottlieb DJ, Hek K, Chen TH, et al. Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. *Mol Psychiatry* 2015;**20**(10):1232–9.

He Y, Jones CR, Fujiki N, et al. The transcriptional repressor DEC2 regulates sleep length in mammals. *Science* 2009;**325**:866–70.

Liblau RS, Vassalli A, Seifinejad A, Tafti M. Hypocretin (orexin) biology and the pathophysiology of narcolepsy with cataplexy. *Lancet Neurol* 2015;**14**(3):318–28.

Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–53.

Ollila HM, Ravel JM, Han F, et al. HLA-DPB1 and HLA class I confer risk of and protection from narcolepsy. *Am J Hum Genet* 2015;**96**: 136–46.

Tafti M, Hor H, Dauvilliers Y, et al. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep* 2014;**37**:19–25.

Toh KL, Jones CR, He Y, et al. An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* 2001;**291**:1040–3.

Varvarigou V, Dahabreh IJ, Malhotra A, et al. A review of genetic association studies of obstructive sleep apnea: field synopsis and meta-analysis. *Sleep* 2011;**34**:1461–8.

Weese-Mayer DE, Rand CM, Berry-Kravis EM, et al. Congenital central hypoventilation syndrome from past to future: model for translational and transitional autonomic medicine. *Pediatr Pulmonol* 2009;**44**:521–35.

A complete reference list can be found online at ExpertConsult.com.

Section 5

Chronobiology

Chapter 32

Introduction: Master Circadian Clock and Master Circadian Rhythm

Fred W. Turek; Phyllis C. Zee

One of the distinguishing characteristics of sleep in animals as diverse as insects, fish, and mammals is that the timing of sleep and wake for the majority of species is rigidly confined to certain times of the day or night. As detailed in a number of chapters in this section, the master biologic clock regulating the timing of sleep and wake in mammals also regulates most if not all 24-hour (i.e., circadian) behavioral, physiologic, and biochemical rhythms. This master circadian clock is located in the bilaterally paired suprachiasmatic nucleus (SCN) in the anterior hypothalamus (Chapters 33 and 34). Although it has recently been discovered that many tissues and organs can generate circadian rhythms in vitro, and therefore independent of the SCN (Chapter 33), the SCN remains at the top of the hierarchy of the mammalian circadian clock system.

Similarly, one could argue that the sleep-wake cycle represents the "master circadian rhythm"; the SCN control of this rhythm in turn controls the timing and expression of a multitude of downstream rhythms. Although the expression of many 24-hour rhythms may be primarily under the control of the circadian clock in the SCN, many rhythms are largely dependent on whether the organism is asleep or awake, regardless of the circadian clock time.[1] Undoubtedly, the expression of most rhythms at the behavioral, physiologic, and biochemical levels is regulated by the integration of inputs from the circadian clock and the sleep-wake state of the animal. Indeed, it can be argued that the entire temporal organization of an organism represents some combined effect of circadian clock inputs and the sleep-wake state of the organism. Thus, the circadian and sleep control centers have evolved together to ensure that the timing of internal events relative to one another and to the external environment is coordinated in such a fashion to maximize the survival of the species.

Of particular note of the "downstream" rhythms regulated by the circadian clock is the feeding rhythm. In the last few years, a number of studies have demonstrated that the timing of food intake can regulate the expression of circadian clock genes in many peripheral tissues,[2] which in turn control the 24-hour rhythm in the expression of many clock-controlled genes (perhaps as many as 10% to 50% of all genes expressed in a given tissue).[3] The many interactions between the systems regulating circadian, sleep, and energy balance, from molecular to behavioral rhythms, have led to much interest in the importance of these linkages to obesity, diabetes, and cardiometabolic disorders.[2]

The need to "rest" and the need to adjust to the daily changes in the physical environment due to celestial mechanics certainly represented two darwinian pressures that guided the evolution of living organisms since life appeared on earth. (At an earlier time in the history of the earth, the solar day is thought to have been 18 hours; with the change to the current 24-hour day occurring over millions of years, there was plenty of time for clock genes to be altered so the molecular circadian cycle stayed in synchrony with the solar day.) As detailed in

the rest of this chapter, the early linkage for the need to rest, and to rest at a specific phase of the daily external environment, may have resulted in the circadian clock and sleep-wake cycle evolving together and being integrated with one another at many different levels of organization.

INTEGRATION OF THE CIRCADIAN CLOCK AND SLEEP-WAKE SYSTEMS

Two chapters in this section deal with the circadian clock system, with a focus on the anatomy of the neural clock system in mammals (Chapter 33) and the physiology of the mammalian circadian clock system (Chapter 34). Chapters that review the molecular and genetic bases for the actual circadian clock core machinery, which has been highly conserved at least from insects to mammals, are included in Section 3 of this volume. Chapters 35 and 37 involve discussions from different vantage points about how the circadian clock system is highly integrated with the sleep-wake regulatory system.

Chapter 37 focuses on how the circadian clock and sleep loss together and independently regulate neurobehavioral performances. Whereas the disruption of circadian rhythms in humans during rapid travel across time zones and in shift workers has been linked to a variety of mental and physical abnormalities for many years (see Section 10), only in the last few years have animal models begun to emerge that are elucidating the extent of the importance of internal circadian synchronization for the health and well-being of the organism as well as for the molecular events that are disrupted at the level of cells and tissues (Chapters 38 and 39). Indeed, there is now considerable evidence that disruption of the circadian clock can exacerbate and perhaps cause a variety of mental and physical disorders, as reviewed in Chapter 39. The final chapter in this section (Chapter 40) reviews how the clock system underlies disorders related to the timing of the sleep-wake cycle that can lead to a disruption of the timing of internal rhythms with one another as well as of the timing of the sleep-wake cycle and other rhythms with the external environment.

In the now classic two-process model of Borbely and colleagues, the timing of sleep and wake is a function of a homeostatic process that defines sleep need as being dependent on the previous amount of sleep and wake (process S) and the circadian clock (process C) that modulates the timing and propensity of sleep (Chapter 36). However, as noted by Buxton and Czeisler, the interactions between the circadian pacemaker and sleep homeostat should not be underestimated, and there is great difficulty in separating these two processes functionally. In addition, recent genetic and anatomic findings (see Section 3 and Chapter 33) also tend to "blur" the distinction between the homeostatic and circadian inputs in the regulation of the sleep-wake cycle. It is noteworthy that at the 2014 European Sleep Research Society meeting in Tallinn, Estonia, there was a symposium entitled "Farewell to the Two-Process Model." Whereas the two-process model has guided and will continue to guide our understanding of sleep regulation mechanisms, there are many new findings that need to be taken into consideration regarding how the circadian, homeostatic, and perhaps other processes underlie the complex regulation of the sleep-wake cycle.

In the two-process model, the electroencephalographic slow wave activity in non–rapid eye movement (NREM) sleep serves as an indicator of sleep homeostasis either under baseline or after sleep deprivation conditions. Whereas there is substantial evidence to indicate that "sleep homeostasis," as defined by the slow wave activity during NREM sleep, is independent mechanistically from the circadian clock, it is not known how the two processes actually contribute to the overall "sleep need" of the organism or what role the circadian clock may play in other homeostatically regulated sleep-wake events, such as total sleep time and NREM or rapid eye movement (REM) sleep time under baseline or sleep deprivation conditions. Indeed, the finding that different alleles in the *per* gene in humans are associated with differences in slow wave sleep emphasize how closely linked the circadian clock and the homeostatic process are in regulating sleep (see Section 3).

REGULATING SLEEP AMOUNT: A HOMEOSTATIC AND A CIRCADIAN INPUT

Many sleep and wake traits are under homeostatic control. That is, the longer one is awake or deprived of specific sleep stages (e.g., REM sleep deprivation), the greater will be the drive to recover the lost sleep or sleep stage. However, this may be true only after short periods of sleep deprivation because during continuous chronic partial sleep deprivations in rodents, there is a loss of the homeostatic recovery sleep increase, as the sleep-wake system appears to change from a homeostatic to an allostatic response system.[4] In addition, the different phases of normal sleep and wake are under circadian control; the clock is doing more than saying wake up or go to sleep at specific times of day. An unanswered question is, What actually regulates the amount of sleep? Or put another way, What are the relative contributions of the homeostatic and circadian process to the actual amount of time a given animal is awake and interacting with the external world or asleep and avoiding the external world?

Studies in laboratory, zoo and, wild animals reveal that sleep times are unrelated to taxonomic classification, and as noted by Siegel,[5] "the range of sleep times of different primates extensively overlaps that of rodents which overlaps that of carnivores, and so on across many orders of mammals." The total sleep time is correlated in a global sense to body size; for example, the opossum sleeps 18 hours, the ferret sleeps 14.4 hours, the cat sleeps 12.5 hours, the dog sleeps 10.1 hours, humans sleep 8 hours, and the elephant sleeps 3 hours.[6] Although on a global level body size (and associated metabolic rate) is inversely related to total sleep time, there appear to be other factors regulating sleep time as indicated by the fact that the smaller mouse sleeps about the same amount as the larger rat.[6,7] Indeed, there is considerable variability in sleep time among strains of mice; strains of mice that are of similar size and of the same species can show total sleep time differences up to as much as 2.5 hours.[7]

In the evolution of life on earth, there has been great pressure for organisms to adapt their lifestyle to the external world and to coordinate the internal temporal environment so as to maximize the chances of survival and to pass on their genetic material to the next generation. Total sleep time could be expected to be part of the survival strategy to ensure that animals are engaged as well as disengaged with the external environment at the appropriate times of day and night. Similarly, as noted at the beginning of this chapter, the sleep-wake

cycle can be considered the master circadian rhythm, as the expression of so many behavioral and physiologic rhythms is tied to the sleep-wake/activity-rest cycle. Thus, the total amount of sleep and wake may have evolved in individual species, in part, to create an internal temporal framework, in conjunction with the circadian clock, to maximize survival and reproduction fitness.

The fact that the circadian clock is a major determiner of the pressure to sleep or the ability to stay awake has mechanistic indications on various neural, genetic, and molecular levels as well as potentially important therapeutic implications for treating sleep-wake disorders. The recent findings in humans and rodents that the circadian clock is regulating not only the timing of sleep and wake but also the propensity to sleep and wake render the circadian clock system a potential target for interventions that could promote sleep and wakefulness that go beyond interventions in use today.

CLINICAL PEARL

Our 24/7 society has led humans to be the only animal species that routinely ignores its biological clock as we are often awake when the clock is telling us to be asleep. Both the chronic disruption of circadian organization and chronic insufficient sleep have been associated with a wide range of mental and physical disorders. Modern medicine is only beginning to recognize that the treatment of many disorders of human health may need to take into account circadian medicine to improve overall 24-hour temporal organization between and within the central nervous system and peripheral tissues.

Selected Readings

Feillet C, van der Horst GT, et al. Coupling between the circadian clock and cell cycle oscillators: implication for healthy cells and malignant growth. *Front Neurol* 2015;**6**:96.

Fuhr L, Abreu M, Pett P, Relógio A. Circadian systems biology: when time matters. *Comput Struct Biotechnol J* 2015;**13**:417–26.

A complete reference list can be found online at ExpertConsult.com.

Anatomy of the Mammalian Circadian System

Joshua J. Gooley; Clifford B. Saper

Chapter 33

Chapter Highlights

- Circadian rhythms in mammals, including sleep-wake cycles, are endogenously driven by the suprachiasmatic nucleus (SCN) in the anterior hypothalamus. The SCN is a master clock that transmits circadian output signals to effector systems to temporally coordinate behavioral and physiologic rhythms with daily changes in the environment.
- The near-24-hour rhythm of neuronal activity in the SCN is normally entrained to the 24-hour light-dark cycle defined by Earth's rotation. Light information is transmitted to the SCN from specialized retinal ganglion cells that contain the photopigment melanopsin. On activation by light, retinal axons release glutamate and pituitary adenylate cyclase–activating polypeptide onto SCN neurons. Clock cells in the SCN also receive dense input from the intergeniculate leaflet in the thalamus and the median raphe nucleus, which play a modulatory role in regulating circadian rhythms.
- The primary neuronal pathway underlying the circadian regulation of sleep-wake cycles involves a dense SCN efferent projection to the adjacent subparaventricular zone, followed by a secondary projection to the dorsomedial hypothalamic nucleus, which projects to other brain regions critical for regulating sleep and wakefulness. The SCN also projects directly and indirectly to the paraventricular hypothalamic nucleus to regulate corticosteroid secretion and synthesis of the hormone melatonin. In addition to regulating circadian cycles of behavior and endocrine function, the SCN plays a hierarchical role in coordinating the timing and function of clocks located in peripheral tissues.
- Understanding the neuroanatomic basis of circadian rhythms is important for developing therapies to improve the quality of sleep-wake cycles, especially in persons with circadian rhythm sleep disorders. This chapter describes the intrinsic organization, inputs, and outputs of the circadian clock in the SCN, with emphasis on the neuronal circuitry and neurotransmitters underlying circadian control of sleep-wake and rest-activity cycles.

Most animals show a pronounced daily rhythm of rest-activity that is synchronized with the solar cycle. To a large degree, the timing of these rhythms is determined by the circadian ("about a day") system. Circadian rhythms persist even in the absence of periodic environmental cues but are normally entrained by the light-dark cycle and feeding schedules. The primary role of the circadian system is to ensure that behavioral and physiologic rhythms are temporally coordinated with daily changes in the environment. By providing an internal representation of day and night, the circadian system anticipates the rising and setting of the sun, hence allowing animals to time appropriately sleep and foraging behavior. The circadian system therefore facilitates adaptation to daily environmental cycles to maintain energy balance, which is thought to increase survival and reproductive fitness. This chapter reviews the organization of the circadian system and the anatomic basis for circadian regulation of sleep-wake cycles and endocrine function.

THE MASTER CIRCADIAN CLOCK IN THE SUPRACHIASMATIC NUCLEUS

The master circadian clock for behavioral rhythms, including the sleep-wake cycle, is located in the suprachiasmatic nucleus (SCN).[1] The SCN is situated in the anterior hypothalamus immediately dorsal to the optic chiasm and lateral to the third ventricle (Figure 33-1, *A*). The circadian rhythm of neural activity in the SCN is generated at the cellular level by a transcriptional-translational molecular feedback mechanism. If the molecular clock is rendered dysfunctional, patterns of rest-activity exhibit irregular and non–24-hour cycles. Similarly, lesions of the SCN or its efferent projections abolish behavioral and endocrine rhythms, demonstrating a critical role for the SCN clock in generating the circadian rhythm of sleep.

Based on neurotransmitter phenotype and afferent-efferent connections, the SCN can be divided into ventrolateral (SCNvl) and dorsomedial (SCNdm) components, commonly referred to as the *core* and *shell*, respectively.[2] The core contains many vasoactive intestinal polypeptide (VIP)-containing neurons (Figure 33-1, *B*), whereas the shell contains a large population of arginine vasopressin (AVP)-containing neurons (Figure 33-1, *C*). In studies of the SCN, the boundaries of the core and shell are often defined by the distribution of VIP and AVP immunoreactivity because these cell groups are conserved in the SCN of rodents, monkeys, and humans. By comparison, the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) is heavily expressed throughout

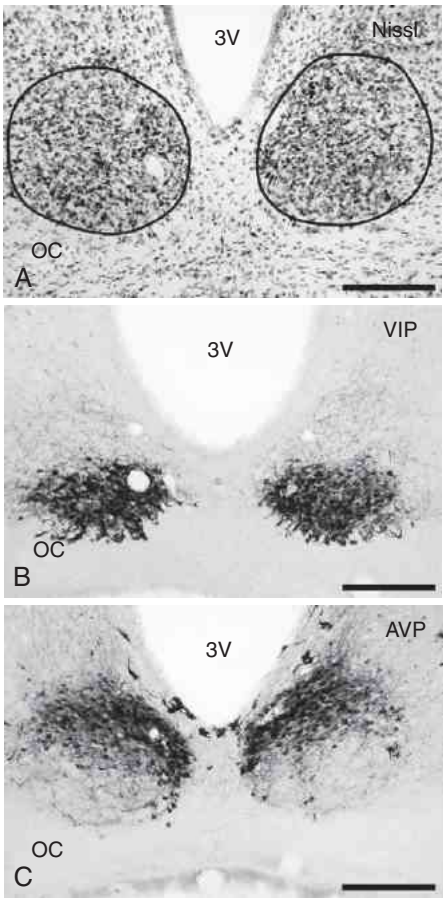


Figure 33-1 The suprachiasmatic nucleus (SCN) is composed of ventrolateral (SCNvl) and dorsomedial (SCNdm) subdivisions. **A,** In Nissl-stained coronal sections in rats, the SCN can be identified by its tightly compacted small-diameter neurons located immediately dorsal to the optic chiasm and lateral to the third ventricle. **B,** VIP-immunoreactive perikarya are found in the SCNvl. **C,** AVP-immunoreactive cell bodies are found in the SCNdm. 3V, Third ventricle; AVP, arginine vasopressin; OC, optic chiasm; VIP, vasoactive intestinal polypeptide. Scale bars equal 200 μm.

both SCN subdivisions, and it is contained in most, if not all, SCN neurons.[3] The core of the SCN in rodents also contains many neurons that express gastrin-releasing peptide (GRP) and smaller numbers expressing neurotensin (NT). In humans, an extensive population of NT-immunoreactive cells is found throughout the SCN,[4] and neuropeptide Y (NPY)-containing neurons are located predominantly in the central part of the nucleus, where they overlap with the distribution of VIP-immunoreactive cells. Smaller subpopulations of neurons containing angiotensin II, enkephalin (ENK), somatostatin, and substance P have also been described in the SCN, but there are species-specific differences in the abundance and distribution of these neurotransmitters.[4] Therefore, the human SCN is likely functionally homologous to the SCN of other mammals, but the relative contributions and combinations of SCN neurotransmitters that regulate circadian rhythms might vary across species.

SCN neurons are highly interconnected, which is presumably important for coupling individual cellular oscillators to produce a coherent neural activity rhythm. The SCN is functionally subdivided, however, such that the SCN core receives most of the retinal input (discussed in greater detail later). Retinal axons form synapses on SCN neurons containing VIP and GRP, and these cells are activated by light, as demonstrated by increased accumulation of the transcription factor c-Fos. VIP and GRP levels in the SCN show a diurnal rhythm in a light-dark cycle, but not in constant darkness, suggesting that the concentration of these neurotransmitters is regulated predominantly by light. VIP-immunoreactive neurons project heavily to both subdivisions within the SCN, consistent with the distribution of $VPAC_2$ receptor (also known as VIP receptor type 2), whereas GRP receptors are found primarily in the SCN shell. A functional role for VIP and GRP in transmitting photic information in the SCN is supported by studies demonstrating that microinjection of either VIP or GRP into the SCN region induces phase resetting of rest-activity rhythms similar to the effects of light.[5] VIP knockout mice exhibit normal diurnal locomotor rhythms in a light-dark cycle, but they show a reduction in the circadian amplitude of wheel-running activity in constant darkness.[6] As in VIP-deficient mice, $VPAC_2$-receptor knockout mice can synchronize their behavior to the light-dark cycle, probably owing to the masking effects of light on behavior, but they have profoundly weakened locomotor activity rhythms in constant darkness.[7] These findings suggest that VIP signaling in the SCN plays a role in coordinating the activity rhythm of SCN neurons.[8,9]

Despite the abundance of AVP and it receptors (V1a and V1b) in the SCN shell, AVP signaling is not required for generating the SCN rhythm or for output of rest-activity cycles. There is robust circadian gene expression of AVP in the SCN, but rats that are AVP deficient show normal behavioral circadian rhythms, and injection of AVP or V1 receptor antagonists into the SCN region does not alter the phase or period of locomotor activity.[10] Nonetheless, a role for AVP signaling in circadian behavior was recently demonstrated in double knockout mice for the V1a and V1b receptors. Mutant $V1a^{-/-}V1b^{-/-}$ mice are resistant to jet lag compared with wild-type mice and can rapidly adapt to a large shift in the timing of the light-dark cycle.[11] It is possible that AVP-mediated communication between SCN neurons limits responses of the clock to external perturbation, and hence large phase shifts are possible when AVP signaling is disrupted.

The output of SCN neurons is thought to be principally inhibitory, based on the observation that most clock cells contain GABA. Histologic and functional studies indicate that GABA acts on both ionotropic $GABA_A$ receptors and metabotropic $GABA_B$ receptors in the SCN. In hypothalamic brain slices, $GABA_A$ receptor agonists inhibit neuronal and metabolic activity of SCN neurons,[12,13] and the $GABA_B$ receptor agonist baclofen inhibits optic nerve–stimulated field potentials.[14] Consistent with these findings, light-induced phase shifts of locomotor activity are blocked by both $GABA_A$ and $GABA_B$ receptor agonists.[15] In dissociated cell culture, daily administration of GABA synchronizes SCN neuronal electrical rhythms,[16] suggesting that GABAergic signaling might also play a role in coordinating the circadian phase of individual SCN cellular oscillators.

SUPRACHIASMATIC NUCLEUS INPUTS

The endogenous circadian rhythm of neuronal activity in the SCN is close to, but not exactly, 24 hours. To entrain to the imposed solar day length, the SCN clock must therefore be reset daily by extrinsic time cues, most notably the light-dark cycle. The SCN rhythm can be further tuned by nonphotic

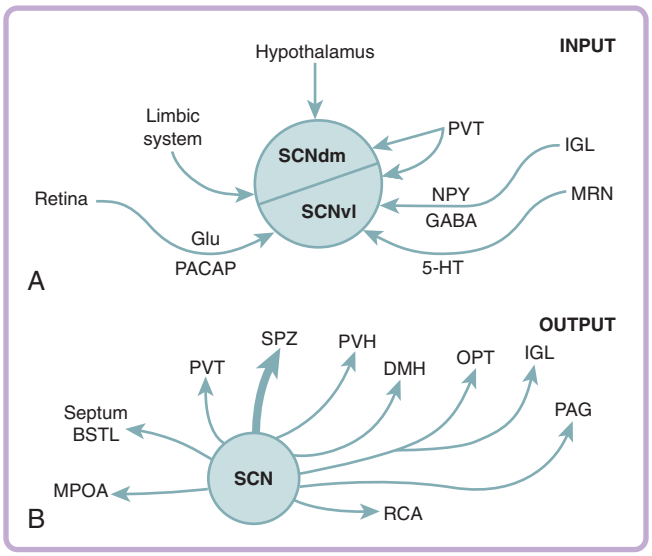


Figure 33-2 Afferent and efferent projections of the suprachiasmatic nucleus (SCN). **A,** SCN afferent projections terminate differentially in the SCNvl and SCNdm. Neurotransmitters contained in these projections are shown next to the *arrows*. **B,** SCN efferent projections are primarily confined to the hypothalamus. 5-HT, 5-hydroxytryptamine (serotonin); BSTL, bed nucleus of the stria terminalis; DMH, dorsomedial hypothalamic nucleus; GABA, gamma-aminobutyric acid; Glu, glutamate; IGL, intergeniculate leaflet; MPOA, medial preoptic area; MRN, median raphe nucleus; NPY, neuropeptide Y; OPT, olivary pretectal nucleus; PACAP, pituitary adenylate cyclase–activating polypeptide; PAG, periaqueductal gray matter; PVH, paraventricular hypothalamic nucleus; PVT, paraventricular thalamic nucleus; RCA, retrochiasmatic area; SCNdm, dorsomedial suprachiasmatic nucleus (shell); SCNvl, ventrolateral suprachiasmatic nucleus (core); SPZ, subparaventricular zone.

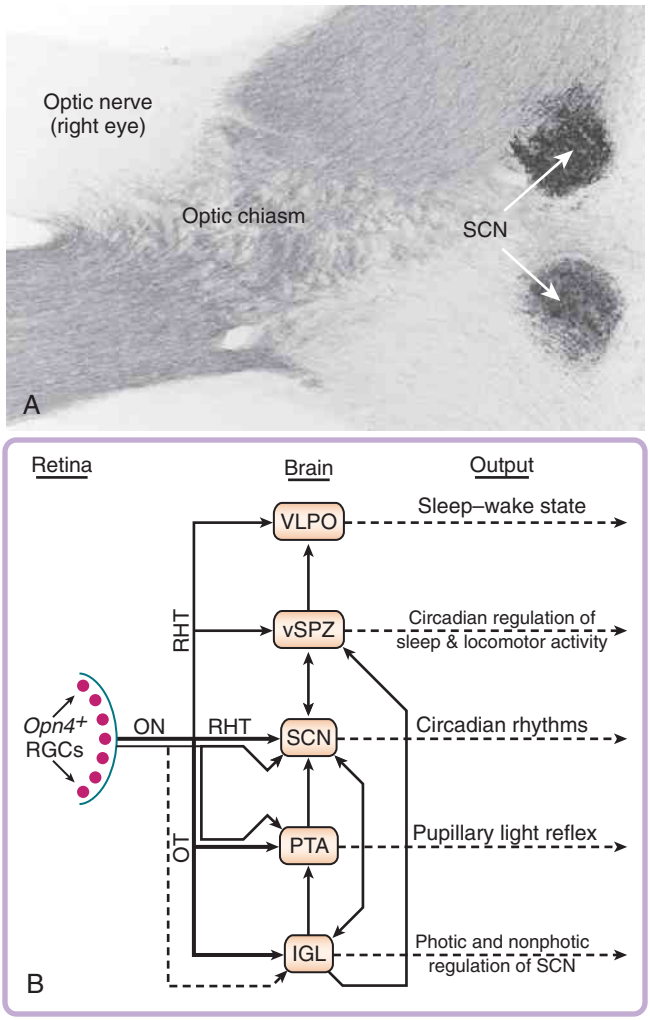


Figure 33-3 Retinal input to the circadian timing system from melanopsin-containing retinal ganglion cells. **A,** Following intraocular injection of cholera toxin B subunit into the left eye, anterogradely labeled axons project bilaterally to the SCN of rats, as shown in a horizontal section through the optic chiasm. **B,** Melanopsin-containing retinal ganglion cells project to brain areas involved in processing nonvisual information, including the SCN and IGL. The *branched solid arrow* to the SCN and PTA indicates collateralized projections, and the *branched dashed arrow* to the SCN and IGL indicates proposed axon collaterals. *Long dashed arrows* indicate physiologic and behavioral outputs of the targeted retinorecipient brain areas. Direct projections between brain areas are shown, but indirect projections are not shown for reasons of clarity. IGL, intergeniculate leaflet; Opn4$^+$ RGCs, melanopsin-positive retinal ganglion cells; ON, optic nerve; OT, optic tract; PTA, pretectal area; RHT, retinohypothalamic tract; SCN, suprachiasmatic nucleus; VLPO, ventrolateral preoptic nucleus; vSPZ, ventral subparaventricular zone. (**B** from Gooley JJ, Lu J, Fischer D, et al. A broad role for melanopsin in nonvisual photoreception. *J Neurosci* 2003;23:7093–7106. Copyright 2003 by the Society for Neuroscience.)

inputs to coordinate behavioral and physiologic patterns with diurnal changes in the environment. The SCN core receives dense input from the retina, intergeniculate leaflet, and midbrain raphe nuclei,[17] whereas the SCN shell primarily receives projections from other hypothalamic areas, basal forebrain, limbic cortex, septal area, and brainstem (Figure 33-2, *A*).[18] Below, we discuss some of the major pathways that regulate the timing of the SCN neural activity rhythm.

Retina

The retinohypothalamic tract (RHT) projects bilaterally to the SCN, with a slight to moderate contralateral predominance in most species (Figure 33-3, *A*). The RHT sends its densest projection to the SCN core where VIP-immunoreactive cells are located, but smaller numbers of RHT axons also reach the SCN shell, subparaventricular zone (SPZ), and other hypothalamic areas. Retinal ganglion cells (RGCs) at the origin of the RHT contain the photopigment melanopsin (OPN4),[19,20] which is preferentially sensitive to short-wavelength (blue) light. Although OPN4-containing RGCs are intrinsically photosensitive, they also receive indirect light information from rod and cone photoreceptors in the outer retina.[21,22] In the absence of rod and cone function, melanopsin-containing RGCs are sufficient to mediate circadian responses to light, as demonstrated in blind humans and sightless mice with intact function of the inner retina.[23,24] Also, in normally sighted individuals, phase shifting and melatonin suppression are most sensitive to short-wavelength light,[25,26] suggesting that melanopsin plays an important role in circadian photoreception. OPN4-deficient mice with intact rod-cone function show moderate deficits in circadian light resetting, but these animals can still entrain to light-dark cycles, suggesting that visual photoreceptors and melanopsin play overlapping roles in circadian photoreception. Consistent with these findings, spectral responses of the circadian system in humans suggest involvement of rod-cone photoreceptors and OPN4, although the relative contribution of these photoreceptor types appears to depend on the irradiance and duration of exposure to light.[27] Based on studies in mice, circadian responses to light are only eliminated when rod-cone and OPN4 signaling pathways are disrupted

simultaneously, or if OPN4-containing RGCs are genetically ablated.[28,29] Hence RGCs that express *Opn4* appear to function as a necessary conduit for light information to reach the circadian clock in the SCN. The OPN4-containing RGCs project not only to the SCN but also to other brain areas that are involved in nonvisual photoreception, including the ventrolateral preoptic nucleus (VLPO), which is a sleep-promoting area in the anterior hypothalamus; the SPZ, which is important for circadian regulation of sleep (discussed later); the olivary pretectal nucleus (OPT), which is part of the afferent pathway mediating the pupillary light reflex; and the intergeniculate leaflet (IGL), which plays a role in modulating circadian rhythms (see Figure 33-3, *B*).[30,31] Each of these retinorecipient brain areas is interconnected with the SCN, suggesting that photic input from OPN4-containing RGCs can modulate sleep and circadian rhythms via multiple routes.

Several lines of evidence suggest that photic information is conveyed to the SCN through the release of excitatory amino acids such as glutamate from RHT terminals. For example, optic nerve stimulation induces the release of glutamate and aspartate in the SCN in brain slice preparations,[32] and glutamate mimics the effects of optic nerve stimulation on the circadian rhythm of SCN neuronal activity in vitro.[33] Microinjection of *N*-methyl-d-aspartate (NMDA) into the SCN region in vivo also mimics the phase-shifting effects of light on rest-activity rhythms.[34] Additionally, receptor subunits from the NMDA-, α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionic acid (AMPA)–, and kainate-preferring classes of ionotropic glutamate receptors are found in the SCN, in addition to subunits from metabotropic glutamate receptors.[35]

OPN4-containing RGCs that give rise to the RHT also express pituitary adenylate cyclase–activating polypeptide (PACAP),[36] which colocalizes with glutamate at RHT terminals. PACAP binds the PACAP-type 1 (PAC_1) receptor and $VPAC_2$ receptor with equal affinity, both of which are expressed in the SCN. Mice deficient in PACAP or PAC_1 receptors show abnormal circadian resetting but can nonetheless entrain to light-dark cycles,[37] suggesting that PACAP signaling plays a partially redundant role with glutamate in photic circadian entrainment.

Light activation of SCN neurons triggers Ca^{2+} influx and activation of intracellular signaling cascades,[38] leading to increased expression of *Period* genes, which is thought to be the molecular mechanism by which light resets circadian rhythms in mammals.[39] Circadian resetting and expression of c-Fos protein are likely regulated in part by the Ca^{2+}-binding proteins calbindin-D28K (CB) and calretinin (CAL). In hamsters, RHT axons form synapses on CB-containing neurons in the SCN core, and light induces c-Fos in these cells.[40] Targeted disruption of the *Calb1* gene in mice reduces the amplitude of circadian behavior[41] and alters circadian resetting responses and clock gene expression.[42] CB- and CAL-immunoreactive neurons have also been described in the SCN of monkeys and humans,[43,44] suggesting involvement of Ca^{2+}-binding proteins in the regulation of circadian behavior in primates.

Intergeniculate Leaflet

In rodents, the IGL is a thin cellular layer sandwiched between the dorsal and ventral subdivisions of the lateral geniculate nucleus (LGN) of the thalamus. The IGL projects heavily to the SCN core through the geniculohypothalamic tract and also sends efferents to hypothalamic areas adjacent to the SCN, including the SPZ and retrochiasmatic area (RCA).[45] Additionally, the IGL sends axons to the dorsomedial hypothalamic nucleus, midline thalamus, pretectum, periaqueductal gray matter, superior colliculus, and lateral and dorsal terminal nuclei of the accessory optic system. The IGL receives dense bilateral projections from OPN4-containing RGCs, the IGL itself, and the RCA, as well as smaller projections from the SCN, locus coeruleus, midbrain raphe nuclei, and brainstem cholinergic nuclei. Unlike the dorsal LGN, which is thought to function primarily as a relay in image-formation visual processes, the IGL has a well-established role in contributing to circadian rhythm regulation.[46] Lesions of the IGL have been reported to result in a broad array of circadian phenotypes, including a slower rate of reentrainment following a shift in the photoperiod, a block in the lengthening of circadian period by constant light, and elimination of circadian resetting induced by introducing a running wheel into the animal's cage.[47] Based on these findings, the IGL is thought to transmit both photic and nonphotic information to the circadian clock. Animals with lesions placed in the region of the IGL are clearly able to entrain to light-dark cycles, however, indicating that the IGL is not necessary for photic circadian entrainment and rhythm generation, but instead plays a modulatory role.

Most IGL neurons contain the inhibitory neurotransmitter GABA, and the IGL is characterized by large populations of NPY- and ENK-containing neurons.[48] In rodents, NPY-containing neurons in the IGL project heavily to the SCN core region and form synapses on VIP-containing neurons. It is currently unclear whether there is an IGL homologue in the human brain, but a population of NPY-containing neurons has been identified in the pregeniculate nucleus bordering the LGN.[49] There is a dense NPY-immunoreactive terminal plexus in the human SCN, but it has yet to be determined whether this originates solely from NPY-containing cells within the boundaries of the nucleus or whether there is an additional source of input from neurons in the pregeniculate nucleus.

In rodents, the effects of NPY and photic stimuli on the circadian pacemaker appear to be mutually inhibitory. Consistent with the effects of NPY on SCN electrical activity in hypothalamic slices, microinjection of NPY into the SCN region induces a phase advance shift in locomotor activity when administered during the biologic daytime, and NPY-induced phase shifts are attenuated by glutamate or light.[47] Conversely, glutamate-induced phase shifts in SCN electrical activity or light-induced phase shifts in locomotor activity are inhibited by NPY. Pharmacologic studies suggest that NPY attenuates light-induced phase shifts through Y1 or Y5 receptors,[50] whereas the Y2 receptor mediates phase resetting of the locomotor activity rhythm.[51] In contrast to Y1/Y5 receptors, however, Y2 receptors have yet to be detected in the SCN using in situ hybridization or immunohistochemistry techniques. Hence additional studies are needed to determine the mechanisms by which NPY induces phase shifts of SCN neural activity.

Midbrain Raphe Nuclei

The SCN receives a dense serotoninergic input from neurons in the median raphe nucleus (MRN) and relatively sparse

input from neurons in the dorsal raphe nucleus (DRN).[52] Serotonin (5-HT)-immunoreactive axonal terminals in the SCN overlap extensively with the terminal fields of the RHT and geniculohypothalamic tract, forming synapses with VIP-containing neurons. In contrast to the MRN, the DRN projects to the IGL and to the SPZ immediately dorsal to the SCN, with few fibers to the SCN itself. The midbrain raphe nuclei are not required for photic circadian entrainment or for maintaining circadian rhythmicity, but 5-HT modulates photic and nonphotic input to the circadian clock. Lesions of the midbrain raphe nuclei have been reported to phase-advance the locomotor activity onset, lengthen the activity phase, induce deficits in rhythmicity in constant light, and reduce the amplitude or precision of circadian rhythms.[53]

Similar to NPY, the effects of 5-HT and photic stimuli on the circadian pacemaker appear to be mutually inhibitory. Phase shifts induced by light, optic nerve stimulation, or glutamate are attenuated by administration of 5-HT during the biologic night, whereas 5-HT-induced phase advance shifts are inhibited by glutamate agonists or optic chiasm stimulation.[54] The 5-HT_{1B} receptor mediates inhibitory effects of 5-HT on light-induced phase shifts,[55] and electron microscopic studies have shown that 5-HT_{1B} receptor immunoreactivity is predominantly found in presynaptic terminals in the SCN, including axonal terminals from the RHT.[56] Consistent with pharmacologic studies conducted in vitro, 5-HT_7 agonists induce phase advances in hamster behavioral circadian rhythms when administered during the biologic daytime,[57,58] and the 5-HT_7 receptor has been localized to dendrites of GABA-,VIP-, and AVP-containing SCN neurons and also to presynaptic axonal terminals.[56]

SUPRACHIASMATIC NUCLEUS OUTPUTS

Considering the important role that the SCN plays in regulating circadian behavior and physiology, the number and density of SCN efferent pathways are surprisingly limited (see Figure 33-2, *B*).[59,60] The SCN projects rostrally to limbic structures, including the lateral septum and bed nucleus of the stria terminalis (BSTL), and to the medial preoptic area (MPOA) and VLPO. The SCN sends projections dorsally to the midline thalamus, including the paraventricular thalamic nucleus (PVT) and paratenial thalamic nuclei, and dorsocaudally to the SPZ, paraventricular hypothalamic nucleus (PVH), and dorsomedial hypothalamic nucleus (DMH). A caudal projection from the SCN terminates in the RCA, and some fibers branch toward the supraoptic nucleus and the lateral hypothalamic area (LHA). The SCN also sends a minor projection laterally to the IGL and small posterior projections to the OPT and central gray matter. Based on retrograde tracing studies in rats, the SCN shell projects more strongly to the MPOA, DMH, and BSTL, whereas the SCN core projects more strongly to the lateral septum and tuberal hypothalamus.[59,61] Both SCN subdivisions project strongly to the SPZ and midline thalamus. In monkeys, VIP-immunoreactive axons from the SCN project rostrally into the MPOA and dorsally into the SPZ and anterior hypothalamic area ventral to the PVH, and small numbers of terminals are also observed in the PVH, LHA, and PVT. A dense VIP-containing projection extends caudally into the RCA, and some fibers continue into the capsule of the ventromedial hypothalamic nucleus, DMH, and dorsal hypothalamic area. In humans, VIP-containing SCN neurons project most heavily into the region just dorsal to the SCN and extending to the area just ventral to the PVH, corresponding to the SPZ in rodents.[4] Similar to that in monkeys, the SCN in humans sends a dense VIP-immunoreactive projection caudally into the RCA. Together, these studies demonstrate that the SCN projects extensively to other hypothalamic regions, and the pattern of projections is generally conserved across mammalian species.

Sleep-Wake Rhythm

The densest output from the SCN is to the SPZ, which is located immediately dorsal to the SCN and extends dorsocaudally ventral to the PVH.[60] Tracing studies have shown that the SCN core projects more strongly to the lateral SPZ, whereas the SCN shell projects more densely to the medial SPZ.[61] Consistent with these results, AVP-immunoreactive axonal terminals are found medially in the SPZ, whereas VIP-containing terminals are found more laterally. The SPZ innervates similar targets to the SCN but in much greater density, suggesting that the SPZ amplifies circadian output from the SCN. Cell-specific lesions of the SPZ eliminate circadian rhythms of sleep, locomotor activity, and core body temperature, suggesting that the SPZ is part of the primary neuronal pathway mediating the output of SCN-generated circadian rhythms.[62] Lesions in the ventral part of the SPZ abolish circadian rhythms of sleep and locomotor activity while having lesser effects on the body temperature rhythm, whereas lesions in the dorsal SPZ reduce the body temperature rhythm while having minimal effects on rhythms of wake-sleep or locomotor activity. The ventral SPZ, in turn, sends a strong projection caudally to the DMH. Cell-specific lesions of the DMH eliminate circadian rhythms of sleep—wake, locomotor activity, feeding, and plasma corticosteroids but not body temperature or plasma melatonin.[63] Hence, the major neuronal pathway mediating the SCN-generated circadian rhythm of sleep is through a first-order projection to the SPZ, followed by a second-order projection to the DMH.

As might be expected, the DMH projects heavily to brain areas involved in regulating sleep and wakefulness. For example, the DMH sends a primarily GABAergic projection to the VLPO. The VLPO contains the neurotransmitters GABA and galanin and is thought to promote sleep through its inhibitory projections to the ascending arousal systems.[64,65] Although the SCN and SPZ also send minor projections directly to the VLPO, the projection originating from the DMH is one of the largest inputs to this area.[66] The DMH also sends a primarily glutamatergic projection to the lateral hypothalamus, which contains wake-promoting neurons, including orexin-expressing neurons.[63] In summary, the DMH receives circadian input from the SCN and the SPZ and projects to the VLPO and LHA, defining a putative pathway for the circadian regulation of sleep and wakefulness (Figure 33-4). The relay of SCN circadian signals in the SPZ and then the DMH might allow for modification of circadian rhythms by other inputs such as food availability, external temperature, or social cues.

Consistent with this hypothesis, cell-specific lesions of the DMH impair the ability of rats to synchronize their sleep-wake, locomotor activity, and body temperature rhythms with restricted daytime feeding.[67] Also, restoring clock gene expression in the DMH of *Bmal1* null mice has been shown to rescue circadian adaptation to restricted feeding.[68] Two later

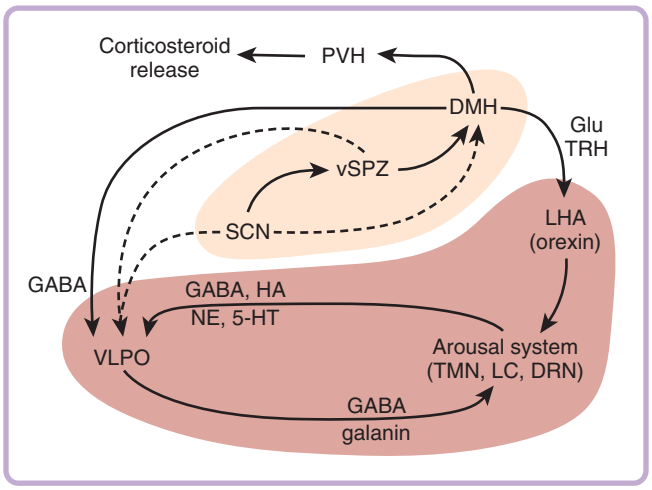


Figure 33-4 Circadian regulation of sleep-wake. The neuronal pathway that regulates circadian sleep-wake cycles (SCN to vSPZ to DMH, indicated in *beige*) directly interacts with arousal- and sleep-promoting brain regions (indicated in *light red*), providing a putative pathway for the circadian regulation of sleep-wake. *Solid arrows* indicate prominent neuronal projections, and *dashed arrows* indicate relatively small projections. Neurotransmitters contained in these projections are shown next to the *arrows*. DMH, dorsomedial hypothalamic nucleus; DRN, dorsal raphe nucleus; GABA, gamma-aminobutyric acid; Glu, glutamate; HA, histamine; 5-HT, 5-hydroxytryptamine (serotonin); LC, locus coeruleus; LHA, lateral hypothalamic area; NE, norepinephrine; PVH, paraventricular hypothalamic nucleus; SCN, suprachiasmatic nucleus; TMN, tuberomammillary nucleus; TRH, thyrotropin-releasing hormone; VLPO, ventrolateral preoptic nucleus; vSPZ, ventral subparaventricular zone.

studies questioned the necessity of clock genes for food entrainment by showing that wheel-running in *Bmal1* or *Per1/2* null mice anticipated meal times when the duration of daily food availability was decreased gradually over the course of days.[69,70] However, food anticipatory activity was not clearly demonstrated at the end of the experiments during continual food deprivation in the absence of light-dark cues,[69] so it could not be concluded that the activity pattern under those conditions was due to a self-sustained circadian clock mechanism. Notably, mice with genetic ablation of orexin neurons show blunted behavioral adaptation to restricted feeding cycles,[71] similar to DMH-lesioned rats, suggesting that the outflow pathway from the DMH to orexin neurons might be important for integrating photic and nonphotic information in order to establish a pattern of sleep-wake behavior that is most adaptive for survival.

Circadian Regulation of Melatonin and Cortisol

Under normally entrained conditions, synthesis of melatonin usually begins a few hours before a person's bedtime. Plasma melatonin levels are highest in the middle of the night and then decrease before usual wake time. The SCN is required for the circadian synthesis and release of melatonin in the pineal gland, which is regulated through a circuitous pathway.[72] A subset of SCN neurons projects directly to dorsal parvocellular neurons in the autonomic subdivision of the PVH. These PVH neurons send an excitatory projection to the sympathetic preganglionic neurons in the intermediolateral cell column in the upper thoracic spinal cord. The intermediolateral cell column sends a cholinergic projection to the superior cervical ganglion, and the superior cervical ganglion postganglionic sympathetic neurons send a noradrenergic projection to the pineal gland. The release of noradrenalin activates α- and β-adrenergic receptors in the pineal gland that stimulate the production of melatonin by controlling the enzyme serotonin *N*-acetyltransferase.

The SCN is most active during the biologic daytime and inhibits tonic activity of PVH neurons. This results in a lower firing rate in the remainder of the pathway, effectively inhibiting melatonin synthesis. Therefore melatonin is produced during the biologic night, when SCN activity is low, and this is a common feature of both diurnal and nocturnal animals. Exposure to light during the biologic night activates SCN neurons through the RHT, resulting in the inhibition of melatonin synthesis through the aforementioned pathway. Similar to circadian light resetting of the melatonin rhythm, light-induced suppression of melatonin synthesis is mediated by OPN4-containing RGCs.[29]

Melatonin can also provide direct feedback to the circadian system via melatonin receptors (MT1 and MT2) located in the SCN. Melatonin acutely inhibits SCN electrical activity, and this response is mediated by the MT1 receptor.[73] The locomotor activity rhythm of rats can be entrained by periodic administration of melatonin as long as the SCN remains intact,[74] suggesting that melatonin receptor signaling in the SCN is required for melatonin-induced entrainment. Melatonin-induced phase shifts persist in MT1 or MT2 receptor knockout mice, suggesting that melatonin receptor subtypes play an overlapping role in this response.[73,75] In humans, daily administration of melatonin can entrain the circadian system of blind persons,[76,77] indicating that melatonin elicits phase shifts of the human circadian clock and can be used to treat circadian rhythm sleep disorders even in the absence of photoreception.

Early lesion studies of the SCN showed loss of the circadian rhythm of adrenal corticosterone, the major glucocorticoid in rodents. In humans, the circadian rhythm of cortisol rises sharply before the waking period, presumably to promote a general state of readiness in anticipation of the myriad stressors and metabolic demands associated with the active phase. SCN neurons project directly and indirectly to the DMH, which sends efferents to the PVH and is critical for expression of the corticosteroid rhythm.[63] Corticotropin-releasing hormone–containing neurons in the PVH project to the median eminence, where corticotropin-releasing hormone is released into the portal circulation and activates the release of adrenocorticotropic hormone from the anterior pituitary gland. The SCN-driven release of adrenocorticotropic hormone into the blood results in rhythmic induction of corticosteroid secretion from the adrenal gland.

Diffusible Suprachiasmatic Nucleus Output Signals

In SCN-lesioned hosts, transplantation of fetal SCN grafts into the third ventricle restores a low-amplitude circadian rest-activity rhythm with the period of the donor animal.[78] In contrast, photic entrainment, reproductive responses to photoperiod, estrous cycles, and corticosteroid and melatonin rhythms are not restored by SCN transplants. Polymer-encapsulated SCN transplants that prevent neuronal communication between host and donor tissue reinstate low levels

of locomotor activity rhythms in SCN-lesioned hosts, suggesting that a diffusible factor can partially reconstitute the rest-activity cycle.[79] The distance of SCN transplants from the normal site of the SCN is an important factor for the recovery of the locomotor rhythm, indicating that a diffusible factor acts locally in a paracrine fashion.

Candidate diffusible mediators have been identified in the SCN that might regulate the output of circadian behavioral rhythms, including transforming growth factor-α, cardiotrophin-like cytokine, and prokineticin 2 (PK2). PK2 is a clock-controlled gene that is rhythmically expressed in the SCN, with higher expression levels observed during the biologic daytime. Intracerebroventricular injection of PK2 inhibits locomotor activity during the biologic night in rats (the active phase),[80] and PK2 null mice show attenuated circadian rhythms of sleep-wake, body temperature, and glucocorticoids.[81] Consistent with a role for PK2 as an SCN output factor, PK2 receptor has been described in many SCN output regions, including the PVH, DMH, PVT, paratenial nucleus, lateral septum, and the SCN itself.[80] Additionally, PK2 receptor knockout mice show reduced circadian expression of rest-activity and body temperature rhythms.[82]

SYNCHRONIZATION OF CENTRAL AND PERIPHERAL OSCILLATORS

The circadian system is thought to be hierarchically organized. The SCN contains a master clock for regulating behavioral rhythms, but circadian clocks are also found in tissues throughout the body. The molecular clock mechanism is preserved across different cell types, and circadian expression of clock genes has been demonstrated in nearly all tissues studied, as well as in many sites in the brain.[83] Notably, damage to SCN neurons causes peripheral clocks to fall out of synchrony, even though individual cells in peripheral tissues can continue to generate circadian rhythms of gene transcription.[84] Under most conditions, the SCN entrains peripheral clocks to ensure coordinated changes in physiology that are appropriately timed to the rest-activity cycle. The pathways by which the SCN entrains clocks in peripheral tissues is not fully understood, but there is evidence that shifts in temperature can reset peripheral, but not SCN, clocks.[85] Thus by controlling the daily cycle of body temperature, the SCN would reset clocks throughout the body and keep them in synchrony. Other aspects of physiology that are directly regulated by the SCN may also contribute to tissue circadian phase. For example, glucocorticoid signaling activates clock gene expression and influences the rate of entrainment of peripheral clocks to feeding cycles.[86,87] The phase of peripheral clocks can become uncoupled from the SCN pacemaker in animals that are given daily restricted access to food, as shown for circadian gene expression in tissues such as the liver, kidney, and heart.[88,89] Entrainment of peripheral clocks by feeding cycles can potentially occur through nutrient-sensing pathways because proteins involved in metabolism and energy balance (e.g., sirtuin 1 and AMP-activated protein kinase) interact with core clock proteins and affect their function.[90] Although it is well established that the circadian system plays an important role in regulating metabolism, recent work suggests that the dietary composition and timing of meals can also affect the molecular clock and clock-controlled metabolic function.[91] This has potential clinical implications for shift workers and other persons who regularly consume meals during the biologic night.

CLINICAL PEARL

The suprachiasmatic nucleus is an internal body clock that allows daily scheduling of wake-sleep, feeding, corticosteroid secretion, and other bodily functions during a normal light-dark cycle. This cycle can be modified by light exposure or by administration of melatonin, which resets the body clock, but feeding schedules can also reset the daily activity cycle by acting on circuits that are downstream of the clock. Thus in patients with difficulty adjusting to shift work, jet lag, or other disruptions of their schedules, in addition to adjusting light exposure and taking melatonin, it is usually possible to adjust to a new schedule more quickly if the meal times, social interactions, and sleep-activity schedule of the new environment are adopted as soon as possible.

SUMMARY

The circadian system plays an integral role in coordinating behavioral and physiologic rhythms. Under normally entrained conditions, exposure to light activates OPN4-containing RGCs, which transmit light information to the core region of the SCN by releasing glutamate and PACAP. The effects of light on SCN neuronal activity are modulated by input from NPY- and GABA-containing neurons in the IGL and serotoninergic input from the MRN. The solar cycle entrains the master clock in the SCN, which is responsible for generating circadian patterns of behavior. The SCN regulates the circadian sleep-wake rhythm through a primary projection to the SPZ, followed by a secondary projection to the DMH. The DMH, in turn, projects to brain areas critical for promoting sleep or wakefulness. A direct projection from the SCN to the PVH mediates rhythmic control of melatonin secretion from the pineal gland, and an indirect projection from the SCN to the PVH through the DMH is critical for the circadian release of corticosteroids. Although SCN neuronal projections are required for photic entrainment and circadian control of endocrine rhythms, SCN-diffusible factors are sufficient to support a weak circadian rest-activity rhythm. The SCN also coordinates the timing of peripheral clocks, which respond to SCN-directed circadian changes in temperature, glucocorticoids, and nutrient-sensing pathways. Through the aforementioned pathways, the circadian system ensures that sleep and other biologic rhythms are timed appropriately with respect to daily changes in the environment.

Selected Readings

Berson DM, Dunn FA, Takao M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* 2002;**295**:1070–3.

Buhr ED, Yoo SH, Takahashi JS. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* 2010;**330**:379–85.

Chou TC, Scammell TE, Gooley JJ, et al. Critical role of dorsomedial hypothalamic nucleus in a wide range of behavioral circadian rhythms. *J Neurosci* 2003;**23**:10691–702.

Jones JR, Tackenberg MC, McMahon DG. Manipulating circadian clock neuron firing rate resets molecular circadian rhythms and behavior. *Nat Neurosci* 2015;**18**:373–5.

Leak RK, Moore RY. Topographic organization of suprachiasmatic nucleus projection neurons. *J Comp Neurol* 2001;**433**:312–34.

Moore RY. The suprachiasmatic nucleus and the circadian timing system. *Prog Mol Biol Transl Sci* 2013;**119**:1–28.
Moore RY, Weis R, Moga MM. Efferent projections of the intergeniculate leaflet and the ventral lateral geniculate nucleus in the rat. *J Comp Neurol* 2000;**420**:398–418.
Morin LP. Neuroanatomy of the extended circadian rhythm system. *Exp Neurol* 2013;**243**:4–20.
Morin LP, Allen CN. The circadian visual system, 2005. *Brain Res Rev* 2006;**51**:1–60.
Saper CB, Lu J, Chou TC, Gooley J. The hypothalamic integrator for circadian rhythms. *Trends Neurosci* 2005;**28**:152–7.
Wang JL, Lim AS, Chiang W-Y, et al. Suprachiasmatic neuron numbers and rest-activity circadian rhythms in older humans. *Ann Neurol* 2015; [in press].

A complete reference list can be found online at ExpertConsult.com.

Physiology of the Mammalian Circadian System

Alan M. Rosenwasser; Fred W. Turek

Chapter 34

Chapter Highlights

- The bilaterally paired suprachiasmatic nuclei (SCNs) of the anterior hypothalamus were first identified in the early 1970s as critical for the normal expression of circadian rhythms, and subsequent research employing a wide variety of approaches has firmly established a role for the SCN as the master circadian pacemaker in mammals. Substantial effort has been exerted to elucidate the intrinsic anatomic organization of the SCN and to map its major afferent and efferent pathways.
- The circadian pacemaker is synchronized ("entrained") by environmental light-dark cycles and other periodic events. Light entrainment depends on a subset of intrinsically photosensitive retinal ganglion cells that express the novel photopigment, melanopsin, and that give rise to the retinal-hypothalamic tract, terminating in the SCN and certain other "nonvisual" brain areas.
- The molecular basis of circadian rhythm generation has been revealed to involve transcription-translation feedback loops comprising a number of circadian "clock genes" and their protein products. Although rhythmic clock gene expression was first identified in SCN tissue, subsequent work has revealed similar molecular clocks in numerous cells and tissues throughout the body.
- The mammalian circadian system is now seen as hierarchically organized and widely distributed, such that the master pacemaker in the SCN synchronizes multiple "downstream" circadian clocks distributed throughout the brain and body, thereby maintaining system-wide temporal synchrony. Failure of internal synchrony appears to be both a cause and an effect of disease-related pathophysiology. Thus the balance between health and disease is strongly dependent on the proper synchrony within and between central and peripheral oscillating systems.

A great deal is now known about the mammalian circadian timing system at the molecular, cellular, neural systems, and behavioral levels. The basic neuroanatomy, neurochemistry, and molecular neurobiology of the circadian pacemaker, its synchronization (entrainment) to the external environment, and its role within the multioscillatory circadian timing system have now been elucidated. Despite the fact that the subtle details of these processes continue to be elucidated, much of the current work on the physiology of the circadian system concerns how the master circadian pacemaker in the hypothalamus receives inputs from the internal and external environments, how the pacemaker interacts with a multitude of peripheral clocks to regulate the diverse spectrum of rhythmic processes, and how these functions contribute to health and disease.

THE SUPRACHIASMATIC NUCLEUS: MASTER CIRCADIAN PACEMAKER

Pacemaker Function of the Suprachiasmatic Nucleus

The realization that molecular circadian clocks exist in cells and tissues throughout the brain and body has revolutionized our understanding of the overall organization of the mammalian circadian timing system. Nevertheless, it remains clear that the hypothalamic suprachiasmatic nucleus (SCN) is the site of a "master" circadian pacemaker, responsible for regulating—directly or indirectly—most, if not all, circadian rhythms.[1-3] Over the years, a variety of studies involving SCN lesions, in vivo and in vitro recordings of SCN neural activity, functional metabolic mapping, fetal tissue transplantation, and molecular rhythm analysis have revealed that the SCN not only is capable of sustained rhythmic activity, both in vitro and in vivo, but also is responsible for maintaining coherent circadian rhythmicity in central and peripheral tissues, as well as in a wide range of physiologic and behavioral processes.[4,5]

Although early mathematical models raised the possibility that the circadian pacemaker could be constructed from an ensemble of coupled, high-frequency (i.e., noncircadian) oscillatory units, it is now clear that the generation of circadian signals by the SCN is fundamentally a cellular process. Studies using a variety of in vitro models, including long-term SCN cell culture,[6] simultaneous recording of multiple single units using multielectrode plates,[7,8] and optical monitoring of calcium flux[9] or gene expression[10] in individual SCN neurons, have now provided compelling evidence that circadian oscillation is indeed a cell-autonomous process expressed in many, but probably not all, individual SCN neurons. Nevertheless, this multitude of cellular circadian oscillators normally

interact to produce coherent circadian patterns of behavior.[11] Although the mechanisms underlying intra-SCN oscillator coupling have not been identified completely, early studies using tetrodotoxin suggested that individual neuronal oscillators in the SCN apparently remain synchronized even in the absence of sodium-dependent action potentials, both in vivo and in SCN slice culture.[12,13] These findings led to the suggestion that gap junctions, glial coupling, calcium-dependent signaling, or local diffusible signals might be responsible for maintaining intercellular oscillator synchrony in the SCN.[14,15] On the other hand, more recent studies using tetrodotoxin have found that blocking of action potentials in SCN slices or in dispersed cell cultures can disrupt intercellular phase synchrony,[16,17] reviving interest in the possible role of synaptic signaling in oscillator coupling. In support of this hypothesis, in vitro and in vivo evidence has accumulated to indicate that both gamma-aminobutyric acid (GABA) and vasoactive intestinal peptide (VIP) release may contribute to coupling among cellular oscillators within the SCN.[18-23] Further, several sources of evidence now indicate that, despite the capacity of individual SCN neurons for autonomous rhythmicity, multicellular network interactions increase the amplitude and number of rhythmic cells detected in culture and contribute to the overall robustness of SCN pacemaker function.[24-27]

Molecular Basis of the Suprachiasmatic Nucleus Pacemaker

A critical role for protein synthesis in the mammalian circadian pacemaker was established in the late 1980s,[28,29] and elucidation of the fundamental molecular-genetic circadian oscillatory mechanism began about 10 years later. Thus the first mammalian circadian clock gene, *Clock*, was identified using a forward-genetics mutagenesis screen,[30,31] and this discovery was followed quickly by the identification of a number of other "core" clock genes, some of which showed clear homology to previously discovered circadian clock genes in the fruit fly (e.g., the *Per* genes).[32] Reciprocally, homologues to several clock genes first identified in mice were later shown to be expressed in flies, and this extreme evolutionary conservation facilitated the analysis of both mammalian and nonmammalian systems. The most well-studied mammalian clock genes include the three *Per* genes (*Per1, Per2, Per3*), two plant cryptochrome gene homologues (*Cry1* and *Cry2*), *Clock, Bmal1* (also known as *Arntl1* and *Mop3*), *CK1ε*, and more recently, *Rev-erba* and *Fxbl3*, all of which are expressed in SCN neurons.[33] As described more fully later, several of these genes code for proteins that interact directly to form interlocking negative and positive transcription-translation feedback loops that define the molecular core of the circadian oscillator, whereas others exert critical posttranslational processing that regulates the period and precision of the molecular clock.[32-34]

The primary negative feedback loop of the molecular clock involves negative regulation of the transcriptional activity of CLOCK/BMAL1 heterodimers by PER and CRY through direct protein-protein interactions (Figure 34-1). In turn, inhibition of CLOCK/BMAL1 activity results in reduced *Per* and *Cry* gene transcription, leading eventually to disinhibition of CLOCK/BMAL1 activity. PER protein is posttranslationally modified by CK1ε (and probably other CK1 isoforms) through phosphorylation, which regulates PER stability and also appears to control its nuclear translocation, thus influencing the period of the clock. More recently, it has been

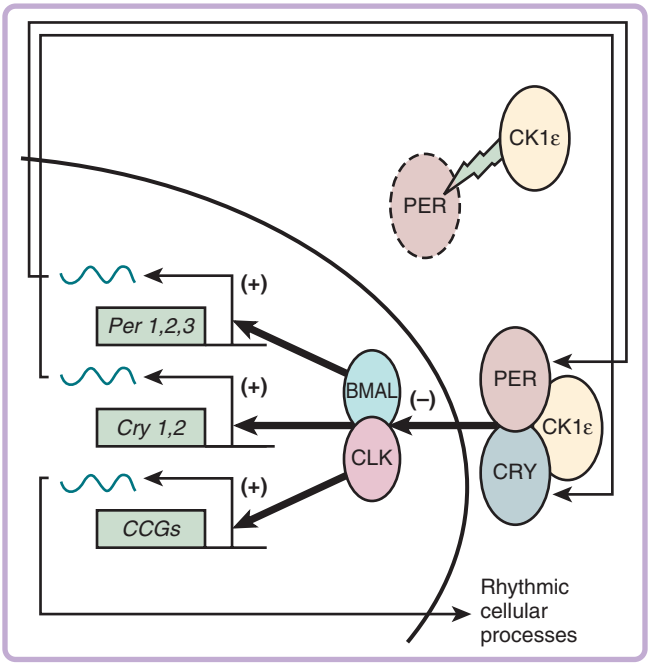


Figure 34-1 Essential elements of the core molecular loop underlying circadian timing at the cellular level in mammals. The transcription factors CLOCK (Clk) and BMAL1 (B) form a protein heterodimer that promotes the transcription of several rhythmically transcribed clock genes, including the *Per* ("period") genes *Per1*, *Per2*, and *Per3*, and the *Cry* (cryptochrome) genes *Cry1* and *Cry2*. CLOCK and BMAL1 also drive transcription of a large number of "clock-controlled genes" (CCGs) that convey the circadian timing signal to a wide variety of cellular processes. The protein products of the *Per* and *Cry* genes dimerize in several different combinations, including both PER-CRY (as shown here) and PER-PER pairings. After nuclear translocation, PER and CRY inhibit the transcriptional effects of CLOCK-BMAL1 through direct protein-protein interactions and thus exert negative autoregulation of their own transcription. This negative feedback results in circadian expression of Per and Cry transcripts. This basic transcription-translation negative feedback loop is modulated by several posttranslational processes. For example, casein kinase 1 epsilon (CK1ε) acts to regulate the nuclear translocation of the PER-CRY dimer and also functions to phosphorylate the Per protein, thereby tagging it for subsequent degradation by ubiquitin. Through these actions, CK1ε plays a critical role in regulating the period of the molecular clock. At the behavioral level, this model predicts (1) the shortening of free-running period seen in *tau*-mutant hamsters carrying a mutation of the *CK1ε* gene, (2) the lengthening of free-running period seen in *Clock*-mutant mice, and (3) the loss of coherent free-running rhythms seen in *Clock* mice and in *Per-* and *Cry*-knockout mice. Note that the molecular clockwork also includes a positive feedback loop, omitted from this diagram, that regulates the availability of BMAL1 and that appears to increase the overall stability of the clock.

discovered that FBXL3 is responsible for degrading the CRY proteins and that multiple *Fxbl3* mutations alter clock period.[35-37] In addition to this primary negative feedback loop, a secondary positive feedback loop regulates *Bmal* transcription through the clock gene *Rev-erba*, which is itself transcriptionally regulated by the CLOCK/BMAL1 complex. These interlocked loops result in rhythmic transcription of specific clock genes in the in vivo[38,39] and in vitro SCN,[10,40,41] and, as discussed later, in many non-SCN tissues as well. Finally, molecular elements of the core oscillator drive rhythmic expression of a large number of clock-controlled genes (i.e., genes that are controlled by, but not part of, the core circadian oscillator loop), which in turn serve as the basis for rhythmic outputs to myriad other cellular processes.[32,33] Ultimately, these cellular-molecular processes are reflected at the

behavioral level, as amply documented by analysis of altered circadian behavioral phenotype in animals bearing specific clock gene mutations.

As described earlier, BMAL1 provides positive transcriptional drive on the *Per* and *Cry* genes, so it is perhaps not surprising that *Bmal1* knockout mice express immediate arhythmicity under free-running conditions.[33] In contrast, the classic *Clock* mutation, which codes for a dominant-negative CLOCK protein, initially lengthens free-running period and results in a gradual loss of rhythmicity under long-term free-running conditions.[41] Surprisingly, however, *Clock* null mice express robust and persisting circadian rhythms, with a modest shortening of circadian period.[42] This is because the CLOCK paralogue, NPAS2, can substitute for CLOCK as a dimerization partner for BMAL1 within the SCN, thus maintaining circadian pacemaker function.[43,44] Regarding the *Per* genes, several different mutation and deletion models have been studied by different laboratories, but in general, *Per1* and *Per2* disruption shortens circadian period and reduces the robustness of free-running rhythms.[33,45] Similarly, *Cry* mutant mice also exhibit alterations in free-running period, whereas *Cry1/Cry2* double-mutants are rendered arrhythmic.[46,47] In contrast to other clock genes, the circadian clock function of *Ck1ε* was discovered by genetic analysis of so-called *tau*-mutant hamsters expressing a spontaneous single-gene mutation that dramatically shortens free-running period.[48] Cloning of the *tau* gene revealed its identity as *Ck1ε*, and subsequent transgenic insertion of this allele into mice recapitulated the hamster short-period phenotype, while deletion of *Ck1ε* in mice lengthened circadian period.[49] Finally, multiple mutations of the *Fxbl3* gene have been shown to lengthen free-running period.[35,36,37] For the molecular clock to drive circadian rhythmicity in physiology and behavior, rhythmic clock gene expression must be linked to intracellular signaling pathways regulating neuronal membrane potential and, ultimately, firing rate.[50] In turn, electrophysiologic rhythmicity in SCN neurons underlies the transmission of efferent signals to the "downstream" extra-SCN targets that more directly regulate neurobehavioral processes. (In addition, as discussed previously, similar mechanisms apparently contribute to intra-SCN oscillator coupling as well.) Remarkably, recent research demonstrates that ionic events at the cell membrane influence the molecular clock through some of the same intracellular signals that convey clock signals to the membrane, and in some cases, these ionic currents may be necessary for self-sustainment of the molecular clock.[50] Such results at a minimum serve to blur the distinction between the clock mechanisms and the so-called hands of the clock.

Finally, in addition to their effects on circadian behavior, circadian clock gene mutations affect sleep-wake homeostasis, as well as several forms of affective behavior, suggesting possible molecular links between the circadian, sleep-regulatory, and motivational systems of the brain.[51-53] Such effects could be due to alterations in SCN efferent signaling to affective systems in the brain as well as to altered clock gene expression within these downstream brain regions. We will return to the health implications of clock gene mutations later.

Functional Architecture of the Suprachiasmatic Nucleus

Traditionally, the SCN has been characterized as comprising distinct ventrolateral and dorsomedial subdivisions.[54] More recently, however, this scheme has been reconceptualized to include SCN "core" and "shell" subnuclei, a concept that may better accommodate species differences in the anatomic distribution of neuropeptides, afferent terminal fields, and gene expression patterns in the SCN[55,56] (Figure 34-2). Although SCN neurons express a large number of neuropeptides, core and shell subnuclei have been most commonly identified by the concentration of arginine vasopressin–positive neurons in the SCN shell and by VIP-positive and gastrin-releasing peptide (GRP)–positive neurons in the SCN core. Beyond this basic organization, however, clear species differences have been noted, even among nocturnal rodents. For example, the hamster SCN core contains a very distinct and compact cluster of photoresponsive calbindin-positive cells, which is

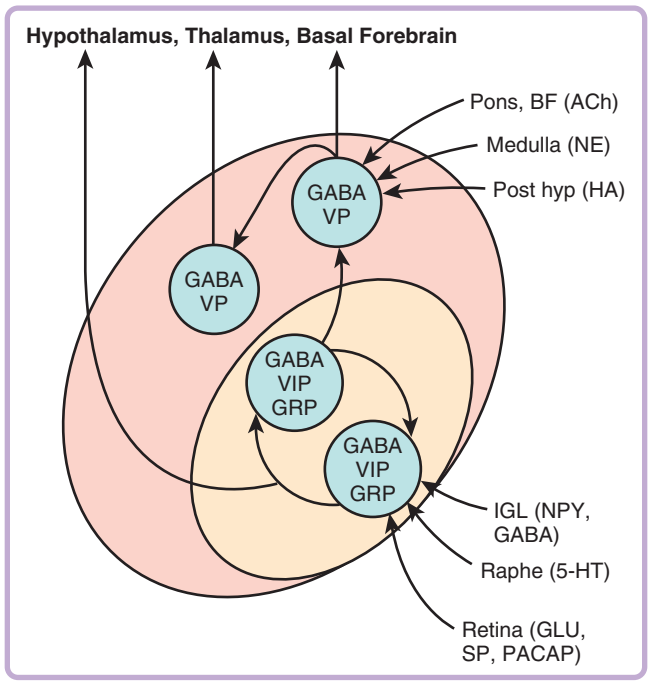


Figure 34-2 Core and shell organization of the suprachiasmatic nucleus (SCN). Most SCN neurons release the inhibitory amino acid transmitter gamma-aminobutyric acid (GABA). In the SCN core (*yellow*), GABA is commonly colocalized with one or more neuropeptides, including vasoactive intestinal polypeptide (VIP) and gastrin-releasing peptide (GRP), whereas neurons of the SCN shell (*orange*) frequently contain GABA colocalized with arginine vasopressin (VP). SCN core neurons project to other core neurons, to SCN shell neurons, and to extra-SCN targets, most prominently in the diencephalon and basal forebrain; SCN shell neurons project to other shell neurons and to extra-SCN targets, but not to SCN core neurons. This anatomic organization implies that the flow of information within the SCN is generally from core to shell. Consistent with this suggestion, the three best-characterized SCN afferent systems, originating in the retina, the intergeniculate leaflet of the thalamus (IGL), and the mesencephalic raphe nuclei, converge in the SCN core. Retinal afferents contain the excitatory amino acid transmitter glutamate (GLU) as well as the neuropeptides substance P (SP) and pituitary adenyl cyclase–activating peptide (PACAP); raphe afferents contain 5-hydroxytryptamine (5-HT; serotonin); and IGL afferents contain neuropeptide Y (NPY) and GABA. Beyond these core afferents, several less well-characterized afferent systems converge in the SCN shell, including acetylcholine (ACh)-containing projections from the basal forebrain (BF) and pons, medullary norepinephrine (NE)-containing projections, and histamine (HA)-containing projections from the posterior hypothalamus (Post hyp). A number of other anatomically identified but functionally uncharacterized SCN afferent systems have been omitted from this figure and are not discussed in the chapter.

absent in the rat.[55] In light of this and other species differences as well as other complexities in SCN organization, it has been argued that the popular distinction between SCN core and shell may be such an extreme oversimplification as to impede understanding of the functional organization of this critical structure.[2,57,58]

Although the specific functions of chemically defined SCN cell populations are not fully known, a reasonable heuristic is that VIP and GRP neurons of the SCN core serve to collate afferent information relevant to pacemaker entrainment, whereas the vasopressinergic (or other) neurons of the SCN shell are primarily responsible for the self-sustaining generation of the circadian timing signal. A preeminent role for the SCN core in pacemaker entrainment is supported by findings that major SCN afferent systems converge in the core subnucleus[54,55]; administration of SCN core peptides such as VIP and GRP can mimic both light-induced phase shifting and *Per* gene expression in the SCN in vivo and in vitro[59-61]; and light-evoked changes in SCN physiology and gene expression spread over time from core to shell.[18,62] Conversely, evidence for a preeminent role of the SCN shell in pacemaking includes findings that the core projects robustly to the shell but not vice versa[55]; spontaneous circadian rhythmicity in neuronal activity, neuropeptide release, and *cfos* and *Per* gene expression are seen more robustly and reliably in the shell than in the core[63-65]; and spontaneous rhythmicity in SCN gene expression appears to flow from the most dorsomedial toward more central-lateral regions over the course of the circadian cycle.[66,67] On the other hand, the view that SCN core and shell underlie completely discrete entrainment and pacemaking functions, respectively, is probably too simplistic because (1) several arousal-related afferents of limbic and brainstem origin target the SCN shell[54-56]; (2) in vitro studies have revealed independent free-running rhythmicity in secretion of core and shell peptides from the same tissue explant[68]; and (3) SCN core and shell can exhibit stable dissociation of rhythmic gene expression in vivo under certain conditions.[69] Together, these findings implicate separate core and shell oscillators. Further, studies using microlesions indicate that the integrity of the SCN core is essential for the maintenance of high-amplitude behavioral and molecular-level rhythmicity, suggesting that rhythmic signals from the core serve a permissive gatelike role in sustaining oscillatory function in the shell.[62,66]

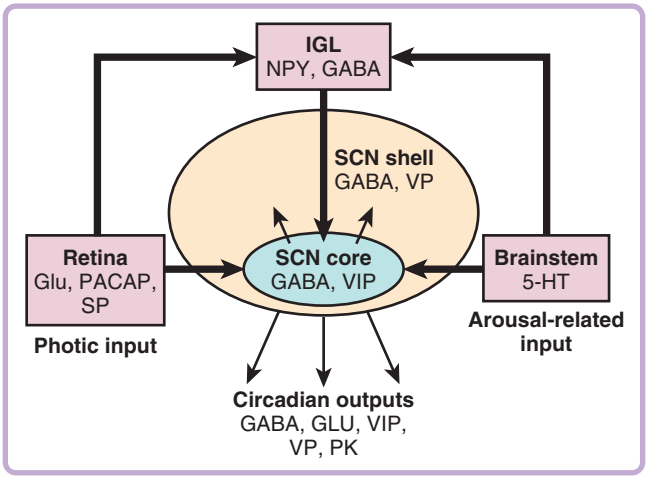


Figure 34-3 Overview of functional neuroanatomic pathways in the mammalian circadian system. Major suprachiasmatic nucleus (SCN) afferent systems originating in the retina and raphe nuclei also target the intergeniculate leaflet of the thalamus (IGL), which in turn projects to the SCN. Retinal projections to the SCN and IGL mediate photic input to the circadian system; raphe projections to the SCN and IGL mediate the effects of certain nonphotic, behavioral state–related signals; and IGL-SCN projections are involved in mediation of both photic and nonphotic signaling to the SCN pacemaker. As described in the text, photic and nonphotic pathways generally interact via both presynaptic and postsynaptic mechanisms to produce mutually antagonistic effects on the circadian pacemaker. Thus photic signals evoke circadian phase shifting during subjective night and antagonize nonphotic phase shifting during subjective day, whereas signals related to arousal and wakefulness evoke phase shifting during subjective day and antagonize photic phase shifting during subjective night. These antagonistic interactions are mediated in part at the level of the SCN, but the scheme presented here suggests that the IGL is also a probable locus for interaction between photic and nonphotic signals—a hypothesis that is largely unexplored. SCN outputs appear to include both neural efferents and secreted paracrine signals. GABA, Gamma-aminobutyric acid; GLU, glutamate; 5-HT, 5-hydroxytryptamine (serotonin); PACAP, pituitary adenyl cyclase–activating peptide; PK, prokineticin; SP, substance P; VIP, vasoactive intestinal polypeptide; NPY, neuropeptide Y.

Light Input to the Suprachiasmatic Nucleus: The Retinohypothalamic Tract

Although stimuli such as temperature, sound, food, and social cues appear to contribute to phase control, environmental light and darkness are the primary cues for circadian entrainment. A specialized retinal projection system, referred to as the *retinohypothalamic tract* (RHT), is both necessary and sufficient for photic entrainment of the circadian pacemaker.[2,70,71] Unlike the classic "image-forming" visual system, the RHT system functions more like a light meter than a camera lens, and it is thus not surprising to find a separate afferent pathway dedicated to circadian entrainment (and other light-dependent functions requiring irradiance coding, such as light-evoked pupillary responses, melatonin suppression, and affective responses).[71] The RHT originates from a distinct subset of retinal ganglion cells, separate from those giving rise to the primary visual pathways,[72] and terminates mainly in the SCN, as well as more sparsely in the anterolateral hypothalamus, subparaventricular zone, and supraoptic region.[73,74] RHT axon collaterals also project to the thalamic intergeniculate leaflet (IGL; Figure 34-3), which, as discussed later, is an important component of the circadian system.

Although nonretinal photoreception plays a major role in circadian entrainment in nonmammalian vertebrates,[75] mammalian photoreception appears to be based entirely on retinal mechanisms. Remarkably, retinally degenerate strains of mice, in which nearly all classic photoreceptors (i.e., rods and cones) are lost by early adulthood, exhibit normal circadian responses to light.[76] More recently, similar findings have been reported in genetically engineered mice with a total developmental absence of both rods and cones, demonstrating conclusively that circadian light entrainment can be mediated by a novel, nonrod, noncone photoreceptor system.[77] Recent studies indicate that the protein, melanopsin, found specifically within the small subset of retinal ganglion cells giving rise to the RHT, serves as a circadian photoreceptor molecule in a novel population of intrinsically photosensitive retinal ganglion cells (ipRGC).[78,79] Surprisingly, however, although melanopsin-containing ganglion cells are sufficient for circadian entrainment, studies with melanopsin knockout mice reveal that, despite parametric changes in photoresponsiveness, this photopigment is not necessary for entrainment.[80,81]

Indeed, entrainment is only fully abolished when both classic and melanopsin-based photoreception is eliminated,[82] revealing a potential degree of functional redundancy in how the SCN receives entrainment information from the photic environment. Although signals derived from both classic and nonclassic (ganglion cell) photoreceptors reach the SCN, rod- and cone-derived entrainment signals appear to be relayed to the SCN through synapses on ipRGCs.[83-86]

RHT terminals release the excitatory amino acid neurotransmitter, glutamate (or aspartate), in response to photic stimulation. Extensive evidence from in vivo and in vitro studies indicates that glutamate acts through both *N*-methyl-d-aspartate (NMDA) and non-NMDA receptors and a variety of intracellular signaling molecules (e.g., Ca^{2+}, nitric oxide, calmodulin/calmodulin kinase, protein kinase C, protein kinase G, cyclic adenosine monophosphate–responsive element-binding protein [CREB], and others),[71,87] and immediate early-response genes including *c-fos*,[88] leading to increased expression of *Per1* and, after a delay, *Per2*.[89,90] The protein products of these genes represent state variables of the molecular oscillator, such that alterations in their transcription levels, when superimposed on the ongoing circadian transcription cycle, correspond functionally to phase shifts of the oscillator (Figure 34-4).

In addition to glutamate, RHT terminals release two identified peptide cotransmitters, substance P (SP) and pituitary adenyl cyclase–activating peptide (PACAP). SP appears to play an important role in RHT transmission because selective SP antagonists block light-induced phase shifting and early-response gene expression in vivo,[91-93] as well as glutamate receptor–mediated phase shifting in vitro.[94] By itself, SP can mimic at least one component of the photic phase-response curve (phase delays during early subjective night) both in vivo and in vitro.[95] Supporting its putative role of a modulatory cotransmitter, the phase-shifting effects of SP appear to depend on SP-evoked glutamate release, and they can be blocked by the NMDA antagonist, MK-801.[94] In contrast, PACAP administration has been reported to either antagonize or mimic the effects of glutamate on circadian phase shifting and *Per* gene expression in vitro, depending on dose and on circadian phase of administration.[96-98] Specifically, when administered at relatively high doses, PACAP blocks the effects of glutamate during subjective night and evokes phase advances during subjective day, but when administered at much lower doses, PACAP actually mimics or potentiates the effects of glutamate on the SCN pacemaker.

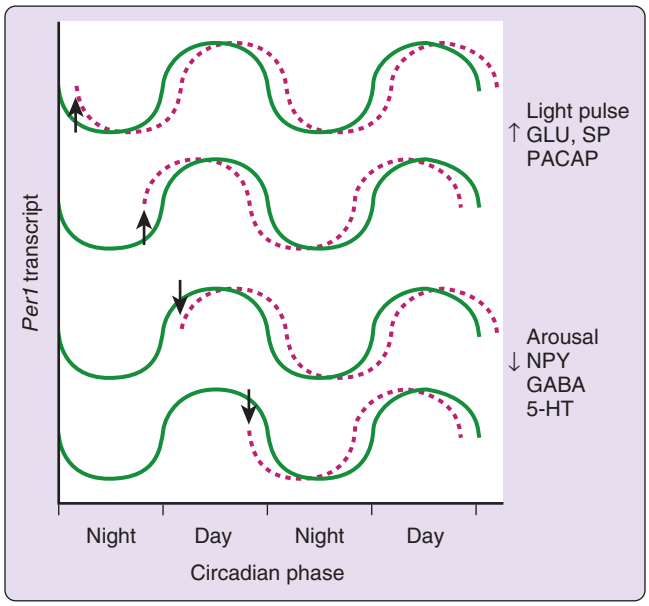


Figure 34-4 A simple qualitative-molecular model for circadian phase shifting by photic and nonphotic signals, and for their mutually antagonistic interaction. In this "phase-only" model, amplitude is fixed and the underlying state variable (here, *Per1* transcript level) can oscillate only within predetermined upper and lower bounds, such that the *Per1* level represents the phase of the molecular oscillator. During the subjective night, *Per1* levels are relatively low (*solid line*), and light pulses (or corresponding neurotransmitters or intracellular messengers) induce an abrupt increase in transcript level (*arrow*). Early in the night, when *Per1* levels are normally decreasing, this increase in transcription essentially forces the oscillator to repeat part of its normal trajectory and is thus equivalent to resetting the oscillator to an earlier phase, resulting in a permanent phase delay (*dashed line*). In contrast, late in the night, *Per1* levels are normally increasing, such that a light-induced increase in transcription forces the oscillator to omit part of its normal trajectory, equivalent to resetting the oscillator to a later phase and resulting in a permanent phase advance. Opposite to light pulses, arousal-related signals (or corresponding neurotransmitters or intracellular messengers) induce abrupt decreases in *Per1* transcription, resulting in phase delays during early subjective day and phase advances during late subjective day. Thus the model predicts that photic and nonphotic phase-response curves (PRCs) should have essentially identical shapes but should be phase-displaced by 180 degrees (12 circadian hours) along the horizontal axis; these predictions are at least roughly consistent with experimental observations.[98] Further, this model accounts for the general insensitivity of the circadian pacemaker to photic phase shifting during midsubjective day and to nonphotic phase shifting during midsubjective night: Because the underlying state variable can vary only within a predetermined range, stimuli that increase *Per1* transcription are ineffective when transcript levels are already maximal, and stimuli that decrease *Per1* transcription are ineffective when transcript levels are already minimal. Nevertheless, despite these periods of insensitivity, nonphotic signals would remain capable of counteracting light-evoked increases in transcription, and photic signals would remain capable of counteracting arousal-evoked decreases in transcription. Finally, the exact waveform and phasing of the photic and nonphotic PRCs would obviously depend on the exact waveform and phasing of the underlying spontaneous transcription cycle, here presented arbitrarily as two interlocking circular arcs centered over midsubjective day and midsubjective night. GABA, Gamma-aminobutyric acid; GLU, glutamate; 5-HT, 5-hydroxytryptamine (serotonin); PACAP, pituitary adenyl cyclase–activating peptide; SP, substance P; NPY, neuropeptide Y.

Other Functional Inputs to the Circadian Clock

An additional major SCN afferent system arises from the IGL, a distinct retinorecipient region of the lateral geniculate complex, intercalated between the dorsal and ventral lateral geniculate nucleus.[99-101] The projection from the IGL to the SCN is referred to as the *geniculohypothalamic tract* (GHT), and GHT neurons release both neuropeptide Y and GABA (see Figure 34-3). Retinal signals are conveyed to the IGL in part by axon collaterals of RHT neurons,[102] and GHT and RHT terminal fields are largely coextensive within the SCN core. It is therefore not surprising that early functional studies emphasized the possible role of the IGL/GHT system in providing a secondary, indirect pathway for photic entrainment of the circadian pacemaker.[100,101] Although lesion studies establish that the IGL is clearly not necessary for photic entrainment, lesions of the IGL/GHT system result in subtle modifications in the ability of light signals to effect phase and period control of the circadian clock.[101] Further, the IGL may have a significant role in entrainment under more naturalistic lighting conditions (e.g., regimens including twilight transitions, seasonally changing photoperiod, or

simulated moonlight), relative to the square-wave light-dark cycles commonly employed in the laboratory.[103,104] In addition to providing a secondary, indirect source of photic signaling to the circadian clock, the IGL also plays a preeminent role in the nonphotic regulation of the circadian system by arousal-related stimuli. Thus IGL lesions abolish the phase-shifting effects of novelty-induced wheel running[105,106] and benzodiazepine administration in hamsters,[107-109] as well as the period-shortening effect of running-wheel access in rats[110] and the entrainment effect of scheduled daily treadmill activity in mice.[111] More recently, evidence has been presented that IGL neurons may be sensitive to metabolic signals and that the GHT may mediate the effects of such signals on the SCN pacemaker.[112,113] This suggestion is of particular interest given the well-known role of neuropeptide Y (NPY) in integrating metabolic and appetite-related signals within other hypothalamic circuits.

Another major SCN afferent system converging mainly on the SCN core originates from the serotonergic midbrain raphe, especially the median raphe nucleus.[114,115] In addition, ascending serotonergic projections originating in the dorsal raphe nucleus innervate the IGL, providing a second potential route for serotonergic regulation of the SCN circadian pacemaker (see Figure 34-3). Extensive evidence has implicated serotonergic projections to the SCN (and IGL) in two distinct functions: (1) modulation of photic effects on the circadian pacemaker during the subjective night,[116] and (2) mediation of nonphotic effects on the pacemaker during subjective day.[117] Thus serotonin depletion potentiates photic phase shifting and impairs the response to nonphotic phase-shifting stimuli.[118,119] Conversely, electrical stimulation of the serotonergic raphe and systemic or intra-SCN administration of serotonergic agonists inhibits photic phase shifting during subjective night and evokes nonphotic phase shifting during subjective day.[120-123] These effects appear to be mediated in part through 5-hydroxytryptamine 1A (5-HT_{1A}) and $5HT_7$ receptors, and studies employing direct intracerebral administration of the serotonin 5-$HT_{1A/7}$ receptor agonist 8-OH-DPAT [8-hydroxy-2-(di-*N*-propylamino)tetralin] have identified several potential loci within the circadian system for these effects, including the SCN, the IGL, and the median and dorsal raphe nuclei.[122,123] Nevertheless, the ability of direct serotonin application to the in vitro SCN to evoke circadian phase shifts indicates that stimulation of intra-SCN serotonin receptors is sufficient to phase-shift the pacemaker.[124] In addition, serotonin inhibits light-induced phase-shifting through $5HT_{1B}$ receptors located presynaptically on RHT terminals.[125,126] Whereas light itself has little or no phase-shifting effect during midsubjective day, light at this phase can block the phase-shifting effects of a serotonin agonist,[127] indicating that serotonergic (nonphotic) and photic (glutamatergic) inputs to the SCN are mutually inhibitory.

Because the phase-shifting effects evoked by serotonergic stimulation closely resemble those seen with other nonphotic phase-shifting stimuli, including novelty-induced activity, sleep deprivation, and benzodiazepine and NPY administration, several studies have directly examined the potential role of serotonergic afferents to the SCN and IGL in mediating the effects of behavioral state on the circadian pacemaker. Thus arousal, wakefulness, and motor activity are all associated with increased forebrain serotonin release,[128-130] and serotonin content in the rat SCN is correlated positively with spontaneous activity level and negatively with free-running period.[131] Indeed, state-dependent variations in serotonin release appear to mediate (1) phase-shifting by activity or sleep deprivation,[129] (2) the effects of activity level on free-running period,[130] (3) entrainment by scheduled daily treadmill activity[111] and restricted daily running wheel access,[132] and (4) activity-dependent inhibition of photic phase shifting.[133]

Several other chemically identified pathways provide afferent inputs to the circadian system, including noradrenergic projections from the locus coeruleus, cholinergic projections from the basal forebrain and pontine tegmentum, and histaminergic projections from the posterior hypothalamus.[1] In addition, noradrenergic and cholinergic projections both innervate the IGL, providing an alternate pathway by which these transmitter systems could alter SCN circadian pacemaker function. Unlike the retinal, geniculate, and raphe projections described previously, which generally form overlapping terminal fields in the SCN core, these afferents target preferentially the SCN shell (see Figure 34-2).[55,56] Although SCN shell afferents are less studied than the SCN core afferents, sufficient data exist to suggest that these SCN shell afferents also contribute to circadian pacemaker regulation.[1] Finally, a recent review concludes that the "extended circadian rhythm system" includes direct monosynaptic projections to the SCN from at least 35 distinct brain areas,[58] revealing enormous potential for pacemaker modulation by a wide range of effective stimuli.

Suprachiasmatic Nucleus Output Pathways

Given the physiologic and behavioral ubiquity of circadian rhythmicity, it may be somewhat surprising that first-order SCN efferents innervate a relatively small number of target areas. On the other hand, these targets, concentrated mainly in the diencephalon and basal forebrain, include well-established relays to the autonomic and neuroendocrine systems, as well as to structures regulating affective, sensory, and motor processes.[134,135] SCN efferents emerge from both the core and shell subnuclei and release a number of neurotransmitters and peptides, including GABA, glutamate, and vasopressin. Remarkably, anatomically distinct populations of SCN neurons appear to innervate specific efferent targets, providing multiple waves of neuronal signals that regulate circadian phase in a target-specific manner.[136]

In addition to neuronal efferents, the SCN appears to regulate at least certain rhythmic processes through diffusible paracrine signals. The presence of a diffusible SCN output signal was first suggested by the finding that complete surgical isolation of the SCN within a "hypothalamic island" abolished SCN-dependent neuroendocrine responses but allowed for persisting locomotor activity rhythms in the same animals.[136] Although this surprising finding stood in conflict with prior studies suggesting that the ability of SCN transplants to restore rhythmicity in SCN-lesioned hosts depends on anatomic integration with the host brain,[137-139] evidence that transplantation of SCN tissue encapsulated within semipermeable capsules could restore locomotor rhythms provided strong confirmation of the paracrine hypothesis.[140] Several diffusible candidate molecules have now been implicated as circadian output signals, including prokineticin-2, tumor necrosis factor-α, and vasopressin.[136]

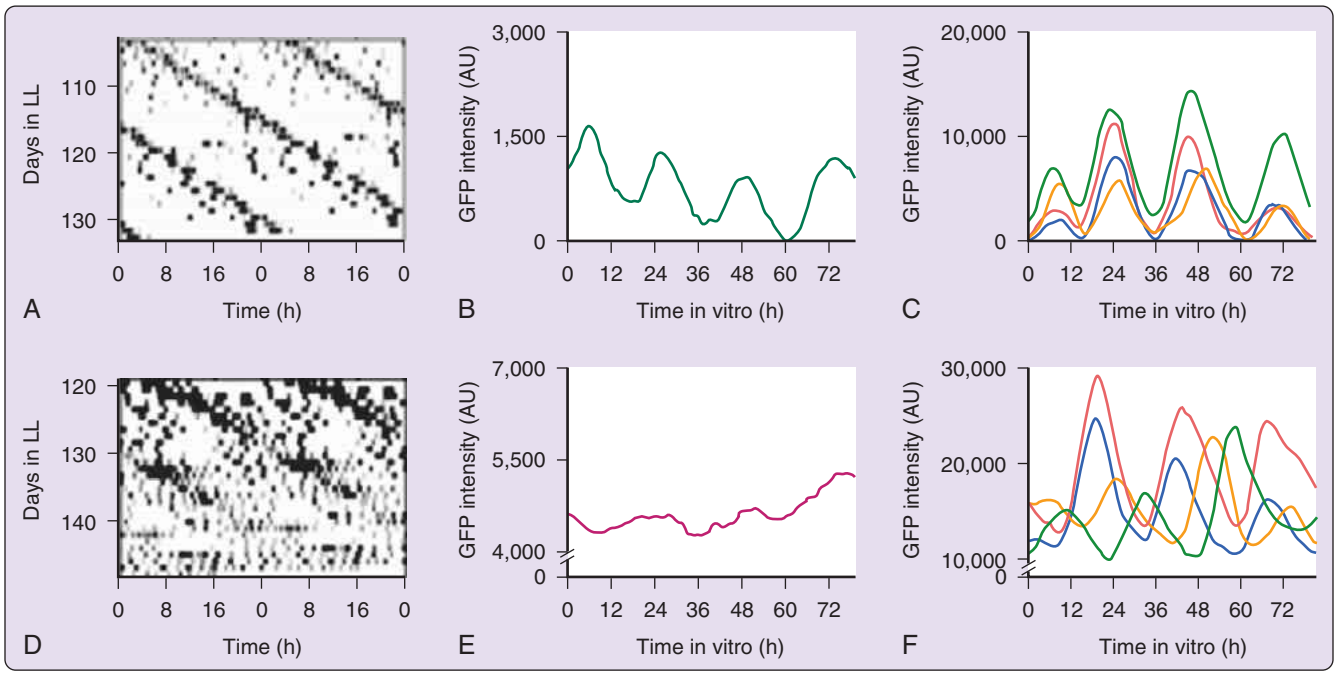


Figure 34-5 Relationship between rhythmicity at the behavioral (**A, D**), tissue (**B, E**), and cellular (**C, F**) levels. These experiments used *Per1:GFP* transgenic mice in which green fluorescent protein (GFP) serves as a real-time optical reporter of *Per1* gene expression, measured by fluorescent emissions and expressed in arbitrary units (AU) of intensity. When maintained in long-term constant light (LL), some animals sustain persisting, coherent rhythmicity at the behavioral level (e.g., **A**). When SCN tissue is cultured from such animals, robust gene expression rhythms are seen at both the whole-SCN (**B**) and individual neuronal (**C**) levels. In contrast, animals that become arrhythmic at the behavioral level (e.g., **D**) also fail to express coherent rhythmicity at the tissue level (**E**). Nevertheless, individual SCN neurons in such cultures continue to show robust gene expression rhythms but fail to show the normal intercellular synchrony as seen in rhythmic animals (**F**). Thus maintenance of normal coupling relationships among potentially autonomous cellular oscillators is essential for the display of coherent rhythmicity at the behavioral and physiological levels. (From Ohta H, Yamazaki S, McMahon DG. Constant light desynchronizes mammalian clock neurons. *Nat Neurosci* 2005;8:267–9.)

MULTIPLE-OSCILLATOR NATURE OF THE CIRCADIAN SYSTEM

The evidence reviewed previously (and in other chapters in this volume) amply demonstrates the critical role of the SCN in circadian regulation and justifies the use of the term *pacemaker* to describe its function. Nevertheless, the circadian system is now known to comprise a multiplicity of circadian oscillators that are distributed throughout the brain and body. As reviewed previously,[141,142] many behavioral and physiologic studies conducted in the premolecular era revealed that circadian systems exhibit several varieties of complex dissociations among multiple rhythmic subcomponents. For example, two or more discrete daily activity epochs may emerge from the single, normally consolidated activity period, a phenomenon known as *splitting*.[143,144] Such phenomena at the behavioral and endocrine level strongly imply the existence of an underlying multioscillatory neurobiologic circadian system. Nevertheless, interest in these complex phenomena was deprioritized for several years, coincident with the ongoing maturation of molecular analyses of the pacemaker mechanism, which focused largely on the SCN (or similar pacemaker-like structures in nonmammalian animals). It is probably fitting, then, that these same molecular approaches, and especially the finding that mammalian clock genes are expressed not only in the SCN but also in other brain regions[145,146] and in many peripheral tissues,[147-149] have spurred renewed interest in the identification and functions of multiple circadian oscillators and how they are organized into an anatomically distributed circadian timing system.

Within the SCN pacemaker, the observation that individual SCN neurons can express cell-autonomous rhythmicity in both molecular and physiologic processes[6-10] demonstrates that the pacemaker is itself composed of numerous, potentially autonomous but normally coupled, cellular circadian oscillators.[26] Further, the appearance of circadian desynchrony at the behavioral level may reflect robustly persisting but uncoupled rhythmicity at the neuronal level[150] (Figure 34-5). SCN cellular oscillators may also form functionally specific multicellular assemblies. For example, spontaneous or forced splitting of circadian activity rhythms into two distinct daily bouts is associated with splitting of SCN neurons into distinct neuronal populations with firing rhythms peaking in antiphase.[151,152] These two functionally distinct neuronal populations are likely to be associated with Pittendrigh's "morning" and "evening" oscillators. In addition, the clock genes *Per1* and *Per2* exhibit a significant degree of functional specialization within the SCN,[153,154] and according to one hypothesis *Per1* and *Per2*

represent state variables of morning and evening oscillators, respectively.[155,156] An important ongoing challenge for circadian biologists is to continue the integration of modern insights into the cellular and molecular nature of the circadian clock with the earlier and equally important era of the field when many of the formal properties and basic principles underlying circadian organization were initially defined.[157,158]

Outside the SCN, recent evidence suggests that several non-SCN neural and neuroendocrine tissues are capable of expressing autonomous (although generally highly damped) circadian oscillations.[146] The first demonstration of autonomous, self-sustained circadian rhythmicity in non-SCN mammalian neural tissue was the finding that cultured hamster retinae display persisting circadian rhythmicity in melatonin secretion.[159] More recently, real-time optical monitoring of *Per* gene (or protein) expression has been employed to demonstrate autonomous tissue-level and cellular-level rhythmicity in non-SCN neural tissue, including the retina,[160] as well as in a wide variety of nonneural tissues collected from *Per-luciferase* or *Per-GFP* transgenic reporter mice.[145-149] Within the central nervous system, such studies have revealed self-sustaining gene expression rhythms in the SCN, neuroendocrine tissues (pineal, pituitary), diencephalic nuclei (e.g., hypothalamic arcuate and paraventricular nuclei, thalamic paraventricular nucleus), and olfactory bulbs. Indeed, an impressive series of experiments have revealed that the olfactory bulb contains a self-sustained circadian clock that is normally entrained by the SCN and that regulates circadian rhythmicity in the functional sensitivity of the olfactory system.[161] Although the relationships between these extra-SCN neural oscillators and the SCN pacemaker have not been fully elucidated, it appears that at least certain types of behavioral rhythm splitting may involve dissociations between intra-SCN and extra-SCN central clocks.[162]

Similar techniques have also been used to reveal rhythmic *Per* expression in a wide variety of peripheral cells and tissues, including liver, lung, kidney, skin, skeletal muscle, heart, and fibroblasts.[147-149,163] The initial studies generally found these peripheral oscillators to be highly damped and dependent on periodic input from the SCN for their continuous expression. In contrast, however, use of a real-time reporter of PER2 protein levels[149] reveals that circadian oscillations in cultured peripheral tissues can in fact persist for many cycles (Figure 34-6). Nevertheless, these tissue-level rhythms depend on

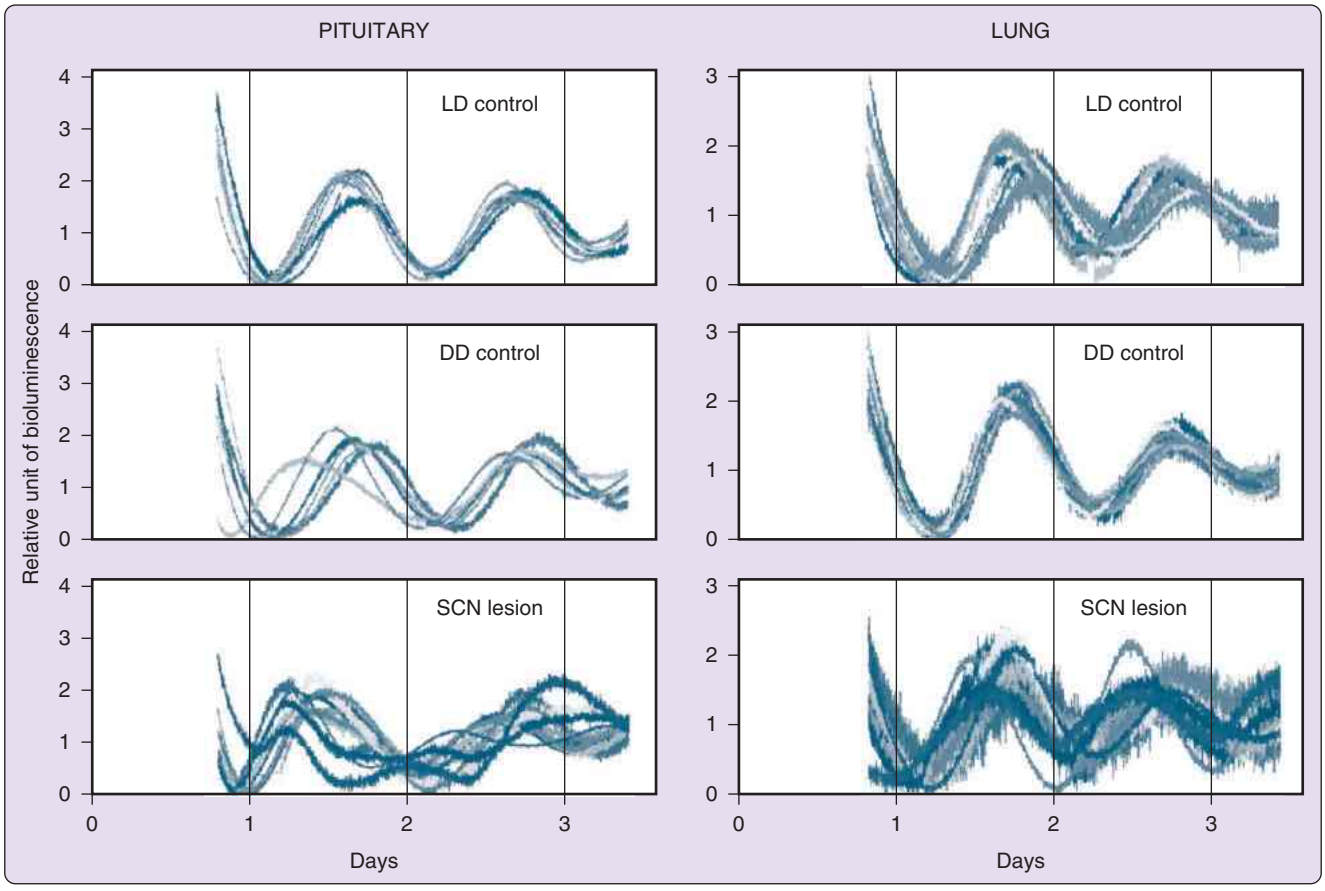


Figure 34-6 Superimposed plots of bioluminescent data from pituitary and lung tissues from individual animals that were intact and maintained on a light-dark cycle (LD controls) or in constant darkness (DD controls), as well as from suprachiasmatic nucleus (SCN)-lesioned animals. The tissue was maintained in vitro and made use of a Period2:Luciferase fusion protein as a real-time reporter of circadian dynamics. The first three cycles in culture are represented; each animal's record is a different shade. Although tissues collected from individual animals on an LD cycle, and relative to activity onset in DD control animals, were in phase with one another, phase desynchronization is evident in individual records of the SCN-lesioned animals for both tissues. (From Yoo SH, Yamazaki S, Lowrey PL, et al. PERIOD2:LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci U S A* 2004;101:5339–46. Copyright [2004] National Academy of Sciences, U.S.A.)

SCN signals for maintaining proper phase relationships among tissue-level clocks[149] Further, as is also true for SCN neurons (see Figure 34-5), apparent damping at the tissue level may obscure robust, persisting rhythmicity at the cellular level.[163] Thus, even if not dependent on the SCN for their sustainability, SCN signals may serve to maintain temporal synchrony among multiple tissues or multiple cells from a given tissue.[150,164,165] Self-sustaining peripheral oscillators may also dissociate from the SCN pacemaker even in intact animals when exposed to certain environmental conditions, such as after light-dark cycle phase shifts (i.e., simulated jet lag)[166] or during restricted feeding schedules, which entrain peripheral but not SCN *Per* gene oscillations.[167] These observations indicate that the SCN pacemaker normally serves to entrain both central and peripheral secondary oscillations generated by a broadly distributed population of autonomous cellular oscillators, but that under certain conditions, these downstream oscillators are capable of adaptive (and possible maladaptive) disengagement from SCN control (Figure 34-7).

With the provocative title, "Circadian lessons from peripheral clocks: Is the time of the mammalian pacemaker up?" Brandstaetter suggested that the elucidation of peripheral circadian clocks might require "the hypothalamic SCN of a rodent . . . to resign from its major function."[168] Nevertheless, to paraphrase the American humorist, Mark Twain, we maintain that it is premature to announce the passing of the SCN pacemaker. For years before the identification of peripheral molecular clocks, the circadian research community speculated that the circadian pacemaker might regulate downstream "slave oscillators." Although it is now clear that many cells, tissues, and organs can produce autonomous circadian rhythms using similar molecular clock machinery as expressed in SCN cells in the absence of the SCN, this does not mean these downstream oscillators are normally *independent* of the SCN. Indeed, what is emerging in the field is the hypothesis that the SCN is still—as we write in late 2014—the "master oscillator" or pacemaker, regulating all (or nearly all) circadian rhythms, whether directly or indirectly. Rhythms are regulated directly when circadian information in the form of neural or neuroendocrine signals from the SCN directly imposes circadian timing on the neural and physiologic systems regulating a particular functional output. However, "indirect" control is also important, as, for example, when the SCN's direct control of behavioral rhythms (e.g., feeding, sleep-wake state, body temperature) alters an organism's metabolic, endocrine, or physiologic processes, which in turn control (entrain) other rhythms. In animals, a broad network of direct and indirect controls ensures that the SCN and other central and peripheral oscillators maintain specific phase relationships, resulting in the overall temporal coordination of the system.

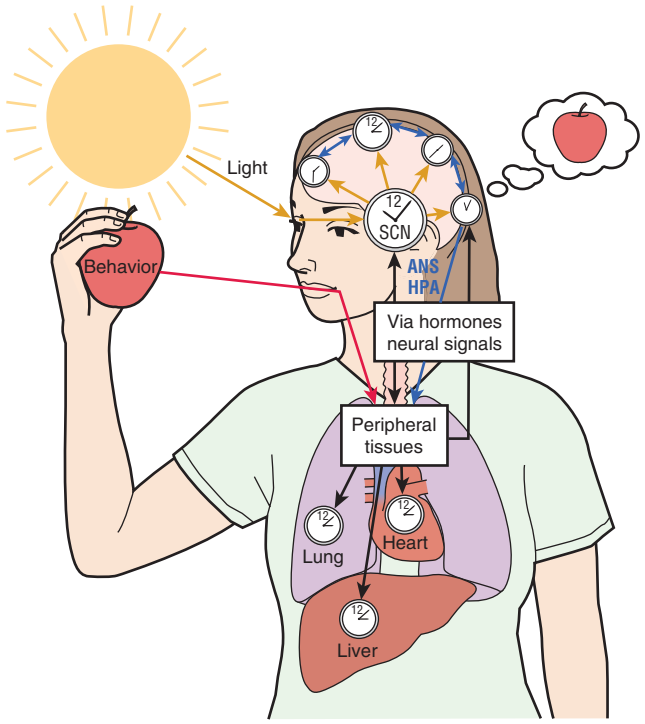


Figure 34-7 The multioscillatory circadian timing system includes a large number of autonomously rhythmic circadian clock cells distributed within both central and peripheral tissues. Molecular feedback loops drive circadian rhythms in gene expression and other cellular processes within individual clock cells (see Figure 34-1); these molecular loops are based on similar, but not necessarily identical, genes and proteins in different tissues. The circadian pacemaker resides in the SCN and is entrained by light-dark cycles and other environmental stimuli. Cellular oscillators within SCN neurons are characterized by relatively strong intercellular coupling, which maintains phase synchrony among individual cells and underlies robustly self-sustaining rhythmicity at the tissue level. Nevertheless, under certain conditions (e.g., constant light), these coupling relationships may be sufficiently weakened such that coherent rhythmicity is lost at the tissue (and behavioral) levels (see Figure 34-5). Cellular oscillators also underlie tissue-level rhythmicity in genomic and physiologic processes in other neural tissues as well as in many peripheral tissues and organs. In general, cellular oscillators outside the SCN appear to be less strongly coupled relative to those within the SCN and thus depend on rhythmic inputs from the SCN to maintain phase synchrony. In the absence of such input from the central SCN pacemaker, a loss of phase synchrony at the cellular level may be reflected in dampening of rhythmicity at the tissue level (see Figure 34-6). Together, SCN and non-SCN central neural oscillators result in rhythmic behavior (such as food intake and motor activity), autonomic nervous system (ANS) function, and hypothalamic-pituitary-adrenal (HPA) axis hormone secretion. These behavioral and physiologic rhythms in turn can give rise to other rhythmic signals (e.g., glucose availability, corticosterone levels, and body temperature) that serve to maintain phase synchrony among peripheral oscillators, probably in a tissue-specific manner. In turn, the activity of peripheral oscillators may give rise to rhythmic signals (e.g., peripheral hormones, autonomic afferents, and metabolic signals) that contribute to the synchronization of the SCN pacemaker and other central oscillators.

THE CIRCADIAN TIMING SYSTEM IN HEALTH AND DISEASE

This new picture of the circadian timing system has raised important questions about the potential adverse health effects that may be associated with a loss of normal synchronization between and among central and peripheral oscillations.[5,169-171] Thus circadian disruption at the molecular and systemic levels has been linked to sleep disorders, obesity and diabetes, heart disease, cancer, and psychiatric disorders. (See Chapters 39 and 40 for more complete discussion of the adverse health effects of disrupting the overall circadian temporal organization.) Indeed, the *Per2* gene appears to play an important role in suppression of mammalian tumorigenesis and may partially mediate the increased cancer risk seen in humans doing shift work and in mice subjected to experimental circadian disruption.[172,173] Given that approximately 5% to 10% of the entire genome is expressed rhythmically in any given tissue or organ,[174,175] many other mechanisms linking circadian synchrony and desynchrony to health and disease will undoubtedly emerge in the next few years.[176]

Although circadian dysregulation may be uncommon in natural animal populations, it certainly occurs quite often in humans, who can override their circadian clock and exert substantial volitional control over their sleep-wake cycles. Under such circumstances, abnormal phase relationships are expressed between sleep-wake behaviors (and other rhythmic processes tightly linked to sleep or wake states) and the circadian clock (and rhythmic processes tightly linked to it). Although the internal desynchronies that occur with jet lag and shift work may be the most dramatic, they are not the only examples of such desynchrony. Indeed, social constraints, work schedules, and the use of artificial lighting may result in the widespread occurrence of "social jet lag" even in people living under relatively stable entrained conditions.[177] Regardless of work or travel schedules, humans in our modern, around-the-clock society are certainly becoming less strictly diurnal, opposing millions of years of evolutionary selection.

CONCLUSIONS

The primary pacemaker for the mammalian circadian system is contained in the SCN, and the mechanisms underlying the pacemaker function of this structure are being elucidated rapidly at the molecular, cellular, and physiologic levels. The SCN contains a large number of normally coupled but potentially autonomous cellular oscillators that generate a circadian time base through the expression of a complex molecular feedback loop. The activity of the core molecular loop results in the circadian expression of a large number of clock-controlled genes, which in turn regulate coordinated circadian rhythms in the metabolism, electrical activity, and neurotransmitter and neuropeptide release of SCN neurons. These processes result in the transmission of circadian timing signals to both passive targets and inherently rhythmic downstream oscillators throughout the brain and periphery.

The core molecular loop is entrained by a number of convergent SCN afferent pathways. Photic signals are transmitted from a specialized set of photoreceptive retinal ganglion cells by a dedicated neural pathway (the RHT) to the SCN, and activity in this pathway results in the release of glutamate as well as multiple peptidergic cotransmitters. Other major SCN afferents arise from the IGL and raphe nuclei, which form terminal fields that largely overlap the RHT terminal field in the SCN core. These afferents serve to regulate photic signaling in the SCN during the subjective night and to mediate the phase-shifting effects of nonphotic stimuli, including behavioral activity and arousal, during the subjective day.

Two revolutionary developments in the circadian clock field—(1) the elucidation of the core molecular clock machinery and (2) the demonstration that most, if not all, tissues and organs can themselves generate the 24-hour molecular clock—have greatly stimulated biomedical research on the importance of internal circadian timing for health and disease. Thus it can be argued that we are at the beginning of a new era in understanding human health and disease. With hundreds and even thousands of genes oscillating in most, if not all, tissues and organs, under the control of both central and local self-sustained circadian oscillations, normal health and well-being undoubtedly depend on the maintenance of proper internal synchronization. Furthermore, internal synchronization (or desynchronization) can occur on two levels, between separate oscillating systems (i.e., among tissue-level clocks within the organism) and within each self-sustained system (i.e., among cellular clocks within individual tissues), and the balance between health and disease is likely to be highly dependent on proper synchrony at both levels.

CLINICAL PEARL

Most, if not all, mammalian organs, tissues, and cells contain the core molecular circadian clock machinery and can produce circadian rhythms in vitro. This greatly heightens the theoretical importance of internal synchronization for normal physiologic function. Abnormal circadian timing between and within tissue and organ systems could be just as important as the overproduction or underproduction of key cellular mediators to overall health. Internal temporal dysfunction may thus be at the root of many physical and mental diseases. Although jet lag and shift work have been the primary ways in which circadian desynchrony were thought to occur, it is probable that circadian desynchrony at the cellular, tissue, and behavioral levels may play much more widespread roles in human medical and psychiatric pathologies.

SUMMARY

The master pacemaker for the mammalian circadian system is contained within the SCN, and the mechanisms underlying the pacemaker function of this structure have been elucidated at the molecular, cellular, and physiologic levels. The SCN contains a large number of coupled cellular oscillators that generate rhythmicity through the expression of a complex molecular feedback loop. The activity of the core molecular loop ultimately regulates SCN electrical activity and neurotransmitter and neuropeptide release, resulting in the transmission of circadian timing signals to downstream oscillators throughout the brain and body. These peripheral oscillator cells also express the core molecular clock mechanism, and hundreds or even thousands of genes exhibit circadian expression patterns under the joint control of central and local circadian oscillations in any given tissue. Thus normal health and well-being depend on the maintenance of proper internal synchronization both within and between tissues throughout brain and body.

ACKNOWLEDGMENTS

Preparation of this review was supported in part by NIH grant P01-AG011412.

Selected Readings

Antle MC, Silver R. Orchestrating time: arrangements of the brain circadian clock. *Trends Neurosci* 2005;**28**:145–51.

Colwell CS. Linking neural activity and molecular oscillations in the SCN. *Nat Rev Neurosci* 2010;**12**:553–69.

Gandhi AV, Mosser EA, Oikonomou G, Prober DA. Melatonin is required for the circadian regulation of sleep. *Neuron* 2015;**85**:1193–9.

Golombek DA, Rosenstein RE. Physiology of circadian entrainment. *Physiol Rev* 2010;**90**:1063–102.

Hastings MH, Reddy AB, Maywood ES. A clockwork web: circadian timing in brain and periphery, in health and disease. *Nat Rev Neurosci* 2003;**4**:649–61.

Ko CH, Takahashi JS. Molecular components of the mammalian circadian clock. *Hum Mol Genet* 2006;**15**:R271–7.

Morin LP. Neuroanatomy of the extended circadian rhythm system. *Exp Neurol* 2013;**243**:4–20.

Schmidt TM, Do MT, Dacey D, et al. Melanopsin-positive intrinsically photosensitive retinal ganglion cells: from form to function. *J Neurosci* 2011;**31**:16094–101.
Silver R, LeSauter J, Tresco PA, Lehman MN. A diffusible coupling signal from the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms. *Nature* 1996;**382**:810–13.
Welsh DK, Takahashi JS, Kay SA. Suprachiasmatic nucleus: cell autonomy and network properties. *Annu Rev Physiol* 2010;**72**:551–77.
Yoo SH, Yamazaki S, Lowrey PL, et al. PERIOD2:LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci U S A* 2004;**101**:5339–46.
Zelinski EL, Deibel SH, McDonald RJ. The trouble with circadian clock dysfunction: multiple deleterious effects on the brain and body. *Neurosci Biobehav Rev* 2014;**40**:80–101.
Zordan MA, Sandrelli F. Circadian clock dysfunction and psychiatric disease: could fruit flies have a say? *Front Neurol* 2015;**6**:80.

A complete reference list can be found online at ExpertConsult.com.

Chapter 35

Human Circadian Timing System and Sleep-Wake Regulation

Charles A. Czeisler; Orfeu M. Buxton

Chapter Highlights

- The circadian pacemaker (or biologic clock) confers endogenous rhythmicity with a period just slightly greater than 24 hours, persists in the absence of periodic changes in the external environment, and has timing or phase relative to the time of day that is genetically determined and influenced by environmental synchronizers.
- Under appropriate conditions, melatonin, body temperature, and many other physiologic processes can be used to assess circadian phase or biologic clock time.
- Although environmental light-dark schedules are the primary circadian synchronizer, other nonphotic stimuli such as exercise can shift circadian phase.
- The circadian pacemaker interacts with sleep-wake regulatory processes to influence many physiologic variables: hormone levels, autonomic nervous system activity, neurobehavioral performance, and the propensity for and timing and internal structure of sleep. Environmental, social, behavioral, and genetic factors, pharmacologic agents, and age influence most elements of this system.
- This chapter emphasizes for the benefit of the student and the practitioner the complexity of interactions of the circadian pacemaker and the sleep homeostat in regulating physiology, with important implications for health, performance, and clinical practice.

Circadian oscillations (or biologic clocks) are phylogenetically ubiquitous, found in species from prokaryotes to humans. Circadian clocks have several defining characteristics: endogenous rhythmicity that persists independent of periodic changes in the external environment, a near-24-hour period (*circadian* from Latin *circa* meaning "about" and *dies* meaning "day"), and the capacity for environmental input to modify or reset the timing or phase of the rhythm.[1,2] We provide an overview of the human circadian timing system and describe how this system interacts with sleep-wake regulatory processes to influence physiologic variables, including hormone levels, autonomic nervous system activity, neurobehavioral performance, and the propensity for, timing, and internal structure of sleep. We consider the influence of episodic and daily recurring behaviors, including sleep itself, on these physiologic variables relative to that of the endogenous circadian pacemaker.

IDENTIFYING THE MAMMALIAN CIRCADIAN PACEMAKER

In mammals, the suprachiasmatic nucleus (SCN) in the anterior hypothalamus is the central neural pacemaker of the circadian timing system. On the basis of careful patient histories characterized by disruptions of sleep-wake timing (e.g., insomnia, reversal of the sleep-wake schedule), Fulton and Bailey[3] postulated in 1929 that a region in the anterior hypothalamus appeared to regulate not the occurrence of sleep but its timing within the 24-hour day. In 1972, the SCN was identified as the site of the mammalian circadian pacemaker.[4,5] Physiologic studies show that multiple distributed circadian oscillators drive daily rhythms in peripheral systems.[6] Molecular research confirms the presence of peripheral clocks that use the same molecular machinery as the central circadian pacemaker in the SCN. Pacemakers like the SCN convey internal synchrony to these distributed oscillators.

INFLUENCE OF SLEEP AND CIRCADIAN RHYTHMS ON HUMAN PHYSIOLOGY

The discovery of the SCN's role as a central circadian pacemaker set the stage for understanding how it drives prominent circadian rhythms in a wide array of physiologic functions in humans synchronized to the 24-hour day and on a normal sleep-wake schedule (Figure 35-1, left column of panels).[7] Core body temperature is lowest and melatonin levels (not shown) are highest[8] during night sleep. Cortisol is low at habitual sleep onset but high at habitual morning wake time. When these endogenous circadian rhythms are entrained or synchronized to the 24-hour day, the temporal profile of each of these parameters exhibits a characteristic fingerprint that results from a combination of drives from the timing of the sleep-wake state to the endogenous circadian pacemaker and responses evoked by other factors such as posture, mood, exercise, and environmental lighting.[8]

To characterize the circadian pacemaker–driven component of a diurnal temporal profile from the effects of sleep-wake state, behavior, posture, and periodic environmental

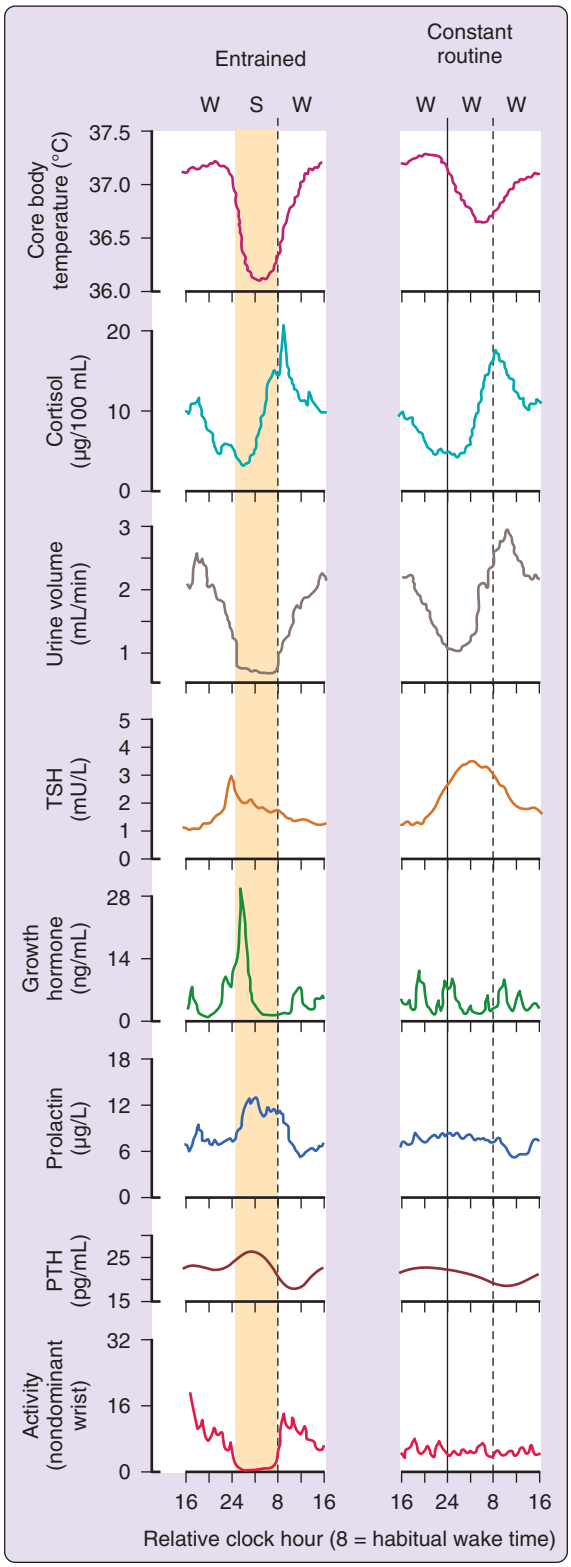


Figure 35-1 Comparison of temporal profiles of an array of physiologic and behavioral variables from participants studied under baseline conditions while maintaining a regular schedule of nocturnal sleep (S) (*yellow shaded area*) and daytime wake (W) at their habitual times (*left column of panels*) compared with profiles from participants under constant-routine conditions while maintaining a schedule of continuous wake in a semirecumbent posture (*right column of panels*). The *vertical dashed line* indicates habitual wake-up time during the week before the study, when participants were required to maintain a regular sleep-wake schedule. All data are from normal young men, 18 to 30 years old, studied under similar conditions. For a given variable, data in the *left panel* are from the same participants as data in the *right panel*; however, not all variables were monitored in the same participants. PTH, Parathyroid hormone; TSH, thyroid-stimulating hormone. (TSH data reproduced with permission from Allan JS, Czeisler CA. Persistence of the circadian thyrotropin rhythm under constant conditions and after light-induced shifts of circadian phase. *J Clin Endocrinol Metab* 1994;79:508–12, copyright The Endocrine Society. Prolactin data reproduced with permission from Waldstreicher J, Duffy JF, Brown EN, et al. Gender differences in the temporal organization of prolactin [PRL] secretion: evidence for a sleep-independent circadian rhythm of circulating PRL levels—a Clinical Research Center study. *J Clin Endocrinol Metab* 1996;81:1483–7, copyright The Endocrine Society. PTH data reproduced with permission from El Hajj Fuleihan G, Klerman EB, Brown EN, et al. Parathyroid hormone circadian rhythm is truly endogenous. *J Clin Endocrinol Metab* 1997;82:281–6, copyright The Endocrine Society.)

stimuli, the constant-routine protocol originally proposed by Mills and colleagues has been extended[9]; participants typically undergo continuous enforced wakefulness throughout day and night in a constant posture at a constant level of minimal physical activity and in constant, relatively dim, ambient illumination.[10] Under such conditions, the temporal profiles of many physiologic variables are significantly altered, and the components of these rhythms that are driven by the endogenous circadian pacemaker can be separated from those that reflect changes in the sleep-wake state, posture, or periodic external environment. Given the influence of posture[11] and the minimal influence of sleep[12] on the endogenous circadian melatonin rhythm, we have sometimes used a constant posture protocol in which participants are maintained in a constant semirecumbent posture in constant dim light but are allowed to sleep at night, so that endogenous circadian melatonin phase can be assessed.

Body temperature declines during sleep,[13-15] as illustrated by the profile of core body temperature recorded during a normal sleep-wake schedule and on a constant routine (see Figure 35-1, right column of panels). Sleep and changes in posture, light intensity, and activity level generate a drop in body temperature relative to wake.[6,15-19] This sleep episode–induced drop in body temperature combines with the circadian-driven decline in body temperature during the biologic night to yield a larger apparent amplitude than that of the endogenous circadian component alone (as measured on constant routine). Urine volume exhibits a robust oscillation under constant-routine conditions that is also influenced by sleep-wake state.[20]

Rhythmicity in some variables appears nearly independent of sleep-wake state. The temporal pattern of the hormone melatonin is relatively unchanged whether a participant is asleep or awake all night on a constant routine, although significant age-dependent effects of sleep and sleep deprivation on melatonin amplitude have been quantified.[12] Posture is reported to influence circulating melatonin concentrations somewhat.[11] Because the endogenous circadian cortisol rhythm is usually at its nadir at the time of habitual sleep onset, the overall profile of cortisol is relatively unchanged whether a person sleeps on a habitual schedule or remains awake all night, although cortisol levels will be elevated if the

person continues to remain awake throughout the following afternoon and evening.[21] However, suppression of plasma cortisol concentrations by deep slow wave sleep is evident whenever sleep onset occurs at the crest of the cortisol rhythm rather than at the nadir.[22]

Several other hormones are sensitive to sleep-wake state. Sleep opposes the circadian rhythm regulating thyroid-stimulating hormone (TSH), inhibiting TSH release during the peak of the endogenous circadian TSH rhythm, which would otherwise occur in the middle of the night.[23-25] Under entrained conditions, nocturnal TSH secretion is blunted by the timing of sleep, such that TSH levels are highest just before sleep onset and continue to be suppressed during the remainder of the sleep episode. This inhibitory effect of sleep on TSH secretion has been closely associated with slow wave sleep[26] and with relative delta power in the sleep electroencephalogram (EEG).[27] Growth hormone, prolactin, and parathyroid hormone levels all show a prominent sleep-dependent increase.[25] For growth hormone, a major sleep-related secretory episode is associated with slow wave sleep[28] and with relative delta power of the sleep EEG,[29] although such associations for prolactin, which remains elevated throughout the sleep episode, are controversial.[30] Interestingly, growth hormone levels blunted by acute sleep deprivation at night are increased during wakefulness the following day in sleep-deprived participants, compensating for the blunting of the major sleep-related pulse, such that average 24-hour levels are similar.[31] After sleep restricted to 4 hours per night from 1:00 to 5:00 AM for 1 week, growth hormone levels were maintained by the combination of a presleep, circadian-related secretory episode, together with a somewhat diminished sleep-related response.[32]

Leptin levels exhibit circadian rhythmicity, although the typical day-night pattern is reflected in the interaction of circadian rhythmicity with energy intake and expenditure[33] and sleep duration.[34] Ghrelin levels exhibit a day-night variation related to energy intake related to the presence of sleep[35] and to sleep duration.[36]

Ultradian variations in the release of renin from the kidney—a key factor in blood pressure control—are closely linked to the timing of the rapid eye movement (REM) and non–REM (NREM) sleep cycle,[37] an association evident even among patients with disturbed sleep, whose plasma renin profiles reflect pathologic changes in sleep structure. Increased relative delta power in the sleep EEG is associated with increased levels of plasma renin activity, whereas decreased slow wave activity is associated with a decrease.[38]

Even in the absence of sleep, prolactin and parathyroid hormone also have an endogenous circadian component that is lowest a few hours after habitual wake-up time,[25] and GH responses to exogenous growth hormone–releasing hormone exhibit a circadian rhythm.[39]

Effects of the interaction between sleep-dependent and circadian factors on these hormones can be important when the sleep-wake schedule is not synchronized with the circadian pacemaker. Shift workers who remain awake throughout the first night shift, for example, will secrete more TSH compared with that released at night under normal sleep-wake conditions during entrainment, owing to the absence of sleep-related suppression of TSH during the nocturnal peak of endogenous TSH secretory drive (see Figure 35-1); such increased secretion is not reversed during subsequent daytime sleep because the endogenous circadian rhythm of TSH secretion is already very low at that time. On the other hand, these workers would be deprived of their normally higher levels of growth hormone, prolactin, and parathyroid hormone during the waking night, although the sleep-related release of these hormones will recur during subsequent daytime sleep. Such alterations of the profiles of a variety of hormones have been documented in laboratory studies of night-shift workers.[40] In a laboratory simulation of circadian misalignment, participants exhibited increased leptin and increased glucose (despite increased insulin) and misalignment of the endogenous circadian rhythm of cortisol secretion with respect to the inverted sleep-wake schedule, along with the expected reduction in sleep efficiency.[41]

The circadian pacemaker significantly influences a variety of neurobehavioral and cognitive functions.[42-45] Under the conditions of the constant routine, participants display a circadian variation in short-term memory, cognitive performance, and alertness that is tightly coupled to the timing of the body temperature rhythm (Figure 35-2).[46] During a constant routine, these cognitive functions tend to be at their nadir shortly after habitual wake-up time due to an interaction between sleep loss and the circadian rhythms of performance.

EFFECTS OF LIGHT ON HUMAN CIRCADIAN RHYTHMS

The light-dark cycle is the primary environmental signal that synchronizes circadian systems in a wide array of species, including humans.[7,47] Nonvisual or non–image-forming retinal photoreception provides input to the circadian system, the pupillary light reflex, and other systems. Direct retinal input travels through the retinohypothalamic tract, a monosynaptic pathway by which information about the environmental light-dark cycle reaches the SCN. Postmortem studies reveal that the human brain contains the same key structural elements—the SCN and retinohypothalamic tract—as that of other mammals.[48] Neuropathologic studies associate damage to these structures with abnormalities in the timing of the sleep-wake cycle and other circadian rhythms.[49-51]

Studies in rodents and humans have shown that the three-cone system and rods, the visual photoreceptors, are not required for transmitting light signals to the circadian system.[7] A distinct set of ganglion cells in the inner retinal layer that project to the SCN are intrinsically photosensitive. Only the ganglion cells that project from retina to SCN selectively contain the vitamin A–based photopigment melanopsin.[52] Blue and short-wavelength green light (about 450 to 500 nm), which matches the sensitivity peak of melanopsin, is the most potent in shifting circadian phase in animals[53,54] and for melatonin suppression and phase-shifting responses in humans.[55] Both daytime and nighttime retinal exposure to such monochromatic blue (460 nm) light significantly improves reaction time, reduces attentional failures, and improves EEG correlates of alertness.[56] The magnitude and duration of the alerting effect of light at night depends on illuminance history and appears to be subject to sensitization and adaptation. The alerting response to light is greater and lasts longer when the light exposure occurred following prior exposure to dim light (1 lux) compared with ordinary indoor light (90 lux).[57] Within this specific set of intrinsically photosensitive retinal

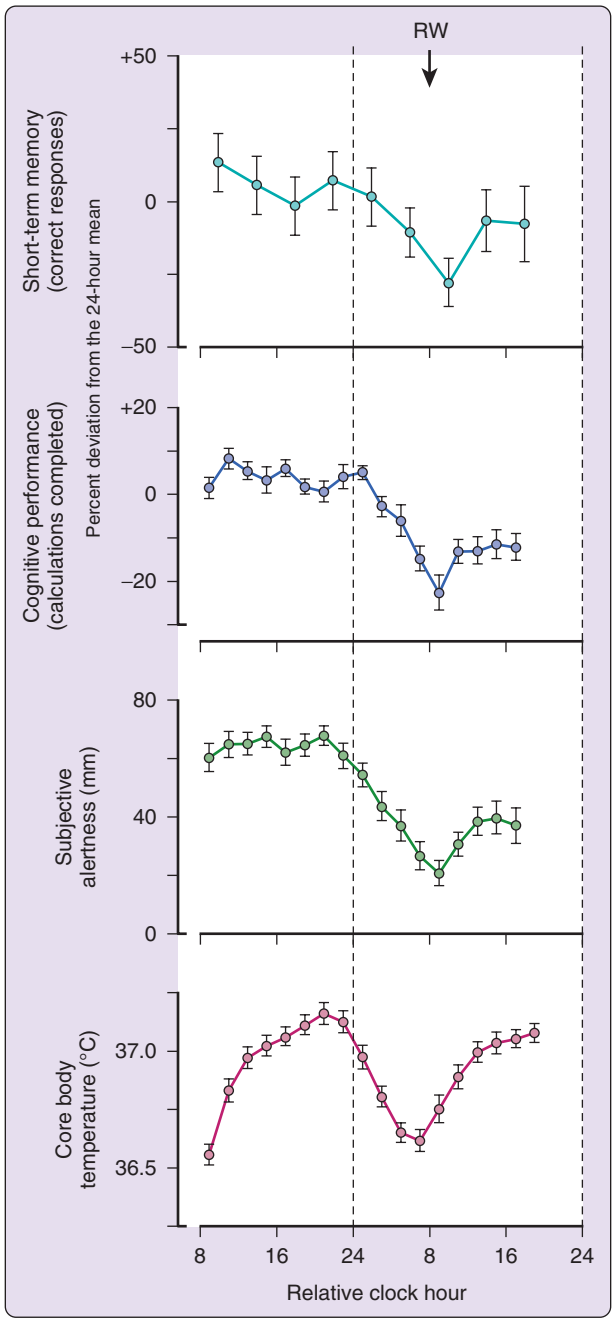


Figure 35-2 Daily patterns of short-term memory, cognitive performance, subjective alertness (mm), and core body temperature (°C) averaged across 18 participants during a 36-hour constant routine. Data collection times are normalized with respect to each participant's regular wakeup time (RW) (assigned a reference value of 8:00 AM and indicated by the *downward arrow*). The extent to which memory and performance scores deviated from the participant's 24-hour mean is averaged across participants. Data are expressed as percentages by which these absolute deviations differed from the participants' overall 24-hour mean score (assigned a reference value of zero). Each point is the centered mean (SEM) of all determinations made across a 2-hour interval for performance, alertness, and temperature and across a 4-hour interval for short-term memory. (Reproduced with permission from Johnson MP, Duffy JF, Dijk D-J, et al. Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *J Sleep Res* 1992;1:24–9.)

ganglion cells, melanopsin is the active photopigment. Rods and cones that synapse onto melanopsin-containing ganglion cells also participate, creating redundancy in circadian photoreception.[58]

A neural output pathway of the SCN passes through the intermediolateral cell column of the upper thoracic spinal cord, to the superior cervical ganglion that provides sympathetic input into the pineal gland. The absence of melatonin in patients who have cervical spinal cord injury is due to disruptions of this neural pathway to the pineal gland[59] and is associated with decreased sleep efficiency.[60]

Photic Suppression of Melatonin Secretion

The neural pathway from the SCN to the pineal provides for the regulation of the pineal output of melatonin by the SCN, including inhibition of melatonin release by retinal light exposure through a retinohypothalamic pathway[61] that can be used as an assay for the functional input of light into the circadian system.[62,63]

Preservation of light-induced melatonin suppression in otherwise totally blind people suffering from severe damage to the outer retina[63,64] led to the discovery that a distinct visual system mediates photic entrainment.[65,66] The nocturnal increase in melatonin is illustrated in Figure 35-3 (upper panel) for a normally sighted participant on a constant routine.[63] During a second peak of melatonin on the next night, a bright light stimulus induced an acute suppression of melatonin levels, which returned to elevated nighttime levels after light exposure was terminated. In the lower panel, bright light still suppresses melatonin even in a totally blind participant with no conscious light perception and a negative electroretinogram. The loss of conscious light perception does not necessarily indicate the loss of photic input to the circadian timing system,[63] although that is the case in most blind individuals without light perception. Two distinct visual systems exist: one for visual perception and a separate non–image-forming visual system for synchronization of the circadian pacemaker in the SCN, alerting input to the sleep switch in the ventrolateral preoptic area, suppression of melatonin secretion, and mediation of the pupillary light reflex.[52,67] Even in visually blind individuals, non–image-forming photoreception through intrinsically photosensitive retinal ganglion cells can trigger some awareness for light, which stimulates higher cognitive brain activity, independent of vision and in the absence of functional impact from the rods and cones, and can engage supplemental brain areas to perform an ongoing cognitive process.[51,68]

In natural-light-only conditions, the internal circadian clock is synchronized to solar time with melatonin onset near sunset and melatonin offset before wake time and after sunrise, at a significantly earlier circadian phase.[69] In contrast, evening reading from an electronic tablet that emits short-wavelength–enriched visible light delays endogenous circadian melatonin phase and the timing of REM sleep and increases evening alertness, sleep latency, and morning sleepiness compared with reading a printed book.[70] Taken together, these findings suggest that artificial light between dusk and dawn alters physiology through the non–image-forming visual system by shifting circadian phase, inhibiting sleep-promoting neurons, activating arousal-promoting orexin neurons in the hypothalamus, and suppressing melatonin. These effects of nocturnal artificial light in turn mask sleepiness, transiently increase

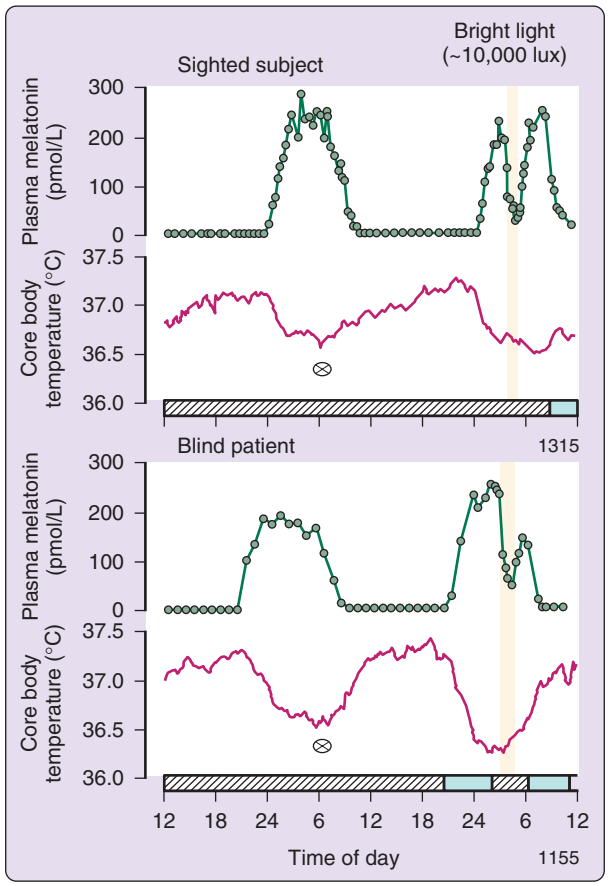


Figure 35-3 Melatonin-suppression test in a healthy sighted participant (*upper panel*) and a blind participant (*lower panel*). In each, plasma melatonin (*upper green traces*) and temperature (*lower red traces*) were measured repeatedly during a constant routine (*hatched bar*) and subsequent episode(s) of sleep (*solid blue bars*). The light intensity was ~10 to 15 lux during the constant routines, less than ~0.02 lux during the sleep episodes, and ~10,000 lux during 90 to 100 minutes of exposure to bright light (*open columns*) 22 to 23 hours after the initial fitted temperature minimum (*encircled Xs*). In both participants plasma melatonin concentrations decreased markedly in response to bright light and increased after the return to dim light. (Reproduced with permission from Czeisler CA, Shanahan TL, Klerman EB, et al. Suppression of melatonin secretion in some blind patients by exposure to bright light. *N Engl J Med* 1995;332:6–11.)

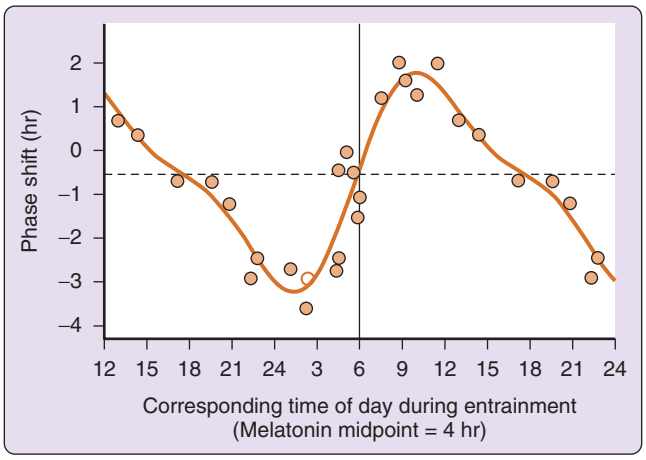


Figure 35-4 Phase-response curve to 6.7-hour light pulses in human participants. Phase shift in hours is plotted for a light pulse centered at different times relative to the initial endogenous circadian phase of the timing of melatonin secretion. By convention, phase advances to an earlier time are depicted as positive numbers, and phase delays as negative numbers. (Reproduced with permission from Khalsa SB, Jewett ME, Cajochen C, Czeisler CA. A phase response curve to single bright light pulses in human subjects. *J Physiol* 2003;549:945–52.)

alertness, and directly interfere with sleep, leading to chronic sleep deficiency.[71]

Human Phase-Response Curves to Light

In circadian biology, the phase-response curve (PRC) is used to characterize the synchronizing effects of light on a circadian pacemaker.[1,10,72,73] To construct a photic PRC, discrete light stimuli are applied systematically over the entire circadian cycle, and the magnitudes of light-induced phase shifts are plotted as a function of circadian phase at which the organism is exposed to the stimuli. In humans, measurement of the phase of endogenous circadian rhythms on a constant routine has been used to estimate both the initial circadian phase of the pacemaker before a stimulus and the final circadian phase after a stimulus, with the difference representing the phase shift.

All circadian systems exhibit a characteristic photic PRC, in which the largest light-induced phase shifts are generated in the biologic night. Phase delays are generated in response to light stimuli late in the biologic day and early in the biologic night, and phase advances are generated from stimuli in the late biologic night and early biologic day.[1] Figure 35-4 illustrates the PRC to a single pulse of light in humans. In humans, phase delays were observed in response to single, 1-hour and 6.7-hour bright light pulses applied before the minimum of the core body temperature cycle, which occurs on average about 2.3 hours before habitual wake-up time. Phase advances were observed when such light pulses were applied after the core body temperature nadir. The resultant human PRCs to single light pulses[74,75] exhibit the classic patterns of light PRCs in many organisms,[1,73] including phase advances and delays, and suggest that appropriate light intensities can shift the phase of the human pacemaker in morning and late afternoon or evening as well as at night. This has important clinical implications, such as use of phototherapy to reset circadian phase in delayed or advanced sleep phase disorder.

Photic Resetting of the Pineal Melatonin Rhythm

Because circadian rhythms are expressed in many physiologic and neurobehavioral variables, the phase of the pacemaker may be estimated by using any of these variables as a marker. In humans, the core body temperature rhythm is often a preferred marker of circadian phase, because it can accurately represent the underlying pacemaker's characteristics under certain conditions. However, melatonin can be an even more precise circadian marker,[76] which is less heavily influenced by sleep and posture.[77] In humans studied during a constant routine, melatonin reflects the phase of the underlying pacemaker following light-induced phase shifts better (less variability) than the endogenous component of the core body temperature rhythm.[77] Both rhythms shift equivalently whether to an earlier or a later hour.[76] Such studies demonstrate that the endogenous circadian melatonin rhythm can be reset to any desired phase within 2 to 3 days by light exposure.[76] Furthermore, photic stimuli designed to suppress the

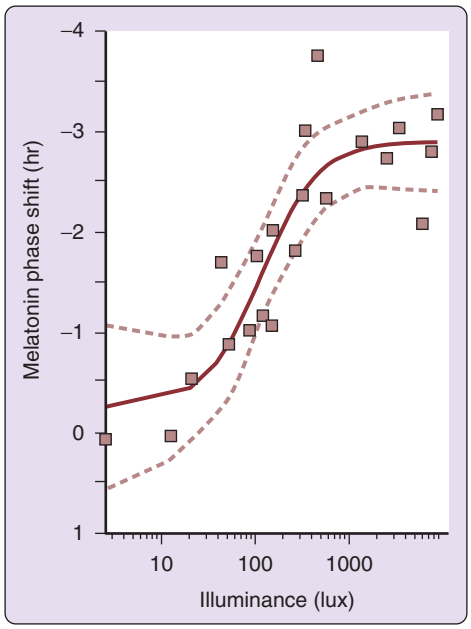


Figure 35-5 Illuminance-response curve of the phase-shifting effect of light on the human circadian pacemaker. Shifts in phase of the melatonin rhythm following the 6.5-hour light pulses, as assessed 1 day after the photic stimulus, are fit with a four-parameter logistic model, using a nonlinear least-squares analysis, that predicts an inflection point of the curve (i.e., sensitivity of the system) at ~120 lux; phase shifts saturate at ~550 lux. Data from individual participants represented by *closed boxes*, model by *solid line*, and 95% confidence intervals by *dashed lines*. (Modified with permission from Zeitzer JM, Dijk D-J, Kronauer RE, et al. Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *J Physiol [Lond]* 2000;526:695–702.)

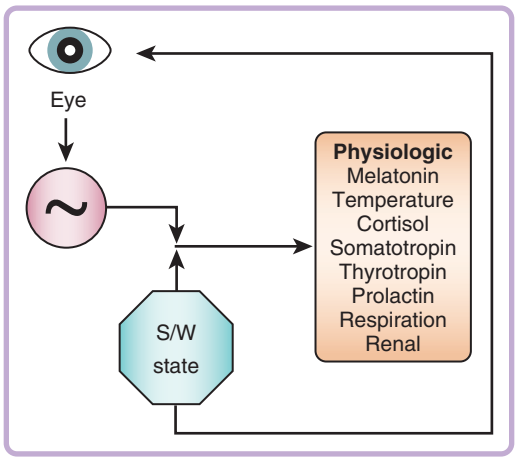


Figure 35-6 Schema illustrating influence of the circadian pacemaker (*circle with oscillator symbol*) and sleep-wake state (*octagon*) on several physiologic variables. Under normal conditions, the circadian pacemaker and sleep-wake state each influence these variables; relative contribution and nature of interaction (i.e., synergistic or oppositional) of each depends on the variable observed. Also illustrated are the influences of environmental illumination on the human circadian clock through the eye and of the sleep-wake state in determining timing of this illumination through behavioral action (i.e., switching off artificial indoor room lights, drawing bedroom window shades at bedtime, eyelid closure during sleep, and eyelid opening during waking).

amplitude of the endogenous circadian temperature cycle also suppress the amplitude of the endogenous circadian melatonin rhythm.[76]

The use of the melatonin rhythm as a circadian marker has additional practical advantages: melatonin in human saliva correlates well with that in plasma, and it allows the evaluation of circadian phase in patients with suspected circadian rhythm disorders or research participants relatively noninvasively.[78]

Human Dose-Response Curve to Circadian Phase-Resetting Effects of Light

In addition to dependence on wavelength and circadian phase, the degree of light-induced phase shift also depends on light stimulus intensity and consecutive days of exposure. Three consecutive daily pulses of light can generate a larger phase shift than a single light pulse. This intensity relationship also applies to brightness or illuminance level of light to which the retina is exposed. Following a 6.5-hour, single bright light stimulus of different intensities at a circadian phase known to generate a phase delay, an increase in resetting response is seen at 50 lux, with maximal slope at 100 lux and maximal shifts by about 550 lux (Figure 35-5).

The observation that ordinary room lighting of about 100 lux with only 1% of the intensity induces 50% of the resetting response to a 10,000 lux stimulus has important implications. We are exposed to bright light for a relatively short time each day,[79,80] but in modern industrialized societies we are exposed to ordinary indoor room light for many hours, a predominance of exposure that may have a greater impact on our circadian system than a few minutes of exposure to bright light. Phase-resetting and melatonin suppression responses to the resetting effects of evening light exposure are dose dependent and nonlinear; shorter light exposures (only 12 minutes long) more efficiently phase-shift the clock, suppress melatonin, and induce alertness than longer durations of retinal light exposure.[81]

Figure 35-6 illustrates the influence of the circadian pacemaker and the sleep-wake state on physiologic variables and the influence of light input through the eye to the circadian pacemaker. The feedback loop from the sleep-wake state to the eye represents the effects of exposure to the environmental light cycle because the sleeping state in humans is usually associated with eyelid closure and self-selected exposure to darkness, achieved by drawing window shades and switching off artificial light sources, whereas the waking state in humans is usually associated with opening of the eyelids and exposing the retina to light through self-selected use of artificial light or exposure to outdoor light during waking hours. Under a strict sleep-wake and light exposure schedule, the pacemaker's timing is consistent from day to day. However, whenever sleep is initiated late or terminated early, or a waking episode occurs within a sleep episode, the associated light exposure can reset the pacemaker. This association between waking and light exposure and the fact that low light intensity has a significant resetting effect on the pacemaker has practical relevance for routine sleep-wake scheduling and for understanding the influence of sleep disruption, which is often associated with light exposure, on circadian phase.

NONPHOTIC CIRCADIAN PHASE RESETTING AND REENTRAINMENT

Nonphotic input to the human circadian system is less well characterized than photic input. Results of early studies, which focused on social cues such as gong sounds or regularly

scheduled performance tests, meals, and bedtimes,[82] were confounded by limitations of phase measures used and self-selected lighting conditions. Exposure of healthy young men to nocturnal exercise of 1 to 3 hours' duration resulted in phase delays in nocturnal melatonin the following day.[83,84] Early evening 1-hour high-intensity exercise (at approximately 6:30 PM) elicited *phase advances* significantly different from the phase delays in response to morning, afternoon, and nocturnal exercise and in no-exercise participants (Figure 35-7). In a study of the facilitation of reentrainment to a delay-shifted sleep-wake episode in extremely dim light (to control for ambient light during exercise), exercise during the biologic night produced phase delays compared with no exercise.[84] Thus appropriately timed nonphotic stimuli such as exercise or other forms of arousal can facilitate adaptation to acute changes in light-dark cycle.

That appropriately timed exposure to exercise also results in phase advances is demonstrated by partial entrainment to a 23.5-hour light-dark and sleep-wake schedule in healthy volunteers exercising at moderate intensity twice daily (midday and late afternoon) over 2 weeks.[85] Participants exercising daily in late afternoon exhibited partial entrainment, advancing on average 10 minutes per day more than nonexercising controls, consistent with phase-advancing effects of late afternoon exercise on the human circadian clock. Given the slightly greater than 24-hour endogenous circadian period of humans and the net daily phase advance required for stable entrainment, evening exercise, particularly repeated daily

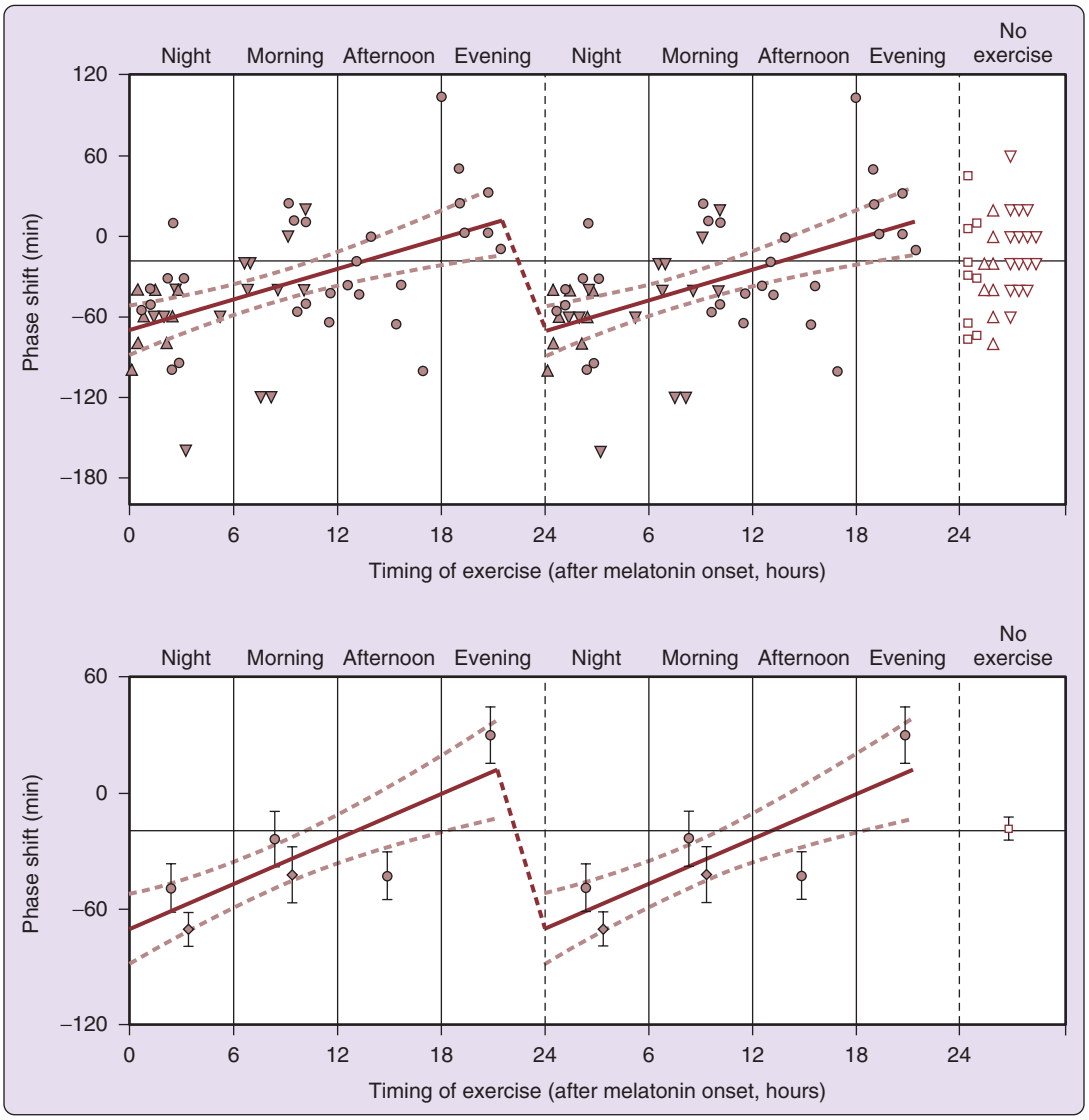


Figure 35-7 Phase-response curves in response to exercise at different circadian times of day. Phase delays were observed in response to nocturnal exercise; phase advances were observed in response to exercise during late afternoon or early evening. *Closed circles* indicate phase shifts in response to high-intensity, 1-hour nocturnal exercise and daytime exercise. *Upward* and *downward triangles* (*top panel*) and *diamonds* (*bottom panel*) indicate phase shifts in response to low-intensity, 3-hour exercise sessions. The *line* indicates a significant relationship between phase shifts and circadian time of exercise ($r^2 = 0.28$, $P = .0003$; slope significantly different from zero). *Dashed curves* indicate 95% confidence intervals of slope of line. (Reproduced with permission from Buxton OM, Lee CW, L'Hermite-Balériaux M, et al. Exercise elicits phase shifts and acute alterations of melatonin levels that vary with circadian phase. *Am J Physiol* 2003;284:R714–24. Copyright 2003 American Physiological Society.)

Figure 35-8 Professor Nathaniel Kleitman (*left*) attends to experimental equipment while fellow research participant Bruce Richardson lies in bed deep within Mammoth Cave in Kentucky where, for the first time, human participants were studied while shielded from periodic environmental changes on Earth's surface. The two pioneers lived on an imposed 28-hour sleep-wake schedule in these quarters from June 4 to July 6, 1938, in an effort to approximate uniform environmental and behavioral conditions, free from the influence of Earth's 24-hour day. In a 60-foot wide chamber free from any external environmental sounds, the temperature remained at 54°F (±1°), humidity approached complete vapor saturation, and darkness was absolute when the artificial light used during waking hours was shut off. The Mammoth Cave Hotel provided daily meals, which were consumed on awakening and after the 7th and 13th hour of each 19-hour waking day. (Photo courtesy of National Park Service, Mammoth Cave National Park, Mammoth, Kentucky; description adapted from Kleitman N. *Sleep and wakefulness*. Chicago: University of Chicago Press; 1963. p. 178–9.)

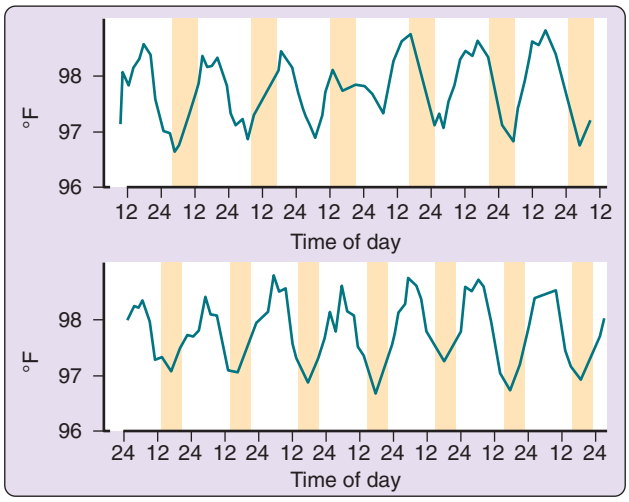


Figure 35-9 Weekly body temperature rhythms of a participant under two different routines of sleep and wake. *Top*, Data from participant K on a 28-hour daily routine of 19 hours' wake and 9 hours' sleep during Professor Kleitman's historic forced-desynchrony protocol (see Figure 35-8). Data based on last 3 weeks in Mammoth Cave. *Bottom*, Laboratory data recorded at the University of Chicago from participant K on his customary daily 24-hour routine of 17 hours' wake and 7 hours' sleep. *Shaded areas* indicate time in bed attempting to sleep. Data are from the 5 weeks after the 24-hour routine of living. Each weekly record shows seven body temperature waves, within minima in *shaded areas* on the customary 24-hour routine but not on the artificial 28-hour sleep-wake schedule. This participant's endogenous circadian temperature cycle maintained a near-24-hour oscillation, despite the scheduled 28-hour length of his sleep-wake cycle. Temperature data from participant R (not shown) appeared to adapt to the non-24-hour routine, something not observed in more recent forced-desynchrony studies. Interindividual differences in the strength of endogenous versus evoked components of the body temperature rhythm might account for what appeared to be circadian adaptation during forced desynchrony in participant R of Kleitman's pioneering experiment. (Figure and parts of legend adapted with permission from Kleitman N. *Sleep and wakefulness*. Chicago: University of Chicago Press; 1963. Copyright 1963 by the University of Chicago.)

exposure, could result in daily phase advances leading to nonphotic entrainment of the human circadian system if the timing and intensity of the exercise were optimized.

INVESTIGATING CIRCADIAN AND SLEEP-WAKE DEPENDENT MODULATION

The Kleitman Protocol

Separation from 24-Hour Environmental and Behavioral Cues

Nathaniel Kleitman was the first investigator to study human circadian rhythms in the absence of periodic 24-hour cues in the external environment (Figure 35-8).[42] Core body temperature records from one of his two participants in Mammoth Cave, Kentucky in 1938, who underwent a 28-hour imposed sleep-wake schedule, were compared with laboratory data collected at the University of Chicago from the same participant living on a 24-hour routine[42] (shown in Figure 35-9). On a 24-hour schedule, there were seven cycles of the body temperature rhythm, as one would expect over the course of a 1-week recording. The week with an imposed 28-hour schedule also has 7 cycles of body temperature rhythm, but only six sleep-wake cycles (see Figure 35-9, upper panel). Despite the confounding effect of sleep on core body temperature, this experimental protocol still separated the influence of timing of the sleep-wake schedule from that of the circadian pacemaker—at least in this participant.

This imposed desynchrony between sleep-wake schedule and output of the circadian pacemaker driving the temperature rhythm occurs when the non–24-hour sleep-wake schedule is outside the range of entrainment or range of capture of the circadian system. This protocol, termed the *forced-desynchrony* protocol, is useful in evaluating the influence of the circadian pacemaker on many physiologic variables because it allows separation of the confounding effect of the sleep-wake schedule from the output of the endogenous circadian pacemaker.[19,42,86] Figure 35-10 illustrates a raster plot of a forced-desynchrony experiment incorporating core body temperature and wake data for a participant living on a 28-hour day in a laboratory shielded from external time cues.[87] The waking episodes in this protocol are 18 hours, 40 minutes, followed by sleep episodes of 9 hours, 20 minutes. Core body temperature exhibited a period of 24.3 hours in this participant and was therefore desynchronized from both the 24-hour day and the timing of the imposed 28-hour sleep-wake schedule.

Separating Circadian Modulation and Sleep-Wake Modulation

The constant-routine protocol does not permit complete and unconfounded separation of the circadian and homeostatic influences on neurobehavioral and physiologic variables.

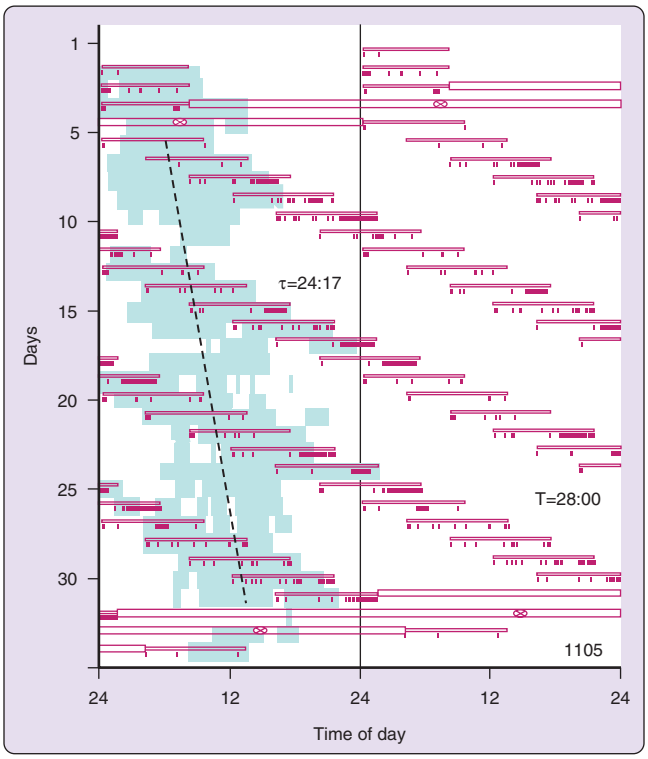


Figure 35-10 Double plot of a 28-hour forced-desynchrony protocol. Successive days are plotted next to and beneath each other. Scheduled sleep episodes are indicated by *narrow open bars*, polysomnographically determined wake within each sleep episode is indicated by *red tick marks* below the *narrow open bars*, and intervals during which core body temperature was below mean are indicated by the *blue area*. Intrinsic temperature cycle of 24.3 hours from this participant's data was estimated by nonparametric spectral analysis of core body temperature data during the forced-desynchrony part of the protocol. The *broken line* indicates the phase of circadian temperature rhythm minimum. The *encircled X* indicates the minimum of endogenous circadian rhythm of core body temperature unmasked by the 40-hour constant-routine protocol (*narrow open bars*). (Reproduced with permission from Dijk D-J, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci Lett* 1994;166:63–8.)

However, in the Kleitman forced-desynchrony protocol, sleep and wake are distributed much more evenly over the entire circadian cycle during the course of the experiment. It is thus possible to average data over either successive circadian cycles or successive sleep-wake episodes to separate these components. Averaging isolates the circadian profile of the variable of interest by removing the contribution of the confounding sleep-wake contribution in the averaging process. Conversely, the temporal contribution of the sleep-wake profile can be isolated from the confounding circadian influence. This averaging process is similar to that of cortical evoked potential recordings that effectively subtract background noise not temporally related to the evoked response.

Neurobehavioral Functions

To understand and predict the time course of neurobehavioral function, we must recognize the influence of the sleep-wake state on what is termed a *sleep homeostat*[19] driving neurobehavioral functions, as is apparent when we examine these variables during a longer course of sleep deprivation when more than one circadian cycle has elapsed.[46] The cyclic influence of the circadian pacemaker on alertness and performance is superimposed on an overall decline in function during the experiment, as described in models incorporating both homeostatic and circadian influences in the regulation of sleep and wake.[88,89] The rate of this performance decline increases sharply under conditions of chronic sleep restriction.

In a 20-hour forced-desynchrony protocol, the temporal profiles of cognitive performance and subjective sleepiness as a function of both circadian phase and time into the scheduled waking day[90] (Figure 35-11) suggest that the overall magnitudes of circadian and wake-dependent drives are similar during a typical waking day. From the timing of circadian and sleep-dependent profiles, we can qualitatively reconstruct their separate contributions to maintenance of alertness and performance over a normal waking day (see Figure 35-11). In the first half of the day following wake time, there is little homeostatic sleep drive because it was discharged by the prior sleep episode, so both alertness and cognitive performance are high. In the latter half of the waking episode, when homeostatic sleep drive would otherwise cause alertness and cognitive performance to decline, the circadian drive rises and opposes that decline, thereby sustaining a high, stable level of alertness throughout the normal waking day. Performance in the 3 hours before the onset of melatonin secretion (i.e., the wake maintenance zone) is significantly improved compared with performance during a 3-hour block earlier in the biologic day, despite a longer time awake. This effect is greater after extended wakefulness (i.e., on day 2 of a circadian rhythm), when homeostatic sleep pressure is high. The wake maintenance zone may therefore contribute to sleep-onset insomnia complaints when sleep timing is highly variable.[91] Remarkably, neurobehavioral performance, as measured by reaction time, can be preserved during this circadian wake maintenance zone even under conditions of chronic sleep restriction.[92]

Sleep and Wake

Similar dynamics apply for reconstructing the respective circadian and homeostatic contributions to sleep and wakefulness. The raster plot of the forced-desynchrony experiment (see Figure 35-10) shows that almost all wakefulness within a scheduled sleep episode occurs when the participant's sleep episode is not in phase with the body temperature nadir,[87] an observation first quantified by Kleitman from his Mammoth Cave data.[42] Averaging polysomnographically recorded sleep data from free-running participants on a self-selected cycle and in an environment free of time cues yields the data in Figure 35-12, showing the temporal profiles of sleep parameters as a function of circadian phase.[93] The circadian contribution to REM sleep timing is robust and exhibits a maximum centered just after the core body temperature nadir.

Figure 35-13 depicts sleep-dependent changes in the propensity for wake as a function of the circadian phase at which a sleep episode is initiated.[19] Under entrained conditions, a consolidated bout of sleep is maintained with minimal wake during the scheduled sleep episode by initiating the sleep episode at the end of the wake maintenance zone. However, when sleep is initiated in the early morning hours (as in a shift worker after the first night shift), a high percentage of time is spent in wake during the latter half of this intended sleep episode. During entrained conditions, homeostatic drive for

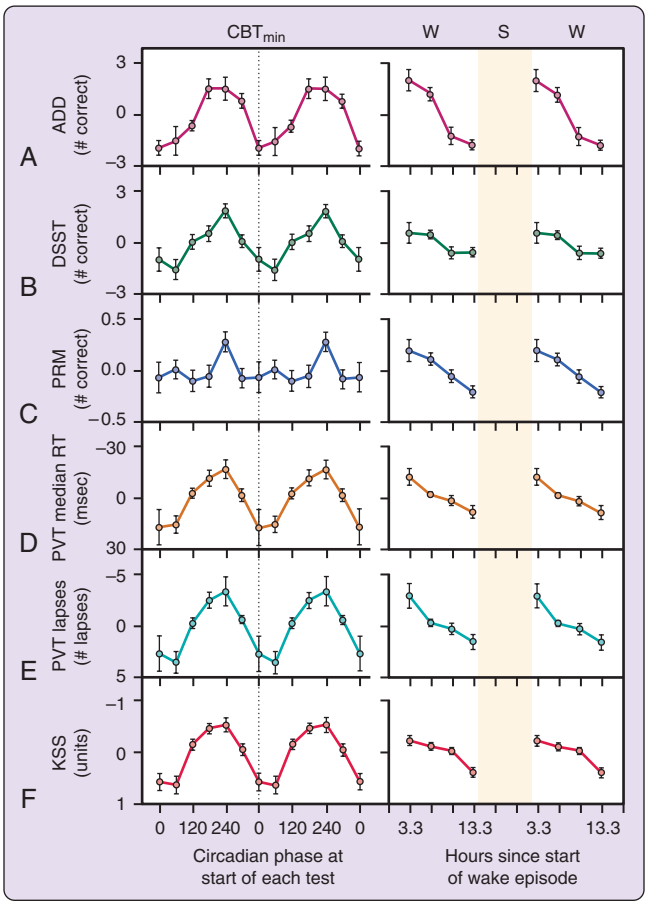


Figure 35-11 Double plots of the main effects of the circadian phase relative to the minimum of core body temperature CBT_{min} (*left panels*) and duration of scheduled wake (*right panels*: w = wake; s = sleep) on neurobehavioral measures. Plotted points show deviation from mean values during forced desynchrony and standard errors of the mean (SEMs). For all panels, values plotted lower in the panel represent impairment of that measure. Addition Task (ADD) **(A)**, Digit Symbol Substitution Task (DSST) **(B)**, and Probed Recall Memory (PRM) **(C)** scores were derived from total correct responses. Psychomotor Vigilance Task (PVT) results represent median reaction time **(D)** and total lapses (**E,** reaction times >500 msec). Karolinska Sleepiness Scale (KSS) scores **(F)** represent responses on this 1-9 Likert-type scale; higher scores represent greater sleepiness. (Reproduced with permission from Wyatt JK, Ritz-De Cecco A, Czeisler CA, Dijk DJ. Circadian temperature and melatonin rhythms, sleep, and neurobehavioral function in humans living on a 20-h day. *Am J Physiol* 1999;277:R1152–63.)

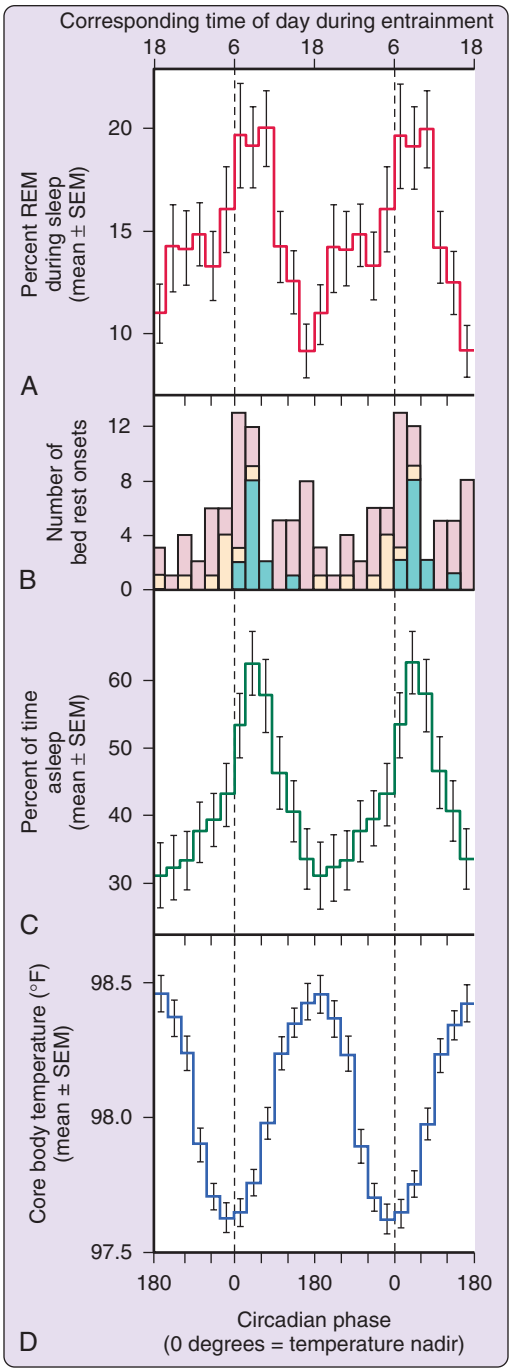


Figure 35-12 Variations in occurrence and internal organization of sleep with circadian temperature cycle phase (94 days of data from four participants). **A,** Percent of REM during sleep; **C,** Percent of time asleep; **D,** Core body temperature. In panel **B,** REM sleep episodes that occurred within 10 minutes after bed rest onset are indicated by *green* areas; those in which REM sleep episodes occurred within 30 minutes after bed rest onset are indicated by *peach* areas. (Reproduced with permission from Czeisler CA, Zimmerman JC, Ronda JM, Moore-Ede MC, et al. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep* 1980;2:329–46.)

sleep is greatest after an extended bout of wake at sleep onset and facilitates sleep in the first half of the night. In the latter half of the sleep episode, as the homeostatic drive declines, the circadian drive for sleep becomes greater, thus maintaining elevated sleep drive through the end of the sleep episode. These two components interact to facilitate consolidated sleep throughout the night.[19,87]

The three-dimensional representation in Figure 35-14 combines the temporal profiles of circadian and sleep-dependent drives to illustrate their respective contributions in maintaining wake.[19,87] A maximum in sleepiness quantified by slow rolling eye movements occurs at the endogenous circadian temperature nadir, which corresponds to occurring just before the habitual wake time. The sleep-dependent contribution exhibits an increasing profile over the wake episode, with the greatest propensity to slow eye movements after 14 hours of wakefulness. The magnitude of the circadian rhythms of sleepiness and performance increases with increasing homeostatic sleep drive. Thus, when increasing homeostatic sleep pressure combines with an adverse circadian phase, the drive for sleep is so great that slow eye movements and lapses of attention often intrude involuntarily during wake. The performance

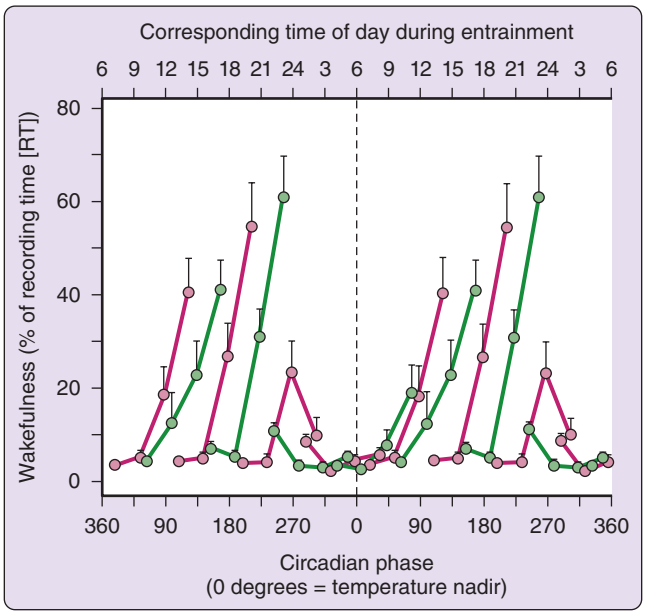


Figure 35-13 Wakefulness during scheduled sleep episodes (expressed as percentage of recording time) as a function of circadian temperature phase. Sleep episodes were assigned to twelve 30-degree bins based on circadian phase at lights out. (Modified with permission from Dijk D-J, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci* 1995;15:3526–38.)

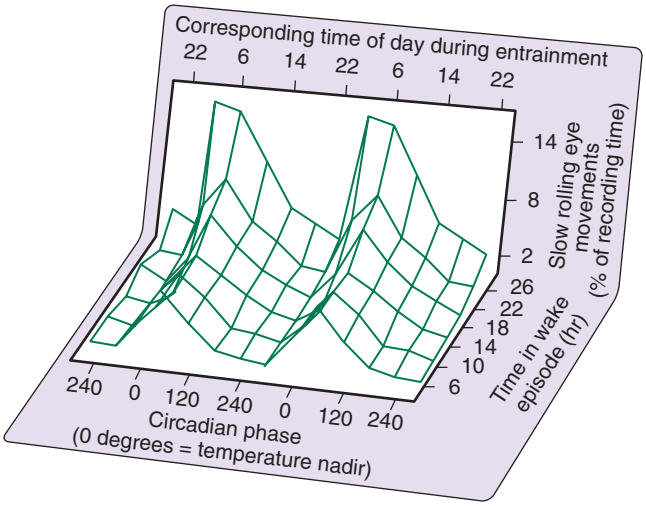


Figure 35-14 Quasi three-dimensional plot of slow rolling eye movements (SREMs) within scheduled wake episodes relative to circadian phase and time elapsed since start of wake episode. Data were assigned to twelve 30-degree circadian-phase bins and six 112-minute time-since-start-of-wake-episode bins. Each point represents SREMs expressed as percentage of recording time in a bin. (Modified with permission from Cajochen C, Wyatt JK, Bonikowska M, et al. Non-linear interaction between circadian and homeostatic modulation of slow eye movements during wakefulness in humans. *J Sleep Res* 2000;9:58.)

impairment is an order of magnitude more severe when extended wakefulness coincides with an adverse circadian phase under conditions of chronic sleep restriction.[92]

The circadian and sleep-wake modulation of sleep-wake propensity and neurobehavioral function are illustrated schematically in Figure 35-15. Experimental evidence indicates that a simple additive model cannot account for variations in alertness and cognitive performance data.[44,46,94] In fact, when

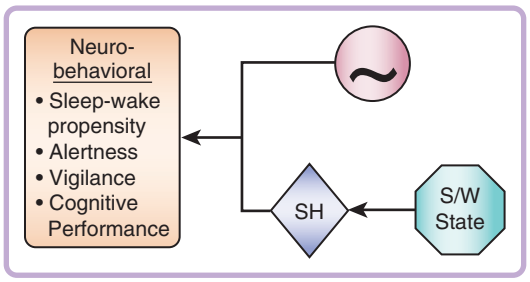


Figure 35-15 Schema illustrating combined influence of circadian clock and sleep-wake state on neurobehavioral variables (sleep-wake propensity, alertness, vigilance, and cognitive performance). Sleep-wake state (S/W State) influence is illustrated by an intermediary of the sleep homeostat (*SH in blue diamond*).

averaged across all circadian phases, there is relatively little circadian variation in various waking neurobehavioral measures in the first few hours of wakefulness, when homeostatic sleep drive is low, and the circadian contribution increases as a function of number of hours awake, suggesting that homeostatic and circadian drives are not independent and further suggesting a nonadditive interaction between homeostatic and circadian systems that drive alertness and cognitive performance.[95] Furthermore, the buildup of the homeostatic drive in response to acute sleep deprivation is distinct from the response to chronic sleep loss.[92]

Internal Sleep Structure

REM sleep propensity varies with circadian phase.[19,93,96] Studies with nap opportunities evenly distributed throughout day and night every 1.5 to 3 hours first established the REM sleep propensity rhythm in a protocol that did not involve concomitant variations in prior wake length.[93,97,98] REM sleep latency, the rate of REM sleep accumulation, REM sleep episode duration, and REM sleep propensity were then shown to vary with phase of the endogenous circadian temperature cycle in free-running participants whose self-selected rest-activity cycle spontaneously desynchronized from the timing of the endogenous circadian temperature cycle (see Figure 35-12).[93,99] The peak of the endogenous circadian rhythm in REM sleep propensity in these participants was just after the nadir of the endogenous component of the circadian temperature cycle, coincident with the circadian peak of sleepiness and sleep propensity (see Figures 35-11 to 35-14).[93,99] During such spontaneous desynchrony, free-running participants who chose to go to bed near the peak of the REM sleep propensity rhythm usually exhibited sleep-onset REM sleep episodes,[93,99] an otherwise rare phenomenon normally diagnostic of narcolepsy. Under these conditions, the density of REMs per minute of REM sleep exhibits a sleep-dependent variation apparently dissociated from the REM sleep propensity rhythm itself.[100]

These findings on the timing of the circadian REM sleep propensity rhythm have since been confirmed and extended with polysomnography data from participants studied in the forced-desynchrony protocol.[19,87] Because sleep episodes in the forced-desynchrony protocol always begin after a fixed duration of enforced wakefulness, the results were less subject to the confounding effects of systematic variations in prior wake durations characteristic of spontaneous desynchrony. Furthermore, because participants were scheduled to remain in bed for a fixed interval on the forced-desynchrony protocol,

results were not confounded by self-selected termination of the sleep episode, although circadian variations in sleep efficiency prevent complete elimination of this confounding factor. Nonetheless, under such conditions, the twofold circadian variation in REM sleep propensity again peaked just after the nadir in the endogenous circadian component of body temperature rhythm, within each one fifth of the scheduled sleep episode, notwithstanding the average sleep-dependent increase in REM sleep propensity. A sleep-dependent increase in REM sleep propensity independent of circadian phase was also quantified. A significant nonadditive interaction between circadian phase and time since the start of the sleep episode was found from the REM sleep data collected during forced desynchrony.[19]

With the forced-desynchrony protocol, significant and substantial circadian and sleep-dependent variations in NREM sleep propensity were also observed, whereas the robust sleep-dependent decline in slow wave activity was associated with only a small but statistically significant variation of slow wave activity as a function of circadian phase.[19] Similar circadian variations in internal sleep structure are documented in a blind patient whose circadian pacemaker was not synchronized to the 24-hour day, despite his decades-long maintenance of a regular sleep-wake schedule.[101] Such blind patients, in essence, live in society on the biologic equivalent of a forced-desynchrony protocol because the 24-hour day is outside the range of entrainment of the circadian pacemaker in such patients.

Quantitative analysis of the sleep EEG reveals circadian variations in EEG activity during NREM and REM sleep,[102] with low-frequency sleep spindle activity in NREM sleep paralleling the endogenous circadian melatonin rhythm. Overall, these data indicate that the timing and internal structure of sleep are profoundly dependent on an interaction between robust circadian and homeostatic regulatory factors, with circadian factors predominant in the regulation of REM sleep and with sleep-dependent factors predominant in the regulation of slow wave sleep.

Potential Feedback Pathways

As is typical of physiologic regulatory systems, feedback pathways play a significant role in this system. The neurobehavioral variables influenced by the circadian pacemaker and the sleep homeostat can influence the sleep-wake state through the influence of wake and sleep propensity on determination of sleep and wake times. For example, a sleep episode is more likely to be initiated following a rise in sleep propensity to a high level during an extended waking episode, and a sleep episode is more likely to be terminated following a decline in sleep propensity to a low level over the course of a sleep episode. This influence on behavior, in turn, influences the level of the sleep homeostat and, because of associated changes in light exposure and activity, it can affect the phase or amplitude (or both) of the circadian pacemaker.

There may be another important feedback pathway in this system. Studies demonstrating that melatonin receptors can be found on cells within the human SCN draw attention to a potential feedback pathway from the pineal gland to the SCN through circulating melatonin. Several physiologic studies suggest that exogenous melatonin has a phase-resetting effect on the human circadian pacemaker, and both a melatonin PRC and dose-dependent phase shifting have been reported. The first such study that has thoroughly controlled for retinal light exposure has revealed that the resetting responses to melatonin are even greater than previously reported.[103] There is also great interest in the potential efficacy of melatonin as a hypnotic because the sleep-promoting effects of exogenous melatonin depend on circadian phase, as established in young adults on a forced-desynchrony protocol.[104]

Examination of the temporal profile of endogenous melatonin secretion during the forced-desynchrony protocol shows a daily circadian increase in melatonin levels coincident with a decrease in wake (Figure 35-16).[102] This melatonin rise might open a gate that allows sleep to occur.[105] These data suggest feedback from the pineal gland to both the circadian pacemaker and neurobehavioral variables involved in regulating the sleep-wake state. Other physiologic systems might also affect sleep-regulating mechanisms, such as an effect of growth hormone on sleep.[106,107]

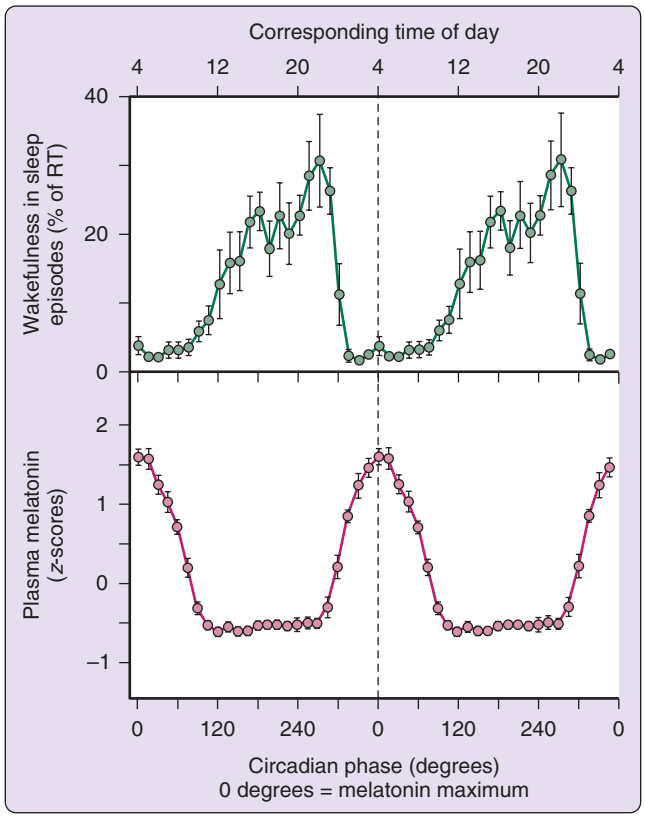


Figure 35-16 Phase relationships between endogenous circadian rhythms of wakefulness and plasma melatonin were assessed during a forced-desynchrony protocol. Data were plotted against the circadian phase of the plasma melatonin rhythm (0 degrees on the lower abscissa scale corresponds to fitted melatonin maximum). To compare with entrained conditions, the upper abscissa scale indicates the approximate clock time corresponding to circadian melatonin rhythm during the first day of the forced-desynchrony protocol (i.e., immediately on release from entrainment). Plasma melatonin data are expressed as *z*-scores to correct for interindividual differences in mean values. Wakefulness is expressed as a percentage of recording time (RT). Data are double plotted (i.e., all data plotted left of the *dashed vertical line* are repeated to the right of the *vertical line*). (Modified from Dijk D-J, Shanahan TL, Duffy JF, et al. Variation of electroencephalographic activity during non–rapid eye movement and rapid eye movement sleep with phase of circadian melatonin rhythm in humans. *J Physiol [Lond]* 1997;505[3]:851–8.)

Intrinsic Period of the Human Circadian Pacemaker

Early human studies were performed in the absence of environmental synchronizers but were confounded by self-selected exposure to ordinary room light, which led to the erroneous conclusion that the average period of the human circadian pacemaker was 25 hours.[47,86] Participants were permitted to illuminate their living quarters while awake and switch off lighting while asleep because these experiments were predicated on the incorrect belief that ordinary room light had no resetting effect on the human circadian system.[108,109] Thus results of these studies were systematically compromised by this retinal light exposure.[47,86] Recognition of this confounding effect led to the use of Kleitman's forced-desynchrony protocol to assess the intrinsic period of the human circadian pacemaker.[47,86,110]

Subsequent studies using the Kleitman forced-desynchrony protocol have controlled the intensity of background illumination and timing of exposure to the light-dark cycle.[47,86] With this forced-desynchrony protocol in participants living in dim light (about 10 to 15 lux) on either a 28-hour or a 20-hour schedule, the average intrinsic period of the circadian pacemaker is estimated to be much closer to 24 hours, with an average period of about 24.15 hours[110,111] rather than 25 hours (Figure 35-17). This holds true for all of the circadian markers tested—core body temperature, melatonin, and cortisol—and is consistent with results of other studies under a variety of protocols.[112-115] The average intrinsic period of the human circadian pacemaker is significantly shorter in women [24.09 ± 0.2 hour (24 hours, 5 minutes ± 12 minutes)] than in men [24.19 ± 0.2 hours (24 hours, 11 minutes ± 12 minutes)].[111] Entrainment studies have functionally confirmed the intrinsic circadian period in humans to be near 24 hours because a weak stimulus (candlelight during the scheduled day and sleep in darkness during the scheduled night) can entrain most people to a light-dark cycle with an imposed period of 24.0 hours but not an imposed period of 23.5 or 24.6 hours.[116]

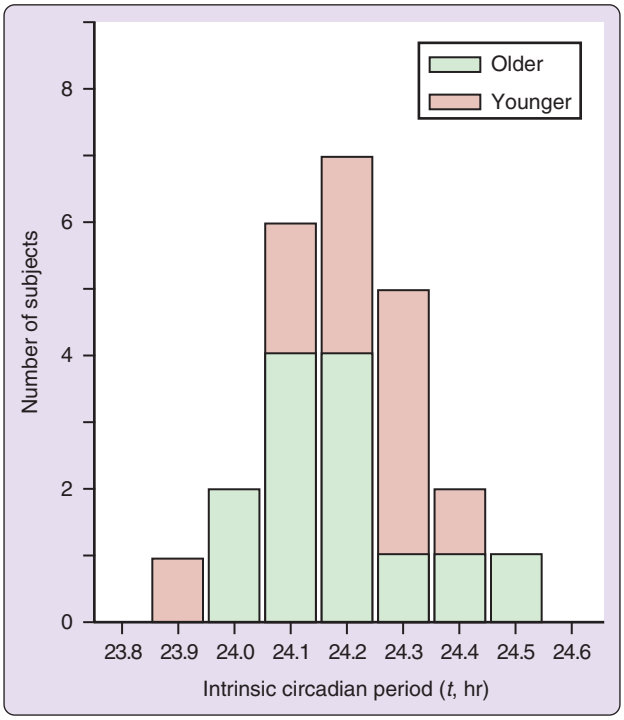


Figure 35-17 Histogram of intrinsic circadian period estimates derived from young and older participants. *Green bars* indicate intrinsic circadian period (*t*) estimates of older participants; *pink bars* indicate young participants. Each participant's estimated intrinsic circadian period is reported as the average of estimated periods from his or her core body temperature, melatonin, and cortisol rhythms. (Reproduced with permission from Czeisler CA, Duffy JF, Shanahan TL, et al. Stability, precision, and near–24-hour period of the human circadian pacemaker. *Science* 1999;284:2177–81.)

AGING AND CIRCADIAN SLEEP-WAKE REGULATION

Aging also has a pervasive influence on many aspects of the circadian and sleep-wake regulating system.[117-123] The prevalence of disrupted sleep complaints is much greater in older than in young people. In fact, 57% of people in the United States older than 65 years complain of at least one chronic sleep problem, 43% complain of difficulty initiating or maintaining sleep, and 19% complain of awakening too early in the morning.[124] Key to examining this question is the extent to which the circadian pacemaker or the sleep homeostat is involved in these age-related changes.

On average, the circadian clock is set to an earlier hour, and the amplitude of some endogenous circadian rhythms is lower in older people than it is in young adults (Figure 35-18).[119,125] However, intrinsic circadian period does not shorten with age in healthy humans.[110,111,126] Importantly, young participants can sleep over a much wider range of circadian phases than older people, who awaken spontaneously at an earlier internal circadian phase.[125,127] Because older people usually awaken at an earlier circadian phase, they are typically exposed to light at an earlier hour; this earlier light exposure, which will reset the circadian pacemaker to an earlier hour, likely accounts for the earlier average entrained circadian phase observed in older people.[125] Remarkably, older participants are much less vulnerable to the adverse effect of sleep loss and misalignment of circadian phase on neurobehavioral performance.[128] At the same time, a common clock gene polymorphism is associated with the timing of activity rhythms and with the time of death, such that rs7221412GG individuals have a mean time of death nearly 7 hours later than rs7221412AA/AG individuals.[129]

INFLUENCE OF SOCIAL AND WORK-RELATED FACTORS

The self-selection of sleep and wake times in humans is another important factor in the sleep-wake regulatory system. Although circadian and homeostatic drives for sleep influence the choice of sleep and wake times through a feedback pathway, social factors (e.g., child care, school and work responsibilities, entertainment, social interaction) and environmental factors (e.g., noise, artificial light, alarm clocks) often override those biologic determinants. In contrast, animal behavior exhibits activity and sleep predictable enough to be used as markers of the biologic time of day. Humans, especially since the advent of alarm clocks and artificial lighting, can and do override the signals from the circadian and sleep-wake regulating system and freely decide to stay awake later than they otherwise would or wake up earlier than they would

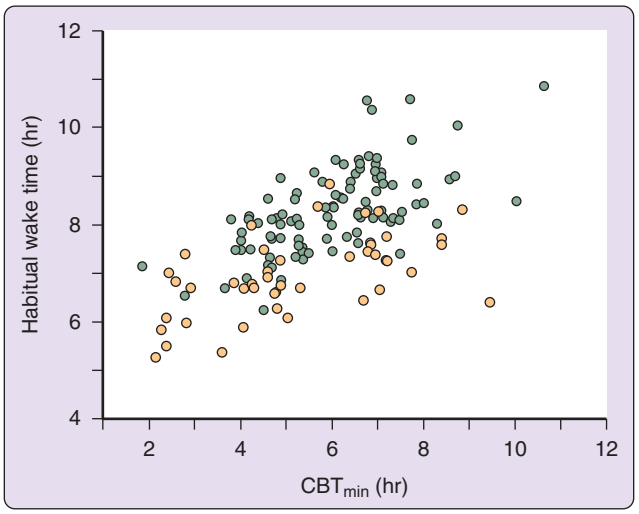


Figure 35-18 Habitual wake time versus endogenous circadian phase of young and older adults. Symbols represent average self-reported wake time from the pre-study week versus the phase of core body temperature minimum (CBT_{min}) for each participant. *Pink circles* indicate older participants ($n = 44$); *green circles* indicate young participants ($n = 101$). Between-groups relationship is significant (slope of older participants, 0.266 ± 0.06; slope of young participants, 0.471 ±0.05). (Reproduced with permission from Duffy JF, Dijk D-J, Klerman EB, Czeisler CA. Later endogenous circadian temperature nadir relative to an earlier wake time in older people. *Am J Physiol* 1998;275:R1478–87.)

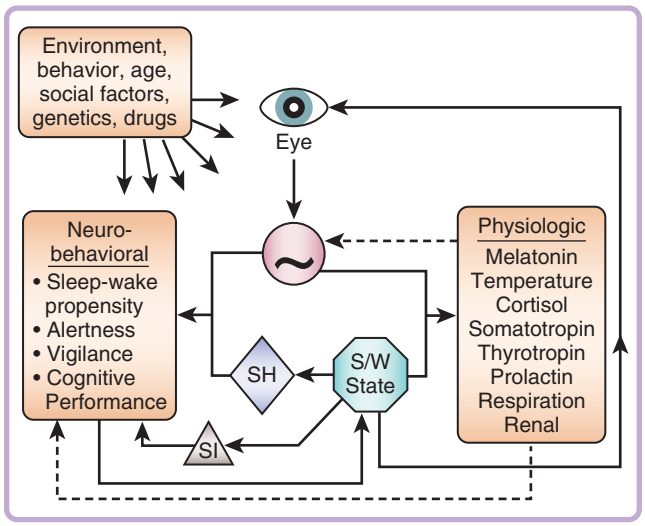


Figure 35-19 Overall schema illustrating potential influence of the circadian clock on neurobehavioral and physiologic variables and of physiologic variables (e.g., melatonin, body temperature) on neurobehavioral variables or circadian clock (*dashed arrows*), feedback influence of variables on sleep-wake state (S/W State) (*solid arrow*), influence of sleep-wake state on variables through sleep homeostat (SH) and sleep inertia (SI) (*solid arrows*). Global influence of environment, behavior, age, social factors, genetics, and drugs on virtually all elements contributing to sleep-wake regulation are also represented.

spontaneously because of job or school requirements or recreational and social events. Thus modern humans may be more sleep deprived than their ancestors.[130] The long-term consequences of this relatively recent trend in industrialized society are unknown. Yet the modern consolidated sleep episode is significantly different from that under more naturalistic conditions in which sleep episode duration was determined by longer length of darkness in the natural winter environment.[69,130,131] The most conspicuous example of this self-selection is rotating shift work, when people choose to work in direct opposition to the modulation of circadian and homeostatic regulatory systems, resulting in internal temporal dissociation, fragmented sleep, and impaired wake. This does have consequences. Resident physicians working extended-duration work shifts (24 to 30 hours) every third day showed progressive deterioration in performance across 3 successive weeks of work in an intensive care unit. Response times deteriorated with time on shift and cumulatively, demonstrating chronic sleep deficiency with each successive extended-duration work shift.[132] Prolonged exposure to prolonged sleep restriction with concurrent circadian disruption in controlled laboratory conditions decreases resting metabolic rate and increases plasma glucose concentrations after a meal, an effect resulting from inadequate pancreatic insulin secretion, suggesting that prolonged sleep restriction with concurrent circadian disruption alters metabolism and could increase the risk for obesity and diabetes.[133] Therefore, in any realistic model of the human circadian and sleep-wake regulatory system used to manage performance and prevent disease, such environmental and societal influences must be recognized.

We can assemble an overall system schema and incorporate the feedback pathway of neurobehavioral variables on sleep-wake state and the putative feedback pathways from melatonin and other physiologic variables onto the circadian pacemaker and neurobehavioral function (Figure 35-19). This schema incorporates the global influence of environmental, social, behavioral, genetic, pharmacologic, and age as factors influencing all elements of this system, together with the decrements in neurobehavioral performance and alertness that immediately follow the sleep-to-wake transition, a phenomenon called *sleep inertia*. The time course of sleep inertia has been shown to persist for up to 2 hours after a long sleep episode, and it is most profound within the first few minutes after awakening.[134,135] Although the final schema incorporates much of what is known about the roles of the circadian pacemaker and sleep homeostat in regulating sleep, it is not intended to completely represent all factors involved in the regulation of sleep, which is beyond the scope of this chapter. Nevertheless, its strength is its use in understanding the interplay between circadian and homeostatic drives and perhaps as a framework for initiating future scientific inquiry.

CLINICAL PEARL

Disruption of the circadian timing system by shift work, travel across time zones, and the activities of our round-the-clock society has an adverse effect on many physiologic variables including hormones, glucose, metabolism, autonomic nervous system activity, the immune system, inflammation, neurobehavioral performance, and the propensity for, timing, and internal structure of sleep. This disruption can result in patient complaints of insomnia or excessive daytime sleepiness and can impair overall health.

SUMMARY

The circadian pacemaker interacts with sleep-wake regulatory processes to influence many physiologic variables: hormone levels, autonomic nervous system activity, neurobehavioral performance, and the propensity for and timing and internal structure of sleep. Environmental, social, behavioral, and genetic factors, pharmacologic agents, and age influence most elements of this system. Under ordinary circumstances, when people sleep at night in darkness and are awake in daylight, it is difficult to distinguish the relative contributions of the sleep homeostat and that of the circadian pacemaker to a given recurrent daily characteristic, symptom, or disorder of sleep or wakefulness (e.g., narcolepsy, delayed sleep phase syndrome). A pathologic event, such as a nocturnal seizure, that occurs the same time each night may be driven by the circadian pacemaker, the sleep homeostat, a specific sleep stage, or some combination of these processes. It is currently possible, although difficult, to experimentally dissociate these factors for research purposes (e.g., using the forced-desynchrony protocol in humans or suprachiasmatic lesions in animals). Clinically feasible techniques, such as measurement of dim-light salivary melatonin onset, can provide useful information about circadian phase.[136] Continued basic and clinical research is needed to assess the impact of the complex interaction of sleep and circadian rhythmicity on sleep disorders, such as insomnia, and overall health and well-being.

ACKNOWLEDGMENTS

Research conducted in the Division of Sleep and Circadian Disorders has been supported by grants NIA-1-R01-AG06072, NIA-P01-AG09975, and NIA-1-U01-12642 from the National Institute on Aging; grant NIMH-1-R01-MH45130 from the National Institute of Mental Health; grants T32-HL007901, R01-HL52992, R01-HL077453 and R01-HL107240, R01-HL095472, R01-HL94654, and U01-HL111478 from the National Heart Lung and Blood Institute; grant R01-GM105018 from the National Institute of General Medical Sciences, grants NASA NCC9-119, NASA NAG9-524 and NASA NAG5-3952, and NASA NNX10AF47G from NASA; the NASA through the NASA Cooperative Agreement NCC 9-58 and HFP01601 with the National Space Biomedical Research Institute; grant FEMA EMW-2010-FP-00521 from the Federal Emergency Management Agency, and grant numbers 1 UL1 RR025758-01 and M01-02635, Harvard Clinical and Translational Science Center from the National Center for Research Resources and by the Brigham and Women's Hospital.

Selected Readings

Anderson C, Sullivan JP, Flynn-Evans EE, et al. Deterioration of neurobehavioral performance in resident physicians during repeated exposure to extended duration work shifts. *Sleep* 2012;**35**(8):1137–46.

Buxton OM, Cain SW, O'Connor SP, et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. *Sci Transl Med* 2012;**4**(129):129–43.

Chang AM, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci U S A* 2015;**112**(4):1232–7.

Czeisler CA. Perspective: casting light on sleep deficiency. *Nature* 2013; **497**(7450):S13.

Dijk DJ, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci Lett* 1994;**166**(1):63–8.

Scheer FA, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc Natl Acad Sci U S A* 2009;**106**(11):4453–8.

Smith-Coggins R, Broderick KB, Marco CA. Night shifts in emergency medicine: the American Board of Emergency Medicine longitudinal study of emergency physicians. *J Emerg Med* 2014;**47**(3):372–8.

Wright KP Jr, McHill AW, Birks BR, et al. Entrainment of the human circadian clock to the natural light-dark cycle. *Curr Biol* 2013;**23**(16): 1554–8.

A complete reference list can be found online at ExpertConsult.com.

Sleep Homeostasis and Models of Sleep Regulation

Peter Achermann; Alexander A. Borbély

Chapter 36

Chapter Highlights

- Sleep homeostasis refers to the aspect of sleep regulation dependent on sleep and waking, as homeostatic mechanisms counteract deviations from an average reference level of sleep. These mechanisms augment sleep propensity when sleep is curtailed or absent, and they reduce sleep propensity in response to excess sleep. In general, homeostasis can be defined as "the coordinated physiological processes, which maintain most of the steady states in the organism."[1,2]
- Non–rapid eye movement sleep is not a homogeneous substate of sleep but can be subdivided according to the predominance of slow waves in the electroencephalogram (EEG). One of the most important functional EEG parameters is referred to as slow wave activity; It is equivalent to delta activity and encompasses components of the EEG signal in the frequency range of approximately 0.5 to 4.5 Hz as obtained by spectral analysis. Under physiologic conditions, this EEG variable can be regarded as an indicator of sleep depth or sleep intensity.[7]
- The two-process model postulates that a homeostatic process rises during waking and declines during sleep and interacts with a circadian process that is not directly dependent of sleep and waking. The model served as a conceptual framework and inspired many experiments to test its predictions.[3,27,83]
- Models may address processes at different levels (from the microscopic [cellular] level to the macroscopic [systemic] level) and at different time scales (from the range of milliseconds or seconds up to hours or days).[74] Many mathematical models of sleep regulation represent extended versions of the two-process model incorporating neurophysiologic processes. Being aware of the power and limitations of models is important for selecting the most appropriate one for the question to be addressed.

Three distinct processes underlie sleep regulation. A homeostatic process, whose level is a function of prior sleep and waking, plays a major role in sleep regulation. Sleep is also modulated by a circadian process, a clocklike mechanism that is independent of prior sleep and waking. An ultradian process occurs within sleep and is represented by the alternation of the two basic sleep states: non–rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep.

One of the main topics of this chapter is sleep homeostasis. *Homeostasis* has been defined as "the coordinated physiological processes, which maintain most of the steady states in the organism."[1] The term *sleep homeostasis*[2] refers to the aspect of sleep regulation dependent on sleep and waking, as homeostatic mechanisms counteract deviations from an average reference level of sleep. These mechanisms augment sleep propensity when sleep is curtailed or absent, and they reduce sleep propensity in response to excess sleep.

The interest in the modeling approach to sleep regulation has increased over the past decades. Models help delineate the processes involved in the regulation of sleep and thereby offer a conceptual framework for the analysis of existing and new data.[3] For reviews of modeling circadian rhythms related to sleep, see Roenneberg and colleagues,[4] Beersma,[5] and Klerman and St Hilaire.[6]

HOMEOSTATIC REGULATION OF SLEEP

Electroencephalographic Slow Wave Activity: A Physiologic Indicator of NREM Sleep Homeostasis

Slow Wave Sleep and Slow Wave Activity

NREM sleep is not a homogeneous substate of sleep but can be subdivided according to the predominance of slow waves in the electroencephalogram (EEG). One of the most important functional EEG parameters is referred to as slow wave activity (SWA). It is equivalent to delta activity and encompasses components of the EEG signal in the frequency range of approximately 0.5 to 4.5 Hz as obtained by spectral analysis.[7]

In addition to delta waves, a low-frequency component with a mean peak value of 0.7 to 0.8 Hz is present in the EEG power spectrum of NREM sleep.[8-10] The typical decline in delta activity from the first to the second NREM sleep episode was not present at frequencies below 2 Hz.[8] This could indicate that the low-frequency component reflects a separate process that differs from the one represented by delta activity. Alternatively, the results could be due to a frequency shift in the course of sleep.[11] The changes of the low-frequency component in response to naps did not differ from those of SWA.[12] Increased sleep pressure resulted in a redistribution of waves

between 0.5 and 2 Hz: The number of waves per minute was reduced below 0.9 Hz but was increased above 1.2 Hz; EEG power was increased only above 1 Hz.[13] Periodicities at even lower frequencies include the recurrence of sleep spindles at 4-second intervals,[8,14] the tendency of slow waves to recur at 20- to 30-second intervals,[8] and a cortical infra slow oscillation (0.02 to 0.2 Hz).[15,16]

Slow Waves and Sleep Intensity

It was recognized as early as 1937 that sleep intensity is reflected by the predominance of slow waves in the sleep EEG.[17] Subsequent studies confirmed that the responsiveness to stimuli decreases as EEG slow waves become more predominant.[18] Under physiologic conditions, this EEG variable can be regarded therefore as an indicator of sleep depth or sleep intensity.

Global Time Course of Slow Wave Activity during Sleep

Slow wave sleep (SWS; the high-intensity part of NREM sleep) appeared to be a good candidate for a physiologic marker of sleep homeostasis. The predominance of SWS in the early part of the sleep episode was documented in several early studies.[18-20] All-night spectral analysis of the sleep EEG made it possible to quantify SWA and to delineate its time course during sleep.[7] Its mean value per cycle plotted for consecutive NREM-REM sleep cycles showed a monotonic decline over the first three cycles. Figure 36-1 (*left*) shows the changes of mean EEG power density over four cycles for the frequency range between 0.25 and 20 Hz. The values of each bin are expressed relative to the reference level of cycle 4 (100%). Note that although the largest changes occur in the low delta range, they encompass frequencies up to the theta band.

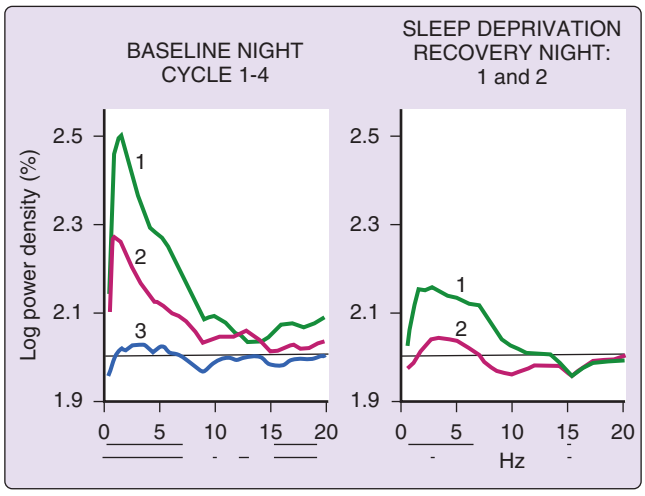


Figure 36-1 *Left,* Changes of relative spectral power density over the first four NREM-REM sleep cycles of a baseline night (N = 8). In each frequency bin the data are expressed relative to the value in the fourth cycle (100%; *horizontal line*). *Right,* Effect of sleep deprivation (40.5 hours waking) on spectra of the sleep electroencephalogram (EEG). In each bin, the values of the first two recovery nights are plotted relative to the baseline night (100%). The *upper* and *lower horizontal bars* below the abscissa indicate for the left part significant differences between cycles 1 and 2, and between cycles 2 and 3, respectively, and for the right part between recovery 1 and baseline, and between recovery 2 and baseline, respectively. (Modified from Borbély AA, Baumann F, Brandeis D, et al. Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephalogr Clin Neurophysiol* 1981;51:483–93.)

Nap Studies

The analysis of daytime naps is useful for assessing the level of SWA after various durations of waking. Naps taken later in the day contained more SWS than naps taken earlier in the day. In a detailed study of daytime naps scheduled at 2-hour intervals throughout the day, direct evidence for a monotonic rise of SWA was obtained.[12,21-23] If naps reverse the rising trend of slow wave propensity, a reduction of SWA in the subsequent nighttime sleep can be expected. This prediction was borne out by the results of several experiments (see Werth et al.[24] for literature references). Furthermore, when the duration of nighttime sleep was shortened, SWA in a subsequent morning nap was enhanced.[25,26]

Effect of Sleep Deprivation

It has repeatedly been shown that partial or total sleep deprivation gives rise to increased SWS in the recovery night (see Borbély[27] for a review of the older literature). Webb and Agnew[20] presented compelling evidence that SWS increases as a function of prior wake duration. The quantitative assessment of SWA using all-night spectral analysis revealed that a night without sleep (i.e., 40.5 hours of wakefulness) resulted in an enhancement of this EEG variable during recovery sleep.[7] Figure 36-1 (*right*) illustrates the changes of power density in the two recovery nights relative to the baseline level (100%). In the first recovery night, the largest increase was present in the low delta range, the part of the spectrum undergoing the largest changes in the course of baseline sleep (see Figure 36-1, *left*).

Figure 36-2 depicts the global trend as well as the ultradian dynamics of SWA over successive NREM-REM sleep cycles. The prolongation of the waking period causes a prominent rise of SWA during recovery sleep. A declining trend over three to four cycles is evident in both records. Note that the peaks are at a steady low level during the last four cycles of recovery sleep.

The enhancement of SWA by sleep deprivation was confirmed in numerous studies[28,29] (for references before 1992, see Borbély and Achermann[30]). The extent of the increase was shown to be a function of the duration of prior waking.[23,31]

Selective Slow Wave Deprivation

The propensity of SWA does not necessarily dissipate during sleep but may persist at an elevated level if SWS is prevented. Thus suppression of slow waves by acoustic stimuli during the first 3 hours of sleep resulted in a prominent rise of SWA after the stimuli were discontinued.[32] In another study, daytime sleep episodes with and without SWS deprivation were compared.[33] The experimental suppression of SWS during an interval corresponding to 90% of the undisturbed episode resulted in an increased accumulation of SWS and an extension of sleep duration. Topographic studies revealed regional differences in the effectiveness of and response to selective SWS deprivation.[34] Taken together the results indicate that slow waves are not merely an epiphenomenon of sleep but instead reflect major sleep-regulating mechanisms. This is supported by the findings that selective slow wave deprivation impairs perceptual[35] and visuomotor learning.[36] In another study, two nights of SWS disruption increased sleepiness but had only minor effects on daytime functioning.[37]

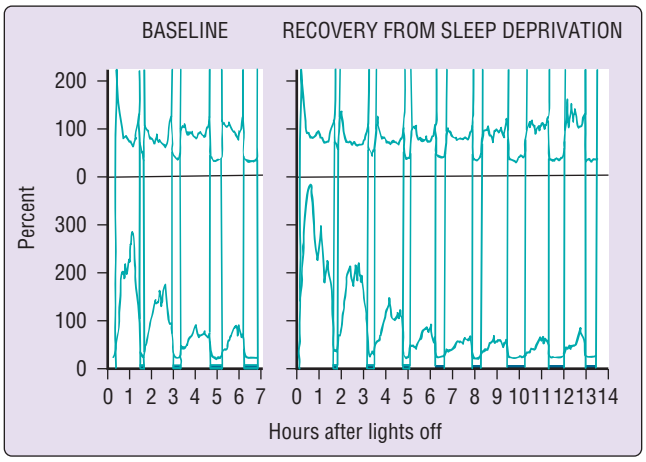


Figure 36-2 Time course of slow wave activity (power in the 0.75- to 4.5-Hz band; *lower curves*) and activity in the spindle frequency range (13.25- to 15.0-Hz; *upper curves*) recorded under baseline conditions and after sleep deprivation (36 hours of wakefulness). NREM sleep episodes were subdivided into 20 equal intervals, and REM sleep episodes were divided into 5 intervals. Mean values per interval were calculated before averaging across subjects (N = 8 except for cycle 8 of recovery sleep where N = 6) and were expressed relative to the mean level in baseline NREM sleep (100%). The mean timing of REM sleep episodes is delimited by *vertical lines* and *horizontal bars* above the abscissa. (Reanalysis of the data from Dijk DJ, Brunner DP, Borbély AA. Time course of EEG power density during long sleep in humans. *Am J Physiol* 1990;258:R650–1, by D. Aeschbach.)

Ultradian Dynamics of Slow Wave Activity and Spindle Frequency Activity

Buildup of Slow Wave Activity within NREM Sleep Episodes

The mean level and the peak of SWA are determined not only by the duration of prior waking and sleep but also by the rise rate within single NREM sleep episodes.[38-40] It is evident from Figure 36-2 that the rise rate of SWA decreases over the first three episodes both under baseline conditions and during recovery from sleep deprivation. In addition, the effect of prolonged waking manifests itself in a steeper buildup (of SWA) within NREM sleep episodes.[28,29,40] A decrease in the slope of single slow waves was associated with a reduction of sleep pressure during sleep in humans and animals,[41,42] a change that could be simulated in a model by reducing synaptic strength.[43] However, amplitude and slope of slow waves were highly correlated in the 0.5- to 2-Hz range.[13]

Slow Wave Activity and Spindle Frequency Activity

The term *spindle frequency activity* (SFA) is used to denote the power in the frequency range of sleep spindles (12 to 15 Hz). There is a close correspondence between this measure and measures based on the occurrence of sleep spindles.[28]

The time courses of SWA and SFA differ in several respects. The global decline of SWA does not occur in the spindle frequency range. Within NREM sleep episodes, SFA shows a bimodal pattern with an initial and a terminal peak. This gives rise to a U-shaped curve within the episode and a partly inverse relationship to SWA[28,44-49] (see Figure 36-2).

Regulation of REM Sleep

The principles underlying REM sleep regulation seem to be more complex than those for NREM sleep regulation. This was evident in a selective REM sleep deprivation experiment,[50] in which two salient observations were made that were difficult to reconcile. On the one hand, REM sleep deprivation necessitated a dramatic rise in the frequency of interventions during the night to prevent this sleep state. On the other hand, there was a modest rise in the number of interventions across the three consecutive deprivation nights, and the 40% REM sleep rebound in the first recovery night by no means compensated for the amount of REM sleep lost. Two hypotheses were advanced. In the first, it is assumed that the homeostatic drive is strong, which is reflected by the dramatic rise in interventions during the deprivation nights. Waking may in part substitute for REM sleep, thereby accounting for the moderate night-to-night increase in interventions and the small REM sleep rebound. According to the second hypothesis, the homeostatic drive for REM sleep is weak, and the rising trend in the number of interventions is attributed to circadian factors as well as to a sleep-dependent disinhibition of REM sleep propensity. This hypothesis could explain the limited savings from one night to the other as well as the modest rebound. For a recent review on possible functions of REM sleep, see Vyazovskiy and Delogu.[51]

NREM versus REM Sleep Homeostasis

Effect of NREM Sleep Pressure on REM Sleep Homeostasis

During recovery from total sleep deprivation, SWS and EEG SWA exhibit an immediate rebound, whereas the increase in REM sleep is delayed to subsequent nights or is not present at all. Selective REM sleep deprivation augments REM sleep pressure, which is manifested by the increasing number of interventions required to prevent REM sleep episodes (for the older literature, see Borbély[27]). However, the occurrence of REM sleep rebound during recovery sleep is smaller than expected on the basis of the deficit.[50,52] This suggests that REM sleep is not as finely regulated as NREM sleep. However, this notion is contradicted by partial sleep deprivation studies. A REM sleep deprivation in the first 5 hours of sleep induced a REM sleep rebound in the subsequent 2.25 hours.[53] A curtailment of sleep duration during two or four nights, which induced a substantial REM sleep deficit, was followed by REM sleep rebound in the two recovery nights.[54,55] In these experiments, the REM sleep rebound occurred at a time when slow wave pressure was either low at the end of sleep[53] or was much less increased than REM sleep pressure.[54,55] These results also suggest that REM sleep is indeed regulated but that the manifestation of REM sleep homeostasis is hampered by an elevated slow wave pressure.

Effect of REM Sleep Pressure on the NREM Sleep Electroencephalogram

In accordance with the notion of a mutual inhibitory interaction of the factors controlling SWA and REM sleep,[27] not only is REM sleep inhibited by slow wave pressure, but SWA is also inhibited by REM sleep pressure. Thus selective REM sleep deprivation led to a significant reduction in the low-frequency activity of the NREM sleep EEG,[53] an observation that was also made in an animal experiment.[56] The rise in REM sleep pressure induced by repeated partial sleep deprivation suppressed the typical low-delta peak in the NREM sleep spectrum.[54,55] However, this effect was not seen after selective REM sleep deprivation.[50]

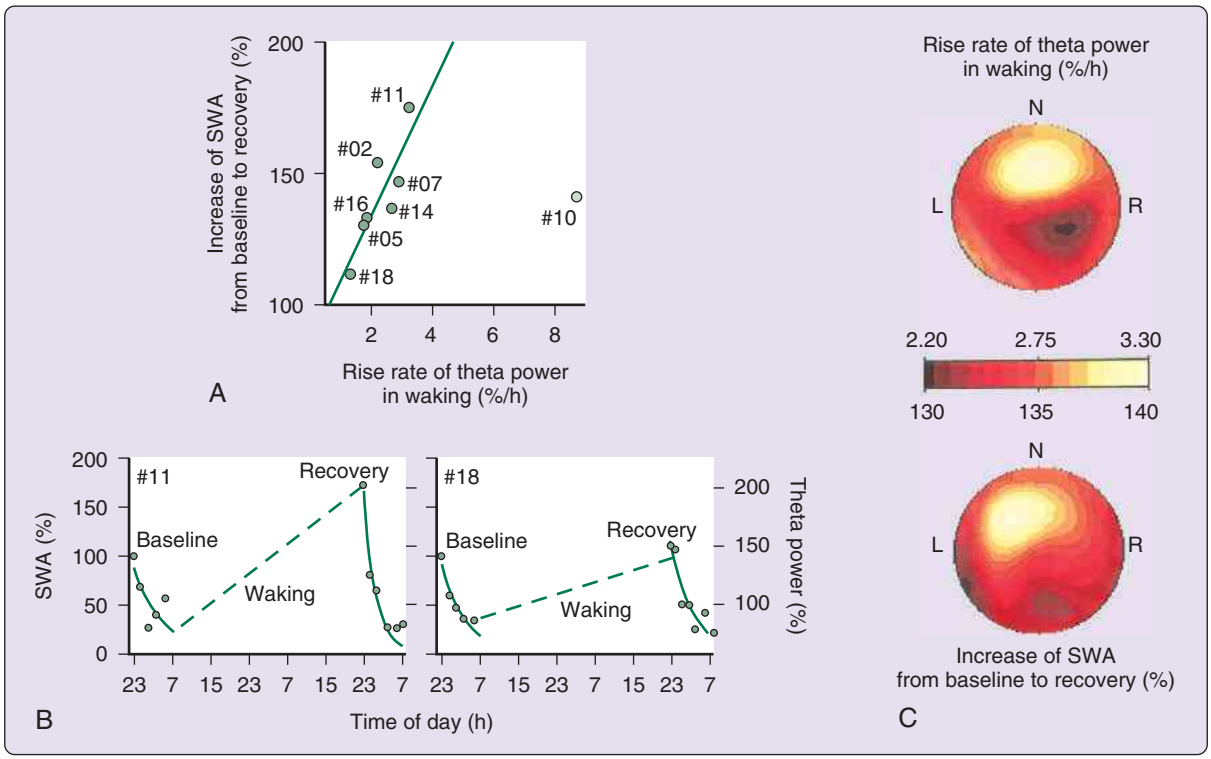


Figure 36-3 **A,** Relationship between homeostatic markers of the sleep electroencephalogram (EEG) and waking EEG. Increase (%) of slow wave activity (SWA; 0.75–4.5 Hz) in the first NREM sleep episode from baseline to recovery sleep is plotted as a function of the rise rate (%/h) of theta power (5.0–8.0 Hz) in waking. The linear regression line fitted through 7 data points is indicated ($r = 0.851$, $r^2 = 0.724$, $P = .015$). Subject no. 10 was excluded from the regression. **B,** Association between rise of SWA in sleep and theta activity in waking illustrated for two subjects. Mean SWA per NREM sleep episode is plotted at the beginning of each episode and expressed relative to the baseline value of the first NREM sleep episode (100%). Exponential functions were fitted through the data points (*solid curves*). The regression line represents theta power in waking (*interrupted line*). **C,** Topographic distribution of the rise rate of theta power (*top*) during waking and of the increase of SWA (*bottom*) in the first NREM sleep episode from baseline to recovery sleep. Maps are based on 27 EEG derivations (average reference; extended 10-20 system). Values are plotted on a color scale at the corresponding position on the planar projection of the hemispheric scalp model. Values between electrodes were linearly interpolated. (Modified from Finelli LA, Baumann H, Borbély AA, Achermann P. Dual electroencephalogram markers of human sleep homeostasis: correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience* 2000;101:523–9.)

Homeostatic Marker in the Waking Electroencephalogram

It had been shown in early studies that power in the theta band (theta activity) and alpha activity of the waking EEG is associated with sleepiness[57,58] and that total[58] or partial sleep deprivation[55] enhanced the power in these frequency bands. A saturating exponential function with a time constant of 18.18 hours was reported to fit the rise of theta activity in the waking EEG.[59] Spectral analysis showed that the largest changes occurred in the theta band (see Borbély and Achermann[60] for review). These undergo a circadian modulation in addition to the changes related to wake time.[61-66] During prolonged wakefulness, subjective sleepiness correlated positively with theta activity with a focus in frontal derivations and negatively with alpha activity at all derivations.[67] A forced desynchrony study with a scheduled waking episode of 28 hours showed a monotonic rise of delta and beta activity in the frontocentral derivation.[65] An analysis of persons subjected to sleep deprivation revealed that the rise rate of theta activity in the waking EEG is correlated with the increase of SWA in the first NREM sleep episode of recovery sleep (Figure 36-3, *A* and *B*).[61] Moreover, both effects were largest in frontal areas (Figure 36-3, *C*). In summary, theta activity in waking and SWA in sleep may be markers of a common homeostatic sleep process.

Independence and Interactions of Homeostatic and Circadian Processes

There is evidence that homeostatic and circadian facets of sleep regulation can be independently manipulated and therefore may be controlled by separate mechanisms. Thus throughout a 72-hour sleep deprivation period, the subjective alertness ratings continued to show a prominent circadian rhythm.[68] Conversely, in a study in which the phase of the circadian process (as indexed by body temperature and plasma melatonin) was shifted by bright light in the morning, the time course of SWA remained unaffected.[69]

A powerful experimental paradigm is the forced desynchrony schedule in which the homeostatic and circadian facet of sleep can be separately analyzed.[70,71] In this long-term protocol, the imposed sleep-wake cycle (e.g., 20 hours or 28 hours) lies outside the range of entrainment of the circadian pacemaker. When this schedule is maintained for 3 weeks,

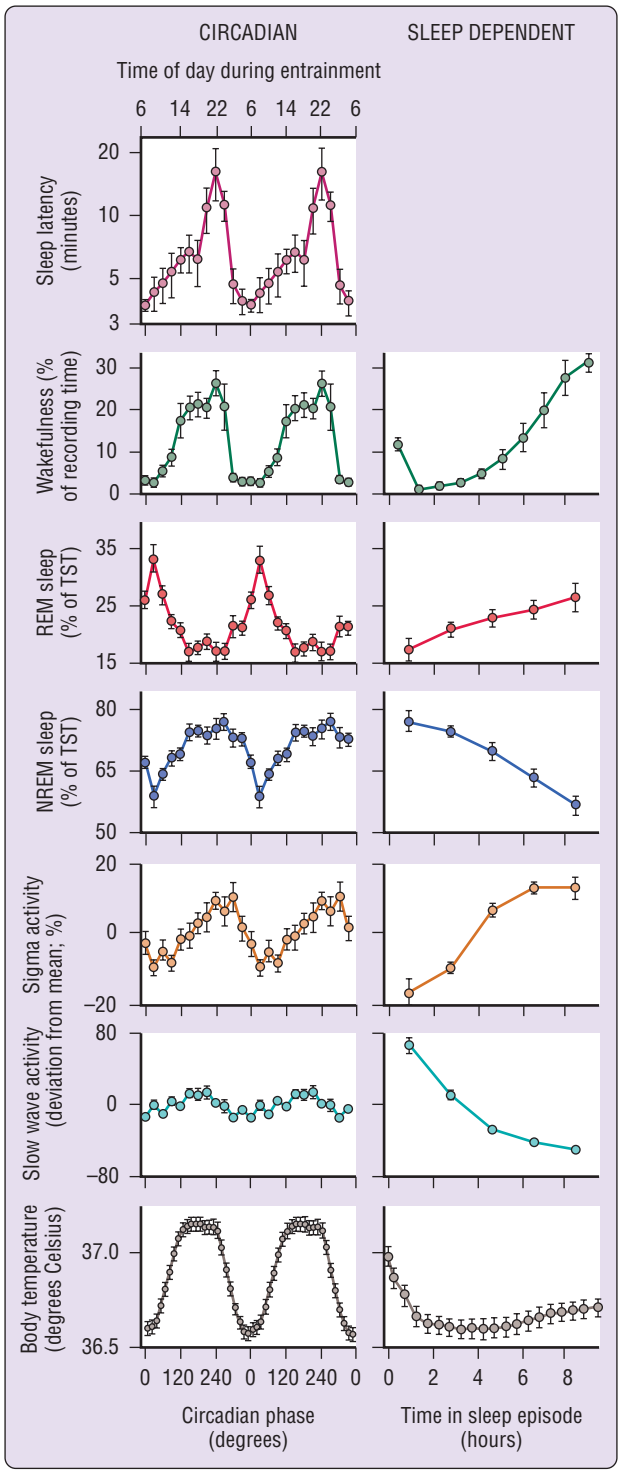


Figure 36-4 Circadian and sleep-dependent or homeostatic factors in sleep regulation. The main effects of circadian phase and sleep homeostasis on sleep were analyzed by aligning the data relative to the circadian component of the body temperature cycle (*left panels*) or the beginning of the sleep opportunity (*right panels*). Slow wave activity shows weak circadian and strong sleep-dependent modulation; sigma (sleep spindle) activity shows strong circadian and sleep-dependent modulation. NREM sleep percentage shows equal circadian and sleep-dependent components. REM sleep percentage shows marked circadian maximum just after the temperature nadir and sleep-dependent increase (disinhibition). Wakefulness in scheduled sleep episodes shows that the circadian drive for wakefulness is maximum 7 to 9 hours before temperature nadir, which is 1 to 3 hours before habitual bedtime; there is a strong wake-dependent increase. Sleep latency shows strong circadian modulation; the longest sleep latencies occur 7 to 9 hours before the body temperature nadir, and the shortest sleep latencies occur at the body temperature nadir. TST, Total sleep time. (Modified from Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves and sleep spindle activity in humans. *J Neurosci* 1995;15:3526–38; and Dijk DJ, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci Lett* 1994;166:63–8. Corresponds to Figure 34-4 of Dijk DJ, Franken P. Interaction of sleep homeostasis and circadian rhythmicity: dependent or independent systems? In: Kryger MH, Roth T, Dement WC, editors. *Principles and practice of sleep medicine.* 4th ed. Philadelphia: Saunders; 2005. p. 418–34.)

sleep occurs at different circadian phases. The contributions of the homeostatic and circadian components can be estimated by folding the data of the variable investigated either at the period of the imposed sleep-wake cycle or at the period of circadian rhythm. Various claims of the two-process model were supported by the results obtained in this paradigm.[70] As shown in Figure 36-4, the variation of SWA is accounted for mainly by homeostatic (i.e., sleep-wake dependent) factors, whereas the percentage of REM sleep, NREM sleep, SFA (sigma activity), and sleep consolidation are determined by both homeostatic and circadian factors. Furthermore, a previously postulated sleep-related disinhibition of REM sleep[27] was confirmed. Severe sleep restriction attenuates the circadian modulation of sleep.[71,72] The role of the metrics used and whether homeostatic and circadian processes exhibit a linear or a nonlinear interaction are still under discussion.[73]

MODELS OF SLEEP REGULATION

Models help delineate the processes involved in the regulation of sleep and thereby offer a conceptual framework for the analysis of existing and new data. In addition, they inspire new experiments to test predictions of the models.

Models may address processes at different levels (from the microscopic [cellular] level to the macroscopic [systemic] level) and at different time scales (from the range of milliseconds or seconds up to hours or days[74]). Being aware of the power and limitations of models is important for selecting the most appropriate one for the question to be addressed.

The term *model* has been loosely applied to hypothetical descriptions of events, features, and processes related to sleep. The following synopsis is restricted to major models that include a mathematical description and address sleep regulation. Their main features are summarized in Tables 36-1 and 36-2. We describe only selectively models addressing the generation of specific features of the sleep EEG and do not include molecular state variables of sleep regulation.[75]

Table 36-1 Two-Process Model and Related Models

Designation	Assumption	Description and Comment
Two-process model[27,78,82,83]	Sleep propensity is determined by a homeostatic process S and circadian process C. The interaction of S and C determines the timing of sleep and waking.	Time course of S derived from EEG slow wave activity; phase position and shape (skewed sine wave) of C derived from sleep duration data obtained at various times of the 24-hr cycle.
Model of ultradian variation of slow wave activity[84-86]	Derived from the two-process model. The level of S determines the buildup rate and the saturation level of slow wave activity within NREM sleep episodes.	In contrast to the original two-process model, the change of S, not the level of S, corresponds to slow wave activity; that is, the decline of S is proportional to the amount of slow wave activity. A REM sleep oscillator triggers the decline of slow wave activity before REM sleep.
Three-process model of sleepiness-alertness regulation[97-99,167]	Sleepiness and alertness are simulated by the combined action of a homeostatic process, a circadian process, and sleep inertia (process W). Extension to include performance, sleep latency, and sleep length. Adaptation of homeostatic process to account for performance prediction in sleep-restriction protocols.	Parameters derived from rated sleepiness during sleep-wake manipulations. Alertness nomogram for sleep-related safety risks.
Interactive mathematical models of alertness and cognitive throughput[100]	Alertness and cognitive throughput are determined by a nonlinear interaction of a homeostatic (H) and a circadian (C) process. Sleep inertia is also included. H shows a sigmoidal decline during waking and a saturating exponential increase during sleep at a rate determined by the circadian phase.	Parameters derived from sleep inertia studies, sleep deprivation studies initiated across all circadian phases, and 28-hr forced desynchrony studies.

Two-Process Model and Related Models

The two-process model and related models are summarized in Table 36-1. Further models inspired by the two-process model addressing performance simulations were discussed in a special issue of *Aviation, Space and Environmental Medicine*[76] and elsewhere.[3,30,60,77-79]

The relationship between SWS and the duration of prior waking has been documented by Webb and Agnew[20] and placed into a theoretical framework by Feinberg.[80] The two-process model, originally proposed to account for sleep regulation in the rat,[2,81] postulates that a homeostatic process (process S) rises during waking and declines during sleep and interacts with a circadian process (process C) that is not directly dependent on sleep and waking. The time course of the homeostatic variable S was derived from EEG SWA (Figure 36-5, *A*). Various aspects of human sleep regulation were addressed in a qualitative version of the two-process model.[27] An elaborated, quantitative version of the model was established in which process S varied between an upper and a lower threshold that are both modulated by a single circadian process.[82,83] This model was able to account for such diverse phenomena as recovery from sleep deprivation, circadian phase dependence of sleep duration, sleep during shift work, sleep fragmentation during continuous bed rest, and internal desynchronization in the absence of time cues.[83]

In a later version of the model (proposed by Beersma et al.[21] and Dijk et al.[32] and formalized by Achermann and Borbély[84,85]), it is the change of S, and not its level, that is proportional to the momentary amount of SWA. The elaborated model addressed not only the global changes of SWA as represented by process S but also the changes within NREM sleep episodes. In general, a close fit was obtained between the simulated and empirical SWA data and their time course (Figure 36-5, *D*). In particular, the occurrence of late SWA peaks during extended sleep could be simulated (Figure 36-5, *C*). The simulations demonstrated that the model accounts in quantitative terms for empirical data and predicts the changes induced by the prolongation of waking or sleep. This version of the model was used to simulate the dynamics of SWA in an experimental protocol with an early evening nap[24] and the effect of changes in REM sleep latency on the time course of SWA.[86] Zavada and colleagues[87] used this model's approach to investigate regional aspects of sleep homeostasis.

The homeostatic process S is modeled by two exponential functions, one describing the rising limb during waking, the other the declining limb during sleep (see Figure 36-5, *A*). The buildup of sleep pressure is approximated by a saturating exponential function (time constant of increase; upper asymptote), its dissipation by a declining exponential function (time constant of decrease; lower asymptote). Thus, process S oscillates between an upper and lower asymptote (relaxation oscillator), its dynamics being determined by the distance between the asymptotes and by the time constants. The time constants show significant interindividual variability[88] and vary also across brain regions.[89] Importantly, homeostasis is mainly reflected in the time constants, and it was hypothesized that a slower buildup may be associated with an increased tolerance to sleep deprivation.[88] A slowing of the buildup of sleep pressure was observed in preschool children in the course of development[90] and during early adolescence.[91] It is important to recognize that the absolute values of SWA are highly variable

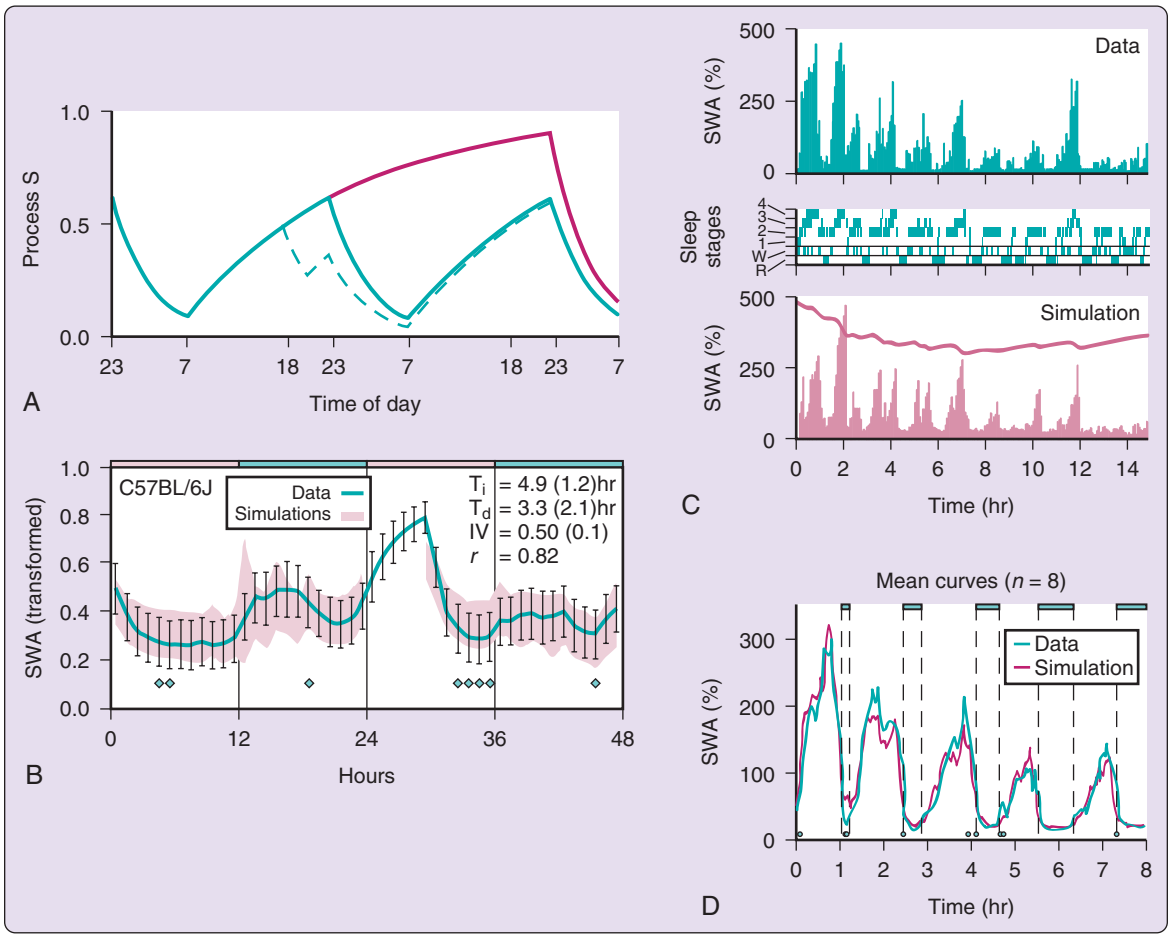


Figure 36-5 Two-process model of sleep regulation. **A,** Simulations of the homeostatic process S according to different experimental conditions. The *green line* indicates baseline condition with an 8-hour sleep episode; the *red line* indicates sleep deprivation (40 hours of wakefulness) and recovery sleep; the *dashed line* indicates effect of a 2-hour daytime nap at 18:00 hours. **B,** Sleep regulation in the mouse. Time course of slow wave activity (SWA) and simulation with the optimized time constants for the increase (Ti) and decrease (Td) and initial value (iV) of process S for C57BL/6J mice (N = 8). *Curves* and *shaded areas* connect 1-hour mean values (±SEM) for 24-hour baseline, 6-hour sleep deprivation, and 18-hour recovery. The close fit between the simulation of process S (*pink areas*) and time course of empirical SWA (*solid line*) indicates that the model can predict SWA from the temporal organization of sleep. *Diamonds* indicate differences between simulation and data ($P < .05$; two-tailed paired t-test). For the comparison between SWA and S, SWA was transformed according to a linear regression. *Inset:* Mean values of Ti, Td, and iV (SEM) and the mean *r* value of the fit between SWA and S. **C,** Empiric SWA (*top*), sleep stages (*center*), and simulation of SWA and process S (*bottom*) of an individual extended baseline sleep episode starting at 00:00 hours (prior waking: 17:00 hr). Empiric and simulated SWA were standardized with respect to the mean value of the first 7 hours of sleep. Values are plotted for 1-minute intervals. **D,** Mean empirical (*green line*) and simulated SWA (*pink line*) (N = 8) of an extended baseline experiment (analysis of first 8 hours). Significant differences are indicated by *open circles* (paired t-test; $P < .05$). *Bars on top* and the *interrupted vertical lines* indicate REM sleep episodes (mean values). (**B** modified from Huber R, Deboer T, Tobler I. Effects of sleep deprivation on sleep and sleep EEG in three mouse strains: empirical data and simulations. *Brain Res* 2000;857:8–19. **C** and **D** modified from Achermann P, Dijk DJ, Brunner DP, Borbély AA. A model of human sleep homeostasis based on EEG slow-wave activity: quantitative comparison of data and simulations. *Brain Res Bull* 1993;31:97–113.)

between individuals and are in a large part dependent on age.[92] Thus absolute levels of SWA are not a measure of sleep pressure or homeostasis per se, but it is the relative change in SWA in response to a challenge that is informative (e.g., for total or partial sleep deprivation, sleep restriction, naps). Furthermore, the distance between the asymptotes may be interpreted as the capacity of the brain to generate slow waves.[88,90,91]

Finally, the data analysis showed that not only the timing of sleep but also the changes in daytime vigilance are governed by the interaction of processes S and C, as simulated by Daan and colleagues.[83] The rising homeostatic sleep pressure during waking seems to be compensated by the declining circadian sleep propensity.[83,93-95] Conversely, during sleep the rising circadian sleep propensity might serve to counteract the declining homeostatic sleep pressure, thereby ensuring the maintenance of sleep.[96]

Based on a similar concept, the changes of subjective sleepiness/alertness ratings were simulated by a combined action of a homeostatic process (S), a circadian process (C), and a process representing sleep inertia (W) (three-process

Table 36-2 Neurophysiologic Network Models of Sleep-Wake Regulation (Brainstem and Hypothalamic Control of Sleep-Wake States)

Designation/Model	Assumption*	Description/Comment
Quantitative model of sleep-wake dynamics based on physiology of the brainstem ascending arousal system Phillips and Robinson[125,132-137,168-171]	Model includes VLPO (where circadian and homeostatic drives enter the system), monoaminergic and cholinergic nuclei of the ascending arousal system, and corresponding interconnections Mutual inhibition between wake-promoting monoaminergic group and sleep-promoting VLPO causes flip-flop behavior REM sleep not incorporated	Human sleep: simulation of basic sleep behavior; effects of sleep deprivation, effect of caffeine, fatigue, impulsive stimuli, internal desynchronization, and shift work Animal sleep: simulation of interspecies differences and unihemispheric sleep
Biologically based mathematical model of the sleep-wake cycle Rempe et al.[126]	Model is based on flip-flop conceptual models for sleep-wake and REM-NREM sleep alterations Includes the sleep-promoting neurons in the VLPO, the wake-promoting monoaminergic cell groups, and orexin neurons Mutual inhibition of REM-on and REM-off populations	Human sleep: simulation of basic sleep patterns; sleep deprivation; role of orexin; and sleep-onset REM sleep
Mathematical model of neuronal regulation of waking, NREM sleep, and REM sleep Kumar et al.[172]	Two distinct flip-flop circuits, one governing wake-NREM sleep transitions (POAH, MRF, CRF, and orexinergic groups) and the other NREM sleep–REM sleep transitions (LDT/PPT and LC) REM sleep generation by presynaptic inhibition of substantia nigra onto REM-off terminals projecting on REM-on neurons. Putative REM sleep homeostasis. Mutual inhibition of REM-on and REM-off populations	Human sleep: simulation of basic sleep patterns; sleep deprivation; and role of orexin (orexinergic neurons stabilize the wake-sleep cycle)
Quartet neural system model of sleep and wakefulness mechanisms Tamakawa et al.[127,173]	Neural system consisting of sleep-active preoptic or anterior hypothalamic neurons (N-R group); wake-active hypothalamic and brainstem neurons (WA group); brainstem neurons (REM group); and basal forebrain, hypothalamic, and brainstem neurons (W-R group) WA neurons have mutual inhibitory couplings with REM and N-R neurons. W-R neurons have mutual excitatory couplings with WA and REM neurons. REM neurons receive unidirectional inhibition from N-R neurons. N-R neurons are activated by sleep-promoting substances Reciprocal interaction between REM-on and REM-off populations	Human sleep: simulation of basic sleep-wakefulness rhythms Rodent sleep: model reproduced sleep and wakefulness patterns of rats; microinjection experiments; circadian modulation
Mathematical model of network dynamics governing mouse sleep-wake behavior Diniz Behn et al.[124,174]	Sleep-wake network composed of coupled relaxation oscillators Network includes wake-, sleep-, and REM sleep–promoting populations Incorporation of orexin signaling Reciprocal interaction between REM-on and REM-off populations	Rodent sleep: simulation of dynamics underlying state transitions
Modeling framework of the sleep-wake regulatory network in the brainstem and hypothalamus Diniz Behn & Booth[128,175-179]	Neurotransmitter-mediated interactions among brainstem and hypothalamic neuronal populations that participate in the transitions between wake, REM sleep, and NREM sleep Reciprocal interaction between REM-on and REM-off populations	Human sleep: simulation of basic sleep behavior and internal desynchronization Rodent sleep: simulation of basic sleep behavior; microinjection experiments; circadian modulation; temporal architecture of sleep patterns

CRF, Caudal reticular formation; LC, locus coeruleus; LDT/PPT, laterodorsal tegmentum/pedunculopontine tegmentum; MRF, midbrain reticular formation; POAH, pre-optic anterior hypothalamus; VLPO, ventrolateral preoptic area.

*Basic assumption in all models is that transitions between sleep or NREM sleep and wake are mediated by mutual inhibition between sleep or NREM sleep–promoting and wake-promoting neuronal populations. Inclusion of homeostatic and circadian components.

model; Table 36-1[97,98]). The model has been expanded[99] to account for discrepancies observed in chronic sleep restriction experiments.

Jewett and Kronauer (Table 36-1[100]) proposed interactive mathematical models of subjective alertness and cognitive throughput in humans. A homeostatic component (H) falls in a sigmoidal manner during waking and rises according to a saturating exponential function during sleep. The rise of H during sleep is determined by the circadian phase. H interacts with a circadian component (C[101]), accounting for the effect of light on the circadian pacemaker. The amplitude of C depends on the level of H. In addition, a sleep inertia component (W) is included. In contrast to the two- and three-process models, a nonlinear interaction is assumed. Whether the interaction is linear or nonlinear is still unresolved.[102,103] A statistical approach to test for nonlinear interactions was proposed[73] and is based on a comparison of model predictions and empiric data.

Experiments with chronic sleep restriction demonstrated that the homeostatic sleep drive is not associated with neurobehavioral performance.[104] The latter showed a cumulative impairment with a considerable degree of individual variability. To account for the slow recovery of neurobehavioral functions after sleep deprivation, the possibility of a second homeostatic process with a very long time constant was considered.[105] McCauley and colleagues[72,106] expanded the two-process model to a broader class of models that have a dynamic repertoire capturing waking neurobehavioral functions across a wide range of wake-sleep schedules. Models of sleep, fatigue, and performance based in large part on the two-process model are reported in a special issue of *Aviation, Space and Environmental Medicine*.[76] For a recent review, see Dawson et al.[107]

Because most of these models are based on average data, predictions of individual performance are limited.[108] A difficulty in addressing individual differences is due to the fact that individual parameters need first to be determined. Two approaches to predict individual performance in sleep deprivation experiments have been proposed.[109,110] Both approaches are based on the two-process model and allow a continuous parameter adaptation in an iterative process as new empiric data become available.

Although the qualitative version of the two-process model had originated from animal data,[2,81] the quantitative version was elaborated on the basis of findings from human studies. In the meantime, quantitative simulations of NREM sleep homeostasis were also performed in rats[111-114] and mice[115,116] (Figure 36-5, *B*). SWA of consecutive 4-second epochs in a 24-hour baseline, a 6-hour sleep deprivation, and 18-hour recovery period[115] served as the database for the simulation in mice. Like in the original human version of the model, process S was assumed to decrease exponentially in NREM and REM sleep and to increase according to a saturating exponential function in waking. Unlike in the human model, S increased also in REM sleep. After optimizing the initial value of S (IV) as well as its time constants (increase Ti; decrease Td), a close fit was obtained between the hourly mean values of SWA in NREM sleep and the prediction of process S (see Figure 36-5, *B*).

Modeling REM Sleep

Whereas many models have focused on NREM sleep homeostasis, on the interaction of NREM and REM sleep, and on the circadian oscillator, REM sleep regulation per se has been largely ignored. This is because of the complex and not yet clearly understood principles underlying REM sleep regulation (see earlier). The NREM-REM sleep cycle has been accounted for by the limit cycle reciprocal interaction model on the basis of interacting neuronal systems (for a elaborated version and for the activation-synthesis model, see Pace-Schott and Hobson[117]). Attempts were made to integrate various concepts into a combined model.[118,119] Saper and coworkers proposed in their flip-flop model a mode of interaction between different neurotransmitter systems that accounts for the sharp state transitions between waking and sleep[120] as well as between NREM and REM sleep.[121,122] Mathematical models incorporating these neurophysiologic aspects are summarized in Table 36-2.

Neurophysiologic Models

Neuronal models may be subdivided into neurophysiologic models based on detailed neuronal architecture (e.g., neurons, synapses) addressing thalamocortical interactions[123] and models that capture essential features based on the dynamics of neuronal populations (approximation of activity of a large ensemble of neurons; see Table 36-2).

Extended versions of the two-process model were derived from neurophysiologic data[124-129] (see Table 36-2). For a recent review and mathematical analysis of the models, see Booth and Diniz Behn.[130] Interactions of neuronal populations in the brainstem and hypothalamus were used to simulate transitions between sleep and waking or between NREM and REM sleep. In addition, a homeostatic and a circadian component was included. To a large degree, the slow dynamics (hours to days) of these models resemble the dynamics of the two-process model.[130,131] The degree of similarity between the two-process model and the Phillips-Robinson model[125] was recently explored.[131] The authors demonstrated that the slow dynamics of the Phillips-Robinson model could be explicitly related to the two-process model providing a neurophysiologic interpretation of the thresholds. This model[125] (see Table 36-2) was applied to simulate sleep fragmentation experiments,[132] differences in mammalian sleep patterns,[133] and subjective fatigue during sleep deprivation.[134] It was refined to simulate effects of caffeine[135] and spontaneous internal desynchronization through feedback of the sleep-wake cycle on the circadian component.[136,137]

For a better understanding of sleep-wake regulation, the development and analysis of models based on neurophysiologic mechanisms are essential. However, these models become more and more complex and comprise a large parameter space with only few parameters that can be determined empirically. Additionally, large numbers of simulations have to be performed to establish the system behavior. Therefore it is difficult to compare the performance of the different models. A further challenge is model validation. What are the relevant end points to investigate? The replication of qualitative patterns of behavioral state transitions is insufficient for parameter estimation of such highly complex models. Thus the application of sophisticated statistical approaches to determine model parameters is needed for further progress.[130]

A large-scale network[123] based on detailed neuronal architecture (65,400 neurons; 4,860,450 connections) has been proposed. It encompasses portions of two visual areas and associated thalamic and reticular thalamic nuclei addressing

the generation of slow oscillations (slower than 1 Hz). The areas are characterized by slow membrane potential fluctuations of cortical neurons, with depolarizing (up-state) and hyperpolarizing (down-state) components alternating with a frequency slower than 1 Hz. The model was used to simulate effects of neuromodulators or changes in synaptic strength[43] in slow waves. Furthermore, in a modified model encompassing three-layered motor cortex, effects of transcranial magnetic stimulation were investigated.[138]

CONCLUSIONS AND PERSPECTIVES

Ever since the first EEG recording by Hans Berger with a Siemens galvanometer, technical advances were instrumental for progress in sleep research. This has been evident in regard to the issues addressed in this chapter. The increasing ease of recording the sleep EEG from multiple derivations, combined with access to computer programs to analyze and visualize the data, has made it possible to explore the sleep process on a regional level. Differences became readily apparent between frontal and occipital derivations.[139,140] Frontal derivations have been shown to exhibit the largest response to changes in sleep pressure in terms of both SWA during sleep and theta activity during waking.[141-144] The topographic analysis of the EEG revealed that the frontal predominance of low-frequency activity is a common feature of NREM sleep, REM sleep, and waking and therefore represents a state-independent trait.[142,145] Thus basic features of the homeostatic process are present throughout the sleep-wake cycle, suggesting that sleep and waking could be seen as part of a continuum.[144]

Regional differences pose a challenge to modeling because differences in the characteristics of the homeostatic process (e.g., time constants, level of asymptote) may reflect different regulatory features of specific neuronal ensembles.[87,89]

The report of unihemispheric deep sleep in the dolphin[146,147] not only constituted evidence for a regional segregation of the sleep process but also raised the question of its functional significance. Hypotheses have been advanced implying that regional increases in neuronal activity and metabolic demand during wakefulness might result in selective changes in EEG synchronization of these neuronal populations during NREM sleep.[148-153]

The theory of a local, use-dependent increase of sleep intensity was tested by investigating whether a local activation of a particular brain region during wakefulness affects the EEG recorded from the same site during sleep. The first positive result consisted of the increase in low-frequency activity over the contralateral somatosensory cortex in the first hour of sleep following a vibratory stimulus to the contralateral hand.[154] This regional use-dependent effect was subsequently confirmed and extended.[155,156] Analogous findings were obtained in the rat and human, in which unilateral sensory or optic stimuli, respectively, during waking caused an interhemispheric shift in low-frequency power in the NREM sleep EEG.[157,158] Conversely, in a human study the selective understimulation of the cortical arm projection area during waking achieved by unilateral arm immobilization induced a reduction of power over the corresponding cortical region during subsequent sleep.[159]

The notion of sleep homeostasis, originally derived from the sleep-wake-dependent changes of the EEG slow waves, was recently expanded to the synaptic level. Tononi and Cirelli[151,152] proposed a synaptic homeostasis hypothesis postulating that synaptic strength is maintained over time by alternating phases of predominant potentiation during waking with phases of predominant depression during sleep. NREM sleep and the typical slow waves would subserve synaptic downscaling and thereby safeguard energy, space, and cellular supplies. The synaptic homeostasis hypothesis has the merit of relating the changes at the level of the EEG to well-known mechanisms at the synaptic level and thereby allowing specific predictions that can be not only simulated,[43] but also tested by electrophysiologic and neurochemical techniques in both humans and animals.[160-163] Vyazovskiy and Harris[153] proposed that periods of reduced synaptic input (so-called off periods or down states) serve prophylactic cellular maintenance. Krueger and collaborators[164-166] also underline the local aspect of sleep. Starting from an early theoretical paper,[149] they view sleep as an emergent property of cortical columns and propose a nonlinear mathematical model to account for the interactions between columns.[150]

In conclusion, models have proved useful for delineating regulating processes underlying such a complex and little-understood phenomenon as sleep and thereby offer a conceptual framework for analyzing existing and new data. The major models have already inspired a considerable number of experiments.

CLINICAL PEARL

The common experience that a good night's sleep dissipates fatigue and tiredness and regenerates energy points to a specific restorative function of sleep that cannot be achieved by merely resting. Sleep homeostasis denotes a basic principle of sleep regulation that can lead to a better understanding of sleep pathologies. Deficient sleep homeostasis might account for the altered sleep architecture in depression, and the transient normalization of sleep propensity can explain the antidepressant effect of sleep deprivation. The elucidation of sleep homeostasis at the cellular and molecular levels could open new avenues for the pharmacologic therapy of sleep disorders.

SUMMARY

Sleep homeostasis denotes a basic principle of sleep regulation. A sleep deficit elicits a compensatory increase in the intensity and duration of sleep, whereas excessive sleep reduces sleep propensity. It is as though sleep pressure is maintained within a range delimited by an upper and lower threshold. Sleep homeostasis is represented in the two-process model of sleep regulation by process S, which increases during waking and declines during sleep. The timing and propensity of sleep are modulated also by a circadian process. EEG SWA serves as an indicator of sleep homeostasis in NREM sleep. The level of SWA, a correlate of sleep intensity, is determined by the duration of prior sleep and waking. In the waking EEG, theta activity shows a rising trend with the progression of wakefulness and represents a marker of process S. However, in contrast to SWA, theta activity undergoes a marked circadian modulation. Advanced versions of the two-process model were applied to simulate the SWA pattern in a variety of experimental schedules. Other models addressing sleep regulation are derived from neurophysiologic data.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation and nano-tera.ch.

Selected Readings

Achermann P, Borbély AA. Low-frequency (<1 Hz) oscillations in the human sleep EEG. *Neuroscience* 1997;**81**:213–22.

Blake H, Gerard RW. Brain potentials during sleep. *Am J Physiol* 1937;**119**: 692–703.

Booth V, Diniz Behn CG. Physiologically-based modeling of sleep-wake regulatory networks. *Math Biosci* 2014;**250**:54–68.

Borbély AA. A two process model of sleep regulation. *Hum Neurobiol* 1982;**1**:195–204.

Borbély AA, Baumann F, Brandeis D, et al. Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephalogr Clin Neurophysiol* 1981;**51**:483–93.

Daan S, Beersma DGM, Borbély AA. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol* 1984;**246**:R161–78.

Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci* 1995;**15**: 3526–38.

Mang GM, Franken P. Genetic dissection of sleep homeostasis. *Curr Top Behav Neurosci* 2015;**25**:25–63.

Olbrich E, Claussen JC, Achermann P. The multiple time scales of sleep dynamics as a challenge for modelling the sleeping brain. *Philos Trans A Math Phys Eng Sci* 2011;**369**:3884–901.

Phillips AJ, Robinson PA. A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *J Biol Rhythms* 2007;**22**:167–79.

Ramakrishnan S, Lu W, Laxminarayan S, et al. Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss? *J Sleep Res* 2015;**24**:262–9.

Skeldon AC, Dijk DJ, Derks G. Mathematical models for sleep-wake dynamics: comparison of the two-process model and a mutual inhibition neuronal model. *PLoS ONE* 2014;**9**:e103877.

Tononi G, Cirelli C. Sleep function and synaptic homeostasis. *Sleep Med Rev* 2006;**10**:49–62.

Van Dongen HPA, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;**26**:117–26.

Vyazovskiy VV, Harris KD. Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nat Rev Neurosci* 2013;**14**: 443–51.

A complete reference list can be found online at ExpertConsult.com.

Chapter
37

Circadian Rhythms in Sleepiness, Alertness, and Performance

Rylie J. Gabehart; Hans P.A. Van Dongen

Chapter Highlights

- Alertness and performance vary across the day, driven by the circadian rhythm of the biologic clock. We tend to be less alert in the early morning and late at night, but the changes in alertness and performance over time also depend on the circumstances.
- A variety of other factors (e.g., activity, posture, caffeine intake) influence the pattern of circadian rhythmicity as observed in alertness and performance measures. Investigating the effect of the biologic clock requires careful control over these so-called masking effects.
- Temporal patterns of alertness and performance also reflect the interaction of the circadian process with a homeostatic process regulating sleep. Accounting for this homeostatic process is critical for understanding and predicting the occurrence of performance deficits across the 24 hours of the day.

Both wakefulness and sleep are modulated by an endogenous, circadian regulating system, the biologic clock, located in the suprachiasmatic nuclei of the anterior hypothalamus. The impact of the biologic clock goes beyond compelling the body to fall asleep and to wake up again. The biologic clock also modulates hour-to-hour waking behavior, as reflected in alertness and performance, generating circadian rhythmicity in almost all neurobehavioral variables investigated.

Before focusing on circadian rhythmicity in waking neurobehavioral functions, it is important to give a brief description of some variables capturing aspects of waking functioning because different definitions can be found in the literature.[1-3] Here, we use the term *sleepiness* for subjective reports of sleepiness or the desire to sleep. In operational settings the term *fatigue* is often used in lieu of sleepiness[4] (but in this chapter, the term is not used). By *alertness* we mean the antonym of sleepiness (although these terms have been differentiated[5]) and the ability to sustain attention. *Performance* refers to cognitive functioning on tasks ranging from psychomotor vigilance and working memory tests to logical reasoning and decision-making tasks. These concepts may be discussed collectively as *neurobehavioral functioning*.

The term *sleepiness* captures the link between alertness and performance during wakefulness on the one hand and sleep on the other hand. The interaction of the circadian and sleep-wake systems in regulating sleepiness, alertness, and performance is described in the second part of this chapter. Sleep propensity[6,7] is not covered in this chapter; the focus here is on effortful cognitive performance and the associated subjective states.

CIRCADIAN RHYTHMS

Self-Report Measures of Sleepiness and Alertness

Many different techniques are available for the detection of circadian rhythmicity in neurobehavioral variables, including a wide array of subjective measures of sleepiness and alertness. These include an array of visual analogue scales (VAS)[8]; Likert-type rating scales such as the Stanford Sleepiness Scale[9] and the Karolinska Sleepiness Scale[10]; and adjective checklists such as the Profile of Mood States.[11] Despite structural differences among these scales, self-report measures of sleepiness and alertness tend to be highly correlated over time. They have been used to index circadian rhythmicity by applying them repeatedly across the day.[12-14]

Subjective measures of sleepiness and alertness are subject to numerous confounding influences, which can "mask" their circadian rhythmicity. *Masking* refers to the evoked effects of noncircadian factors on measurements of circadian rhythmicity. The context in which such measurements are taken (i.e., the environmental and experimental conditions) is a major source of masking effects.[15] Masking can alter or obscure a circadian rhythm or create the appearance of a circadian rhythm where there is none. Masking factors affecting sleepiness and alertness may include the following: demand characteristics of the experiment,[16] distractions by environmental stimuli and noise,[17] boredom and motivational factors,[18-20] stimulation,[21] stress,[22] food intake,[23,24] posture and activity,[25,26] ambient temperature,[19] lighting conditions,[27,28] and drug intake (e.g., caffeine).[29]

Physical, mental, and social activities can be masking factors for circadian rhythms in subjective sleepiness and alertness as well. For example, subjects report feeling less alert after performing a challenging cognitive task than before performing said task.[30] In general, prior activity can influence subjective estimates, and it can interact with circadian effects if not properly controlled when measuring rhythmicity in subjective states. Sleep and wakefulness can also be considered masking factors when measuring circadian rhythmicity in neurobehavioral variables. Sleep and sleep loss have significant effects on alertness and performance, as discussed in the second half of this chapter.

A variety of factors also affect the biologic clock itself, changing its timing by causing phase advances or delays. Such

factors are called *zeitgebers* (time givers or time cues) and include exercise, social cues, food intake, sleep, and especially light exposure.[31-34] Their effects typically vary depending on the timing of exposure relative to the circadian cycle of the biologic clock, making it difficult to account for them. As such, it is important to control zeitgeber exposure in studies designed to measure circadian rhythms in neurobehavioral variables. Zeitgebers are discussed further in other chapters in this volume.

Cognitive Performance

Rather than relying on subjective measures, many studies of circadian rhythms have used objective performance measures. For example, studies have employed search-and-detection tasks[35] and simple and choice reaction time tasks[36] to obtain objective measures of circadian variation in cognitive performance. In many tasks, the speed or accuracy of responses to a series of stimuli are analyzed. The sensitivity of the performance metrics in such tasks depends on speed and accuracy tradeoffs (which in turn depend on the demand characteristics of the experiment); on task duration; on the stimulus density (i.e., the number of stimuli presented per unit time); and on whether the task is subject paced or rather experimenter or computer paced.[37-39]

Many performance outputs have been considered, including signal detection,[40] simple sorting,[41] logical reasoning,[42] memory access,[43] meter reading accuracy,[44] and school performance.[45] Furthermore, a number of subcomponents of cognition and cognitive processes involved in task performance can be distinguished, such as sensory input, stimulus encoding, working memory updating, and motor response.[46,47] A variety of tasks have been used to study circadian variation in these different aspects of performance. Some studies concluded that different tasks[48,49] and different task outcomes[50,51] may yield different peak phases of circadian rhythmicity. This has led to speculation that there are many different circadian rhythms and multiple different clock mechanisms controlling them.[52,53]

Under strictly controlled laboratory conditions, however, most of the intertask differences disappear.[54,55] As illustrated in Figure 37-1, it can generally be stated that under such conditions, the circadian rhythms of neurobehavioral performance variables covary with subjective sleepiness. Furthermore, these rhythms reflect the circadian rhythm of core body temperature (CBT), a conventional marker of the biologic clock. High and low CBT roughly correspond to good and poor performance, respectively.[55-57] However, there is a phase difference, such that neurobehavioral variables exhibit their average minimum approximately between 3.0 and 4.5 hours after the time of the body temperature minimum.

This phase delay contradicts the common belief that alertness and performance are worst at the body temperature minimum. Although body temperature predominantly reflects the endogenous biologic clock, neurobehavioral functions are also affected by homeostatic pressure for sleep, which builds up over time awake (as discussed in the second half of this chapter) and contributes to the phase delay. Thus neurobehavioral functions generally show a circadian decline at night, as is observed in CBT, but they continue their decline after CBT begins to rise, making the subsequent 2- to 6-hour period (i.e., clock time approximately 6 AM to 10 AM) a zone of maximum vulnerability to degraded alertness and performance failure.

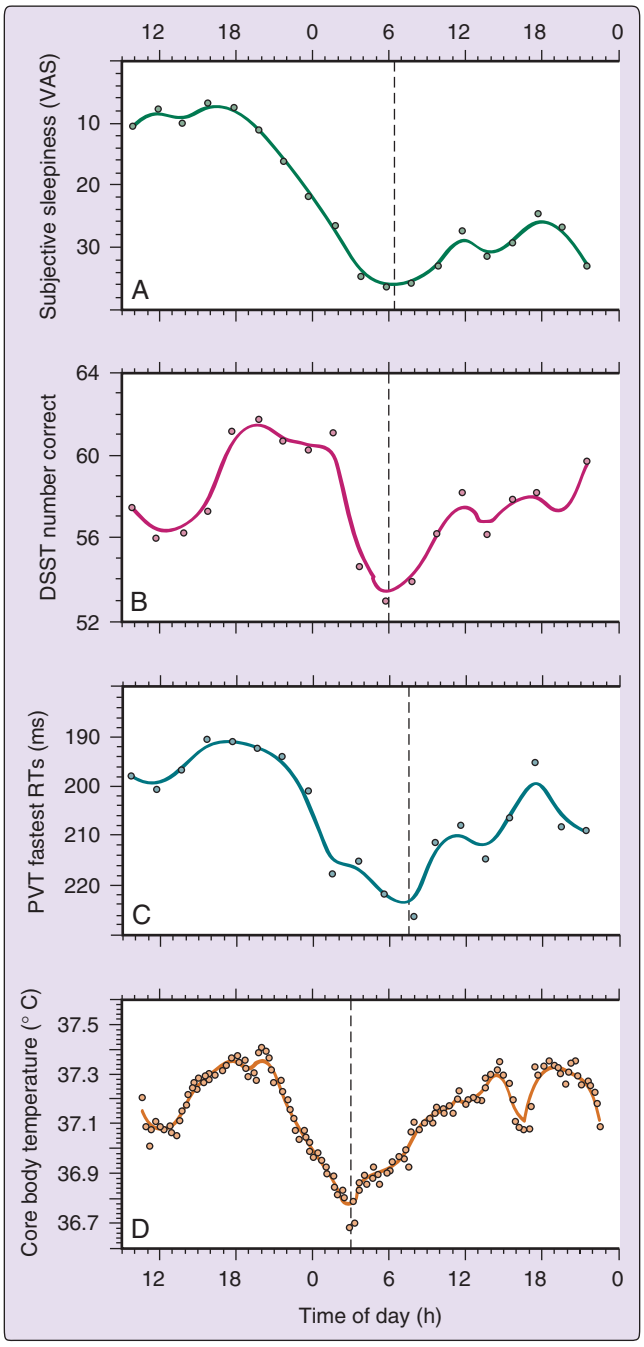


Figure 37-1 Covariation of circadian changes in neurobehavioral variables and core body temperature. **A,** Subjective sleepiness (scale reversed) as assessed by visual analogue scale (VAS).[8] **B,** Cognitive performance as assessed by the digit symbol substitution task (DSST).[129,155] **C,** The 10% fastest reaction times (RTs; scale reversed) as assessed by the psychomotor vigilance test (PVT).[119] **D,** Core body temperature as assessed using a rectal probe. Data shown are the mean values from five subjects who remained awake in dim light, in bed, in a constant routine protocol, for 36 consecutive hours (a distance-weighted least-squares function was fitted to each variable). The circadian trough in each variable is marked by a *vertical broken line.*

In comparison with subjective sleepiness and alertness, assessment of circadian rhythmicity in cognitive performance tends to be complicated.[58] On some tasks, subjects may change their performance strategy on a task over time, for example, by invoking subvocalization in a rhythmic, circadian pattern.[59]

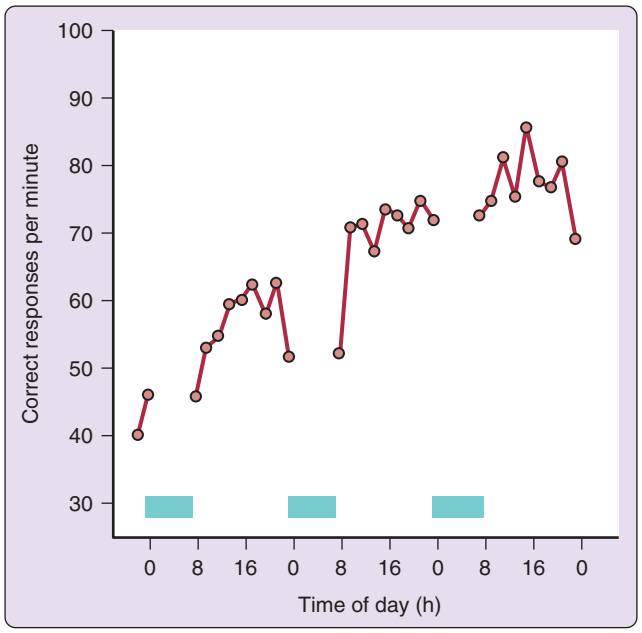


Figure 37-2 Practice effect on the serial addition-subtraction test. Data shown are mean cognitive throughput (correct responses per minute) of 29 subjects tested every 2 hours from 7:30 AM until 11:30 PM each day over a 3-day period. In the serial addition-subtraction test, subjects are presented with a rapid sequence of two single digits (0–9) followed by an operator (+ or –). They are instructed to enter only the least-significant single digit of the algebraic sum, unless the result is negative, in which case 10 is to be added to the answer first.[156] The *solid bars* indicate 8-hour sleep periods (from 11:30 PM until 7:30 AM).

Thus it can be difficult to distinguish the circadian rhythm in task performance per se from that of simultaneous changes in performance strategy. The same issue applies to compensatory effort (i.e., increased effort to keep up performance). The effect of compensatory effort may be especially notable if subjects are informed about their results during a performance task (i.e., performance feedback).[60] Differences among people in aptitude for a task may also confound the assessment of circadian rhythmicity in cognitive performance. This issue can be avoided by means of within-subject study designs and analyses.

Another complicating factor for the assessment of circadian rhythmicity in cognitive performance is the practice effect (or learning curve). This is illustrated in Figure 37-2, which shows performance on a serial addition-subtraction task improving substantially across 3 consecutive days (note the doubling of mean correct responses within 30 test bouts). This practice effect dominates the performance profile, obscuring circadian changes within days. The practice effect contaminates most cognitive performance tasks, and it is difficult to dissociate from the circadian rhythm in performance. This problem may be circumvented, however, by training subjects to asymptotic performance levels before attempting to assess circadian rhythmicity in cognitive performance.

Many of the variables that mask circadian rhythmicity in subjective estimates of sleepiness and alertness (e.g., demand characteristics, distractions, motivation, posture, ambient temperature, lighting conditions; see earlier) may also mask circadian variation in performance. The effects of masking include distortion of the magnitude of circadian variation, shifting of the timing of observed circadian rhythmicity, and changes in the shape of the circadian profile; even total concealment of the circadian rhythm is possible (see Figure 37-2, where sleep prevents measurement of nocturnal performance and the practice effect obscures circadian variation in diurnal performance). Thus it is difficult to extract meaningful information about the amplitude (magnitude) and phase (timing) of the circadian rhythm in performance measures without understanding the masking effects that influence these variables.

Physiologic Measures

The circadian rhythm in cognitive performance reflects functional changes in the brain over time of day. Evoked or event-related potentials (ERPs)—waves (peaks and troughs) in the electroencephalogram (EEG) produced by the brain in response to a stimulus—have been used to measure alertness and cognitive performance. Typically, many ERP measurements are needed (i.e., many stimuli must be presented) to average out the background EEG. Therefore ERPs are usually recorded during repetitive search-and-detect and reaction time tasks. Diurnal changes in the amplitude and the location of ERP waves have been interpreted as reflecting circadian variations in alertness.[61,62] Hemispheric differences have been detected,[63] suggesting separate circadian rhythms for the left and right hemispheres. However, masking from a variety of sources presents a problem in the interpretation of ERP data.[64]

Changes in the background EEG during wakefulness have also been associated with circadian variation in alertness. Despite difficulties in the recording and analysis of the waking EEG,[65,66] the amounts of theta and alpha activity (i.e., EEG activity in the frequency bands from 4 to 8 Hz and from 8 to 12 Hz, respectively) in the resting EEG with eyes held either open or closed (to avoid artifacts from blinking) have been related to alertness level.[67-69] However, significant effects in the EEG occur primarily when alertness is much lower than what is normally encountered at the trough of circadian variation—such as when subjects are sleep deprived.[70]

The EEG has also been used to measure the latency to fall asleep—a measure of sleep propensity—at various times of day. These sleep latency tests include the Multiple Sleep Latency Test[6] and the Maintenance of Wakefulness Test[71] and variations of these paradigms. Sleep latency tests are covered in other chapters in this volume.

Slow eyelid closures have been found to be systematically related to sleepiness.[72,73] Similar findings have been reported for slow-rolling eye movements and other ocular variables.[10,74-79] Pupil diameter is related to autonomic tone (i.e., pupils dilate with greater sympathetic dominance), which covaries with sleep pressure.[80] Pupillometry may therefore also yield estimates of sleepiness,[81] but only if environmental light and other sources of error variance are strictly controlled. The various ocular measures of sleepiness have been investigated primarily under conditions of considerable sleep loss, when the observed effects are more substantial than across the circadian cycle. The same applies for cardiovascular measures that have been found to correlate with sleepiness, such as heart rate variability.[82] Whether ocular and cardiovascular measures can be employed reliably for detecting circadian fluctuations in sleepiness remains to be determined.

Interindividual Differences

Interindividual differences in circadian phase[83] and amplitude[84] are reported throughout the literature. Interindividual

differences in the free-running circadian period have been reported as well,[85] although it should be emphasized that under normal, entrained circumstances the circadian period equals 24 hour.[86] Interindividual differences in circadian variables have been linked to development[87,88] and aging[89-91] and to genetic factors,[92] which are discussed in other chapters in this volume.

Morningness-eveningness (i.e., the tendency to be an early "lark" or a late "owl") is a well-known, phenotypic aspect of interindividual variation in circadian rhythmicity.[93] Morningness-eveningness is commonly measured with questionnaires that ask about a person's preferred timing of sleep and daily activities.[94,95] Laboratory studies have shown that morning- and evening-type individuals differ in the phase (timing) of the endogenous circadian rhythms in CBT[83] and circulating levels of melatonin.[96] This difference is echoed in the diurnal course of their neurobehavioral functioning[97,98]—some people are more alert and perform better in the morning hours, whereas others are at their best later in the day.

In some individuals, there appears to be an afternoon dip in the circadian profiles of CBT and neurobehavioral variables,[99] which is referred to as the midafternoon, siesta, postprandial, or postlunch dip. This phenomenon has been observed in field studies[44] and in controlled laboratory experiments[100] and is thought to be endogenous and independent of food intake. Epidemiologic analyses of human performance (e.g., road accident rates across the 24 hours of the day[101]) seem to support the existence of a midafternoon dip (Figure 37-3). However, proper interpretation of such data requires accounting for temporal variation in exposure, that is, differential amounts of activity contributed by varying numbers of people over time,[102] which is complicated. More direct evidence for the existence of a midafternoon dip in some individuals comes from studies on sleep propensity[103,104] and on the natural timing of daytime naps.[105] This suggests that interindividual differences in sleep-wake parameters may play a role.

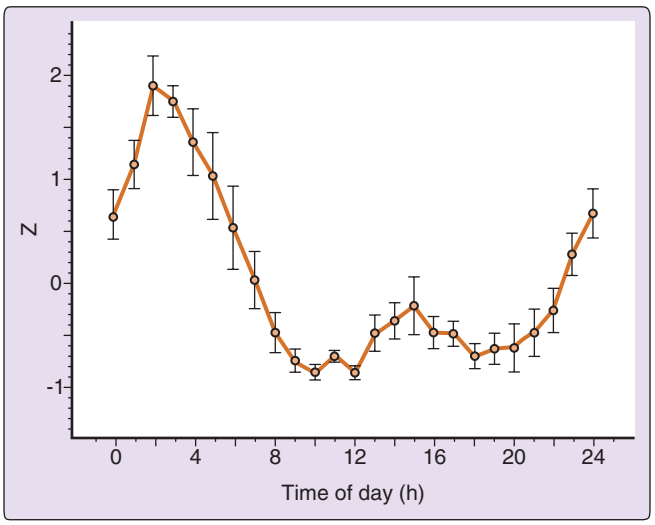


Figure 37-3 Circadian rhythm in road traffic accident risk. Data shown are mean (with standard error) over six published studies after Z transformation. (From Folkard S. Black times: temporal determinants of transport safety. *Accid Anal Prev* 1997;29:417–30, with permission.)

CIRCADIAN RHYTHMICITY VERSUS SLEEP-WAKE CYCLES

Sleep Deprivation

Considerable research effort has been devoted to unmasking circadian rhythms, that is, eliminating sources of extraneous variance to expose the endogenous circadian rhythms in variables of interest, including alertness and cognitive performance. The constant routine procedure[106,107] is generally regarded as the gold standard for measuring unmasked circadian rhythms. By keeping subjects awake with a fixed posture in a constant laboratory environment for at least 24 hours, circadian rhythms in a variety of physiologic and neurobehavioral variables can be recorded without confounds. Indeed, for CBT and melatonin, the circadian rhythm is believed to be free of masking effects when measured under constant routine.

The elimination of sleep (i.e., sleep deprivation) and the stimulation required to sustain wakefulness constitute masking factors for neurobehavioral variables. In constant routine experiments, these masking effects are evident in subjective and objective measures of alertness.[83,108] Figure 37-1 shows the somewhat increased sleepiness and reduced cognitive performance after 30 hours awake under constant routine compared with the values of these variables 24 hours earlier (i.e., at the same circadian time but without sleep deprivation).

Typically, superimposed on the circadian rhythm in a neurobehavioral variable, there is a progressive change across time spent awake.[109,110] When total sleep deprivation is continued for several days (whether in a constant routine procedure or an experimental design involving ambulation), the detrimental effects on alertness and performance increase. Although the circadian rhythm can thus be exposed,[111] it is overlaid by a continuing change reflecting a buildup of pressure for sleep.[112]

This is illustrated in Figure 37-4, which shows lapses of attention on a psychomotor vigilance test (PVT)[113] during a 16-hour baseline day and a 64-hour period of sleep deprivation. As seen in the figure, lapses on the PVT are relatively rare during the baseline day and during the first 16 hours of the sleep deprivation period, both of which fall on the diurnal portion of the circadian cycle and have little pressure for sleep. However, after the first 16 hours of the sleep deprivation period, lapses are clearly evident, indicating a substantial increase in neurobehavioral dysfunction. There is a steady rise in lapses across days, modulated by a pronounced circadian rhythm—the combined effect roughly takes the form of a staircase function.

Neurobehavioral functions that show circadian variation also appear to respond to sleep loss, and vice versa. The interaction of the circadian rhythm with the effect of sleep deprivation, which is nonlinear,[114,115] makes it difficult to dissociate the two effects in constant routine and sleep deprivation experiments (although it is possible by focusing on interindividual differences[116]). However, a reasonable separation of the two effects can be achieved with other experimental designs, as discussed later.

It is noteworthy that performance impairment during the circadian trough and while sleep deprived is associated with increased moment-to-moment variability in brain functioning,[117,118] of which lapses of attention on the PVT are a sensitive measure.[119] The state instability hypothesis posits that the moment-to-moment variability is caused by sleep-initiating

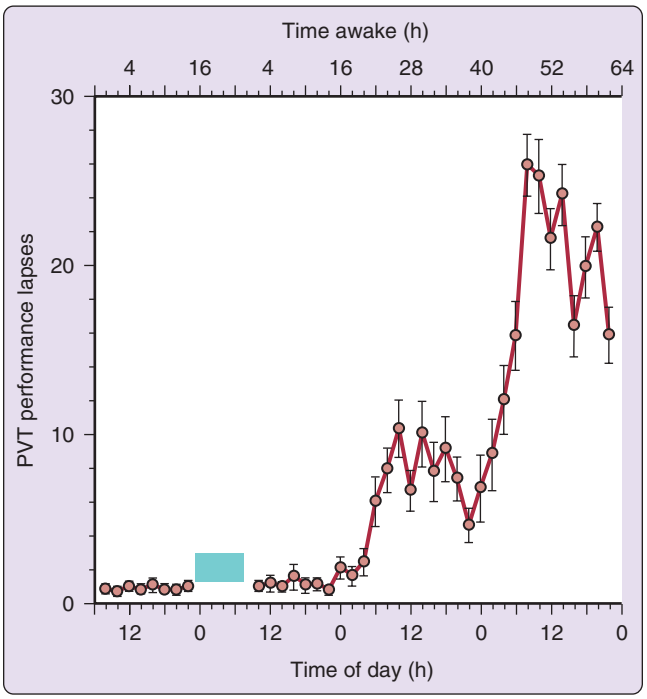


Figure 37-4 Performance lapses on the psychomotor vigilance test (PVT) during a 16-hour baseline day and a 64-hour period of sleep deprivation. Data shown are mean (with standard error) number of lapses (reaction times longer than 500 ms) for 24 subjects tested every 2 hours on the 10-minute PVT. This simple reaction time task requires subjects to respond as quickly as possible to a stimulus that appears on a display at random intervals of 2 to 10 seconds.[119] The *solid bar* indicates an 8-hour sleep period (from 11:30 PM until 7:30 AM) between the baseline day and the sleep deprivation period.

mechanisms interfering with sustained wakefulness, making cognitive performance unstable and dependent on compensatory mechanisms such as increased effort to perform.[120] According to local sleep theory,[121] the state instability may occur specifically in brain networks involved in performance of the cognitive task at hand.[122]

Sleep-Wake Regulation

The observed superposition of circadian modulation of alertness and performance on monotonic change during sleep deprivation (see Figure 37-4) has prompted efforts to mathematically model the regulatory processes involved. The two-process model of sleep regulation has been applied to describe the temporal profiles of sleep[123,124] as well as waking alertness and performance.[124,125] The model consists of a homeostatic process (process S) and a circadian process (process C), which combine to determine the onset and offset of sleep. The two processes together also drive waking neurobehavioral functioning.

The homeostatic process represents a drive for sleep that increases during wakefulness and decreases during sleep. When the "homeostat" increases above a certain threshold, sleep is triggered; when it decreases below another threshold, wakefulness is invoked. The circadian process represents daily oscillatory modulation of the two thresholds.[123] An alternative (but equivalent) view is that the circadian process promotes wakefulness to counteract the homeostatic drive for sleep.[126,127] The two-process model is discussed more extensively in another chapter in this volume.

The circadian and homeostatic processes interact to determine waking neurobehavioral functions as expressed in alertness and performance.[110,114,128] This is clearly seen in prolonged sleep deprivation experiments (see Figure 37-4). For alertness and performance, sleep and sleep loss are not only masking factors but also dynamic biologic forces that interact with the circadian system.

Under conditions of chronic sleep restriction—that is, daily curtailment of sleep (rather than total sleep deprivation)—there is a progressive buildup of pressure for sleep across days that is not predicted by the two-process model.[129] It has been hypothesized that this reflects allostatic adjustment of the set point of the homeostatic process.[130] The buildup of sleep pressure across days notwithstanding, the circadian process partially protects the afternoon and early evening from neurobehavioral impairment,[131] as illustrated in Figure 37-5 (see also Figure 37-1). The period when this is most noticeable, a window of several hours preceding habitual bedtime, is known as the "forbidden zone for sleep"[132] or "wake maintenance zone."[133]

Forced Desynchrony and Ultradian Days

The forced desynchrony protocol[127,134,135] is an experimental procedure designed to dissociate the effects of the circadian and homeostatic processes. In this protocol, a subject stays in an environmentally and temporally isolated laboratory in which the sleep-wake cycle is scheduled to deviate substantially from the normal 24-hour day (e.g., 20-hour and 28-hour sleep-wake cycles have been used). The biologic clock is unable to synchronize to such a schedule. The subject therefore experiences two distinct influences simultaneously: the schedule of predetermined sleep and wake times controlling the homeostatic process, and the rhythm of the subject's unsynchronized (i.e., free-running) circadian process.

Neurobehavioral variables can be recorded during the subject's waking periods of this experimental design. By folding the data over either the free-running circadian cycle or the imposed sleep-wake cycle, the influence of the other cycle can be averaged out. In this manner, the effects of the circadian and homeostatic processes on the recorded variables are separated. As expected, forced desynchrony studies have found that both the circadian and homeostatic processes influence alertness and performance. The interaction of the two processes appears to be oppositional during natural diurnal wake periods (from about 7 AM until 11 PM) such that a relatively stable level of alertness and performance can be maintained throughout the day.[127] This explains why studies of alertness and performance tend to find little temporal variation during the waking portion of a normal day (see Figure 37-4).

Another way to dissociate the circadian and homeostatic processes is through study designs with very short (i.e., ultradian) sleep-wake cycles. Such paradigms seek to redistribute the opportunities for sleep and wakefulness across the natural 24-hour day to sample waking behavior across the circadian cycle without significantly curtailing the total amount of sleep. Studies have been done with a 7-/13-minute sleep-wake schedule,[7] which alternately allows subjects to sleep for 7 minutes and forces them to stay awake for 13 minute; with a 90-minute day schedule,[136] which alternately permits subjects to sleep for 30 minutes and forces them to stay awake for 60 minutes; and with a 3-hour ultra-short sleep-wake schedule,[137] which alternates 1 hour of sleep and 2 hours of wake periods.

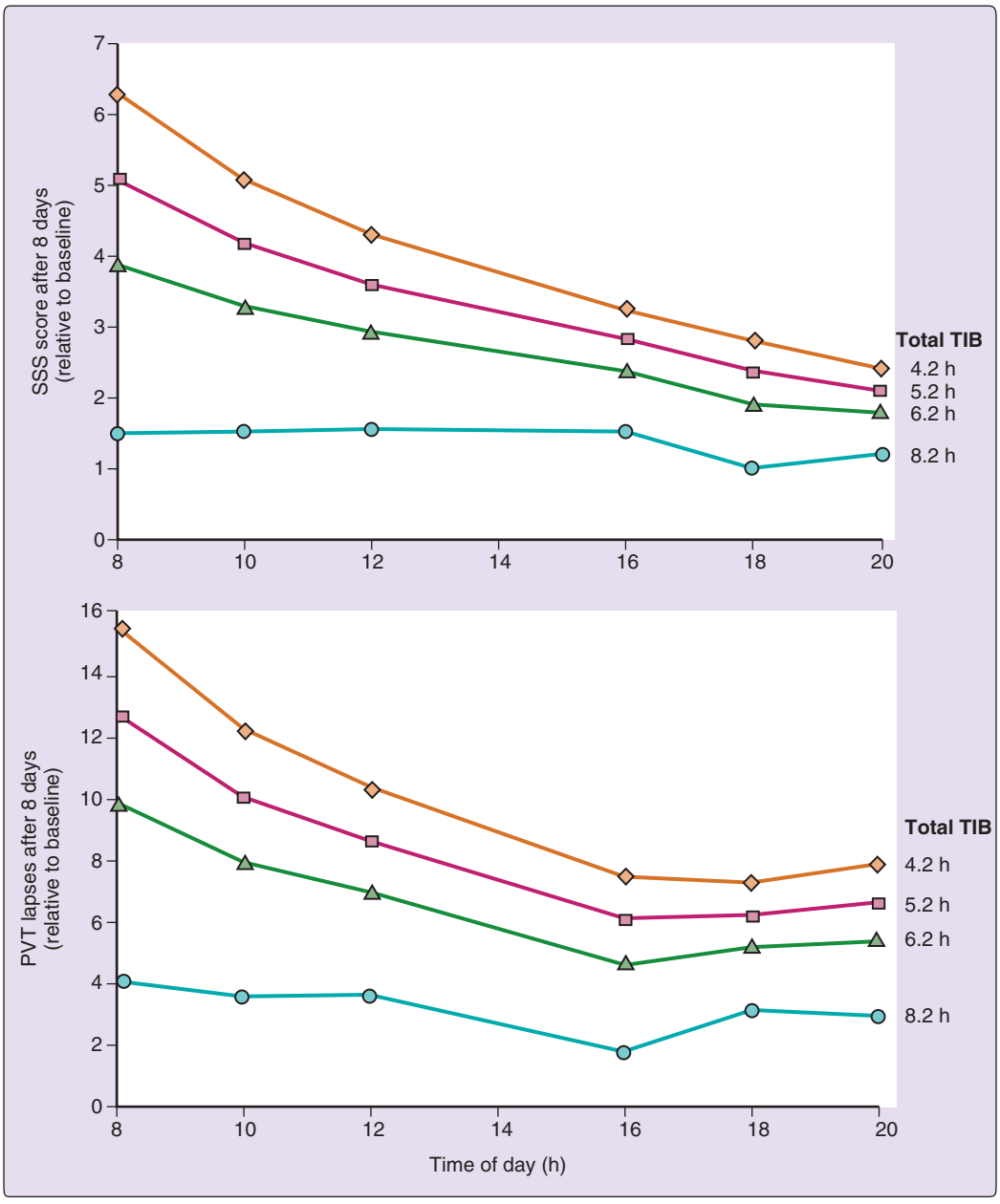


Figure 37-5 Sleepiness and performance under conditions of chronic sleep restriction as a function of time of day. Data shown are mean subjective sleepiness score on the Stanford Sleepiness Scale[9] (SSS; *top panel*) and mean number of lapses (reaction times longer than 500 ms) on the 10-minute PVT[119] (*bottom panel*) after 8 days of nocturnal sleep restriction with or without a daytime nap. Total time in bed (TIB) per day was 4.2 hours, 5.2 hours, 6.2 hours (sleep restriction conditions), or 8.2 hours (control condition); 90 subjects were each randomized to one of these conditions. (From Mollicone DJ, Van Dongen HPA, Rogers NL, et al. Time of day effects on neurobehavioral performance during chronic sleep restriction. *Aviat Space Environ Med* 2010;81:735–44, with permission.)

With respect to objective measures, studies with very short sleep-wake cycles have focused primarily on sleep propensity. However, cognitive performance was assessed in an experiment using the 7-/13-minute sleep-wake schedule and in an experiment employing the 3-hour ultrashort sleep-wake schedule. Robust circadian rhythms emerged for response times on a choice reaction time task[138] and on an abbreviated (5 minutes) version of the PVT.[137]

Subjective sleepiness scores were recorded in the 7-/13-minute sleep-wake schedule,[139] the 90-minute day schedule,[136] and the 3-hour ultrashort sleep-wake schedule,[137] and also showed clear circadian rhythms. However, after a 24-hour period of 7-/13-minute sleep-wake cycles (but not after 24 hours on the schedules with 90-minute or 3-hour cycles), the level of subjective sleepiness was elevated with respect to the initial level 24 hours earlier. This suggests that across 24 hours of the 7-/13-minute sleep-wake schedule, the recovery potential of the sleep obtained—which itself is modulated by the circadian process—may have been insufficient.

All things considered, the separation of circadian and homeostatic influences on neurobehavioral variables presents a conceptual, experimental, and mathematical problem. The

interaction of the two processes has been found to be nonlinear.[114,115] It is therefore difficult, if not impossible, to quantify the relative importance of the two influences on neurobehavioral functions, even in forced desynchrony and ultradian day experiments. Moreover, the relative contributions of the two processes may vary across different experimental conditions[54,114] and among subjects.[116,140]

Sleep inertia is yet another problem that may interfere with the assessment of alertness and performance in circadian studies and with the dissociation of the circadian and homeostatic processes. Sleep inertia refers to the cognitive performance impairment, feeling of disorientation, grogginess, and tendency to return to sleep experienced immediately after awakening.[141] Sleep inertia may affect alertness and performance on every artificial day of a forced desynchrony study or ultradian sleep-wake cycle study. The circadian and homeostatic processes appear to interact with sleep inertia,[142-145] varying the impact of sleep inertia across the artificial days in these study designs and making it difficult to account for its effects.

Circadian Regulation of Alertness and Performance in Context

Figure 37-6 shows a conceptual schematic of how the circadian drive for wakefulness, the homeostatic drive for sleep, the sleep inertia effect, and various internal states and external circumstances simultaneously affect neurobehavioral functioning.

As illustrated in the upper part of the figure, wakefulness typically begins with rapidly dissipating sleep inertia, which suppresses neurobehavioral functioning for a brief period after awakening. The homeostatic drive for sleep accumulates throughout wakefulness and progressively downregulates neurobehavioral performance and alertness. Unlike the circadian process, which is limited in amplitude, the homeostatic drive for sleep may accumulate far beyond the levels typically encountered in a 24-hour day (indicated in Figure 37-6 by the increasing density of downward arrows).

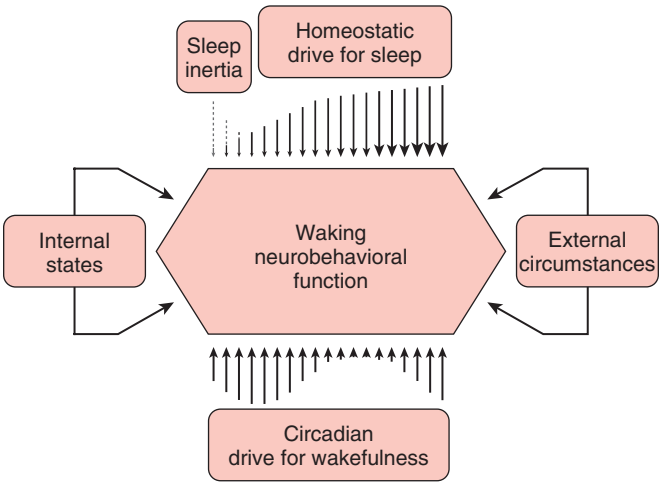


Figure 37-6 Schematic representation of the conceptual interplay of circadian and homeostatic processes and other factors in the regulation of neurobehavioral functioning. Sleep inertia and the homeostatic process degrade performance. Sleep inertia dissipates rapidly after awakening, whereas the homeostatic drive for sleep builds up progressively over time awake. The circadian process provides an oscillatory countereffect by promoting wakefulness during the day and withdrawing the effect at night. A wide range of internal states and external circumstances modulate waking neurobehavioral functions. Their effects are typically transient and may improve or degrade neurobehavioral functioning in interaction with the homeostatic and circadian processes.

In opposition to these suppressing influences on performance and alertness is the endogenous circadian rhythm of the biologic clock, as illustrated in the bottom part of the figure. The circadian process modulates performance and alertness by promoting wakefulness. The improvement in waking neurobehavioral functions by the circadian drive for wakefulness is an oscillatory process, which periodically involves robust opposition to the homeostatic process alternated with withdrawal of the circadian drive.

Modulators of neurobehavioral functions other than the sleep and circadian drives are subsumed in Figure 37-6 under the broad categories of internal states and external circumstances. They may include wake-promoting factors—endogenous (e.g., anxiety) or exogenous (e.g., caffeine intake)—that counteract the homeostatic drive for sleep. They may also include sleep-promoting factors—endogenous (e.g., immune response-induced) or exogenous (e.g., rhythmic motion)—that oppose the circadian drive for wakefulness, either directly or indirectly by exposing the homeostatic drive for sleep.

The neurobiologic substrates of these exogenous and endogenous factors are diverse. Although common in the real world, they are considered masking factors in most laboratory experiments. However, with regard to the regulation of alertness and performance, they cannot be regarded as mere confounds that should be eliminated or controlled. Although the effects of internal states and external circumstances tend to be transient, they are an integral part of the regulation of neurobehavioral functions and the interaction of individuals with their environment.[93]

Understanding the complexity of circadian rhythmicity in neurobehavioral functions is important when the sleep-wake rhythm is misaligned, as is the case during night-shift work,[146] or when the circadian rhythm is misaligned, as is the case after transmeridian flights.[147] In such situations, the circadian and homeostatic processes are not properly synchronized, and their interaction degrades alertness and performance. This problem is compounded when sleep is lost chronically,[129,148-150] putting individuals at increased risk for accidents.[151-154]

CLINICAL PEARL

Clinicians should recognize that 24-hour profiles of alertness and performance combine the effects of endogenous circadian rhythmicity, homeostatic regulation of sleep, sleep inertia, and a variety of endogenous and exogenous "masking" influences. Distinguishing these factors and accounting for interindividual differences is important for diagnosis and treatment of sleep disorders involving excessive daytime sleepiness or circadian dysregulation.[3]

SUMMARY

The biologic clock drives circadian rhythms and regulates changes in behavior over the 24 hours of the day. There are circadian rhythms in almost all variables describing alertness and performance. People tend to be less alert in the early morning and late at night, but it also depends on the

circumstances. A variety of factors (e.g., activity, posture, light exposure) can mask circadian rhythms. Even with masking influences experimentally controlled, measurements of the endogenous circadian rhythmicity in alertness and performance still reflect the interaction of the biologic clock with the homeostatic regulation of sleep. It has been argued that certain masking factors (e.g., sensory stimulation, body movement) are an integral part of the mechanisms regulating waking neurobehavioral functions. Accounting for their interactions with the biologic clock helps to explain or predict the occurrence of cognitive performance deficits across the circadian cycle.

Selected Readings

Daan S, Beersma DGM, Borbély AA. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol* 1984;**246**:R161–78.

Dijk DJ, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci Lett* 1994;**166**:63–8.

Duffy JF, Czeisler CA. Age-related change in the relationship between circadian period, circadian phase, and diurnal preference in humans. *Neurosci Lett* 2001;**318**:117–20.

Facer-Childs E, Brandstaetter R. The impact of circadian phenotype and time since awakening on diurnal performance in athletes. *Curr Biol* 2015;**25**:518–22.

Folkard S, Åkerstedt T. Trends in the risk of accidents and injuries and their implications for models of fatigue and performance. *Aviat Space Environ Med* 2004;**75**:A161–7.

Johns MW. Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *J Sleep Res* 2002;**11**:61–7.

Johnson MP, Duffy JF, Dijk DJ, et al. Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *J Sleep Res* 1992;**1**:24–9.

Kerkhof GA. Inter-individual differences in the human circadian system: a review. *Biol Psychol* 1985;**20**:83–112.

Lavie P. Ultrashort sleep-waking schedule. III. "Gates" and "forbidden zones" for sleep. *Electroencephalogr Clin Neurophysiol* 1986;**63**:414–25.

Minors DS, Waterhouse JM. Anchor sleep as a synchronizer of rhythms on abnormal routines. *Int J Chronobiol* 1981;**7**:165–88.

Mollicone DJ, Van Dongen HPA, Rogers NL, et al. Time of day effects on neurobehavioral performance during chronic sleep restriction. *Aviat Space Environ Med* 2010;**81**:735–44.

Whitney P, Hinson JM, Jackson ML, Van Dongen HPA. Feedback blunting: total sleep deprivation impairs decision making that requires updating based on feedback. *Sleep* 2015;**38**:[in press].

A complete reference list can be found online at ExpertConsult.com.

Chapter
38

Central and Peripheral Circadian Clocks

Edith Grosbellet; Etienne Challet

Chapter Highlights

- The master clock in the suprachiasmatic nucleus (SCN) controls the sleep-wake cycle and hormonal rhythms as well as a multitude of other circadian rhythms.
- The suprachiasmatic clock conducts the multitude of brain and peripheral clocks to ensure circadian temporal organization and its adjustment to the daily variations of the environment.
- Rhythmic signals from the SCN couple the master clock to secondary brain and peripheral clocks through behavioral, nervous, and neurohumoral pathways. Endocrine rhythms (e.g., pineal melatonin and adrenal glucocorticoids) distribute internal temporal messages within the body.
- Light perceived by the retina is the most potent synchronizer of the master clock in the SCN, whereas most brain and peripheral clocks can be shifted as a function of meal time as well as the timing of sleep.
- Circadian clocks and intracellular metabolism are tightly and reciprocally connected.

THE CENTRAL CIRCADIAN CLOCK

Self-Sustained Oscillations

In mammals, the central clock is located in the suprachiasmatic nucleus (SCN) of the hypothalamus (Figure 38-1). See Chapter 27 for the genetics and genomics of circadian clocks. The SCN controls most circadian rhythms in behavior (e.g., sleep-wake cycle) and physiology (e.g., hormonal rhythms). The SCN consists of a heterogeneous population of neuronal and glial cells distributed in two anatomic subdivisions: a ventral "core" region, receiving retinal input, and a dorsal "shell" region, receiving dense input from the core. A majority of core neurons express vasoactive intestinal polypeptide (VIP), and fewer express gastrin-releasing peptide. In contrast, most shell neurons express vasopressin. Most neuropeptides colocalize with gamma-aminobutyric acid (GABA), and most synapses between SCN neurons are GABAergic.[1]

When physically isolated, either in vitro or in vivo, the SCN generates pronounced circadian rhythms of electrical activity.[2,3] Furthermore, recordings of firing rate in a dissociated culture of SCN neurons revealed the presence of single-cell circadian pacemakers, expressing circadian oscillations with different phases and periods.[4] SCN neurons are intrinsic but unstable oscillators, which need network interactions to stabilize their otherwise noisy cycling.[5] As a multicellular structure, the SCN provides a more precise and robust output than single SCN neurons.[6,7] Neuronal firing, chemical, and electrical (gap junctions) synapses are required for circadian coupling within the SCN in vivo.[1] The molecular clock machinery involves 24-hour oscillations of core clock components called *clock genes* (i.e., genes whose protein products are necessary for generating circadian rhythmicity within individual cells).[8] Within the molecular clockwork, two transcription factors, CLOCK and BMAL1, dimerize and bind to E-box sequences in the promoters of other clock genes, *Period 1 to 3 (Per1, Per2, Per3)* and *Cryptochrome 1* and *2 (Cry1, Cry2)* to activate their transcription. After reaching a given concentration in the cytoplasm, PERs and CRYs dimerize and translocate into the nucleus to inhibit their own transcription mediated by CLOCK/BMAL1 dimers. Auxiliary loops are composed of the circadian nuclear receptors REV-ERBα-β and RORα-β-γ, which modulate negatively and positively the circadian oscillations of *Bmal1* expression, respectively.[8,9] The SCN clock machinery also modulates the expression of numerous clock-controlled genes (e.g., vasopressin), which constitute circadian outputs that provide either local or distributed timing signals.[10] Many levels of regulation are involved in the proper functioning of the circadian clock, including transcriptional, posttranscriptional, and posttranslational mechanisms. Transcription rates can be affected by the recruitment of different transcription factors or by modifications within the chromatin structure. Of note, CLOCK possesses a histone-acetylase activity, essential for rescuing circadian rhythmicity in *Clock* mutant cells, revealing a crucial role of chromatin remodeling in clock mechanisms (Figure 38-1).[11] Moreover, posttranscriptional mechanisms regulate the stability and translation levels of messenger RNA (mRNA),[12] whereas posttranslational mechanisms, such as phosphorylations, control the targeting of clock proteins for proteosomal degradation.[13] All together, these regulations provide robust oscillations, resilient to large fluctuations in temperature and overall transcriptions.[14]

Photic Entrainment of the Master Clock

The daily synchronization of a self-sustained oscillator by an external signal (zeitgeber) is called *entrainment.* The most potent synchronizer of the central clock in mammals is light perceived by the retina. In the absence of environmental inputs, the clock free-runs with a period close to, but not exactly, 24 hours. The synchronizing effects of light depend on when the clock receives light cues, defining a so-called phase-response curve. In a nocturnal rodent, light exposure in

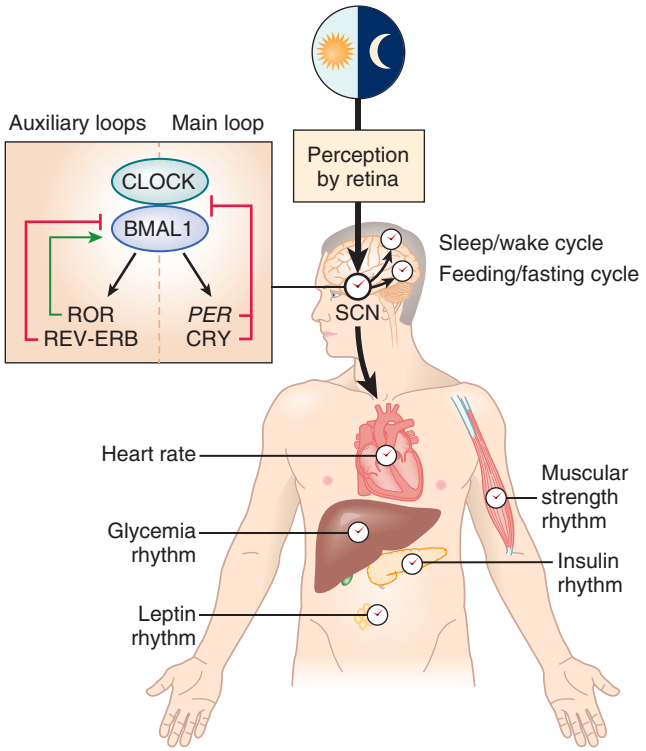


Figure 38-1 Hierarchic organization of the circadian system. The master clock located in the suprachiasmatic nucleus (SCN) synchronizes a network of brain and peripheral clocks, leading to circadian rhythms of physiologic, metabolic, and hormonal parameters. The molecular clockwork relies on transcriptional-translational feedback loops. The main loop involves CLOCK-BMAL1 stimulating the transcription of *Per* and *Cry* genes, which in turn inhibit the transcriptional activity of CLOCK-BMAL1. In the auxiliary loops, after stimulation of their transcription by CLOCK-BMAL1, ROR and REV-ERB stimulate and inhibit the transcription of *Bmal1*, respectively. The auxiliary loops help stabilize the 24-hour oscillations of clock proteins. BMAL1, Brain-muscle-arnt-like protein; CLOCK, circadian locomotor output cycles kaput; Cry, Cryptochrome; Per, Period; REV-ERB, reverse viral erythroblastic oncogene product; ROR, retinoic acid receptor–related orphan nuclear receptor.

early and late subjective night produces phase delays and advances, respectively, whereas light during the middle of the subjective day has no shifting effect on the central clock.[15] Interestingly, the same phase-response curve is obtained in diurnal rodents, whose clock gene expression and temporal patterns of SCN activity are in phase (in astronomical or circadian times) with nocturnal rodents, despite opposite sleep-wake cycles.[16,17] This strongly suggests that the distinction between nocturnal and diurnal animals relies on mechanisms that operate downstream of the SCN clock.

Light intensity is detected in the retina by classic photoreceptors, namely rods and cones, and by intrinsically photosensitive ganglion cells containing the photopigment melanopsin, highly responsive to blue light.[18] The axons of these ganglion cells constitute the retinohypothalamic tract and project monosynaptically to the SCN core, where they release mainly glutamate and pituitary adenylate cyclase–activating protein (PACAP).[19] Importantly, the retinohypothalamic tract projects not only to the SCN but also to the intergeniculate leaflet (IGL) of the thalamus. From the IGL, the geniculohypothalamic tract projects to the SCN and can thus indirectly convey light information by releasing neuropeptide Y, GABA, and enkephalin.[20] A few other structures could also convey indirect light information to the SCN, such as the basal forebrain nuclei that send cholinergic projections to the SCN.[21]

In response to light, glutamate and PACAP are released from retinohypothalamic terminals and bind to their receptors expressed in ventral SCN neurons.[15] This downstream signaling induces acute expression of clock genes *Per1* and *Per2*, in addition to several immediate early response genes such as *c-fos*,[22-24] which are induced by light only at night (i.e., during the photosensitive phase of the SCN). Antisense oligonucleotides against *Per1* or *Per2* inhibit light-induced phase shifts, highlighting the importance of these clock genes in the photic resetting processes.[25-27] Stimulation of glutamate and PACAP receptors induce Ca^{2+} influx, activating several kinase pathways.[28] The most extensively studied pathway involves ERK (extracellular signal–regulated kinase), which phosphorylates cycle adenosine monophosphate (cAMP) response element–binding (CREB) protein. CREB stimulates *Per1* and *Per2* transcription by binding to a CRE element in their promoter regions. Because only the SCN core receives photic inputs, synchronization of the shell to light is mediated through core-to-shell projections involving GABA, nitric oxide, and VIP signaling.[15,29,30]

In addition to its phase-shifting effects, light also modulates daily rhythmicity by direct, clock-independent responses to light. For example, bright light at night has an immediate inhibitory effect on physical activity and promotes sleep in nocturnal rodents, whereas it enhances alertness and sustained attention in (diurnal) humans.[31,32] Furthermore, light at night inhibits melatonin secretion.[33] Light also influences other peripheral functions, such as heart rate, blood glucose, and glucocorticoids.[34-37] These direct effects of light could be conveyed successively by the SCN clock, the subparaventricular hypothalamic region, and the sympathetic nervous system.[34,36] Because melanopsin-containing ganglion cells project to several brain targets beyond the SCN, such as the previously mentioned subparaventricular hypothalamic zone,[19] light cues can also bypass the SCN and reach directly this hypothalamic region that could, in turn, relay photic signals to peripheral organs through the sympathetic pathways. This alternative mechanism is supported by the properties of light-induced release of glucocorticoids.[37]

Nonphotic Phase Shifting of the Master Clock

Even if the light is the most important zeitgeber, the environment provides numerous other temporal cycling cues (e.g., temperature, food availability, social interactions) called *nonphotic* (i.e., different from light) *synchronizers*. One of the best studied nonphotic factors is exercise, voluntary or forced. Transient hyperactivity in nocturnal rodents typically causes phase advances of their locomotor activity rhythm if it occurs during the subjective day, corresponding to their normal resting period.[38-40] In humans, exercise in evening and late night leads to phase advances and delays, respectively.[41] Metabolic cues serve as other nonphotic signals that could affect or entrain the SCN clock. They are discussed at the end of the chapter. At least two major input pathways are considered to transmit nonphotic messages to the SCN: the geniculohypothalamic fibers from the IGL and the serotonergic input from the midbrain raphe nuclei, also projecting to the IGL.[20,42] Stimulation of neuropeptide Y or serotonin receptors in the SCN activates kinase-mediated phosphorylation events,[43,44]

leading to a reduction in *Per1* and *Per2* mRNA levels.[45,46] Injections of antisense oligonucleotides against *Per1* in the SCN region produce nonphotic-like phase advances of the rest-activity rhythm.[47] Most nonphotic and photic stimuli interact with one another, usually in opposite directions: one stimulus inhibiting the phase-shifting effect of the other one.[48] The IGL, which receives both photic and nonphotic cues, may be involved in integrating these conflicting signals before they reach the SCN.[20]

Outputs from the Master Clock

Because the SCN clock does not generate peripheral rhythms, but controls their timing, it is considered as a conductor within the multioscillatory circadian network. Rhythmic signals from the SCN are distributed to the brain and the entire body by two main pathways: (1) release of neurotransmitters and neuropeptides from terminals of SCN efferents, a pathway critical for controlling hormonal rhythms[49]; and (2) a neurohumoral pathway involving secretion of diffusible output signals regulating preferentially the rest-activity rhythm.[50] The SCNs project most densely to the medial hypothalamus, in particular to the subparaventricular zone. These SCN efferents are mainly GABAergic, although also glutamatergic and neuropeptidergic.[51] For instance, preautonomic neurons in the paraventricular hypothalamic nucleus (PVN) and arousal-promoting orexin (hypocretin) neurons in the lateral hypothalamus are controlled by glutamatergic and GABAergic inputs from the SCN. Neurosecretory corticotropin-releasing factor neurons in PVN, involved in the hypothalamic-pituitary-adrenal axis, are inhibited by vasopressinergic inputs from SCN.[51] Furthermore, GABAergic and vasopressinergic SCN neurons project to dorsomedial hypothalamic nucleus, a key site for the integration of circadian timing into numerous physiologic processes.[51] Moreover, dopaminergic neurons of the oscillator in the tuberoinfundibular nucleus that express rhythmic mRNA of *Per1* and *Per2* are directly regulated by VIPergic projections from the SCN.[52,53] The second possible mode of transmission for circadian cues from the SCN is the secretion of molecules into the extracellular space and cerebrospinal fluid. The existence of such a neurohumoral pathway was first supported by the fact that SCN grafts, encapsulated to prevent axonal growth and sprouting, are still capable of restoring behavioral rhythmicity in otherwise arrhythmic SCN-lesioned rodents.[50] Thus the rest-activity rhythm is thought to be regulated preferentially by diffusible output signals. Three candidate diffusible factors synthesized in the SCN have been proposed to modulate the timing of locomotor behavior: transforming growth factor-α, prokineticin 2, and cardiotropin-like cytokine.[54-56] These three molecules contribute to the suppression of locomotor activity, although possible stimulatory factors have not yet been identified.

BRAIN AND PERIPHERAL CIRCADIAN CLOCKS

Extra–Suprachiasmatic Nucleus Brain Clocks

The development of transgenic animals containing luciferase reporter driven by promoters of clock genes was particularly instrumental to reveal that the core clock mechanism described in the SCN is present in almost all brain regions and peripheral (i.e., outside the brain) tissues studied so far.[57,58] Many brain areas exhibit daily oscillations of clock genes.[59,60] Retina and olfactory bulbs are the only extra-SCN brain clocks with very strong oscillatory capacities that have been identified so far.[61,62] Other brain areas, such as arcuate nucleus and dorsomedial hypothalamic nucleus, two structures of the mediobasal hypothalamus involved in feeding and energy metabolism, are capable of self-sustained oscillations for several cycles when isolated in vitro. Cells of these oscillators exhibit independent circadian rhythms, but are weakly coupled, and their synchronization requires daily inputs.[63,64] Strikingly, the timing of clock gene oscillations and the rhythm of electrical activity in the secondary brain clocks of nocturnal rodents differ from the SCN in most cases. In the SCN, electrical activity and *Per1* expression peak during (subjective) daytime, while the corresponding peaks occur at night (i.e., during the active period) in extra-SCN oscillators.[63,65] In sharp contrast, the bed nucleus of stria terminalis, a basal forebrain structure modulating a wide range of physiologic and motivational processes, displays electrical activity in phase with the SCN, suggesting a strong coupling between the two structures.[65] Furthermore, several brain structures, such as the ventromedial hypothalamic nucleus, display a daily rhythmicity dependent on timed inputs because these structures become arrhythmic as soon as they are isolated in vitro.[64] Of note, the core clock mechanisms in brain oscillators seem very close to those identified in the SCN. In the forebrain, however, NPAS2 replaces CLOCK as a partner of BMAL1 in the positive loop.[66]

Clocks in Peripheral Tissues

Circadian transcriptome profiling studies in the master and peripheral clocks reveal that about 10% of a tissue's transcriptome has a circadian pattern of expression.[67,68] Most peripheral cells contain the molecular clock machinery.[69,70] Bioluminescent constructs (e.g., luciferase under the control of *Per* promoter) have allowed for real-time visualization of oscillations in clock genes, both in vitro and in vivo.[57,58,71] Peripheral clocks, such as liver explants, can generate a number of circadian cycles of *Per2*-luciferase expression.[58] With in vivo conditions, the SCN participate in the phase coherence between hepatocytes.[72] How peripheral oscillators are entrained by neuronal, endocrine, and behavioral signals coming from the SCN is described in a later section.

Cultured fibroblasts have been an in vitro model of choice to study the molecular regulation of the circadian clocks.[69,70] Similar to the master clock, cultured fibroblasts are resilient to large changes in temperature and overall transcription rates.[14] Activation of a multitude of intracellular pathways affects their clockwork.[73] Cultured fibroblasts do not generate metabolic rhythmicity as assessed by 2-deoxyglucose uptake, despite synchronized oscillations of clock genes. Sustained oscillations of 2-deoxyglucose uptake, however, can be produced in fibroblasts if they are cocultured (without physical contact) with immortalized SCN cells.[74]

In the liver, clock-controlled genes encode key enzymes involved in hepatic metabolism of fatty acids, cholesterol, bile acids, amino acids, and xenobiotics.[75-77] Specific inactivation of *Bmal1* in the liver of mice (L-*Bmal1*$^{-/-}$) disrupts rhythmic expression of glucose regulatory genes and glucose metabolism, including circulating glucose levels. L-*Bmal1*$^{-/-}$ mice are mildly hypoglycemic during the resting phase, suggesting that the hepatic clock drives a daily rhythm of hepatic glucose export counterbalancing the brain-driven fasting-feeding cycle.[78]

The adipose tissue also exhibits robust oscillations of core clock components, controlling the circadian expression of many transcription factors.[79,80] The lipoprotein lipase (i.e., an extracellular lipase synthesized in adipocytes and hydrolyzing triacylglycerols from circulating lipoproteins) displays a rhythmic activity in adipose tissue, suggesting that the adipose clock is somehow involved in lipid metabolism.[81] Moreover, the adipose tissue secretes several hormones termed *adipokines*, including leptin and adiponectin, involved in the regulation of energy balance. Several adipokine genes show a rhythmic expression in mouse adipose tissue.[79] Circulating levels of leptin display clear diurnal variations in both rodents and humans.[82-84] Moreover, leptin secretion was shown to be rhythmic in cultured adipocytes, suggesting that rhythmic synthesis or secretion of this adipokine may be under the control of the adipose clock.[85]

All together, the results show that peripheral cells, such as fibroblasts, hepatocytes, or adipocytes, fulfill the usual criteria to consider them as peripheral cellular clocks. In most peripheral tissues, however, neighboring cellular clocks fail to maintain phase coherence, in contrast to the strong intercellular coupling in the SCN. Another functional difference with the SCN is the fact that CLOCK is indispensable for circadian gene expression in peripheral tissues, at least the liver and lung.[86] Moreover, there can be specific differences in the characteristics of the clockwork according to the cell type in a given tissue, as exemplified in the skin.[87]

Molecular Links between Core Clock Components and Metabolism

Several transcriptional networks connect the core clock mechanisms with intracellular metabolic pathways. These interactions involve, among others, a number of nuclear receptors, including REV-ERBs and RORs (i.e., circadian components defining auxiliary loops in the clock mechanism) and peroxisome proliferator-activated receptors (PPARs), which are transcription factors activated by fatty acids. In the skeletal muscle, RORα directly regulates genes involved in fatty acid metabolism.[88] Moreover, REV-ERBα plays a pivotal role at the interface between the liver clock and lipid metabolism. Genetic loss and gain of function experiments showed that REV-ERBα participates in the circadian modulation of sterol regulatory element-binding protein activity and its target genes, which are involved in cholesterol and lipid metabolism.[89] REV-ERBα also plays a pivotal role in the daily variations of fuel utilization.[90] *Pparα* is rhythmically expressed in tissues with high rates of fatty acid oxidation, such as the muscles, heart, or liver, and is strongly involved in lipoprotein and lipid metabolism.[91] *Pparα* is a clock-controlled gene whose activation involves CLOCK and BMAL1, which in turn can activate *Bmal1* transcription.[92,93] Thus PPARα provides a close link between circadian clocks and lipid metabolism in peripheral tissues, in particular in the liver. PPARα can also directly activate *Rev-erbα* expression, and PER2 is able to recruit PPARα and REV-ERBα to modulate *Bmal1* expression, highlighting further reciprocal interactions between clock components and metabolism.[94,95] Additionally, a critical role has been demonstrated for PGC-1α, a coactivator of PPAR (PPAR coactivator-1α), which stimulates *Bmal1* expression through coactivation of RORs. Because PGC-1α is a metabolic regulator sensitive to various signals, including nutritional status and temperature, it could be a key component in the coupling of metabolism to clocks.[96] Furthermore, lipidomic profiling in adipose tissue indicates that PER2 is implicated in normal lipid metabolism. This effect is mediated by PPARγ, a major regulator of adipogenesis and lipid metabolism in adipose tissue, whose transcriptional activity is directly repressed by PER2.[97]

Interactions between the circadian clocks and cellular metabolism also involve cellular energy sensors such as sirtuin1 (SIRT1) and AMP-activated protein kinase (AMPK). SIRT1 catalyzes NAD^+-dependent deacetylation of various substrates. By deacetylating histones, SIRT1 participates in chromatin condensation and thus epigenetic silencing. SIRT1 also contributes to multiple metabolic pathways, such as gluconeogenesis, lipid metabolism, insulin secretion, and mitochondrial activity. SIRT1 plays a crucial role in the life-span extension associated with calorie restriction, in part through the SCN clock.[98] SIRT1 influences the transcription of several clock genes and promotes deacetylation and degradation of PER2, thus modulating the timing of the core clock loops.[99,100] Inhibition of SIRT1 leads to circadian disturbances and to the acetylation of BMAL1 and histone H3, both substrates of the acetylase function of CLOCK.[100]

Besides SIRT1, AMPK is another important metabolic fuel gauge, sensing changes in the intracellular AMP/ATP ratio. AMPK integrates nutritional and hormonal signals in peripheral tissues and the hypothalamus, where it mediates cellular effects of adipokines (e.g., leptin) in regulating glucose and lipid homeostasis. Like SIRT1, AMPK responds to low energy levels.[101] In mouse skeletal muscle, AMPK enhances SIRT1 activity by increasing cellular NAD^+ levels, resulting in the deacetylation and modulation of the activity of SIRT1 targets, such as PGC-1α,[102] which in turn affects the circadian clock. AMPK also has direct actions on the clock machinery. AMPK phosphorylates not only the clock protein CRY, but also casein kinase 1ε, leading to subsequent degradation of PER2 and phase shifts of peripheral oscillations.[103,104]

The connections between the cellular metabolism and the clockwork have been intensively investigated in peripheral tissues. Although much less is known, close mechanisms may be found in central extra-SCN oscillators. In particular, AMPK is a potent regulator of energy balance within the hypothalamus. For example, leptin inhibits specifically AMPK in arcuate nucleus and PVN, and this inhibition is required to mediate anorexigenic and weight loss effects of leptin.[101] AMPK signaling is thus a likely route through which circadian and feeding signals are integrated in the hypothalamus.[105]

COUPLING BETWEEN CENTRAL AND PERIPHERAL CLOCKS

Entrainment of Peripheral Clocks by Nervous Outputs of the Suprachiasmatic Nucleus

It is now well established that the SCN clock controls timing in peripheral organs through the sympathetic pathways.[51,106] To illustrate this, the liver and the white adipose tissue have been chosen as two representative examples in this section. The next section will consider two other peripheral organs, namely the pineal and adrenal glands, whose timing is also controlled by the SCN and that share the feature of delivering hormonal timing signals.

The liver plays a pivotal role in glycemic regulation as a site of glucose uptake and a major source of glucose production.[107] The daily rhythmicity of plasma glucose, peaking before activity onset in rats, is not a passive response to food intake.[108] A functional liver clock is important for glucose metabolism, as was previously mentioned,[78] but is not sufficient because SCN-lesioned rats lose their daily rhythmicity of blood glucose.[108] Retrograde tracing studies from the liver revealed projections through both sympathetic and parasympathetic components of the autonomous nervous system to third-order neurons in the SCN. Moreover, the glucose rhythm can be abolished by inactivation of either the sympathetic or parasympathetic inputs, underlying the importance of balanced inputs from the autonomous nervous system. Rhythmic GABAergic input from the SCN is considered to inhibit the sympathetic and parasympathetic preautonomic neurons of the PVN, predominantly during the day. By contrast, glutamatergic projections from the SCN stimulate sympathetic preautonomic neurons of the PVN. Thus the entrainment of circadian glucose rhythm is performed by the SCN, fine-tuning the balance between both branches of the autonomous nervous system that innervate the liver clock.[107] Moreover, hypothalamic expression of orexin exhibits a diurnal rhythm entrained by GABAergic inputs from the SCN. Besides its role in behavioral activation, orexin is also a key regulator of plasma glucose in rats, modulating in particular the daily peak at dusk through the sympathetic nervous system.[109]

The rich innervation of adipose tissue by sympathetic fibers is well known, and their activation enhances lipolysis. Parasympathetic innervation of white adipose tissue has been shown more recently.[110] As for the liver, the SCN controls both branches of the autonomous nervous system that innervate adipose tissues, thus modulating circadian rhythmicity of metabolic and endocrine outputs of the adipose clock. For instance, activity of the hormone-sensitive lipase exhibits a daily rhythmicity that is modified by adipose denervation.[110] Moreover, the leptin rhythm is under the control of both the local adipose and SCN clocks because lesions of the SCN abolish the daily rhythm of plasma leptin in rats.[82,85] Through modulation of the autonomic innervations of liver and adipose clocks, the SCN therefore controls circadian rhythmicity of metabolites (carbohydrates and lipids) and metabolic hormones (e.g., leptin). Some peripheral clocks, however, do not respond directly to neural cues, and rhythmic hormones (glucocorticoids and melatonin) may additionally transmit timing signals from the SCN to a variety of peripheral organs that express glucocorticoid or melatonin receptors.

Entrainment of Peripheral Clocks by Suprachiasmatic Nucleus–Controlled Hormonal Outputs

Melatonin and glucocorticoids are two hormones with time-giving properties because their rhythmic release is tightly controlled by the SCN through nervous pathways, and when released, these endocrine messages can in turn affect or even entrain peripheral clocks (Figure 38-2).

Melatonin synthesized in the pineal gland is best known as a transducer of the photoperiodic information into neuroendocrine changes, through the duration of its nocturnal peak.[111,112] The daily high-amplitude rhythm of melatonin also has a circadian role. Melatonin synthesized from tryptophan is always secreted during the dark phase in both nocturnal and diurnal mammals.[113] The release of melatonin is driven by the SCN clock through a multisynaptic pathway, including the PVN, the intermediolateral cell column of the spinal cord, and the superior cervical ganglions that send sympathetic fibers which release noradrenalin in the vicinity of the pinealocytes. This noradrenergic release triggers the nocturnal synthesis of melatonin.[114] The daily rhythm of nocturnal melatonin provides temporal signals to a multitude of tissues expressing melatonin receptors, including peripheral organs and the SCN (see later). In isolated adipocytes, rhythmic melatonin mimicking a biologic night triggers expression of clock genes, such as *Per1* and *Clock*, and stimulates a lipogenic response.[115] Melatonin modulates clock gene expression in adrenal explants.[116] Furthermore, in the pars tuberalis of the adenohypophysis, rhythmic oscillations of clock genes (i.e., *Cry1* and *Per1*) are driven by the daily rhythm of melatonin.[117] The nocturnal rhythm of endogenous melatonin is a reliable phase marker of the SCN clock because it is relatively impervious to most internal and external disturbances, except bright light exposure at night, which immediately inhibits its synthesis, therefore blunting its temporal message.[33]

Glucocorticoids (corticosterone in rats and mice; cortisol in humans) show a strong daily rhythm, peaking systematically at about wake-up time (dawn and dusk in nocturnal rodents and humans, respectively). This daily peak results from the adrenal clock, which is controlled by SCN cues through the hypothalamic-pituitary-adrenal axis and sympathetic fibers, the latter modulating the sensitivity of the adrenal

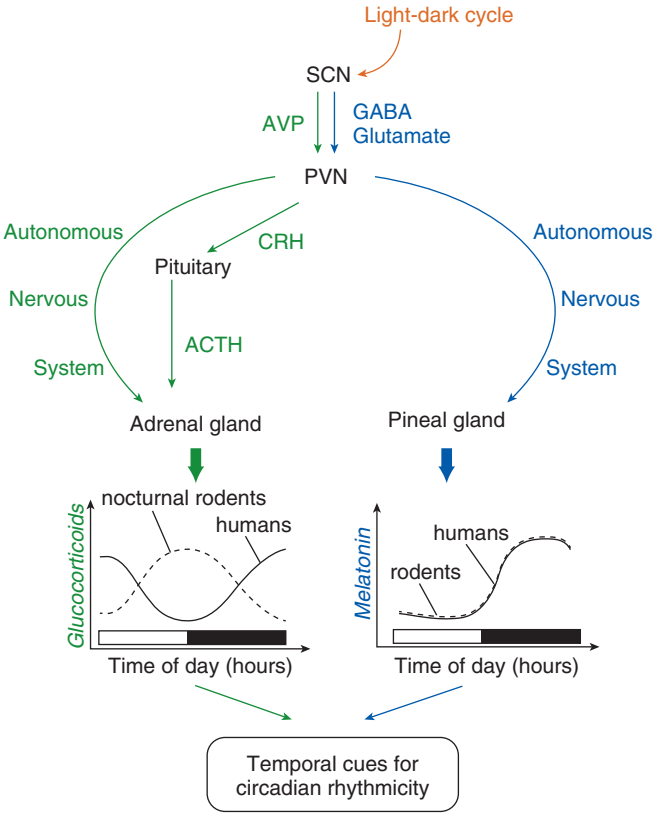


Figure 38-2 Rhythmic secretion of glucocorticoids and melatonin is driven by the suprachiasmatic nuclei (SCN) and acts in turn as temporal cues for circadian rhythmicity. AVP, Arginine-vasopressin; ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; GABA, gamma-aminobutyric acid; PVN, paraventricular nucleus.

glands to ACTH.[118-120] The glucocorticoid nuclear receptors are expressed in most cell types in periphery and the brain, with the notable exception of adult SCN cells.[121] Dexamethasone, a glucocorticoid receptor agonist, activates *Per1* expression and synchronizes rat fibroblasts in vitro. Moreover, dexamethasone produces in vivo phase shifts of peripheral clocks (liver, kidney, and heart), but not of the SCN clock.[122] Activity of glucocorticoid receptors is directly modulated by the clockwork because their transcription can be repressed by CRYs and they can be acetylated by CLOCK.[123,124] In extra-SCN brain structures, such as the bed nucleus of stria terminalis and central amygdala, oscillations of the clock protein PER2 disappear after adrenalectomy and are restored by rhythmic supply of corticosterone through drinking water.[125] In midbrain raphe nuclei that do not express clock genes, rhythmic corticosterone also drives the daily rhythm of tryptophan hydroxylase mRNA, a limiting enzyme for synthesis of serotonin.[126] Thus daily variations of circadian glucocorticoids possess resetting and time-giving properties for central and peripheral structures. It should be emphasized, however, that the circadian rhythm of glucocorticoids can be blunted or markedly modified by environmental conditions, including stressful events, light, and feeding. Acute stress leads to ACTH-induced release of glucocorticoids that is not necessarily in phase with the circadian pattern.[119] Also, light exposure at night induces *Per1* gene expression in the adrenal gland and corticosterone release by activation of sympathetic fibers independently of the hypothalamic-pituitary-adrenal axis.[36] In addition, restricted feeding triggers an anticipatory rise in circulating glucocorticoids before food access. This anticipatory peak is ACTH independent and distinct from the circadian rhythm of glucocorticoids controlled by the SCN clock.[127] It should be noted therefore that circulating glucocorticoids can influence circadian clocks by conveying various temporal signals, only some of them being strictly dependent on the SCN.

Feedback of Peripheral Hormonal Signals to the Suprachiasmatic Nucleus

As detailed previously, in undisturbed conditions, the circadian rhythms of plasma melatonin and glucocorticoids are tightly controlled by the SCN clock. The presence of both MT1 and MT2 receptors in the SCN suggests that melatonin may have feedback effects on the master clock.[128] Daily injections or perfusions of supraphysiologic doses of melatonin entrain the free-running activity of rats in constant darkness, when injections occur at the subjective dusk.[129,130] In vitro application of melatonin on cultured SCN explants induces two distinct effects. First, melatonin acutely inhibits neuronal firing.[131] Second, melatonin shifts the circadian rhythm of electrical activity of SCN neurons in a time-dependent manner.[132] The acute inhibitory effect seems to be mediated by MT1 receptors, whereas the phase-resetting effect may rely on MT2 receptor signaling.[133-135]

Glucocorticoids are not expected to feed back directly to the SCN because their receptors are not expressed in sizeable amount within adult SCN cells.[121] Furthermore, as previously mentioned, the glucocorticoid agonist dexamethasone can induce phase shifts of clock gene expression in peripheral clocks, but not in SCN neurons.[122] However, glucocorticoids modulate the daily synchronization of the SCN to light, as evidenced by faster reentrainment to a new light-dark cycle in adrenalectomized rodents.[136,137] The indirect feedback of glucocorticoids to the SCN is thought to be mediated through serotonergic projections from midbrain raphe.[136] Such a feedback would prevent uncoordinated resetting of the circadian system (e.g., in response to sporadic light exposure), and thus serves as a protection from zeitgeber noise.[137] Together, melatonin and glucocorticoid rhythms appear to stabilize the functioning of the circadian system.

ADJUSTING CLOCKS WITH FEEDING

Extra–Suprachiasmatic Nucleus Clocks Are Entrained by Feeding Time

Among the different ways used by the SCN to synchronize peripheral clocks, the feeding rhythm is a strong zeitgeber for many tissues. In normal conditions, food intake takes place during the active period. Restricted feeding in nocturnal rodents (i.e., when food access is limited to few hours during daytime, a time when nocturnal rodents usually rest) inverts the phase of gene expression in peripheral organs within about a week, thereby uncoupling peripheral clocks from the SCN that remain phase-locked to the light-dark cycle.[138,139] The synchronization velocity is tissue specific. Indeed, food-induced phase resetting proceeds faster in liver than in kidney, heart, or pancreas, with large phase shifts within 2 days of altered feeding schedule. Furthermore, there are peripheral clocks, such as the submaxillary salivary glands, that fail to entrain to restricted feeding (Figure 38-3).[106]

In the brain, food restriction entrains the activity of a number of, but not all, oscillating structures outside the SCN. For example, the multineuronal activity in the lateral hypothalamus of rats under restricted feeding shows a peak entrained to the time of feeding.[140] Moreover, daily patterns of *Per1* and *Per2* mRNA in the cerebral cortex, striatum, and PVN from mice entrained to restricted feeding also show a phase shift with peaks around mealtime, as opposed to the nocturnal peaks of expression in animals fed ad libitum.[59,60] Other structures, such as the hippocampus, display small or no phase change in patterns of clock gene expression in response to feeding time.[59] Nevertheless, in total the data suggest that most secondary clocks within and outside the brain are affected by restricted feeding schedules.

Under restriction feeding, several behavioral and physiologic functions become entrained to the availability of food. More specifically, body temperature and plasma corticosterone rise before food access in phase with behavioral activation, called *food-anticipatory activity,* thought to be the laboratory equivalent of food-seeking behavior in the wild. This rhythmic bout of activity is still expressed in SCN-lesioned animals and is considered a behavioral output of a food-entrainable clock.[141] The precise location of the food-entrainable clock (outside the SCN) and its mechanisms have been the subject of much debate and controversy. Most experimental arguments support the current view that the food-entrainable clock is a network of coupled neural structures, likely involving mediobasal hypothalamic nuclei, which interact together to provide timing and behavioral entrainment of feeding.[142]

Possible Mechanisms of Entrainment of Extra–Suprachiasmatic Nucleus Clocks by Food

The nature of signals that arise from feeding and that entrain peripheral clocks has been an area of intense investigation.

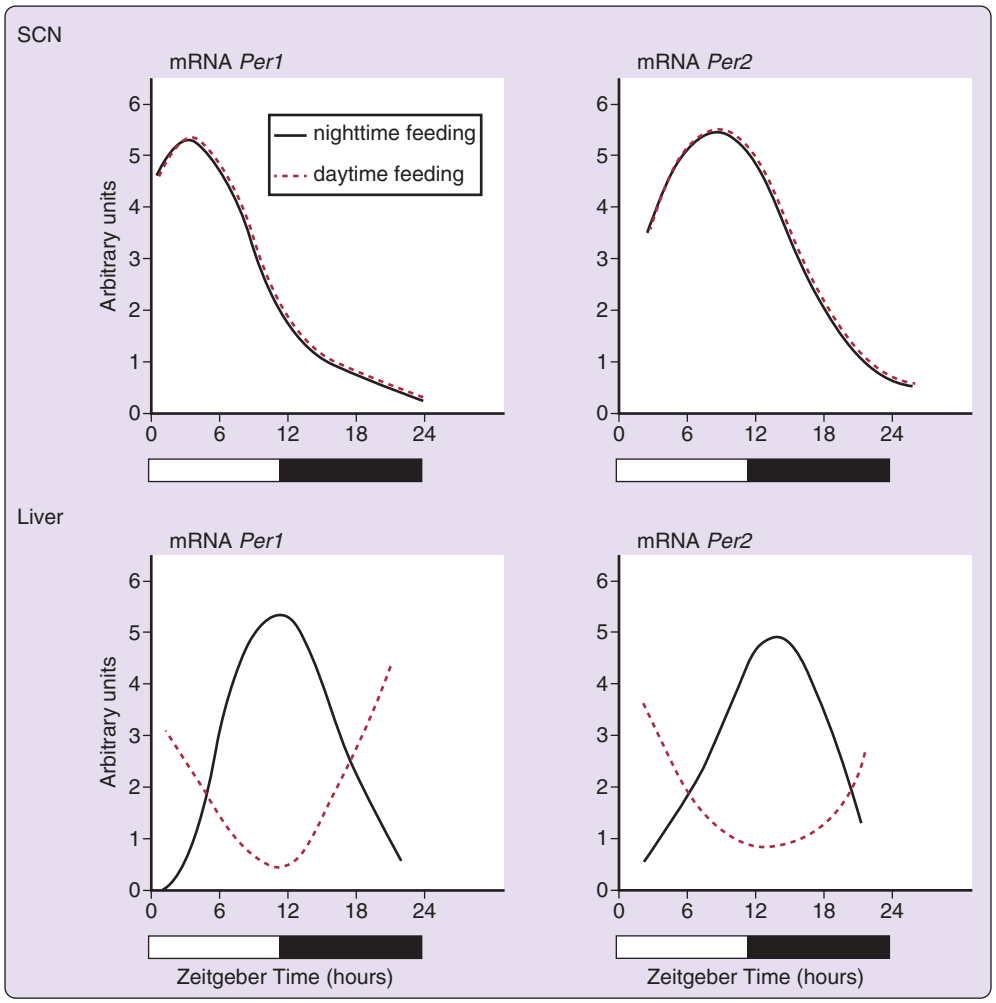


Figure 38-3 Daily expression of *Per1* and *Per2* in the liver and in the suprachiasmatic nucleus (SCN) of mice fed at nighttime (*solid lines*) or at daytime (*dotted lines*). Zeitgeber time 0 is defined as time of lights on. Feeding time affects daily rhythmicity of clock gene expression in peripheral clocks, but not in the SCN. Timed changes in metabolism can thus lead to an uncoupling of peripheral oscillators from the central pacemaker. (Data from Damiola F, Le Minh N, Preitner N, et al. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev* 2000;14:2950–61.)

Feeding cues include a number of parameters, including food absorption, postprandial increase in temperature, secretion of metabolic hormones, food-derived metabolites, and changes in the energetic status of cells.

Variations in temperature are known to entrain behavioral rhythms of heterotherms, such as *Drosophila*.[143] Moreover, temperature fluctuations mimicking body temperature rhythms sustain previously induced oscillations in cultured rat fibroblasts. Inverted environmental temperature cycles in vivo reverse circadian rhythms of clock genes (*Per2* and *Cry1*) in the mouse liver without affecting the SCN.[144] Thus diet-induced thermogenesis during the postprandial period could be an entraining pathway from the feeding-fasting cycle in homeothermic organisms. The mechanisms of entrainment by temperature involve heat-shock factor 1 (HSF1), as revealed in vitro by genetic loss of function experiments and pharmacologic antagonists of heat-shock signaling pathway.[145,146] Hepatic HSF1 exhibits a highly rhythmic activity, which drives the expression of heat-shock proteins in liver.[147] Therefore HSF1 could be a key component linking temperature fluctuations and the phase of molecular clocks.

Anorexigenic (insulin) and orexigenic hormones (ghrelin) may participate in the entrainment of peripheral clocks by feeding. On one hand, insulin that rises in the plasma during the postprandial period causes an acute induction of *Per1* mRNA levels in cultured rat fibroblasts.[73] Moreover, insulin triggers upregulation of *Per2* mRNA and downregulation of *Rev-erbα* mRNA in the liver, thus mimicking the effects of refeeding after fasting.[148] The insulin-dependent phase shifts of peripheral clocks involve PI3K- and MAPK-mediated signaling pathways.[149] Feeding-induced insulin secretion may thus be a critical step in feeding-induced entrainment of the liver clock. On the other hand, released ghrelin from the stomach during fasting is thought to signal hunger state to the brain. Cerebral activation of ghrelin signaling has been implicated in the bout of activity, so-called food-anticipatory activity, that the animals express before timed food access.[150,151]

Another hormone linked to meal entrainment is corticosterone. An anticipatory rise of plasma corticosterone is induced by restricted feeding schedules,[152] and corticosterone is known to entrain peripheral clocks (see earlier).[122] However, corticosterone injections fail to mimic the phase-shifting effects of

feeding in rats.[139] Gene expression rhythms in the liver of adrenalectomized or glucocorticoid receptor–deficient mice are still entrained by restricted feeding. Glucocorticoid signaling may actually provide resetting cues conflicting with feeding synchronizers because food-induced phase shifts of the liver are actually faster in the absence of glucocorticoids.[153] The net effect of these synchronizing interactions depends on the tissue because the liver clock appears to be more sensitive to feeding cues, whereas the lung and kidney clocks are more easily reset by glucocorticoids.[154]

Fascinatingly, the application of glucose in the culture medium of rat fibroblasts causes a downregulation of *Per1* and *Per2* mRNA levels and induces rhythmic expression patterns of numerous genes, including transcription factors. The decrease of *Per1* and *Per2* mRNA levels by glucose seems indirect and mediated by glucose metabolism (i.e., involving transcriptional regulators) rather than by glucose.[155] Many nuclear receptors, such as PPARs, contribute to the daily variations of lipid and glucose metabolism. As a reminder, PPARs are ligand-dependent transcription factors activated by fatty acids (released, for instance, during the fasting state). In addition to their role as mediators of metabolism, PPARs interact with clock components (see earlier).[93] Thus both plasma fatty acids and glucose, two major circulating metabolites, are potential mediators by which food-related cues entrain peripheral clocks.

Another possibility for feeding entrainment is that clock proteins, such as CLOCK (or its paralogue NPAS2), BMAL1, or PERs, sense directly food-related signals. These proteins all contain a PAS domain, which detects redox state (i.e., the reduced or oxidized environment within the cell), reflecting the energy status. Redox signals are transduced by PAS domains, which modulate the functional state of the protein.[156] For example, DNA-binding activity of the dimers CLOCK/BMAL1 and NPAS2/BMAL1 is altered by cellular redox status. The reduced forms of the nicotinamide adenine dinucleotide, NADH and NADPH, activate DNA binding of CLOCK (or NPAS2)/BMAL1, whereas the oxidized forms, NAD^+ and $NADP^+$, inhibit DNA binding. The NAD(P)H/$NAD(P)^+$ ratio is closely tied to mitochondrial activity, and the switch between activation and inhibition of DNA binding is very sensitive, providing a rapid mechanism that could convey changes in fuel availability to the cellular clocks.[66] Therefore, even if feeding can be viewed as a dominant entraining factor for peripheral clocks, the underlying mechanisms are complex because multiple signals seem to be involved at different levels of the circadian system.

Effect of Nutritional Cues on the Central Clock

Although most peripheral clocks are highly sensitive to the synchronizing effects of feeding time and food-related cues, the SCN clock seems impervious to them, provided that animals are exposed to a light-dark cycle and ingest enough daily energy.[138,139] However, the SCN can respond to nutritional cues under specific caloric conditions. Indeed, rats under a light-dark cycle and entrained to timed hypocaloric feeding display phase advances of daily rhythms of locomotor activity, body temperature, and pineal melatonin.[157] Entrainment to a light-dark cycle is also altered in mice submitted to a timed calorie restriction. In addition to the phase-advanced rest-activity rhythm (i.e., nocturnal mice becoming partially diurnal), expression of clock proteins and the clock-controlled factor *Vasopressin* is phase-advanced in the SCN, and the circadian responses of the SCN to light are altered.[158,159] In addition, circadian phase-shift responses to light are reduced in animals under low glucose availability.[160] All together, these results challenge the idea that the SCN is impervious to any nutritional cues. Notably, the reward aspect of food also seems important because in mice housed in constant dark conditions, the SCN entrains to rhythmic access to palatable food (chocolate) given in addition to regular food pellets available ad libitum.[161]

Little is known about the pathways conveying metabolic signals to the SCN. One could first imagine a direct effect of timed calorie restriction, changing the redox status of SCN cells, as in peripheral clocks (see previous paragraph), and affecting subsequently the SCN molecular clockwork. Self-sustained redox cycles have been identified in the SCN cells, where they regulate neuronal activity.[162]

In addition, receptors of feeding-related hormones such as insulin, ghrelin, and leptin are present in the SCN and metabolic hypothalamus (i.e., nuclei of hypothalamus involved in the regulation of energy balance).[163-165] When isolated in vitro, the SCN clock can be phase-advanced by leptin or ghrelin.[166,167] Moreover, leptin modulates firing rates of SCN neurons in hypothalamic slices.[168] However, insulin applied in vitro inhibits the firing rate of the SCN neurons during the subjective day.[169] These hormones are thus possible candidates for conveying metabolic information to the SCN, either directly or indirectly. Indeed, besides a direct modulation on the SCN clock, another possibility would involve relays from brain structures sensitive to nutrients. As previously discussed, nonphotic cues can be conveyed to the SCN by projections from raphe nuclei and IGL. Experiments of lesions suggest an involvement of the IGL in the transmission of metabolic information to the SCN.[170,171] Moreover, orexigenic and anorexigenic neurons in the hypothalamus that control feeding behavior respond to fluctuations in circulating nutrient (e.g., glucose, fatty acids, amino acids) levels that reflect the nutritional status.[172] Because the SCN receives numerous projections from various hypothalamic nuclei, the metabolic region of the hypothalamus could integrate and transmits information from circulating nutrients to the SCN. For example, the mediobasal hypothalamus may be involved in mediating the behavioral phase advance produced by timed calorie restriction.[173,174]

CONCLUSIONS

This overview of the central and peripheral clocks shows the hierarchic organization of the mammalian circadian system at the top of which is the SCN. Because light is the most potent synchronizer of this central clock, a proper temporal organization is normally achieved under a light-dark cycle. When synchronized to light, the SCN controls behavioral (i.e., sleep-wake and feeding-fasting cycles) and physiologic rhythms (e.g., body temperature and plasma melatonin and glucocorticoids) that will, in turn, reinforce the robustness of daily rhythmicity by sending internal timing cues. Forced or voluntary feeding limited to the usual rest period is a potent timer of peripheral oscillations that disturbs internal coupling between the various clocks, depending on their sensitivities to meal resetting. As a consequence of the close connections between cellular clocks and intracellular metabolism, genetic

clock disruptions affect metabolism in rodents. Moreover, chronic alterations in the main zeitgebers of the circadian system, such as bright light exposure during the subjective night (chronic jet lag, shift work) and meal times, lead to circadian desynchronization, with detrimental consequences on metabolic health.

CLINICAL PEARL

As a consequence of the reciprocal connections between circadian clocks and intracellular metabolism, situations of circadian misalignment such as night-eating syndrome, shift work, and chronic jet lag, disrupt sleep homeostasis and increase metabolic risk factors, including obesity, impaired glucose tolerance, and hypertension.

SUMMARY

The master clock, located in the SCN of the hypothalamus, synchronizes a multitude of brain and peripheral clocks. These clocks allow organisms, tissues, and cells to anticipate ongoing changes and optimize the efficiency of a given function at the expected time of daily occurrence. Brain and peripheral clocks also segregate incompatible behaviors (e.g., sleep and feeding) and chemically incompatible reactions (e.g., hepatic gluconeogenesis and glycolysis), respectively. Rhythmic signals from the SCN couple the master clock to secondary brain and peripheral clocks through behavioral, nervous, and neurohumoral pathways. Endocrine rhythms (i.e., pineal melatonin and adrenal glucocorticoids) distribute internal temporal messages within the body.

Light perceived by the retina is the most potent synchronizer of the central clock in the SCN. A few other stimuli, different from light, called nonphotic cues (e.g., exercise, arousal) can phase-shift the SCN, especially when the photic synchronizer is weak or absent. Clocks in peripheral tissues share many molecular properties with the SCN. Besides the stronger intercellular coupling in the SCN, another functional difference is the high sensitivity of most brain and peripheral clocks to the synchronizing effects of feeding as opposed to the relative resistance of the SCN. Cellular clocks and intracellular metabolism are tightly and reciprocally connected. Even if the SCN is not shifted by meal time, metabolic cues can affect SCN clockwork and modulate its synchronization to light.

Selected Readings

Bechtold DA, Loudon AS. Hypothalamic clocks and rhythms in feeding behaviour. *Trends Neurosci* 2013;**36**:74–82.

Bedont JL, Blackshaw S. Constructing the suprachiasmatic nucleus: a watchmaker's perspective on the central clockworks. *Frontiers in Systems Neuroscience* 2015;**9**:74.

Dibner C, Schibler U, Albrecht U. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu Rev Physiol* 2010;**72**:517–49.

Dickmeis T. Glucocorticoids and the circadian clock. *J Endocrinol* 2009;**200**:3–22.

Gerhart-Hines Z, Lazar MA. Circadian metabolism in the light of evolution. *Endocr Rev* 2015;**36**:289–304.

Golombek DA, Rosenstein RE. Physiology of circadian entrainment. *Physiol Rev* 2010;**90**:1063–102.

Guilding C, Hughes AT, Brown TM, et al. A riot of rhythms: neuronal and glial circadian oscillators in the mediobasal hypothalamus. *Mol Brain* 2009;**2**:28.

Kalsbeek A, Yi CX, La Fleur SE, Fliers E. The hypothalamic clock and its control of glucose homeostasis. *Trends Endocrinol Metab* 2010;**21**:402–10.

Mistlberger RE. Neurobiology of food anticipatory circadian rhythms. *Physiol Behav* 2011;**104**:535–45.

Panda S, Antoch MP, Miller BH, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 2002;**109**:307–20.

Pevet P, Challet E. Melatonin: both master clock output and internal time-giver in the circadian clocks network. *J Physiol (Paris)* 2011;**105**:170–82.

Tsang AH, Barclay JL, Oster H. Interactions between endocrine and circadian systems. *J Mol Endocrinol* 2014;**52**:R1–16.

A complete reference list can be found online at ExpertConsult.com.

Circadian Dysregulation in Mental and Physical Health

Chapter 39

Sabra M. Abbott; Roneil G. Malkani; Phyllis C. Zee

Chapter Highlights

- Circadian rhythms are a key component of human health, and dysregulation has been shown to contribute to the pathogenesis, expression, and severity of medical and psychiatric disorders.
- Evidence indicates that dysregulation of circadian clock function at the molecular, cellular, and systems levels in central or peripheral tissues or misalignment because of environmental and behavioral disruptions due to shift work and social jet lag increases the risk for adverse health outcomes.
- Approaches to enhance circadian function and achieve proper alignment between endogenous central and peripheral rhythms with the external environment have the potential for significant positive effects on human health.

Circadian rhythms are a fundamental property of all biologic processes and are essential for the coordination of physiologic and behavioral functions with the 24-hour environment. Therefore disruption to the circadian system can result in a broad range of adverse mental and physical health consequences. The primary mammalian circadian pacemaker is located within the suprachiasmatic nucleus (SCN) in the hypothalamus. The inherent circadian rhythmicity is genetically regulated and is maintained through a transcription-translation feedback loop of core circadian clock genes. These include positive elements: *CLOCK* and *BMAL1* (*ARNTL*); negative elements: Period *(PER)* and Cryptochrome (*CRY*); and regulatory elements including casein kinase (*CK*) Iδ and Iε, glycogen synthase kinase (*GSK*), retinoic acid receptor–related orphan receptor-α (*RORA*), and nuclear receptor subfamily 1, group D, member 1 (*NR1D1) (Rev-Erbα)* (further details available in Chapter 27). These core clock genes are also present in peripheral tissues, from liver to heart to the immune system. The SCN is thought to serve as the master pacemaker, coordinating the timing of these peripheral tissue clocks. Polymorphisms in these core clock genes have been associated with a wide variety of disorders, although data are often limited by small sample sizes or population-specific effects.

Melatonin is a key marker of circadian function. Melatonin is secreted by the pineal gland, and levels peak in the middle of the night. Light suppresses the normally observed nocturnal rise in melatonin. Melatonin can be measured in the blood, saliva, and urine, and both the timing and amplitude of secretion can be evaluated as measures of circadian alignment.[1]

Classically it was thought that circadian disturbances were simply a consequence of underlying disease, but it is increasingly appreciated that dysregulation at a number of different levels, from a mismatch between environmental and internal time, to direct dysfunction at the level of the SCN, to dysregulation in the coordination between peripheral clocks, may play a role in the expression and development of disease (Figure 39-1).

Although the full extent of the effect of circadian dysregulation on human diseases is beyond the scope of this chapter, we will focus on several examples of disorders that highlight how circadian dysregulation can occur at the environmental, physiologic, and molecular levels. One of the most common causes of circadian dysregulation comes from environmental disruption, easily recognized in shift work and jet lag. In addition, circadian rhythm changes can also serve as one of the early indicators of cardiometabolic and neurologic disorders, as is illustrated in diabetes, Alzheimer disease (AD), and Parkinson disease (PD). Finally, at the molecular level direct changes in the core clock genes have been associated with an increased risk for mental and physical disorders, with the largest number of studies to date in mood disorders, although the strength of many of these studies is limited. Overall, however, understanding the influence of circadian clock function in disease expression and treatment represents a novel approach for clinical medicine.

CIRCADIAN MISALIGNMENT

Shift Work and Social Jet Lag

Shift work is a major source of circadian dysregulation because workers are required to be awake during times when they would normally sleep. Although workers can adapt to the change in timing, many live in a chronic state of varying degrees of circadian misalignment, being awake at night on workdays and during the daytime on days off because of social and family responsibilities. In addition to typical shift work, a large number of individuals are required to wake up much earlier on workdays than they would prefer to, then sleeping in on nonworkdays. The mismatch between work and nonwork sleep times is referred to as *social jet lag*. There is growing evidence that the circadian misalignment associated with shift work or social jet lag contributes to the development of disease, with the strongest evidence for an increased risk for cancer and cardiometabolic disorders.

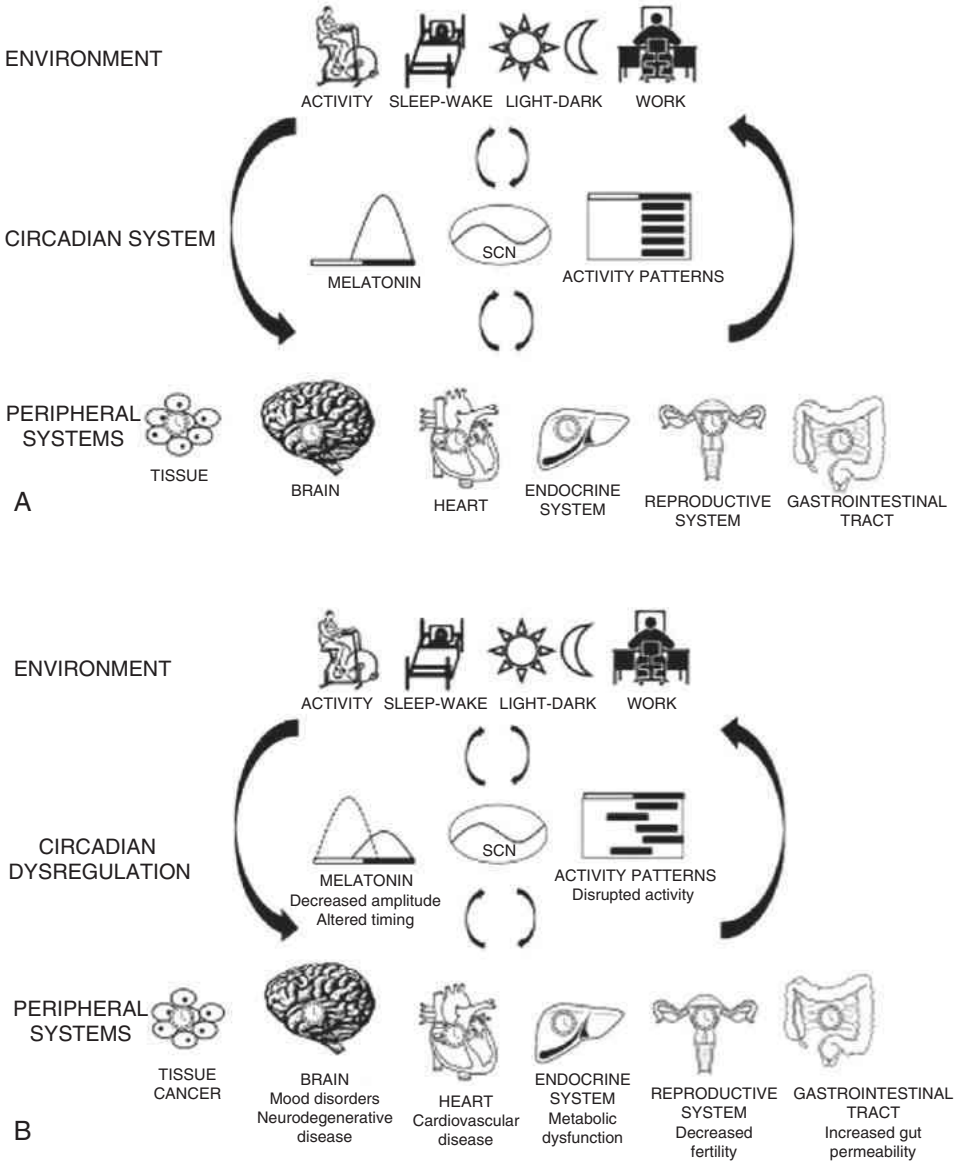


Figure 39-1 Normal circadian regulation **(A)** and dysregulation **(B)**. Circadian dysregulation can occur at multiple levels. The primary circadian pacemaker is located in the hypothalamus in the suprachiasmatic nucleus (SCN). The SCN maintains rhythmicity through a transcription-translation feedback loop of core clock genes, of which various mutations can contribute to human disease. Environmental time cues from the light-dark cycle help to regulate the circadian clock and can be disrupted in shift work. Finally, peripheral organ systems have circadian activity that must be appropriately aligned with the SCN and other peripheral tissues for proper function.

Shift Work and Cancer

The initial evidence for an increased cancer risk among shift workers came from several case–control studies of women with breast cancer. A study of Norwegian telegraph operators found an increased odds ratio for breast cancer among women older than 50 years and those who were exposed to night shift work for more than 2.7 years.[2] A case–control study in Seattle found the risk for breast cancer was significantly increased in woman who worked the "graveyard" shift,[3] whereas a case–control study in France evaluated women with newly diagnosed breast cancer for occupational history, and the odds ratio for developing breast cancer was 1.35 for women who had ever worked overnight shifts. This risk was highest among women who had worked night shifts for more than 4 years before their first full-term pregnancy.[4] A Danish study looking at female military employees showed an odds ratio for developing breast cancer of 1.4 among women who had ever worked night shifts, with the effect being most pronounced among woman with a combination of a strong morning chronotype preference and high exposure to night shift work.[5] Similar results have also been shown in a case–control study of Norwegian nurses.[6] The Nurses' Health Study found that a cohort of women who worked on the night shift for 30 or more years had a relative risk of 1.36 for developing breast cancer,[7] whereas a cohort of 4036 Swedish women found a hazard ratio of 2.02 for developing breast cancer if working shifts that included night work.[8] A recent meta-analysis of 28 studies evaluating circadian disruption and breast cancer risk

found that the relative risk for breast cancer was 1.19 for shift work.[9] Shift work has also been associated with an increased risk for prostate cancer,[10] ovarian cancer,[11,12] lung cancer,[13] and non-Hodgkin lymphoma.[14] Overall these results prompted the International Agency for the Research on Cancer in 2007 to conclude that shift work involving circadian disruption is probably carcinogenic (group 2A).[15] However, several studies have not found an association between shift work and breast cancer, including a cohort of 73,029 Chinese women surveyed regarding lifetime shift work exposure and breast cancer.[16] Interestingly the Electromagnetic Fields and Breast Cancer on Long Island Study did not show a direct association with shift work, but did find an increased incidence of breast cancer in women with greater exposure to light at night.[17]

Mechanisms by which light at night may contribute to cancer include an increased degree of circadian disruption or suppression of the normal nocturnal pineal melatonin secretion. Because melatonin can have tumor suppressant properties in animals,[18] low nocturnal melatonin levels have been proposed to increase the risk for cancer. In support of this hypothesis, studies have found significantly lower levels of 6-sulfatoxymelatonin, the urinary metabolite of melatonin, in both female[19] and male[20] night shift workers and lower amplitude of melatonin secretion over a 24-hour period in shift workers compared with nonshift workers.[21] However, a direct causal relationship between melatonin levels and cancer has not been established.

In human breast cancer cells, melatonin has been demonstrated to bind to MT-1 receptors, suppressing cyclic adenosine monophosphate and blocking linoleic acid uptake and conversion to 13-hydroxyoctadecadienoic acid (13-HODE). 13-HODE normally activates growth factor pathways, so downregulation by melatonin can decrease cell proliferation.[22] In addition, recent in vitro work has shown that melatonin induces expression of PER2.[23] These results suggest that shift workers with increased evening light exposure, resulting in increased melatonin suppression, would have decreased expression of PER2. Mice deficient in mPER2 show an increased incidence of cancer,[24] whereas overexpression of mPER2 induces cancer cell apoptosis.[25] PER2 downregulates the protein β-catenin, which regulates cyclin D. In the absence of PER2, cyclin D increases, resulting in increased cell proliferation.[26] Melatonin has also been demonstrated to activate glycogen synthase kinase 3b (GSK3b), which normally inhibits the epithelial to mesenchymal transition that can promote cancer progression; however, in the absence of melatonin, this process is upregulated, resulting in increased cancer progression.[27]

In support of the circadian disruption hypothesis, there have been multiple studies showing that impaired circadian rhythmicity, seen either in flattening of the diurnal cortisol rhythm or decreased amplitude of the daily rest activity pattern, corresponds with a worse prognosis in lung and colorectal cancer, respectively[28,29]; however, these studies describe correlations, so it is not known whether circadian disruption is the cause of or a result of worsening prognosis. To further evaluate the role of circadian disruption in malignancy, several studies have evaluated the possible role of circadian clock genes. Recent studies have looked specifically at the *hPER2* mutation found in familial advanced sleep-wake phase disorder (ASWPD) and have shown that this mutation results in accelerated tumorigenesis in mice[30]; however, there are currently no reports of an increased incidence of cancer in humans with ASWPD.

Several small studies have evaluated other circadian clock genes for their possible role in carcinogenesis. Recent analysis has shown a decreased expression of hPER3 in individuals with more advanced head and neck cancer,[31] and low hPER3 has also been found in patients with recurrent breast cancer.[32] Specific polymorphisms in *CRY2* have been significantly associated with increased breast cancer risk in women with estrogen and progesterone receptor–negative breast tumors.[33] In chronic lymphocytic leukemia, elevated expression of CRY1 is associated with higher risk disease.[34] Polymorphisms in the *CLOCK* gene have been associated with improved survival in colorectal cancer, as well as reduced risk for breast cancer.[35,36] Evaluating subjects from the Nurses' Health Study cohort demonstrated that polymorphisms in *NPAS2* were associated with increased breast cancer risk, particularly among shift workers,[37] whereas a study of Norwegian nurses demonstrated that polymorphisms in *ARNTL*, *AANAT,* and *RORB* increased the risk for breast cancer in shift workers.[38] In addition, polymorphisms in the *TIMELESS* gene have also been associated with a reduced risk for breast cancer.[24] Although these findings simply reflect association, it is intriguing and quite plausible that circadian clock dysregulation at least in part confers an increased risk for cancer.

Finally, along with the potential increased risk for cancer, there is growing evidence that the presence of circadian dysregulation may also negatively affect treatment response. In one study evaluating sleep-log data and chronotype, those reporting "going to bed at preferred bedtime" had a significantly longer disease-free interval (81.9 vs. 46.9 months) compared with those who were going to bed later or earlier than the preferred bedtime.[39] Evidence for one potential mechanism for the difference in disease progression comes from rat studies, which demonstrated that light exposure at night, promoting circadian disruption, resulted in both increased tumor growth and resistance to tamoxifen, a common medication used in the treatment of breast cancer.[40]

Shift Work and Metabolic Syndrome

There is substantial evidence that exposure to night shift work, rotating shift work, and even social jet lag is associated with an increased risk for metabolic syndrome and its varying components. Metabolic syndrome describes a group of risk factors, including central obesity, elevated blood pressure, elevated triglycerides, elevated fasting glucose, and low high-density lipoprotein (HDL) cholesterol, which increase the risk for cardiovascular disease and diabetes.

One of the first studies to evaluate this relationship included 27,485 workers in Sweden and found an increase in obesity, elevated triglycerides, decreased HDL, and impaired glucose tolerance among shift workers.[41] It has also been found that shift work is associated with an increase in systolic[42] and diastolic blood pressure[43] and with a loss of the normal nocturnal dip in blood pressure.[44] In addition, prospective trials have found that both men and women working rotating shifts have an accelerated development of metabolic syndrome compared with day workers.[45,46] One question that arises is whether these risks are primarily due to social and behavioral differences between shift workers and day workers. However, in a Finnish study, even after controlling for education, smoking, physical activity, alcohol, and insomnia, the

increased prevalence of metabolic syndrome among shift workers persisted.[47]

There are several potential causes of this increased risk. Certainly one component is sleep restriction. Studies have shown that night and rotating shift workers have a decrease in total sleep time ranging from 1 to 4 hours compared with controls,[48,49] and short sleep duration has been associated with many of these risk factors. Under laboratory conditions looking at healthy control individuals, the combination of 3 weeks of sleep restriction and circadian disruption resulted in decreased resting metabolic rate and increased postprandial plasma glucose due to decreased insulin secretion.[50]

In addition to sleep deprivation, shift workers get increased light exposure at night. In animal studies light at night results in increased weight and reduced glucose tolerance.[51] These effects may in part be due to the light-induced suppression of nocturnal melatonin, although of note, not all mouse strains secrete melatonin,[52] and the pattern of melatonin release in the mouse strain used in this study is unknown. However, in support of this concept in humans, decreased morning urinary melatonin levels have been associated with an increased risk for hypertension[53] and diabetes.[54]

CIRCADIAN DYSREGULATION AND PHYSICAL HEALTH

Outside of environmental and genetic disruption, circadian dysregulation has also been demonstrated to play an important role in a variety of other physical disorders from cardiovascular to respiratory disease to reproductive health.

Cardiometabolic Dysfunction

Although the greatest evidence for the role of circadian dysregulation in cardiovascular disease comes from shift work studies, there are many other ways in which the circadian system has been demonstrated to play a role. Major cardiovascular events, including myocardial infarction, sudden cardiac death, and stroke, all have a peak occurrence in the morning,[55-57] which appears to be regulated by circadian factors. Recent research indicates that this may be due to circadian alterations in clotting, affecting both platelet activation[58] and prothrombotic factors.[59] Alteration in the timing of these factors in relation to underlying activity could presumably result in further increased cardiovascular disease risk.

Recent data have demonstrated that even among day shift workers, greater degrees of social jet leg have been associated with an increase in body mass index.[60] In addition, having a later chronotype as determined by midsleep point is associated with poorer glycemic control in type 2 diabetes.[61] Those with later chronotypes are more likely to suffer from social jet lag, again suggesting that it is this mismatch between the internal circadian system and the external environment that is contributing to the increased metabolic risk.

Low nocturnal melatonin levels, either due to light at night, circadian dysregulation, or overall low secretion, also seem to play a role in cardiometabolic disease risk. It is known that loss of normal melatonin secretion in rats through pinealectomy can induce hypertension.[62] In humans, studying a small group of elderly patients with hypertension comparing those who lack the usual nocturnal dip in blood pressure (nondippers) with those who do exhibit a normal dip (dippers), it was found that the total 24-hour urinary melatonin level was the same between groups. However, the nondippers were noted to lack the usually observed nocturnal surge in melatonin,[63] suggesting that the timing may be more important than the absolute level of melatonin. Exogenous melatonin has been demonstrated to act directly on the vascular system as a vasodilator.[64] In addition, it has been postulated that through its antioxidant properties, melatonin decreases oxidative injury and endothelial dysfunction,[65] which regulate vascular resistance and blood pressure.

At a genetic level polymorphisms in the gene coding for the MT2 melatonin receptor have been associated with increased fasting plasma glucose and risk for type 2 diabetes.[66,67] The MT2 receptor is found on pancreatic islet cells and may mediate the inhibition of insulin secretion by melatonin.[68] This association persists in a meta-analysis of 23 studies, although of note the data seem to be stronger in white people than East or South Asians.[69] In mice, mutations in the *CLOCK* gene, resulting in circadian disruption, have been associated with symptoms of metabolic syndrome.[70] Smaller human studies have shown that polymorphisms in the *CLOCK* gene are associated with both the development of metabolic syndrome[71] and the response to dietary interventions to treat metabolic syndrome,[72] whereas polymorphisms in *ARNTL* are associated with hypertension and diabetes.[73] In addition, *NR1D1,* another core clock gene, seems to be involved in lipid and glucose metabolism,[74] although further research in this area is needed.

Pulmonary Disease

From a respiratory standpoint it has long been recognized that asthma attacks tend to worsen overnight.[75] Forced desynchrony protocols in healthy individuals have demonstrated a circadian pattern in pulmonary function, specifically forced expiratory volume in 1 section (FEV_1) and the ratio of FEV_1 to forced vital capacity, independent of sleep state.[76] More recently it has been found that evening types are more likely to have nocturnal asthma than intermediate types,[77] again suggesting a circadian component to the illness, as well as a potential role for circadian dysregulation as a contributing factor, although more research in this area is needed.

Other Disorders

Although there is still only a limited amount of human data in this area, there is growing evidence from animal studies that circadian dysregulation can also contribute to multiple other disease pathologies. In rodent studies, circadian dysregulation has been demonstrated to be associated with increased risk for steatohepatitis, gut leakiness, and fetal loss.[78,79] However, this is clearly an area in which additional human research is needed.

CIRCADIAN REGULATION AND NEUROLOGIC DISORDERS

Many common neurologic disorders exhibit circadian patterns in their occurrence and severity, although it can sometimes be difficult to separate the circadian pattern of symptoms from the underlying sleep-wake cycle. Both seizure and headache onset often peak at specific times of day, as detailed later. In addition, melatonin can be an effective intervention for many neurodevelopmental disorders, again suggesting a circadian component to these illnesses. Autism has been associated with a decreased amplitude of melatonin secretion,[80] and sleep

duration and sleep latency as well as daytime behavior significantly improve following administration of nocturnal melatonin (0.75 to 15 mg).[81] Other less common neurodevelopmental disorders also are associated with abnormal melatonin profiles and can improve with supplemental melatonin. Smith-Magenis syndrome is a neurodevelopmental disorder associated with an inverted melatonin rhythm, resulting in daytime sleepiness and nighttime insomnia,[82] whereas Angelman syndrome is associated with low melatonin levels, and sleep initiation and maintenance improve with supplemental melatonin (2.5 to 5 mg).[83]

Different seizure types have varying patterns of expression, and these also can vary between children and adults. In children being monitored on an inpatient epilepsy unit, tonic and tonic-clonic seizure occurred more frequently during sleep, whereas absence, atonic, and myoclonic seizures occurred primarily during wakefulness, and clonic seizures and epileptic spasms had two peaks, during the early morning and afternoon.[84] In adults monitored outside of the hospital, frontal lobe seizures occurred more frequently during the early morning, whereas temporal lobe seizures occurred more frequently during the afternoon.[85] To try to separate the degree to which these difference are dependent on sleep state versus circadian time, five subjects with epilepsy underwent a forced desynchrony protocol. Overall, interictal epileptiform discharges were most common coming out of non–rapid eye movement sleep; however, in addition there did seem to be a circadian pattern to interictal epileptiform discharge frequency, although the specific pattern varied between subjects.[86] However, this was a very small study looking primarily at subjects with idiopathic generalized epilepsy, so it is hard to generalize these findings at this point.

Headaches also have both strong circadian and sleep-wake components. Cluster headaches often reoccur at the same time of day, and there is functional magnetic resonance imaging evidence for hypothalamic activation during cluster headache onset, suggesting involvement of the SCN.[87] In addition, individuals with cluster headaches have both an advance and a decrease in amplitude of melatonin secretion.[88] Interestingly, hypothalamic deep brain stimulation has been an effective treatment for some individuals with refractory cluster headaches.[89] Migraine headaches tend to more frequently begin during the morning,[90] and more recently it was demonstrated that the mutation responsible for ASWPD is also associated with familial hemiplegic migraine, again suggesting a circadian component.[91]

Circadian Dysregulation in Neurodegenerative Disease

Neurodegenerative disorders are frequently associated with disruptions in circadian timing. This is most commonly seen in the alterations in the sleep-wake cycle, with "sundowning" representing a frequent problem for both patients and caregivers. However, as we will illustrate in more detail with AD and PD, circadian dysregulation may also be one of the earliest indications of the presence of a neurodegenerative disease, and circadian-based interventions may improve the overall disease course.

Alzheimer Disease

AD is the most common form of dementia in adults. The primary clinical manifestation is progressive cognitive deterioration. The pathologic hallmarks include deposition of β-amyloid plaques and hyperphosphorylated tau protein neurofibrillary tangles.[92]

Circadian dysfunction in the sleep-wake rhythm has long been recognized in AD. Several cross-sectional studies, although small, have shown alterations in the sleep-wake rhythm in demented patients. Using wrist actigraphy monitoring, patients with AD appear to have less overall activity, dampened amplitude, and increased fragmentation of the rest-activity rhythm, leading to more nocturnal wake time and daytime napping.[93-95] The degree of reduction in amplitude of the rest-activity rhythm correlates with the severity of AD.[95] The degree of daytime sleepiness based on the Multiple Sleep Latency Test is apparent even in mild to moderate AD and correlates with the degree of cognitive dysfunction.[93] Greater interdaily instability of the rest-activity cycle is associated with cognitive impairment and depression,[96] whereas decreased daytime activity predicts worsening cognition, functional and social impairment, and greater apathy.[94,96] Furthermore, the rest-activity rhythm may be predictive of the risk for dementia. In one large study of 1282 women, those in the lowest quartile for amplitude and robustness and those with a later timing of the rest-activity rhythm had higher risk for developing mild cognitive impairment or dementia over 4.9 years of follow-up.[97]

There are several potential mechanisms for the circadian disruption in AD. Age-associated reduction in exposure to synchronizing agents, such as light and structured social and physical activities, have been shown to contribute to circadian dysfunction. Actigraphy data have demonstrated that patients with severe dementia spend less time exposed to bright light and also have a blunted amplitude of daily activity compared with nondemented nursing home patients.[98] Increased physical activity has been associated with a decrease in dementia-related mortality, independent of age, gender, and cardiovascular risk factors.[99] However, regardless of the total amount of daily activity, greater fragmentation of rest-activity patterns has been associated with decreased cognitive performance.[100] Interventions that increase the amplitude of social and physical activity rhythms as well as daytime light exposure have been shown to decrease daytime sleep and in some cases also improve nocturnal sleep quality.[101] For example, structured social activity for 1 to 2 hours daily for 21 days in 147 nursing home residents with dementia improved consolidation of nighttime sleep and reduced daytime sleepiness.[102] Several small studies (n = 14 to 133) incorporating physical and social activity as part of a multicomponent program have shown some benefits in sleep or circadian rhythms for nursing home residents.[103-105] A larger clinical trial (n = 173) across several nursing homes did show decreases in daytime sleep with these interventions, but no improvement in nighttime sleep measures, thought to result in part from an inability to fully decrease nocturnal noise.[106]

Alzheimer tangles have been seen in the SCN,[107-110] and the degree of pathology is associated with reduced amplitudes in activity, core body temperature rhythms,[111] and melatonin secretion.[112-114] However, the sleep fragmentation seen with AD appears to be more strongly correlated with loss of galanin-containing neurons in the intermediate nucleus, rather than the loss of vasoactive intestinal peptide– or arginine vasopressin–immunoreactive cells in the SCN.[115]

Decreased melatonin secretion can be found even in early pathologic stages of AD and correlates with severity of

pathology.[116] This low melatonin secretion may play a role in the expression of AD. In vitro, melatonin inhibits formation of pathologic β-amyloid fibrils, tau protein phosphorylation, and neurotoxicity.[117,118] In transgenic animal models of AD, melatonin administration inhibits the development of Alzheimer pathology[119-122] and may reduce deterioration of learning and memory in most studies performed.[119,121] One study failed to show changes in cortical levels of β-amyloid plaques following melatonin administration, but melatonin was given when the mice were older, after amyloid plaques typically appear, so the intervention may have been too late.[123]

Another potential mechanism of cognitive dysfunction in AD may be circadian desynchrony of rhythms in brain networks.[124] The hippocampus and other brain regions involved in AD have semiautonomous clocks,[124] and their circadian oscillations differ in their dependence on the SCN.[125] Finally, recent data from transgenic mouse model of AD show that β-amyloid has a diurnal rhythm, increasing during wakefulness and decreasing during sleep.[126] During sleep, toxic products such as β-amyloid are cleared from the brain,[127] whereas β-amyloid continues to increase during sleep deprivation.[126] It remains unclear whether the accumulation and clearance of β-amyloid are directly influenced by the circadian system.

Because of the findings of circadian disruption in AD, many treatment strategies have focused on strengthening circadian signals, primarily focusing on using either melatonin, light, or a combination of the two. Treatment with melatonin has had mixed effects. On the one hand several small or uncontrolled studies using melatonin in older adults have been reported to improve sundowning,[128] agitation,[129] nocturnal activity,[113,130] and the amplitude of the rest-activity rhythm.[131] However, only two of these studies specifically examined AD. One large study of 157 people with AD did not confirm an objective benefit from melatonin administration by actigraphy, although improvements in caregiver ratings of the participants' sleep quality were noted.[132] Bright light therapy appears to more consistently improve sleep and circadian dysfunction in AD. Morning bright light therapy (usually 2500 lux or more) increases nighttime sleep and daytime wakefulness and also improves evening agitation.[133] Evening light exposure improves consolidation of the rest-activity rhythm[133]; even 30 lux of blue light improves sleep efficiency.[134] Light exposure given continuously throughout the day increases total sleep time, improves interdaily stability of the rest-activity rhythm, improves mood, and even slightly attenuates cognitive deterioration when given long term.[133,135] A 24-hour lighting scheme may be helpful to improve the sleep-wake rhythm, decrease agitation, and reduce falls at night.[136] This plan consists of high amount of light exposure during the day and low amount at night, with the exception of nightlights to reduce falls if awake at night. Finally, a combination of light and melatonin (5 mg) demonstrated a significant improvement in daytime activity and an overall strengthening of the rest-activity pattern compared with light alone.[137]

Parkinson Disease

PD is a neurodegenerative disorder with clinical manifestations of bradykinesia, rigidity, and tremor. The pathologic hallmark in PD is the accumulation of Lewy bodies, which are intraneuronal α-synuclein inclusions, in the dopamine-containing nigrostriatal neurons. Sleep and wake disturbances, particularly sleep fragmentation and hypersomnia, are reported in about 60% to 90% of patients with PD.[138,139] Recent evidence from a transgenic mouse model for PD demonstrated increased fragmentation of the rest-activity pattern and decreased firing of the SCN neurons early in the disease course, suggesting that circadian dysregulation may underlie these sleep disturbances and may even contribute to the pathophysiology of PD.[140]

In human studies, circadian patterns of activity and hormone release also appear to be altered in PD. The few studies of the rest-activity rhythm in PD using wrist actigraphy show reduction in overall activity, increased nocturnal activity, and increased daytime sleep, resulting in an overall decreased amplitude.[141-143] In addition, patients with hallucinations had decreased day-to-day stability of the rest-activity rhythm and lower relative activity amplitude than those without hallucinations.[142]

Melatonin rhythms are also affected in PD. Earlier studies suggested that people with PD on dopaminergic treatment have a phase advance in melatonin rhythms compared with untreated patients.[144,145] More recent studies have not been able to confirm this but did demonstrate that individuals with PD on dopaminergic therapy had a larger phase angle between salivary DLMO and habitual sleep-onset time compared with untreated patients and controls.[146] In addition, patients with PD had a lower amplitude of daily melatonin secretion compared with controls, with the lowest amplitudes seen in those with complaints of daytime sleepiness (Figure 39-2).[147] Smaller studies have also shown disturbances in the circadian rhythms of blood pressure,[148] core body temperature,[149] and cortisol rhythm.[150]

Potential mechanisms underlying these circadian disturbances may include alteration of exposure to zeitgebers or degeneration of afferents to the SCN (e.g., retina), the SCN itself, or downstream efferents. People with PD may get less light exposure during the day,[151] although this has not been clearly demonstrated. In addition, age-related meiosis and cataracts may lead to decreased retinal illumination.[152] Dopamine, the primary neurotransmitters affected in PD, enhances melanopsin expression in retinal ganglion cells that project to the SCN[153] and may influence the ability of light to reset retinal PER2 rhythms.[154] Alterations in dopamine function, from either neurodegeneration or dopaminergic therapy, may therefore influence the light response. The SCN itself may also be affected in PD. In a mouse model of PD, decreased firing of SCN neurons has been observed, and this may impair the ability of the SCN to drive peripheral rhythms.[155] There is also some evidence that areas downstream of the SCN that play an important role in regulating circadian rhythms may also be disrupted, such as striatal neurons in which dopamine influences expression of circadian genes *PER1* and *PER2*.[156]

Because of the apparent circadian disruption noted in this disease, there have been some efforts to look at circadian-based interventions, but data examining therapies such as exogenous melatonin or bright light therapy to improve circadian rhythms are limited. Melatonin (3 to 5 mg) at bedtime can improve subjective sleep quality,[157] subjective sleep quantity, and daytime sleepiness.[157,158] However, high doses (50 mg) result in only a small objective improvement in sleep duration.[158] Light therapy for about 1 hour before bedtime

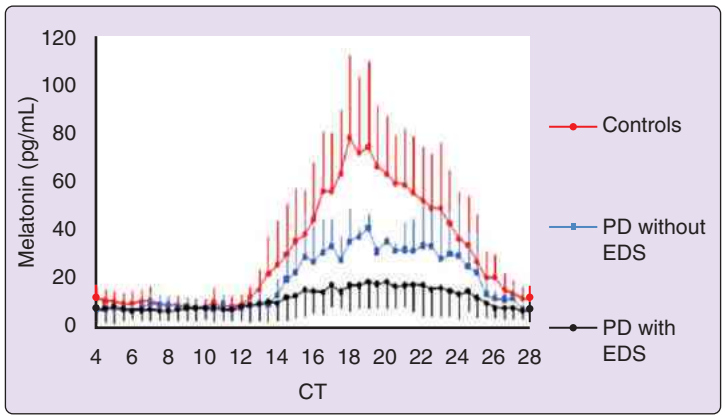


Figure 39-2 The amplitude of daily melatonin release is decreased in Parkinson disease (PD) compared with controls, and a greater decrease is seen in PD individuals with excessive daytime sleepiness (EDS) compared with those without EDS. CT, Circadian time.

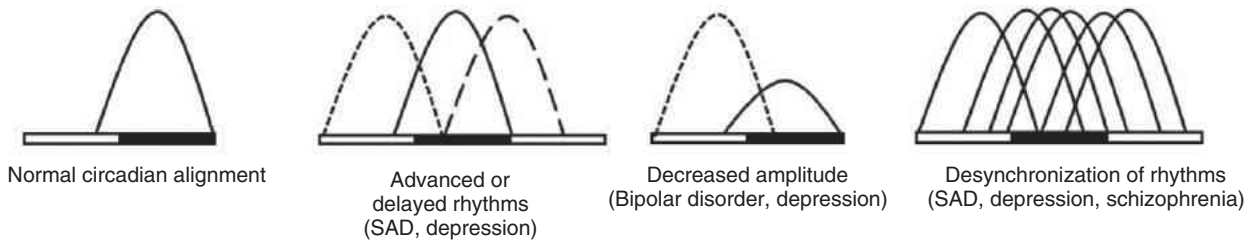


Figure 39-3 Circadian dysregulation can be associated with mood disorders in multiple ways. The phase of the rhythm can be shifted earlier (depression) or later (seasonal affective disorder [SAD]). The overall amplitude of the rhythm can be decreased, which can be seen in symptomatic bipolar patients, or there can be internal desynchrony, in which rhythms are no longer appropriately aligned.

can improve not only sleep onset latency and sleep efficiency but also motor symptoms and mood.[159,160] A controlled prospective study of morning bright light therapy (7500 lux for 30 minutes) in PD also improved subjective motor symptoms and mood, but it did not improve objective motor symptoms.[161]

CIRCADIAN DYSREGULATION AND PSYCHIATRIC DISORDERS

Circadian Dysregulation in the Pathophysiology of Psychiatric Disorders

Many psychiatric disorders have been noted to have a component of circadian dysregulation associated with them, with the strongest evidence for mood disorders, such as seasonal affective disorder (SAD), depression, and bipolar disorder. However, there is also evidence for disruption in the circadian activity pattern in schizophrenia[162] and a greater prevalence of delayed sleep phase disorder in patients with obsessive-compulsive disorder.[163,164] The observed circadian changes generally present as either a shift in the circadian phase, a decrease in circadian amplitude, or a misalignment between rhythms, most frequently seen in the relationship between sleep and the 24-hour patterns of hormone release (Figure 39-3).

SAD consists of depressive symptoms that are primarily present in the late fall and winter, during a period of shorter days when there is decreased exposure to natural bright light as well as a decrease in the intensity of light that is available. It was theorized that the decrease in exposure to an advancing light signal in the morning facilitated delay in the timing of circadian rhythms. Based on this theory, a small trial was performed using morning bright-light exposure as a treatment for SAD, demonstrating a positive response in seven of eight subjects.[165] Interestingly, symptoms appeared to correlate with the initial degree of misalignment between the timing of the dim-light melatonin onset and the midsleep point, and normalization of this difference improved treatment response.[166] These findings suggest that a shift in overall timing of melatonin release, resulting in misalignment between the melatonin and sleep-wake rhythms, contributes to the pathophysiology of this disorder. More recent data suggest that in addition to these factors, impaired light input to the circadian clock may also play a role. Light normally reaches the circadian system through melanopsin-containing retinal ganglion cells. Using a measure of pupillary light constriction as a surrogate marker for melanopsin cell function, it has been demonstrated that individuals with SAD have a decreased sensitivity to light compared with controls,[167] suggesting that it is not only a lack of exposure to light but also a decreased ability to respond to light that contributes to the underlying pathology.

Depression, on the other hand, is often associated with early morning awakenings and decreased rapid eye movement

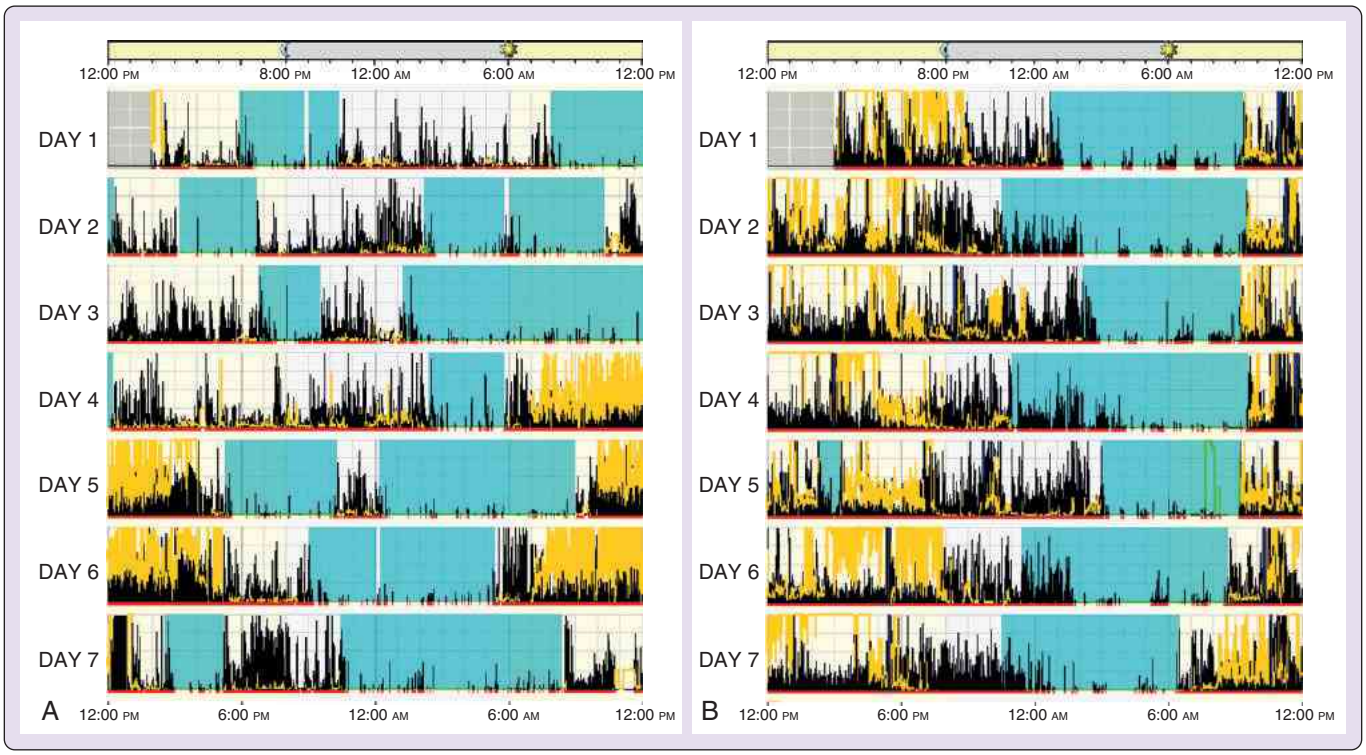


Figure 39-4 Sample actigraphy data from a subject with bipolar disorder during a period of depression **(A)** and during a period of stable mood **(B)**. The total activity and overall amplitude of the daily activity rhythm are decreased during the symptomatic period.

sleep latency, leading to early suggestions of the phase advance hypothesis[168] for the pathophysiology of depression. Supporting this is evidence that 24-hour patterns of cortisol and norepinephrine release are advanced in individuals with depression, and the amplitude of cortisol release is also decreased.[169] However, these effects seem to be age dependent, with younger individuals exhibiting greater phase delays in sleep-wake patterns and older individuals showing primarily a decrease in amplitude of the daily pattern of activity.[170] Things are further complicated by the high comorbidity of mood disorders in individuals with delayed sleep-wake phase disorder, suggesting that symptoms are not always due to a phase advance in circadian rhythms.[171,172] Overall these results suggest that the problem may be more complicated than simply an advance or delay in the circadian rhythm, leading to more recent research focused not just on the phase of circadian rhythms with respect to the environment but also on the alignment between rhythms, examining the relationship between hormone release, core body temperature, and the sleep-wake cycle, as well as the overall amplitude. Data are limited, but it appears that a shorter phase angle between either DLMO[173] or core body temperature[174] and the sleep midpoint may be associated with greater depression. Further research is needed in this area.

Patients with bipolar disorder exhibit recurring cycles of depression and mania or hypomania suggesting an underlying disorder of timing, and there is evidence of greater eveningness preference[175] among individuals with bipolar disorder. In addition, studies have demonstrated decreased stability of the sleep-wake pattern and decreased amplitude of activity rhythms, particularly among symptomatic individuals (Figure 39-4).[176] Data regarding abnormalities in circadian phase or alignment in these individuals are extremely limited to date, and this is an area in need of further research.

Given the findings of circadian dysregulation in psychiatric disorders, there has been extensive work performed looking for associations between polymorphisms in known clock genes or alterations in clock gene expression in psychiatric disorders. However, any potential associations to date have been limited by small sample sizes or poor replication across different ethnic groups. There is some evidence that *CRY1* polymorphisms are associated with major depression in both European[177] and Chinese[178] cohorts, whereas *CRY2* variants are associated with dysthymia.[179] The T3111 polymorphism of the *CLOCK* gene has previously been demonstrated to be associated with morningness-eveningness preference,[180] and a small study demonstrated an association between *CLOCK* polymorphisms and bipolar disorder.[177] However, replicable results have not been found in studies looking at depression,[181] SAD,[182] or schizophrenia.[183-185] A polymorphic repeat domain of *hPER3* has previously been associated with diurnal preference[186] but has not shown consistent associations with psychiatric disorders across different ethnic backgrounds.[184,187-189] Furthermore, associations have not been found between other circadian clock genes, such as *TIMELESS*, *CKIε*, and *δ* polymorphisms with depression, schizophrenia, or bipolar disorders.[190,191]

To try to address the limitations of small samples sizes and geographically limited populations, two large genome-wide association studies have been performed, but again results are not consistent. One study found several associations with bipolar disorder, depression, and schizophrenia,[192] but these

findings were not replicable in a second study, thought to be related to analyzing the data as a pooled data set rather than as individual genome-wide association studies.[193] Alternatively, rather than looking for associations with individual polymorphisms, another group evaluated whether the overall circadian pattern of gene expression differs between affected and control individuals. By performing time-of-death analysis of postmortem brains in individuals with depression and controls, Li and colleagues demonstrated that individuals with depression had alterations in peak timing of core clock gene expression, as well as disrupted phase relationships between gene expression patterns, resulting in weaker overall amplitudes of cyclical patterns.[194] However, this does not distinguish whether the altered expression patterns are a cause or an effect of the underlying disorder. Although circadian rhythms are often disrupted in psychiatric disorders, the exact biologic mechanisms underlying this relationship are unclear.

Chronobiology as Treatment for Psychiatric Disorders

There is growing evidence that treatments based on addressing circadian dysregulation can be effective adjunctive treatment for treating several psychiatric disorders. The best studied and most widely accepted of these treatments is bright-light therapy for the treatment of SAD. SAD symptoms are greater in individuals with a larger degree of circadian misalignment, and response to treatment corresponds with the degree of realignment.[166] In depression light therapy is generally not considered first-line treatment, but it can be successfully used as an adjunctive therapy for patients with drug-resistant depression or for individuals for whom medications may be undesirable (e.g., depression).[195] Finally, for patients with bipolar disorder, interpersonal and social rhythm therapy is a successful treatment strategy that, along with standard medication therapy, focuses on behavioral strategies to address patients' sleep and circadian vulnerabilities.[196] Overall the evidence indicates that circadian dysregulation plays an important role in disorders of mental health, and circadian-based approaches can be used to enhance treatment effectiveness.

CLINICAL PEARLS

- Circadian dysregulation is increasingly being recognized as an important component of disease risk.
- Assessing circadian timing and alignment in all patients, particularly those with cardiometabolic, neurologic, and psychiatric disorders, should be an important element of the clinical history.
- Circadian-based interventions have the potential to decrease disease risk and improve treatment outcomes.

SUMMARY

Disrupted circadian rhythms, either intrinsically due to genetic mutations or alterations in the SCN and its afferent and efferent pathways or extrinsically due to social and work demands, can affect the expression and development of many of the most common and challenging disorders in our modern society, including heart disease, diabetes, neuropsychiatric disorders, and neurodegenerative disorders. Exciting recent evidence points to circadian dysregulation as an early biomarker of neurodegeneration, as is seen in AD and PD. Together these findings suggest increasing importance of biologic timing in sleep medicine and in medicine as a whole.

Selected Readings

Bunney BG, Li JZ, Walsh DM, et al. Circadian dysregulation of clock genes: clues to rapid treatments in major depressive disorder. *Mol Psychiatry* 2015;**20**:48–55.

Buxton OM, Cain SW, O'Connor SP, et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. *Sci Transl Med* 2012;**4**:129ra143.

Dallaspezia S, Benedetti F. Chronobiology of bipolar disorder: therapeutic implication. *Curr Psychiatry Rep* 2015;**17**:606.

Depner CM, Stothard ER, Wright KP Jr. Metabolic consequences of sleep and circadian disorders. *Curr Diab Rep* 2014;**14**:507.

Fonken LK, Workman JL, Walton JC, et al. Light at night increases body mass by shifting the time of food intake. *Proc Natl Acad Sci U S A* 2010;**107**:18664–9.

Li JZ, Bunney BG, Meng F, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc Natl Acad Sci U S A* 2013;**110**:9950–5.

Litinski M, Scheer FA, Shea SA. Influence of the circadian system on disease severity. *Sleep Med Clin* 2009;**4**:143–63.

Morris CJ, Yang JN, Garcia JI, et al. Endogenous circadian system and circadian misalignment impact glucose tolerance via separate mechanisms in humans. *Proc Natl Acad Sci U S A* 2015;**112**:E2225–34.

Muller JE, Stone PH, Turi ZG, et al. Circadian variation in the frequency of onset of acute myocardial infarction. *N Engl J Med* 1985;**313**:1315–22.

Schernhammer ES, Laden F, Speizer FE, et al. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *J Natl Cancer Inst* 2001;**93**:1563–8.

Smolensky MH, Portaluppi F, Manfredini R, et al. Diurnal and twenty-four hour patterning of human diseases: acute and chronic common and uncommon medical conditions. *Sleep Med Rev* 2015;**21**:12–22.

Videnovic A, Noble C, Reid KJ, et al. Circadian melatonin rhythm and excessive daytime sleepiness in Parkinson disease. *JAMA Neurol* 2014;**71**:463–9.

Zee PC, Turek FW. Respect the clock. *Sleep Med Rev* 2013;**17**:395–7.

A complete reference list can be found online at ExpertConsult.com.

Chapter
40

Circadian Disorders of the Sleep-Wake Cycle

Sabra M. Abbott; Kathryn J. Reid; Phyllis C. Zee

Chapter Highlights

- Circadian rhythms are regulated by the suprachiasmatic nucleus in the hypothalamus.
- Circadian rhythm sleep-wake disorders can result from (1) changes in the light-dark cycle in relation to the internal clock, (2) changes in the internal clock in relation to the light dark cycle, and (3) dysfunction in the clock mechanism.
- Circadian rhythm sleep-wake disorders can be treated with a combination of behavioral interventions and carefully timed light and melatonin.

The sleep-wake cycle is generated by a complex interaction of endogenous circadian and sleep homeostatic (need for sleep increases as a function of prior wakefulness) processes, as well as social and environmental factors. Physiologic sleepiness and alertness not only vary with prior waking duration but also exhibit circadian variation. In humans, daily variation in physiologic sleep tendency reveals a biphasic circadian rhythm of wake and sleep propensity,[1,2] with a midday increase in sleep tendency occurring at about 2 to 4 PM, followed by a robust decrease in sleep tendency and increase in alertness that lasts through the early to middle evening hours. A primary role of the circadian clock is to promote wakefulness during the day and thus also to facilitate the consolidation of sleep during the nighttime hours.[1,3-5] In humans, this interaction between circadian and homeostatic processes results in approximately 16 hours of wakefulness and 8 hours of nocturnal sleep.

The timing and duration of the sleep-wake cycle depends on the synchronization of the endogenous circadian clock with the external physical light-dark (LD) cycle, as well as social or professional demands. Circadian rhythm sleep-wake disorders (CRSWDs) arise when there is either disruption of this internal timing mechanism or a misalignment between the timing of the circadian clock and the 24-hour social and physical environments. In this chapter, the nomenclature and diagnostic criteria for CRSWD and its subtypes are based on the criteria published in the *International Classification of Sleep Disorders,* third edition. These dyssynchronous states fall into three categories: (1) the terrestrial LD cycle may change relative to circadian timekeeping (shift work and jet lag), (2) circadian timekeeping may change relative to the terrestrial LD cycle (delayed sleep-wake phase disorder [DSWPD], advanced sleep-wake phase disorder [ASWPD], or non-24-hour sleep-wake rhythm disorder [N24SWD]), or (3) dysfunction in clock mechanisms (irregular sleep-wake rhythm [ISWRD]). The first circumstance occurs in the presence of a normal circadian timekeeping system and is generally self-limited or resolves with environmental change. The second circumstance is believed to occur because of chronic alteration in the circadian system, resulting in the inability of the circadian pacemaker to achieve a conventional phase relation with the external world, and the third is primarily due to dysfunction of the central clock or its afferent-efferent pathways. This chapter focuses on the second and third group of disorders. Because circadian variation in wakefulness and sleep propensity is the most apparent of the many behavioral and physiologic outputs of the circadian pacemaker, it is not surprising that the most apparent circadian rhythm disorders to be recognized involve the sleep-wake cycle.[6]

Disruptions in the timing of sleep and wakefulness are often associated with symptoms of difficulty falling or staying asleep or excessive sleepiness that cause patients to seek medical attention. Thus, in clinical practice, CRSWD is often underrecognized, yet should be considered in the differential diagnosis of any patient presenting with symptoms of insomnia or hypersomnia. A multimodal treatment approach of behavioral or pharmacologic strategies aimed to improve circadian function and alignment of circadian rhythms to the 24-hour environment is often necessary to consolidate sleep and improve daytime function in patients with CRSWD. Growing knowledge about how the circadian system responds to photic as well as to nonphotic entraining agents is increasing the number of practical therapies that can be used in "real life" clinical settings.

REGULATION AND ENTRAINMENT OF CIRCADIAN RHYTHMS

The suprachiasmatic nucleus (SCN), in the anterior hypothalamus, is the central pacemaker responsible for the generation of circadian rhythmicity in mammals.[7] Animals and humans removed from the external LD cycle and other time cues (zeitgebers) exhibit a self-sustaining cycle of sleep and wakefulness as well as many other physiologic and hormonal rhythms.

The endogenous frequency of this cycle of oscillation or free-running period is largely genetically determined.[8] The basic molecular mechanism by which SCN neurons generate and maintain a self-sustaining rhythm is through an autoregulatory feedback loop in which oscillating circadian gene products regulate their own expression through a complex system of transcription, translation, and posttranslational processes.[9] In the mouse[10] and hamster,[11] genes have been identified that lengthen[10] and shorten[11,12] the free-running period. The mammalian circadian period is generally slightly longer than 24 hours in diurnal animals and slightly shorter than 24 hours in nocturnal animals. In humans, the average circadian period has been estimated to be approximately 24.18 hours[13] and must therefore be synchronized or entrained on a regular basis to the 24-hour terrestrial day by external influences.

Entrainment by Light

Light is the major external time cue in mammals. Light reaches the SCN by afferent projections from the retina through the retinohypothalamic tract.[7] Recent evidence indicates that the primary circadian photoreceptors are the melanopsin-containing retinal ganglion cells, which in turn send photic information through projections to the SCN.[14,15]

Although circadian rhythms can be entrained to LD cycles that are not exactly 24 hours in duration, entrainment is restricted to cycles with periods that are "close" to 24 hours in duration.[16] The range of entrainment can vary from species to species and is dependent on the experimental conditions (e.g., intensity of LD cycle, whether period of LD cycle is changed gradually or rapidly), but in general, animals do not entrain readily to LD cycles that are more than a few hours shorter or longer than the period of the endogenous circadian rhythm. If the period of the LD cycle is too short or long for entrainment to occur, the circadian rhythm will free-run with a period of the endogenous pacemaker.

One of the most widely used methods to examine how the LD cycle influences the circadian system has been to expose animals and humans maintained in constant conditions to pulses of light. The effects of the light pulse on a phase reference point of a circadian rhythm (e.g., onset of melatonin, minimum of body temperature) in subsequent cycles is then determined. The direction and magnitude of the phase shifts are strongly dependent on the circadian time at which the light pulse occurs. A phase-response curve (PRC) is a plot of the magnitude and direction of the time shift induced by an environmental perturbation as a function of the circadian time at which the perturbation is given. Light pulses presented near the onset of the subjective night (part of the circadian cycle that occurs during the dark or nighttime) delay circadian rhythms, whereas light pulses presented in the late subjective night or early subjective day (part of the circadian cycle that occurs during the light or daytime) phase-advance circadian rhythms. Extensive studies have demonstrated that the LD cycle can entrain circadian rhythms and that bright light can be used to manipulate human rhythms under a variety of experimental conditions.[17,18] Although bright light (intensities approximating sunlight) is a very strong and reliable entraining agent for the circadian system,[18] there is evidence that lower intensities, such as those encountered in ordinary room lighting, can also affect the timing of human circadian rhythms.[19] The recent discovery that, in humans, both light-induced phase shifts and melatonin suppression are most sensitive to short wavelength light of approximately 460 nm[20,21] provides an exciting new avenue for the development of light therapies to treat circadian rhythm sleep disorders.

Entrainment by Nonphotic Signals

The role of activity and social cues as synchronizing agents for the human circadian system has been recognized since the early 1970s. Studies by Aschoff and colleagues showed that scheduled bedtimes, mealtimes, and various timed social cues were able to entrain circadian rhythms.[22] More recent studies indicate that sleep and social schedules may also phase-shift the circadian clock.[23] In addition, physical exercise during the night can produce a phase delay in human circadian rhythms,[24,25] whereas exercise during the morning can accelerate entrainment following a phase advance in the sleep-wake cycle.[26] These findings indicate that scheduled social and physical activity programs may also be useful strategies for the treatment of circadian rhythms sleep disorders.

Melatonin

Melatonin is an important modulator of circadian rhythms[27] and alters the timing of circadian rhythms in animals[28] and humans.[29] The circadian rhythm of melatonin production and release is controlled by the SCN through an indirect pathway, a noradrenergic synapse from the superior cervical ganglion to the pineal gland.[30] The PRC for melatonin in humans indicates that administration of exogenous melatonin to humans in the early evening advances the phase of circadian rhythms, whereas administration in the early morning delays the phase, with the strongest phase-shifting effects of exogenous melatonin occurring during the evening, just preceding the increase in endogenous melatonin levels.[31-33] In addition to its phase resetting properties, evidence supports a role for melatonin in sleep modulation by increasing evening sleep propensity and reducing core body temperature.[27] Melatonin has been shown to decrease the firing rate of SCN neurons,[34] and it has been proposed that by its inhibition of SCN firing, the increase of melatonin in the evening creates a sleep-permissive state.

The potential importance of melatonin for regulating the sleep-wake cycle has led to interest in its use for treatment of insomnia and circadian rhythm sleep disorders. Indeed, there is good evidence that melatonin may be effective for entraining circadian rhythms in blind people with free-running sleep-wake cycles[35] and phase-advancing circadian rhythms in people with delayed sleep phase disorder.[36] Another potential use for melatonin has been to maintain consolidation of sleep during the early morning hours in elderly people.[37]

DELAYED SLEEP-WAKE PHASE DISORDER, DELAYED SLEEP PHASE TYPE, DELAYED SLEEP PHASE SYNDROME

Clinical Features

DSWPD is characterized by sleep onset and wake times that are usually delayed more than 2 hours and often up to 3 to 6 hours relative to conventional sleep-wake times (Figure 40-1). The typical patient finds it difficult to initiate sleep before 2 to 6 AM and, when free of societal constraints, prefers wake times of 10 AM to 1 PM. Sleep itself is reported to be normal

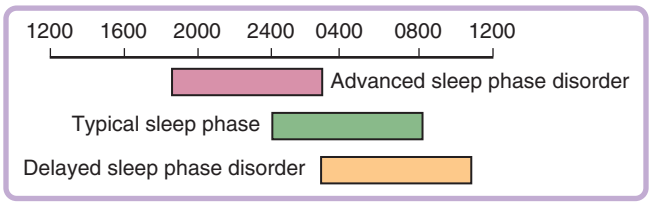


Figure 40-1 Schematic representation of the temporal distribution of sleep and wake in patients with circadian rhythm sleep disorders. Patients with advanced sleep phase disorder typically complain of evening sleepiness and either early-morning awakening or sleep disruption. Patients with delayed sleep phase disorder complain of difficulty initiating sleep, usually before 2 AM, and have difficulty awakening in the morning.

for age.[38,39] These symptoms are chronic, usually of at least 3 months—and quite often of many years duration. The clinical picture may be similar to sleep-onset insomnia. Patients are unable to advance their sleep times despite repeated attempts and may report a history of prolonged sedative-hypnotic drug use, bedtime use of alcohol, behavioral interventions, or psychotherapy.[40] Patients often report feeling most alert in the late evening and score highly as evening types on a morningness-eveningness scale.[41] Enforced "conventional" wake times may result in chronically insufficient sleep and excessive daytime sleepiness. Sleepiness is greatest in the morning and lessens as the circadian drive for wakefulness peaks in the late afternoon. In adolescents the syndrome may be associated with daytime irritability and poor school performance,[42] whereas in adulthood, the syndrome may be associated with impaired job performance and associated financial difficulty, as well as marital problems.[43] DSWPD may be mistaken for depression, in which the sleep-wake cycle may also be delayed (or advanced). Several studies, generally from psychiatric clinics, have emphasized an association of DSWPD with mood, obsessive-compulsive symptoms, and personality disorders.[43-46]

Epidemiology

DSWPD has been reported as young as preadolescence and beyond 60 years of age.[43] Although the actual prevalence of DSWPD in the general population is not well characterized, evidence from one population-based study indicates a prevalence of 0.17%.[47] DSWPD is reported to be more common in adolescents and young adults, with a reported prevalence of 3.3% to 7%,[48,49] whereas in middle-aged adults, the prevalence may be one tenth of that, or 0.7%.[50] In a group of New Zealand adults the prevalence was found to be 1.51% to 8.90%, depending on the definition used.[51] In a sleep disorders clinic, 6.7%[39] to 16%[44] of patients seen for a primary complaint of insomnia were determined to have DSWPD. There are no known gender differences in prevalence.

Pathogenesis

It has been pointed out that the tendency for late sleeping is not simply a function of the circadian drive for wakefulness interacting with the sleep homeostat but is analogous to eating or other behaviors that are mandated by physiology but overlaid by varying individual emotional, social, and medical states.[44] Although the exact cause of DSWPD is not known, there have been several mechanisms proposed, including both behavioral and physiologic factors. Behavioral preference may play a major role in some cases of DSWPD, particularly when bed times and rise times are not enforced.

In adolescence, for example, the biologic delay in the timing of circadian rhythms[52] is likely exacerbated by late evening activities, such as doing homework, watching TV, and using the Internet.[53] Other factors may include use of caffeine to combat sleepiness during normal waking hours. Staying up late and waking up late in the morning or afternoon may result in an abnormal relationship between the endogenous circadian rhythm and the sleep homeostatic process that regulates sleep and wakefulness. Evidence also shows that under certain conditions (a background of dim light), ambient artificial light (as low as 100 lux) at night may be of sufficient intensity to affect circadian timing.[19] Therefore light exposure later in the evening may also perpetuate and exacerbate the phase delay. Furthermore, late wake times will delay exposure to light in the morning and may prevent active advancement of the circadian clock, allowing it to drift to a new phase relation with external clock time. Studies in adolescents with subclinical DSWPD have found that simply following a fixed advanced sleep-wake schedule can result in significant advances in salivary melatonin rhythms,[54] suggesting a strong behavioral component.

For many individuals with DSWPD, however, symptoms often persist despite attempts to structure sleep and wake times, resulting in severe social or professional consequences[40,55,56] and suggesting that behavioral factors alone do not fully explain this disorder. There is considerable evidence that DSWPD is the result of alterations in the endogenous circadian system. For example, many physiologic markers of circadian phase persist in a delayed pattern despite enforced sleep-wake times.[40] There is also evidence that some individuals with DSWPD have a hypersensitivity to nighttime suppression of melatonin by bright light.[57] Reduced sensitivity of the oscillator to photic entrainment (i.e., a reduction in the amplitude of the advance portion of the PRC to light) has also been hypothesized, as has a prolonged free-running period length of the circadian cycle.[40,58,59] Furthermore, the duration and timing of environmental light and dark exposure may play a role in the expression of the DSWPD phenotype. For instance, the prevalence of DSWPD may be increased at extreme latitudes.[60] DSWPD has also been reported following minor traumatic brain injury.[61] There is also evidence for a genetic basis to DSWPD. In some cases the syndrome may be familial, presenting with an autosomal dominant mode of inheritance.[43,62] Further support for a genetic basis for DSWPD derives from reports of polymorphisms in circadian genes such as, *hPer3*, arylalkylamine *N*-acetyltransferase, human leukocyte antigen, and *Clock* that are associated with diurnal preference and DSWPD.[63-67]

Although it is commonly accepted that DSWPD is predominantly a result of alterations of circadian timing, there is recent evidence that alterations in the homeostatic regulation of sleep may play an important role as well.[68,69] Polysomnography (PSG) recordings of sleep in both adults and adolescents with DSWPD have shown that sleep architecture is not disrupted after the initiation of sleep when subjects are allowed to sleep until their desired wake times,[42,43,70,71] although individuals with DSWPD have been shown to have increased difficulty awakening from rapid eye movement (REM) sleep compared with controls.[72] Following 24 hours of sleep deprivation, subjects with

DSWPD, compared with controls, show a decreased ability to compensate for sleep loss during the subjective day and the first hours of subjective night.[68,73] Therefore it is likely that both alterations in circadian timing and impaired sleep recovery contribute to symptoms of insomnia and excessive sleepiness in DSWPD.

Diagnosis

The diagnosis of DSWPD is usually made on the basis of the patient's history of chronic or recurrent complaint of symptoms of insomnia due to a stable delay in the timing of the major sleep and wake period.[74] The sleep disturbance is associated with impairment of social, occupational, or other areas of functioning. In addition, sleep log or actigraphy monitoring should be performed for at least 7 days, but preferably 14 days, to demonstrate a stable delay in the timing of the habitual sleep period. Actigraphy is a practical tool for assessing sleep-wake cycles relative to clock time and has become more widely available clinically[75] (Figure 40-2). A morningness-eveningness questionnaire, such as the Horne-Ostberg[41] or the Munich chronotype[76] questionnaire, may be useful in confirming the patient's circadian preference but is not required to make the diagnosis.[77,78]

To make the diagnosis, medical, mental, or sleep disorders that may cause alterations in the sleep-wake cycle, insomnia, or excessive sleepiness should be excluded or be adequately treated. In adolescents, social maladjustment, family dysfunction, school avoidance, and affective disorders should be considered in the differential diagnosis. Nocturnal PSG is not necessary to establish the diagnosis but should be performed when another primary sleep disorder such as sleep apnea or parasomnia is suspected. When performed during conventional sleep laboratory hours, PSG often shows a prolonged sleep-onset latency as well as prolonged REM latency and may sometimes, in conjunction with an antecedent sleep log, be a clue to the diagnosis.

The use of other physiologic markers of circadian timing, such as continuous recording of body temperature[79] or dim light plasma melatonin onset (DLMO),[80] may also aid in determining the phase relation of circadian and terrestrial time, although routine clinical availability remains limited. DLMO is probably the most useful marker for circadian pacemaker output.[81,82] Individuals with DSWPD usually have DLMO times that occur after 10 PM.[80,83,84] (Figure 40-3). Determination of DLMO can be made by measurements of melatonin from plasma or saliva. Commercially available salivary determination of DLMO for clinical use may be feasible in the near future.

Treatment

The goal of therapy is to align the timing of the circadian clock with the desired 24-hour LD cycle. Richardson[6] has noted that treatment of DSWPD is the same whether it is the result of a primary behavioral or a primary physiologic process. Adherence to good sleep hygiene principles and identification and treatment of comorbid medical and psychiatric disorders are essential components in the management of DSWPD. In addition, chronotherapy, light therapy, and pharmacologic agents, such as melatonin, have been shown to be useful treatments for patients with DSWPD.

Figure 40-2 Representative rest-activity cycles recorded with wrist-activity monitoring of patients with circadian rhythm sleep disorders. The *black bars* indicate activity levels recorded at the nondominant wrist. Delayed sleep-wake phase disorder with sleep-onset times of approximately 3 to 4 AM and wake time of noon (*top panel*). Advanced sleep-wake phase disorder with sleep onset between 8 and 10 PM and wake time of 4 to 5 AM (*bottom panel*).

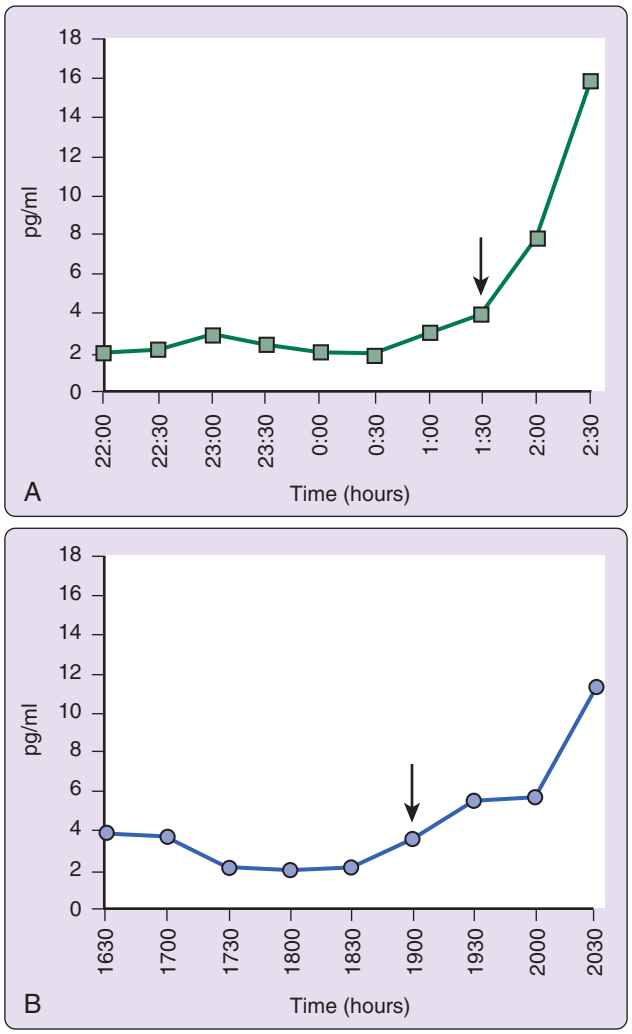


Figure 40-3 Dim light melatonin profiles of an individual with delayed sleep-wake phase disorder with a dim light melatonin onset (DLMO) at 1:30 AM (**A**) and an individual with advanced sleep-wake phase disorder with at DLMO at 7 PM (**B**). The *arrow* indicates the DLMO calculated as 2 standard deviations above the mean of baseline.

Chronotherapy

The use of chronotherapy has been successful in a small group of patients in a laboratory setting.[40] Chronotherapy requires a successive delay of sleep times by 3 hours daily over a 5- to 6-day period until the desired sleep time is achieved. This shift is followed by rigid adherence to a set sleep-wake schedule and good sleep hygiene practices. However, outside the laboratory setting, the potential confounding effects of light exposure at the wrong circadian time may limit the effectiveness and practicality of this approach.[6] In addition, caution must be maintained because there are case reports of individuals with severe DSWPD developing a non-24-hour sleep-wake pattern following chronotherapy.[85,86]

Light

As described earlier, light plays a major role in resetting the human circadian pacemaker.[17,18,87] Therapy with bright light in the morning (the advance portion of the human PRC) should advance the phase of circadian rhythms in DSWPD and may be more practical than chronotherapy.[88,89] Although a number of reports of successful application of bright light therapy in DSWPD exist,[90-93] large, randomized, placebo-controlled studies to determine the intensity, duration, and overall effectiveness are still needed. Rosenthal and colleagues found that 2 hours of bright light exposure (2500 lux) in the morning, together with light restriction in the evening, successfully phase-advanced (by 1.4 hours) circadian rhythms of core body temperature and multiple sleep latencies in 20 patients chosen prospectively after meeting clinical criteria for DSWPD.[93] In contrast, a retrospective report from a referral sleep clinic found that only 7 of 20 patients with DSWPD treated with bright light alone were able to entrain reliably to a desired sleep schedule.

The human PRC to a single 3-hour bright light pulse suggests that a light pulse given slightly before the time of body temperature minimum will result in a maximal phase delay, whereas a pulse given slightly after the minimum will cause a maximal phase advance (each about 2 hours).[87] When light pulses over three successive cycles are used, larger shifts (4 to 7 hours) can be produced. Because body temperature minimum is not routinely measured clinically, light therapy is usually timed using sleep logs to estimate the patient's endogenous circadian phase. A light pulse of 1 to 2 hours' duration and between 2500 and 10,000 lux is usually administered toward the end of the sleep-wake cycle. Because the portion of the PRC at which the greatest phase advance can be achieved occurs during sleeping hours, light is usually given immediately on awakening in the morning, which results in a smaller phase advance. It should be noted, however, that in severely delayed individuals the sleep-wake cycle may not necessarily correlate with circadian phase; therefore early morning light could in theory be inadvertently given on the delay portion of the PRC, worsening the problem. Regenstein and Pavlova reported a patient who slept later after receiving light exposure at 6 AM.[44] Another factor that may limit the practicality of bright light treatment is that many individuals with DSWPD may find it difficult to wake in time for administration of bright light therapy.[44]

Although the timing, intensity, and duration of light exposure have not been fully established in clinical practice, there is sufficient evidence to support the efficacy of bright light therapy for the treatment of DSWPD. The practice parameters established by the American Academy of Sleep Medicine considered light therapy as a treatment guideline for DSWPD.[77]

Melatonin

Administration of exogenous melatonin also shifts the phase of the endogenous circadian clock. It should be noted that the PRC for melatonin is nearly opposite the PRC for light exposure: melatonin delays circadian rhythms when administered in the morning and advances them when administered in the afternoon or early evening.[31,33] In a randomized, double-blind, placebo-controlled crossover study of eight patients with DSWPD, 5 mg of melatonin administered at 10 PM resulted in a phase advance in all subjects, with a mean advance of sleep onset time of 82 minutes and of wake time of 117 minutes.[94] On stopping melatonin, all patients reverted to their previous sleep-wake cycle within 2 to 3 days. A study to determine the optimal timing and dose of melatonin to treat DSWPD indicates that melatonin given about 6 hours before

the DLMO resulted in the largest phase advance and that there was no significant difference in the effect of either a 0.3-mg or 3-mg dose.[95] A more recent PRC specifically examining the effects of lower doses of melatonin (0.5 mg) demonstrated maximal phase advances when administered 2 to 4 hours before DLMO.[96] The physiologic (phase-shifting) dose of melatonin is approximately 0.1 to 0.5 mg or one tenth to one fiftieth of commercially available preparations.[97] Side effects of melatonin at these dose ranges are minimal, although sedative effects occur at higher (80 mg) doses. Melatonin has also been used successfully in several recent studies in children or adolescents with a combination of delayed sleep phase and other comorbid conditions such as attention deficit disorder and neurodevelopmental disabilities.[98-100]

Although several studies have demonstrated the potential effectiveness of melatonin administered in the evening,[36,94,95,99,101,102] the relatively small number of clinical studies, together with the variability in dose and time of administration, have been limiting factors in the development of a standardized approach for treatment with melatonin. The most recent practice parameters established by the American Academy of Sleep Medicine considered melatonin as a treatment guideline for DSWPD.[77]

More recently, several studies have evaluated the use of a combination of bright light and melatonin. A combination of bright light on awakening and melatonin 3 mg taken 8 hours before bedtime resulted in significantly larger advances (1.04 hours) than either melatonin (0.72 hours) or light (0.31 hours) alone.[103] Using a combination of bright light (10,000 lux) for 30 to 45 minutes on awakening in conjunction with 3 mg of melatonin taken 8 hours before sleep midpoint resulted in significant improvement in daytime sleepiness, fatigue, and cognitive functioning.[104]

In summary, the clinical approach to a patient with DSWPD should initially include assessment of circadian sleep phase by sleep diary or actigraphy measures for a period of at least 7 days but ideally 14 days.[75,77] Behavioral interventions such as a structured sleep-wake schedule, good sleep hygiene practices, and avoidance of exposure to bright light in the evening should be prescribed for all patients.[77] In addition, exposure to bright light in the morning (1 to 2 hours shortly after awakening) or administration of melatonin in the evening (5 to 6 hours before habitual sleep time) can advance the timing of the sleep-wake cycle. It is important to note that melatonin has not been approved by the U.S. Food and Drug Administration for this indication.

ADVANCED SLEEP-WAKE PHASE DISORDER, ADVANCED SLEEP PHASE TYPE, ADVANCED SLEEP PHASE SYNDROME

Clinical Features

ASWPD is characterized by habitual and involuntary sleep and wake times that are usually more than 2 hours earlier than societal averages (see Figure 40-1). Sleep itself is normal for age. Individuals frequently complain of persistent and often irresistible sleepiness in the late afternoon or early evening, often preventing their participation in desired evening activities. Because their circadian drive for wakefulness begins to rise prematurely, they may complain of involuntary early morning awakening (2 to 5 AM), which occurs even if sleep onset is voluntarily delayed. Because of professional or social obligations, later bedtimes can lead to chronically insufficient sleep and excessive daytime sleepiness. In general, individuals with ASWPD have less difficulty adjusting to the earlier schedule than those with DSWPD because societal constraints on sleep time are less rigid than on wake time. Individuals with ASWPD may gravitate to professions that are in phase with their endogenous circadian clock. Because of the symptoms of early morning awakening, a diagnosis of depression may be erroneously made.

Epidemiology

ASWPD is less frequently reported than DSWPD, possibly because affected individuals may not perceive it to be pathologic. The actual prevalence of ASWPD in the general population is unknown but is thought to be rare.[47,105] The prevalence in middle-aged adults has been estimated to be about 1%,[50] and it increases with age. Both genders appear to be equally affected.

Pathogenesis

As with DSWPD, the etiology of ASWPD is not well understood, although several hypotheses have been proposed. An abnormal PRC to light exhibiting an increased area under the advance portion of the curve could in theory result in a persistent phase advance. ASWPD may also be the result of a shortened endogenous period. Evidence of a shortened free-running period of less than 24 hours has been demonstrated in a 66-year-old woman with advanced sleep and wake times, with intact or even enhanced responsiveness to photic entrainment,[17] and in a single member of a familial case of ASWPD.[106]

Several familial cases of ASWPD have been reported in the literature.[106-109] These families show a clear autosomal dominant mode of inheritance of ASWPD. Genetic analysis of these familial cases indicates that there is heterogeneity of this disorder between and even within large families. Gene polymorphisms have been identified in the circadian clock gene *hPer2* in a large family with advanced sleep phase,[110] and in another family a missense mutation in CKI-δ has been reported.[111] More recently, several studies of adults born prematurely have demonstrated significant advances in the sleep-wake cycle in preterm infants with low birth weight.[112,113]

Diagnosis

A diagnosis of ASWPD is made primarily on the basis of the clinical history. Individuals have a chronic or recurrent complaint of difficulty staying awake until the desired time in the evening and inability to remain asleep until the desired and socially acceptable time for awakening. The sleep disturbance is associated with impairment of social, occupational, or other areas of functioning.[74] When allowed to choose their preferred schedule, patients will exhibit normal sleep quality and duration for age and maintain an advanced but stable phase of entrainment to the 24-hour day. Sleep log or actigraphy monitoring for at least 7 days and preferably 14 days should be performed to confirm a stable advance in the timing of the habitual sleep period.[74] The use of actigraphy, if available, is frequently helpful (see Figure 40-2).

Other medical, mental, or sleep disorders that may cause alterations in the sleep-wake cycle, insomnia, or excessive

sleepiness should be excluded or adequately treated. Major affective disorders should be carefully excluded. PSG is not required for the diagnosis, but in some patients, it may be necessary to evaluate for sleep-disordered breathing, periodic limb movements, or other causes of sleep disruption. PSG should ideally be performed during the patient's normal sleep period. If it is carried out at conventional laboratory hours, a shortened or normal sleep-onset latency, early REM, and early wake time may be seen. Therefore it is important to consider depression, narcolepsy, or other disorders in the differential diagnosis.[105,114]

When the diagnosis is in question, additional physiologic measures of circadian timing, such as continuous ambulatory monitoring of body temperature or collection of salivary melatonin samples to determine DLMO (see Figure 40-3), may be clinically useful to confirm the advance in circadian phase. Patients with ASWPD have been reported to have a DLMO[107,108] and core body temperature minimum[106] that are advanced several hours compared with controls. A morningness-eveningness scale, such as the Horne Ostberg or Munich chronotype questionnaire to gauge the patient's best time of performance, is also useful.[41,115]

Treatment

There are several therapeutic approaches used to treat ASWPD, each with practical limitations. A chronotherapeutic approach—advancing bedtime by 3 hours every 2 days until the desired bedtime was reached—has been reported,[105] although relapse occurred quickly.[114] Bright light therapy during the delay portion of the PRC (early evening) is usually tried, although data on its efficacy in ASWPD are limited. Bright light from 7 to 9 PM in elderly subjects with sleep maintenance complaints resulted in a phase delay and reduced awakenings.[116] The recent practice parameters established by the American Academy of Sleep Medicine considered bright light as a treatment option for ASWPD.[77]

Melatonin given in the early morning, usually on awakening, could in theory result in a phase delay according to data from the PRC to melatonin.[31,97] However, evidence for its effectiveness or safety in the treatment of patients with ASWPD is generally lacking.[78,117] It should be noted that the sedating effects of melatonin, which can be variable in patients, may limit its usefulness in this regard.

NON–24-HOUR SLEEP-WAKE DISORDER, FREE-RUNNING DISORDER, NONENTRAINED TYPE, HYPERNYCHTHERMAL SYNDROME

Clinical Features

N24SWD is thought to be the result of a circadian pacemaker that has no stable phase relation to the 24-hour LD cycle. Because most individuals must maintain a regular sleep-wake schedule, the clinical picture is that of periodically recurring problems with sleep initiation, sleep maintenance, and rising as the circadian cycle of wakefulness and sleep propensity moves in and out of synchrony with a fixed sleep period time.[118] Without social constraints, sleep onset and wake times are often successively delayed each day (Figure 40-4). This is analogous to the free-running state created when all zeitgebers are removed.[85,119] Because the duration and quality of sleep depend on when it occurs in relation to the circadian cycle,[3] phase "jumps" between two physiologically permissive periods for sustained sleep can be observed.[120-122]

Epidemiology

N24SWD is rare in sighted people, occurring most often in totally blind individuals.[118,119,123-126] In one series, 50% of totally blind persons had free-running plasma melatonin rhythms[125]; in another, 73% were not entrained to a 24-hour sleep-wake rhythm.[127] In a large study of 1073 blind individuals, N24SWD was estimated to occur in 18% of those who were totally blind and 13% of those who were almost blind.[128] The circadian pattern of melatonin production was analyzed in a large cohort of blind women, and in those without light perception only 37% were normally entrained, whereas among those with light perception 69% were normally entrained.[129] A few cases of free-running disorder have also been reported in sighted individuals.[85,130,131]

Pathogenesis

The etiology in blind people is most likely either reduced or absence of the entraining effects of light, although nonphotic time cues, such as an externally imposed 24-hour sleep-wake cycle and social activity, also appear insufficient to entrain some individuals.[118] The melatonin rhythm may be damped[132] or nonexistent[133] or may be normal but delayed.[85,124] Coexistent mental disability, which could make it difficult to process

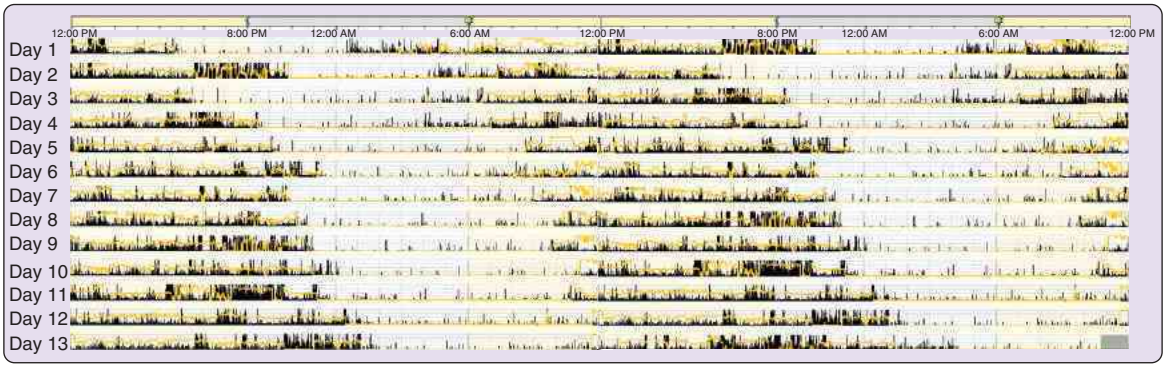


Figure 40-4 Rest-activity cycles recorded with wrist-activity monitoring of a sighted individual with non–24 hour sleep-wake disorder. The *black bars* indicate activity levels recorded at the nondominant wrist. Note that sleep onset is later on each consecutive day and that sleep is initiated from anytime between 5 PM and 1 AM for the 13 days represented in this actogram.

social time cues, may contribute to the symptoms in some individuals.[123] In some cases, totally blind persons without conscious perception of light nevertheless exhibit normal suppression of melatonin when exposed to very bright light and do not appear to have sleep difficulties.[127]

The etiology in sighted individuals is unknown. It has been postulated that sighted individuals may have a reduced sensitivity to the phase-resetting effects of light[85] and may have an increased incidence of psychiatric conditions[131] such as depression or certain personality disorders, which could precipitate the development of the syndrome by changing or removing social time cues.[85] In addition a recent study of six sighted individuals with N24SWD demonstrated that the intrinsic period was significantly longer than in controls.[134]

Nonentrained sleep-wake cycles have developed after chronotherapy for apparent delayed sleep phase type,[85,86] prompting the proposal that such therapy could prolong the free-running period to the point at which it becomes nonentrainable to a 24-hour LD cycle.[86] However, free-running periods, which are too short (<23 hours) or too long (>27 hours)[18] for stable entrainment to a 24-hour cycle, have never been demonstrated in humans. Because persons with DSWPD tend to receive more light exposure in the delay (evening) portion of their PRCs than during the advance portion (early morning), progressive phase delays may sometimes be observed and mistaken for a nonentrained pattern.[6,85]

Diagnosis

The diagnosis of N24SWD is made primarily by a clinical history of insomnia or excessive sleepiness related to abnormal synchronization between the 24-hour LD cycle and the endogenous circadian rhythm of sleep and wake propensity.[74] The pattern of sleep and wake times typically delay each day with a period longer than 24 hours. The pattern is present for at least 1 month and can be confirmed by sleep diary or actigraphy for at least 7 days (but preferably longer) to establish the progressive daily drift in the timing of sleep and wake times (see Figure 40-4). Close analysis of the sleep-wake cycle may reveal two distinct sleep-wake cycle periods, alternation between which can be manifested by phase jumps.[121,122] The sleep disturbance should be accompanied by impairment of social, occupational, or other areas of functioning.

It is important to exclude or adequately treat other medical, mental, or sleep disorders that may cause alterations in the sleep-wake cycle, insomnia, or excessive sleepiness. If the diagnosis is in question, PSG may be useful to evaluate for other types of sleep disorders. PSG, when performed at the appropriate circadian time, is usually normal.[85] Overriding behavioral factors predisposing to irregular sleep-wake cycles (substance abuse, dementia, personality, or affective disorders) should also be considered in the evaluation.

Treatment

Melatonin receptor agonists have been used as the initial treatment of choice in blind and sighted individuals with N24SWD.[29,85,119,123,124,126,135-138] For example, administration of melatonin is started when the patient's free-running period approaches the normal or desired phase (i.e., sleep onset times of 10 to 11 PM). Doses sufficient for phase shifting (0.1 to 0.5 mg) are then given at 8 to 9 PM, or near the expected time of DLMO.[85,126] Initiating evening dosing when the free-running period is not in the "normal" phase could result in an inappropriate delay or advance of circadian phase and prolong the time to entrainment. More recently the melatonin agonist tasimelteon (20 mg) has been approved for use in the treatment of N24SWD, given in a similar manner as melatonin at a fixed time every night.[139] Common side effects include sedation, headache, elevated liver enzymes, and nightmares. Bright light entrainment is an option in sighted individuals or in blind individuals who exhibit intact photic suppression of melatonin.[140] Vitamin B_{12} has been anecdotally reported to be effective in treating nonentrained type,[141,142] although the mechanism is unknown. Benzodiazepines have not been systematically studied, although several anecdotal reports of their possible effectiveness exist.[141,142] Entrainment by nonphotic stimuli (e.g., structured social cues) alone has not been successful. The recent practice parameters established by the American Academy of Sleep Medicine considered bright light and melatonin administration as options for the treatment of N24SWD and in blind individuals recommended melatonin administration as a treatment guideline.[77]

IRREGULAR SLEEP-WAKE RHYTHM DISORDER, IRREGULAR SLEEP-WAKE TYPE, IRREGULAR SLEEP-WAKE DISORDER

Clinical Features

ISWRD is characterized by the absence of a well-defined circadian sleep-wake cycle. There is typically no major sleep period. Rather, patients present with three or more sleep episodes of varying length during a 24-hour period. Diagnosis of this disorder requires a complaint of insomnia or excessive sleepiness associated with multiple irregular sleep bouts or naps during a 24-hour period.[74] Despite irregular and fragmented sleep periods, the total sleep time per 24-hour period is usually normal for age.

Epidemiology

The prevalence of ISWRD in the general population is unknown, but it is estimated to be rare.[143] There is no evidence of gender differences. ISWRD is most commonly seen in association with age related comorbid neurologic disorders such as dementia and Alzheimer disease.[78] It is also seen in patients with brain injury and in children with mental disability.

Pathogenesis

ISWRD is thought to result from dysfunction of the central processes responsible for the generation of circadian rhythms[144,145] or reduced exposure to bright light and regular social schedules. A recent case report demonstrated decreased amplitude of core body temperature rhythmicity in two individuals with ISWRD.[146] In institutionalized elderly patients and those with dementia, these factors are thought to contribute to the development and maintenance of irregular sleep-wake patterns.[147,148] Even when controlling for the level of dementia, lower daytime light levels have been associated with an increase in nighttime awakenings.[149] Evidence suggests that both dysfunction of circadian regulation and reduced exposure to environmental signals are likely involved in the etiology of irregular or arrhythmic sleep and wake patterns.

Diagnosis

In addition to a clinical history, continuous monitoring of sleep and wake activity with actigraphy or a sleep log for a

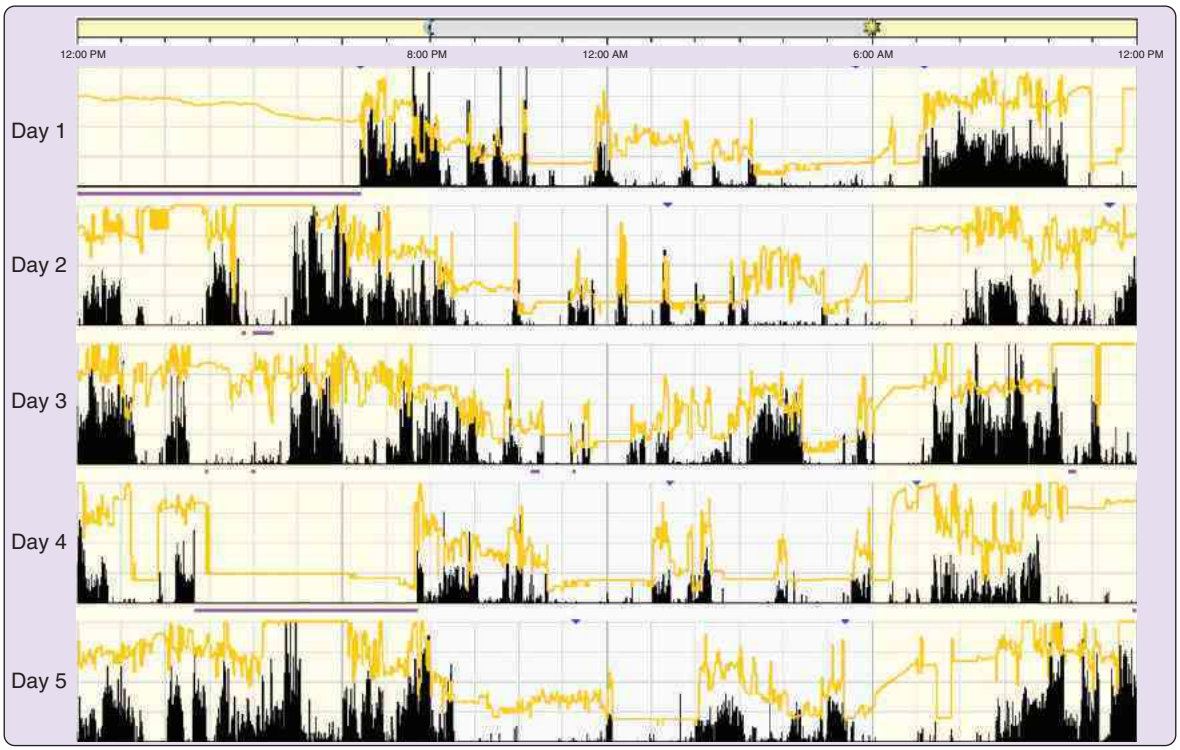


Figure 40-5 Rest-activity cycle recorded with wrist-activity monitoring of an older adult with irregular sleep-wake rhythm disorder. The *black bars* indicate activity levels, and the *yellow bars* indicate the level of ambient light exposure recorded at the nondominant wrist. Note the lack of a discernible circadian sleep and wake rhythm. Sleep is characterized by nocturnal fragmentation and multiple short periods of sleep and wake across the entire 24-hour day.

minimum of 2 weeks is useful diagnostically. Actigraphic recordings show disturbed or low-amplitude circadian rhythm and loss of the normal diurnal sleep-wake pattern with at least three distinct sleep periods per 24 hours (Figure 40-5). This disturbed sleep-wake pattern should be associated with a chronic complaint of insomnia (usually sleep maintenance) and excessive daytime sleepiness. ISWRD should be distinguished from poor sleep hygiene and voluntary maintenance of irregular sleep schedules as seen with shift work.[74]

Treatment

Clinical management aims to improve the amplitude of circadian rhythms and their alignment with the external environment. Increasing exposure to synchronizing agents, such as bright light,[150-153] and structured social and physical activities[154] have been used to consolidate sleep-wake cycles. Mixed-modality therapies combining bright light exposure, physical activity, and other behavioral elements are indicated in the treatment of ISWRD, as outlined in the recent parameters established by the American Academy of Sleep Medicine.[77] Recent randomized controlled studies using mixed-modality therapies have shown increases in the robustness of the rest-activity rhythm and reductions in nighttime awakening following treatments that include increasing light levels, increasing light in combination with evening melatonin administration, implementing measures to keep nursing home residents out of bed during the day, structuring physical activity, instituting a bedtime routine, and taking measures to reduce nighttime noise and light in residents' rooms.[155-158] Another study using a combination of bright light, chronotherapy, vitamin B_{12}, and hypnotics was successful in 45% of patients with irregular sleep cycles.[143]

Evening melatonin administration has been used successfully to improve disturbed sleep-wake patterns in children with mental disability.[159-161] In this population melatonin is recommended as a treatment option in the recent practice parameters established by the American Academy of Sleep Medicine.[77] Despite the potential utility of both behavioral and pharmacologic interventions, treatment may be difficult and outcomes variable.

CLINICAL PEARLS

Circadian rhythm sleep disorders should be considered in the differential diagnosis of patients presenting with symptoms of insomnia or excessive sleepiness. Furthermore, CRSWDs, such as DSWPD, may be comorbid with other types of sleep disorders, making the diagnosis and treatment even more challenging. For effective management of CRSWD, one must obtain as accurate a measure of circadian timing as possible. Sleep-onset time (determined by sleep diary or actigraphy) can be useful to determine circadian phase (DLMO occurs approximately 2 hours before sleep onset) in the clinical setting.

Behavioral interventions such as sleep hygiene, particularly enforcement of stable sleep and wake times, exposure to light at the correct time of the day, and avoidance of exposure at the wrong time of the day is the basic approach for all patients. For the treatment of DSWPD and N24SWD, the use of melatonin may be useful (see specific sections for details). However, the use of melatonin for the treatment of CRSWD has not been approved by the U.S. Food and Drug Administration, and vascular and endocrine adverse effects need to be taken into account, particularly in patients who are at increased risk.

SUMMARY

Disorders of the sleep-wake cycle attributed to the disruption of the circadian timing system are characterized by an abnormal temporal distribution of the major sleep period within the 24-hour day. Although there is evidence that many of these disorders are the result of alterations in the circadian clock, more studies are needed to confirm this theory. The effect of these disorders is probably larger than estimated in terms of numbers, misdiagnoses, and health consequences. Most sleep clinics do not yet provide specific diagnostic tools to assess circadian rhythm profiles. Furthermore, many of the proposed diagnostic tools and therapies, including light, are often considered experimental by the health insurance industry. Application of our expanding knowledge of basic human circadian and sleep physiology to clinical practice remains an important challenge.

ACKNOWLEDGMENTS

We acknowledge the contribution of the late Steven K. Baker to the original version of this chapter. Support for this manuscript was provided by RO1HL69988 and 5P01AG011412-17.

Selected Readings

Burgess HJ, Revell VL, Eastman CI. A three pulse phase response curve to three milligrams of melatonin in humans. *J Physiol* 2008;**586**(2):639–47.

Czeisler CA, Allan JS, Strogatz SH, et al. Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. *Science* 1986;**233**:667–71.

Flynn-Evans EE, Tabandeh H, Skene DJ, Lockley SW. Circadian rhythm disorders and melatonin production in 127 blind women with and without light perception. *J Biol Rhythms* 2014;**29**:215–24.

Freedman MS, Lucas RJ, Soni B, et al. Regulation of mammalian circadian behavior by non-rod, non-cone, ocular photoreceptors. *Science* 1999;**284**:502–4.

Gandhi AV, Mosser EA, Oikonomou G, Prober DA. Melatonin is required for the circadian regulation of sleep. *Neuron* 2015;**85**:1193–9.

Jones CR, Campbell SS, Zone SE, et al. Familial advanced sleep-phase syndrome: a short-period circadian rhythm variant in humans. *Nat Med* 1999;**5**:1062–5.

Kitamura S, Hida A, Enomoto M, et al. Intrinsic circadian period of sighted patients with circadian rhythm sleep disorder, free-running type. *Biol Psychiatry* 2013;**73**(1):63–9.

Minors DS, Waterhouse JM, Wirz-Justice A. A human phase-response curve to light. *Neurosci Lett* 1991;**133**(1):36–40.

Morgenthaler TI, Lee-Chiong T, Alessi C, et al. Practice parameters for the clinical evaluation and treatment of circadian rhythm sleep disorders: an American Academy of Sleep Medicine report. *Sleep* 2007;**30**:1445–59.

Neubauer DN. Tasimelteon for the treatment of non-24-hour sleep-wake disorder. *Drugs Today (Barc)* 2015;**51**:29–35.

Sack RL, Auckley D, Auger RR, et al. Circadian rhythm sleep disorders: part II, advanced sleep phase disorder, delayed sleep phase disorder, free-running disorder, and irregular sleep-wake rhythm. An American Academy of Sleep Medicine review. *Sleep* 2007;**30**:1484–501.

Sack RL, Lewy AJ, Blood ML, et al. Melatonin administration to blind people: Phase advances and entrainment. *J Biol Rhythms* 1991;**6**:249–61.

Wilhelmsen-Langeland A, Saxvig IW, Pallesen S, et al. A randomized controlled trial with bright light and melatonin for the treatment of delayed sleep phase disorder: effects on subjective and objective sleepiness and cognitive function. *J Biol Rhythms* 2013;**28**:306–21.

A complete reference list can be found online at ExpertConsult.com.

Section 6

Pharmacology

Chapter 41

Hypnotic Medications: Mechanisms of Action and Pharmacologic Effects

Thomas S. Kilduff; Wallace B. Mendelson

Chapter Highlights

- Most currently available hypnotics induce sleep by acting at various moieties of the gamma-aminobutyric acid A ($GABA_A$)–benzodiazepine receptor complex. This receptor is a member of the ligand-gated receptor family and is composed of several subunits, each of which appears in multiple isoforms.
- Microinjection studies indicate that the neuroanatomic sites of action of hypnotics from a variety of pharmacologic classes include the preoptic area and specific hypothalamic nuclei. These brain areas express GABA receptors and also contain GABAergic neurons.
- Connections between the preoptic area and the forebrain and brainstem and the integrative nature of the anterior hypothalamus help explain the mechanism by which sleep—and drugs that affect sleep—interact with a variety of physiologic systems.
- Novel hypnotic agents directed toward the hypocretin-orexin receptors have recently become available for clinical use.

Approximately 10% to 15% of the U.S. population suffer from chronic insomnia.[1] The clinically used sedative-hypnotics in the United States currently include five benzodiazepines available in both proprietary and generic forms, three nonbenzodiazepine gamma-aminobutyric acid (GABA) agonists (zolpidem, zaleplon, and eszopiclone), and two melatonin receptor agonists (ramelteon) available as prescription agents (Table 41-1). This market resulted in more than 60 million prescriptions in 2011, a growth of 20% over the previous 5 years,[2] and approximately $3.7 billion in sales.[3] During the 15-year period from 1994 to 2007, prescriptions for nonbenzodiazepine hypnotics increased 30-fold.[4] The most widely prescribed medication for insomnia has typically been the antidepressant trazodone.[5-7]

Although these are large numbers, one could make a case that, in the population at large, prescription hypnotics are taken relatively infrequently by insomniacs. In a survey of more than 7000 enrollees in five large health maintenance organizations,[8] patients with insomnia were divided into those with only sleep complaints (Level 1) and those who believed that their sleep difficulties had an adverse effect on their daytime functioning (Level 2). The percentage of Level 1 insomniac patients taking prescription and nonprescription hypnotics was 5.5% and 11.2%; comparable values for Level 2 insomniac patients were 11.6% and 21.4%, respectively. Many patients, of course, receive other classes of prescription drugs given for purposes of helping sleep. One study of primary care patients in a health maintenance organization indicated that 13% of insomniacs who were not considered to have affective disorder were receiving antidepressant medications such as trazodone.[9] Many people self-medicate with alcohol; a telephone survey of approximately 1000 adults

Table 41-1 Pharmacokinetic Properties and Dosages of Some Hypnotic Drugs Used in the Treatment of Insomnia

	Half-Life (h)	T_{MAX} (h)*	Pharmacologically Active Metabolites	Dose (mg)
Benzodiazepine Hypnotics[†]				
Quazepam (Doral)	48–120	2–3	*N*-desalkyl (flurazepam)	7.5–15
Flurazepam (Dalmane)	48–120	1.5–4.5	*N*-desalkyl (flurazepam)	15–30
Triazolam (Halcion)	2–6	1–2	None	0.125–0.25
Estazolam (ProSom)	8–24	1.5–2	None	1–2
Temazepam[‡] (Restoril)	8–20	1–2	None	15–30
Loprazolam[§] (Dormonoct)	4.6–11.4		None	1–2
Flunitrazepam[§] (Rohypnol)	10.7–20.3		*N*-desmethyl (flunitrazepam)	0.5–1
Lormetazepam[§] (Loramet)	7.9–11.4		None	1–2 (Elderly: 0.5–1)
Nitrazepam[§] (Alodorm)	25–35		None	5–10
Nonbenzodiazepine Hypnotics				
Zolpidem: Oral tablet	2.5 (1.4–4.5)	1.6 (0.5–1.5)	None	5 (age >65 yrs) 5–10 (age <65 yr)
Zolpidem: Extended release (Ambien CR)	2.8 (1.6–4.5)	1.5 (1.5–2.0)	None	6.25–12.5
Zolpidem: Sublingual (Intermezzo)	2.5 (1.4–3.6)	0.6 (0.6–1.3)		women: 1.75; men: 3.5
Zolpidem: Sublingual (Edluar)	2.7 (1.5–6.7)	1.4 (0.5–3.0)		5–10
Zolpidem: Oral spray (Zolpimist)	2.8 (1.7–8.4)	0.9		10
Zopiclone[§] (Imovane)	5–6		None	3.75 (age >65 yr) 7.5 (age <65 yrs)
Zaleplon (Sonata)	1 (0.8–1.3)	1 (0.5–2)	None	5–10
Eszopiclone (Lunesta)	6 (5–8)	1.5 (0.5–2)	None	2–3 (age <65 yr); 1–2 (age >65 yr)
Ramelteon	1–2.6[¶]	0.75 (0.5–1.5)	M-II	8
Doxepin	15 (10–30)	3.5 (1.5–4)		3–6 (Silenor) 10–50[**] (generic)
Suvorexant	12	0.5–6.0	None	10–20
Nonhypnotics Sometimes Used to Aid Sleep[‖]				
Clonazepam (Klonopin)	20–40	1–2.5	4-Amino derivative	0.5–1[**]
Diazepam (Valium)	30–100		*N*-desmethyl	2–10[**]
Chlordiazepoxide (Librium)	24–28		*N*-desmethyl (chlordiazepoxide, demoxepam, oxazepam)	10–25[**]
Alprazolam	6–20	0.6–1.4		0.5–1[**]
Lorazepam	10–20	0.7–1		0.5–1[**]
Quetiapine (Seroquel)	6	1–2	Quetiapine sulfoxide	25–50[**]
Trazodone (Desyrel)	9 (7–15)	1–2	*m*-CPP	25–50[**]

*Data from Smith CM, Reynard AM. *Essentials of pharmacology*. Philadelphia: Saunders; 1995. p. 228; Buysse DJ. Insomnia. *JAMA* 2013;309:706–16; and other sources.

[†]Citations for kinetic information of benzodiazepines are found in Maczaj.[72]

[‡]Originally formulated as a hard capsule in the United States, concerns with kinetics and efficacy led to reformulation of the preparation to a soft gelatin capsule with comparable characteristics of other marketed benzodiazepines of its class.[73,74]

[§]Not available in the United States.

[¶]Active metabolite has half-life of 2 to 5 hours.

[‖]Drugs that do not have U.S. Food and Drug Administration (FDA) indications for aiding sleep.

**Because there is not an FDA indication for sleep, there is no FDA recommended dose for this purpose. Doses stated here are approximations of those often used in clinical practice.

m-CPP, m-Chlorophenylpiperazine.

from the general population found that 10% had done so in the past year.[10] A survey in the Detroit area reported that 13% had used alcohol to aid sleep during the past year, whereas 18% used medication (either prescription or over-the-counter), and 5% had used both.[11] Thus alcohol use to help sleep is common, even though it exposes the patient to the risk for ethanol dependence and is relatively ineffective. Although orally administered ethanol has some sleep-inducing properties, it often promotes sleep disturbance as the night progresses.

MECHANISM OF ACTION

$GABA_A$ Agonists: Benzodiazepines and Nonbenzodiazepines

GABA is the main inhibitory neurotransmitter and is produced by 15% to 20% of all neurons. GABA exerts its effects within the central nervous system (CNS) by acting at two types of receptors. The $GABA_A$ receptor, often referred to as the $GABA_A$-benzodiazepine receptor complex, is part of the ligand-gated ion channel family that also includes the glycine, $5HT_3$, neuronal nicotinic acetylcholine, and other receptors.[12] In contrast, the $GABA_B$ receptor is a member of the G protein-coupled receptor family. Although some hypnotic development activity has been directed toward the $GABA_B$ receptor (e.g., gamma-hydroxybutyrate, used in the treatment of narcolepsy, acts through the $GABA_B$ receptor), the $GABA_A$–benzodiazepine receptor complex may be the most intensely studied receptor for all of neurotherapeutic development.

Since the late 1970s, the pharmacologic effects of GABA and the benzodiazepines have been recognized to result from saturable, stereospecific binding to a specific recognition site.[13,14] The $GABA_A$-benzodiazepine receptor is a pentameric transmembrane structure that is composed of combinations of different subunits that, together, form a benzodiazepine recognition site, a $GABA_A$ recognition site, and a chloride ionophore (Figure 41-1). In humans, there are six alpha, three beta, three gamma, three rho, and one each of the delta, epsilon, theta, and pi subunits.[15,16] The binding site for benzodiazepines appears to be located at the junction of the alpha and gamma subunits, whereas the GABA binding site is at the junction of the alpha and beta subunits. The alpha-1 subunit appears to mediate both sedation and memory effects of agonists.[17]

The most common pentameric subunit combinations are the type I (alpha-1, beta-2, gamma-2) and type II (alpha-3, beta-2, gamma-2), but the total number of receptor isoforms is unclear.[15] Type I represents roughly 40% of $GABA_A$ receptors and is found in most areas of the brain, including on interneurons of the hippocampus and cortex. Type II is perhaps half as common and is found in spinal cord motoneurons and hippocampal pyramidal neurons. Most traditional benzodiazepine hypnotics bind to both types. Some of the nonbenzodiazepine agents, including zolpidem and zaleplon, bind with relatively greater specificity to the type I receptor. This selective binding is thought to underlie specific pharmacologic properties. Indeed, the hypnotic effects of zolpidem have been attributed to the alpha-1 subunit.[18]

Benzodiazepine agonist binding to its specific site results in a net flux of negatively charged chloride ions through the chloride ionophore and hyperpolarization of membrane potential so that the neuron is less likely to produce an action potential. This inhibitory mechanism is very potent because the $GABA_A$ receptor is the most widespread receptor at inhibitory synapses, comprising up to 30% of all synapses in the CNS.

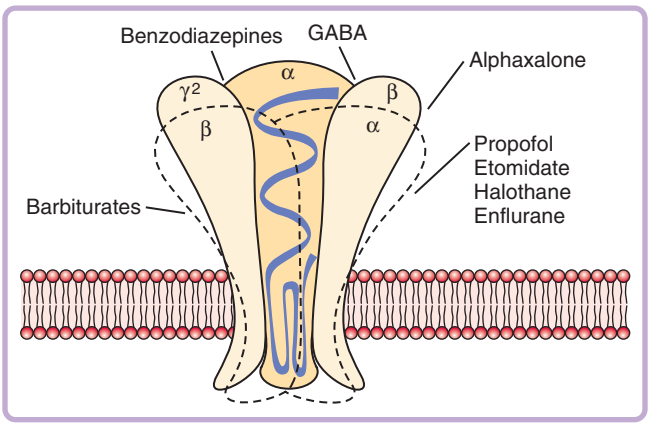


Figure 41-1 Schematic representation of the $GABA_A$-benzodiazepine receptor complex, and possible sites of action of sleep-inducing drugs. (From Rudolph U, Mohler H: Analysis of GABA-A receptor function and dissection of the pharmacology of benzodiazepines and general anesthetics through mouse genetics. *Annu Rev Pharmacol Toxicol* 2004;44:475–98.)

The original characterization of the BZ receptor complex indicated that it mediates the anxiolytic, muscle relaxant and anticonvulsant effects of benzodiazepines.[13,14] A series of studies also showed that this receptor complex mediates sleep-inducing effects. Some β-carboline compounds that act as BZ receptor inverse agonists increase wakefulness and, at low doses, block sleep induction by benzodiazepines.[19] The binding of benzodiazepines is stereospecific, such that an enantiomeric form (the B_{10} compounds) have opposite effects, in which one enantiomer induces sleep whereas the other promotes wakefulness.[19]

Studies of the $GABA_A$-benzodiazepine receptor complex have provided insight into a long-standing mystery—how medications of many different pharmacologic classes can produce sleep: nonbenzodiazepine agents such as zolpidem and zaleplon bind to a subclass of benzodiazepine recognition sites, ethanol has profound effects on chloride channel function, and barbiturates bind to yet another site on the receptor complex. Barbiturates can cause chloride channels to open for prolonged periods,[20] whereas benzodiazepines may increase the frequency of opening.[21] Similarly, the active metabolite of chloral hydrate[12] and the anesthetic propofol[22] modulate $GABA_A$ receptor function. Several lines of evidence have also suggested that the hypnotic effects of benzodiazepines may involve presynaptic effects mediated by voltage-dependent calcium channels. The calcium-channel blocker nifedipine, for instance, can block the sleep-inducing effects of microinjections of triazolam into the medial preoptic area,[23] suggesting that modulation of voltage-gated calcium channels may be another mechanism to induce sleep.

Neuroanatomic Substrates for the Effects of Hypnotic Medications

In contrast to our understanding of the interaction of benzodiazepine agonists at a molecular level, the anatomic sites at which they act to induce sleep are less well understood. The

most parsimonious hypothesis would posit that hypnotics act at loci thought to be important in sleep regulation. Accordingly, when the BZ triazolam was microinjected into many areas (e.g., the locus coeruleus, the gigantocellular tegmental fields, the basomedial nucleus of the amygdala, and the ventrolateral preoptic area), there was either no effect on sleep or increased wakefulness (the dorsal raphe nuclei).[24] In contrast, microinjections into the medial preoptic area (MPA) consistently enhanced sleep. These results were anatomically specific because injections into nearby structures (the lateral preoptic area, the horizontal limb of the diagonal band of Broca) had no effect.

The basal forebrain, MPA, and anterior hypothalamus have long been thought to have an important role in sleep regulation.[25-27] The MPA is a complex structure that receives afferents from many forebrain and brainstem regions. Among these are projections from various areas of the hypothalamus, as well as serotonergic fibers from the dorsal raphe nuclei and noradrenergic projections from the locus coeruleus.[28] GABA is ubiquitous throughout the hypothalamus, and its synthetic enzyme is found in high concentrations in the preoptic area.[29] $GABA_A$-benzodiazepine receptors occur in significant concentrations,[30] and the expression of different receptor subunits are under homeostatic as well as circadian regulation.[31] The preoptic area also contains cells that are responsive to temperature, osmolarity, glucose, and steroids,[32] and receives afferents from various sensory systems.[28] Neurons in both the median preoptic[33,34] and ventrolateral preoptic area[35,36] have elevated discharge rates and accumulate Fos during sleep; basal forebrain neurons also increase firing during non–rapid eye movement (NREM) sleep compared with waking ("sleep-active neurons").[37] Neurons in these regions are GABAergic and inhibit posterior hypothalamic wake-promoting areas.[38,39] The preoptic area and the MPA, in particular, appear to have a role in coordinating various systems involved in reproductive and homeostatic functions.[28] Given the interactions of sleep with cardiovascular, thermoregulatory, endocrine, and sensory systems, the preoptic area is likely involved in integrating sleep with other regulated physiologic functions. Microinjection of triazolam into the perifornical hypothalamus reduces sleep latency and increases total sleep time in rats, raising the possibility that BZ hypnotics reduce wakefulness either directly through the perifornical area or indirectly through modulation of GABAergic outputs from the preoptic area.[40] Thus the preoptic area is likely among the loci at which benzodiazepine agonists act to induce sleep. Microinjections of pentobarbital and other GABA receptor agonists into the mesopontine tegmental area induced an anesthesia-like state, raising the possibility that this area may be involved in drug-induced loss of consciousness.[41]

The Hypocretin-Orexin System

Since the link was made between the sleep disorder narcolepsy and the hypocretin-orexin system in the late 1990s,[42-45] there has been considerable interest in this neuropeptidergic system as a mechanism of arousal. Hypocretin cell bodies are found in a very discrete region of the perifornical and lateral hypothalamus,[46,47] but these cells send widespread projections throughout the CNS[48,49] to innervate brain regions that express the orexin receptor-1 (OX1R) or orexin receptor-2 (OX2R).[50] Among these projection sites are wake-promoting monoaminergic and cholinergic groups in the brainstem, hypothalamus, and basal forebrain.[48,49] This system has been the subject of intense research focus in the sleep field in the past 15 years, as reviewed elsewhere.[51-55] Whereas microinjection of the hypocretin-orexin peptides promotes wakefulness in virtually every region studied,[56-59] receptor antagonists promote sleep.[60-64] Whether the sleep-inducing effects are mediated through OX2R[61,62,65] or through both receptors[61,63,64] has been controversial. The effects of orexin-A on wakefulness and NREM sleep were significantly attenuated in receptor knockout mice, with substantially larger attenuation in *OX2R*$^{-/-}$ mice than in *OX1R*$^{-/-}$ mice.[66] In mid-2014, the U.S. Food and Drug Administration (FDA) approved the dual orexin receptor antagonist suvorexant for treatment of insomnia.

Other Neurotransmitters

Microinjections of pentobarbital,[67] ethanol,[68] adenosine,[69] and propofol[70] into the MPA also enhance sleep. Because adenosine is known to accumulate extracellularly in the basal forebrain with prolonged wakefulness,[71] adenosinergic modulation of cortically-projecting basal forebrain cholinergic neurons may be the mechanism by which the propensity to sleep is enhanced by prolonged wakefulness.[72] Similarly, some of the sleep-promoting effects of antihistamines and tricyclic antidepressants (many of which have significant anticholinergic properties) may be mediated by alteration of basal forebrain cholinergic neuron activity. The wakefulness-promoting histaminergic system, whose cell bodies reside in the tuberomammillary nucleus of the posterior hypothalamus, sends efferents to the preoptic area[73] as well as to the perifornical area[74,75] and the cortex. Thus the sedative effects of antihistamines and some tricyclic antidepressants may be mediated by altering the influence of the tuberomammillary nuclei on these brain regions. Microinjection of the GABA receptor antagonist gabazine into the tuberomammillary nuclei inhibits the effects of centrally administered GABAergic agents, suggesting that this area may be involved in their pharmacologic actions.[76] Attention has recently been focused on the melanin-concentrating hormone neurons of the lateral hypothalamus and midbrain, most of which are GABAergic. Optogenetic stimulation of these cells either induces sleep with short latency or facilitates the transition from NREM to rapid eye movement (REM) sleep.[77-79]

CLINICAL EFFECTS

At the clinical level, less is understood about the mechanisms by which hypnotics act. One possibility is that hypnotics may also alter the *perception* of sleep and wakefulness.[80] This notion stems from the classic observation by Rechtschaffen that poor sleepers, when experimentally awakened early in stage 2 sleep, tend to report that they had been awake, whereas good sleepers tend to report that they had been asleep. Later studies[81,82] replicated this observation and demonstrated that, after administration of triazolam, flurazepam, or zolpidem, insomniacs were more likely to report that they believed they had been asleep compared with when they were given placebo. In contrast, this effect was not evident when flurazepam or zolpidem were given to normal subjects.[83] One interpretation of these data is that hypnotics such as triazolam and zolpidem may correct a misperception of sleep in some insomniac patients, such that their experience of whether they are awake or asleep becomes more like that of good sleepers.

Other Sleep-Inducing Agents

Most over-the-counter hypnotics are first-generation antihistamines such as diphenhydramine and doxylamine. Although they possess to varying degrees other properties, including anticholinergic effects, their common underlying mechanism is inhibition of the histamine-1 (H_1) receptor. Later generations of antihistamines such as fexofenadine enter the CNS less readily and thus produce less sedation.[84] Tolerance to daytime sleepiness appears to develop rapidly, in about 4 days.[85] As discussed earlier, antihistamines may produce drowsiness by inhibiting the wake-promoting histaminergic pathways originating in the tuberomammillary nucleus of the posterior hypothalamus. Some tricyclic antidepressants, notably doxepin, have significant antihistaminergic (H_1 and H_2) properties as well as effects on other neurotransmitter systems, including antagonism at α_1-adrenoreceptors and muscarinic cholinergic receptors as well as binding to serotonin 2a and 2c receptor subtypes. However, these compounds were not developed for insomnia, have long half-lives, and consequently, result in daytime sedation

Although there is debate about the efficacy and range of side effects of melatonin as a hypnotic,[86] agonists of melatonin MT1 and MT2 receptor subtypes are on the market in the United States (ramelteon and tasimelteon), and agomelatine has been approved in Europe. Melatonin and melatonin agonists are thought to affect the circadian system through melatonin receptors in the suprachiasmatic nucleus, but the mechanism by which this would alter sleep is not well understood. It has been suggested that binding of agonists to the melatonin MT1 receptor decreases the waking signal from the suprachiasmatic nucleus. One possibility is that sleep-promoting effects of agonists at melatonin receptors are at least in part GABAergic. Melatonin administration is known to raise GABA concentrations in the rat hypothalamus as well as 3H-diazepam binding in the forebrain.[87] Similarly, decreases in motor activity produced by melatonin in the hamster are prevented by the benzodiazepine receptor blocker flumazenil.[88] Microinjections of melatonin into the MPA have been reported to enhance sleep in the rat, suggesting that its site of action may be similar to that of benzodiazepines, barbiturates, adenosine, and ethanol, as described previously.[89]

Although the topic of anesthetics is beyond the scope of this chapter, we will briefly mention the intravenous agent propofol, which has made it much more practicable to induce anesthesia for prolonged periods in intensive care unit settings. Although its neurochemical mechanism of action has not been fully elucidated, propofol is known to interact with the $GABA_A$-benzodiazepine receptor complex, with resultant decreases in acetylcholine release from the frontal cortex and hippocampus, and to also increase functional activity of dopamine and serotonin in the cortex. Interestingly, microinjection of propofol into the MPA of rats induces sleep, and this effect is blocked by the benzodiazepine receptor blocker flumazenil.[70]

PHARMACOKINETICS

Benzodiazepines

With the possible exception of temazepam, the benzodiazepines used as hypnotics are rapidly and completely absorbed, with most achieving peak plasma levels in 1 to 1.5 hours. Some, notably flurazepam, are detectable primarily in the form of their active metabolites, and many display kinetics that reflect enterohepatic circulation. Generally, oral administration is more reliable and complete than intramuscular injection. Although there is significant protein binding, there are few or no cases in which displacement of other protein-bound drugs is clinically relevant. Most are highly lipophilic and rapidly enter the CNS where concentrations reflect unbound drug in plasma. The older, longer acting agents are metabolized to active compounds, and the shorter-acting agents such as triazolam are broken down into inactive substances (see Table 41-1; Video 41-1). The elimination half-lives vary widely, from the relatively short-acting triazolam, to intermediate agents such as temazepam, to long-acting substances such as flurazepam (see Table 41-1). Accumulation of the longer elimination half-life agents when taken nightly has significant bearing on a major clinical issue—the appearance of daytime residual sedation (Figure 41-2). Because of the lipophilicity, many are rapidly redistributed, which may play an important role in the decline in CNS effects as they are metabolized. Another implication of the high lipophilicity is that the volume of distribution is often increased in older adults (who tend to have a higher ratio of lipid to muscle), resulting in an increased half-life. Most are broken down by hepatic microsomal systems and excreted as conjugated glucuronides. There is no stimulation of the hepatic microsomal systems and hence no enhancement of the rate of breakdown of other drugs that undergo the same metabolic processes. Metabolism of these compounds may be inhibited by use of other drugs such as cimetidine and some steroids and may be accelerated in people who smoke.

The Z Drugs

Zolpidem was the first of the "Z drugs" (zolpidem, zaleplon, zopiclone, eszopiclone), nonbenzodiazepines that bind to various subtypes of the $GABA_A$ receptor. In general, their clinical effects are similar to those of benzodiazepines; however, on the basis of animal studies, it has been argued that they have a wider separation in doses producing hypnotic

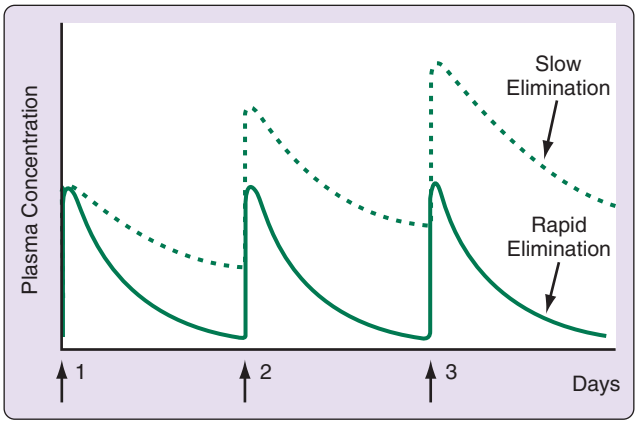


Figure 41-2 A hypnotic with a relatively long elimination half-life (e.g., >24 hours) will accumulate during nightly use, in contrast to an agent with relatively short elimination half-life (e.g., 6 hours). (From Nicholson AN: Hypnotics: clinical pharmacology and therapeutics. In: Kryger MH, Roth T, Dement WC, editors. *Principles and practice of sleep medicine*. 2nd ed. Philadelphia: Saunders; 1989. p. 355–63.)

to nonhypnotic effects.[90] Several studies have suggested that they have a lower risk for dependence and misuse than the benzodiazepines. A German prescription event study, for instance, described a rate of misuse about one third that of benzodiazepines.[91] It has been speculated that this may be due to lower affinity for the alpha-2 subtype, which may be related to abuse potential. In the United States, the Z drugs are considered Class IV Drug Enforcement Administration (DEA) restricted agents, the same classification as the benzodiazepines.

Insomniac patients have been shown in several studies to have higher general health care use and costs than noninsomniac patients. One study has suggested that, during the first 6 months after initiation of treatment of insomnia with the newer nonbenzodiazepine hypnotics, general health care costs show a relative decline compared with those of untreated insomniacs.[92] The broader issue of the interaction of insomnia and its treatment with other illnesses is addressed in a review.[93]

Zolpidem

Zolpidem, an imidazopyridine compound with relative selectivity for the type I $GABA_A$-benzodiazepine receptor, is rapidly absorbed; because of first-pass metabolism, it has a bioavailability of 67% after oral administration of doses up to 20 mg.[94] Peak concentrations are reached after 1.6 hours. Total protein binding is approximately 92%. Absorption is slightly decreased when taken on a full stomach. It has no pharmacologically active metabolites and is eliminated primarily by renal excretion. In the United States, it is also marketed as a coated two-layer tablet in which the inner layer has a more extended release time (Ambien CR).

Zopiclone and Eszopiclone

Zopiclone, which is on the market in Europe and Asia but not the United States, is a cyclopyrrolone that acts at the $GABA_A$-benzodiazepine receptor complex, but possibly at a different binding domain or by producing different conformational changes than the benzodiazepines.[95] It is rapidly absorbed, with peak plasma concentrations occurring in 0.5 to 2 hours. Bioavailability is about 80%, implying that the first-pass effect is relatively small. It is very lipophilic and enters rapidly into the CNS. Protein binding is approximately 45%. It has two major metabolites, the *N*-oxide, which has lower pharmacologic activity, and the inactive *N*-desmethyl derivative, which along with various minor metabolites is excreted primarily by the kidneys and lungs.

Eszopiclone, the S-isomer of racemic zopiclone, is marketed in the United States as Lunesta. Peak concentrations are achieved in 1 hour with an elimination half-life of approximately 6 hours. It is weakly bound to protein and metabolized through oxidation and demethylation; its biotransformation is attributed to CYP3A4 and CYP2E1.[96] Eszopiclone has been found to have effectiveness without tolerance for at least 6 months[97] and was the first hypnotic medication approved by the FDA without a limit on duration of administration.

Zaleplon

A pyrazolopyrimidine that binds selectively to the type I benzodiazepine receptor, zaleplon is rapidly absorbed after oral administration, with peak concentrations reached in 1 hour and an elimination half-life of approximately 1 hour. Protein binding is approximately 60%. It is rapidly metabolized to inactive forms, with approximately 71% of labeled compound recovered in the urine and 17% in feces. Recommendations for administration in the United States include taking it immediately before bedtime or after the patient has gone to bed and has experienced difficulty falling asleep. In the latter case, it should be taken at least 4 hours before time of arising to avoid any possible memory difficulties and residual sedation.

Ramelteon

Ramelteon is a potent agonist at the melatonin type I and II receptors with negligible affinity for the $GABA_A$, dopamine, serotonin, or muscarinic cholinergic receptors. It reaches peak plasma concentrations in 0.75 to 0.94 minutes and has an elimination half-life of 1 to 2.6 hours; an active hydroxylated metabolite (M-II) has a slightly longer half-life of 2 to 5 hours. It is metabolized in the liver, primarily with CYP1A2 and to a lesser degree CYP2C9 and CYP3A4. Its mechanism of action is not fully understood, but it has been hypothesized that binding at melatonin receptor subtypes in the suprachiasmatic nucleus attenuates the waking signal. Clinically, its efficacy is primarily in reducing sleep latency, with little effect on awakenings during the night. It has essentially no dependence-producing effects, and in the United States it is not a DEA-restricted drug.

PHARMACOLOGIC PROPERTIES

Although the hypnotic benzodiazepines are given for purposes of aiding sleep, they share a spectrum of pharmacologic properties with agents given as daytime sedatives or anxiolytics; indeed, some authors have suggested that the designation of some benzodiazepines as hypnotics is as much a marketing plan as a pharmacologic decision. Among their effects are anxiolytic, myorelaxant, and anticonvulsant properties. Many, particularly the longer-acting agents, have mild respiratory depressant properties,[98,99] which are much less evident in shorter-acting agents. There is some evidence that triazolam may improve sleep-disturbed respiration in central sleep apnea,[100] although whether this is related to direct respiratory effects or secondary to decreasing the number of arousals (and hence postarousal respiratory pauses) during sleep is not clear. Even for the longer-acting agents, however, respiratory depression is much milder than those of older hypnotics such as the barbiturates. In practical terms, there is no significant effect in patients with normal ventilation, although it may become evident when there is preexisting compromised respiration such as in patients with chronic obstructive pulmonary disease or persons with unrecognized sleep-disordered breathing. The newer nonbenzodiazepines appear to have very few respiratory effects. One study of clinically used doses of zaleplon, for instance, found few effects on measures of sleep-disturbed respiration in patients with mild to moderate obstructive sleep apnea on continuous positive air pressure.[101] In most cases, the preceding generalizations in this section apply to zolpidem as well. There is some evidence from animal studies that there is a greater separation of hypnotic and sedative doses for zolpidem. In humans, zolpidem has been reported to lack respiratory depressant properties up to doses of 10 mg in healthy normal individuals and to exhibit very mild inhibition of mean inspiratory drive at 20 mg.[102] Zopiclone in therapeutic doses appears to have no significant effect on sleep-disordered

breathing in patients with chronic obstructive pulmonary disease.[103]

In general, adverse reactions to hypnotics are relatively rare and mild; one review of 3 years' experience in a 1000-bed teaching hospital found the median rate of reported adverse events to be 0.01% (1 in 10,000) doses administered and ran as high as 0.05%.[41] The rate for triazolam was 0.02%.[104] Although effective, benzodiazepine compounds manifested other problems: rebound insomnia on cessation and occasionally a state of "confusional arousal" in which the patient may have been acting out unaware, half asleep but moving around, not unlike sleepwalking, which has led to legal cases. Consequently, triazolam was withdrawn from many markets or the dose decreased. Even nonbenzodiazepines that act at $GABA_A$ receptors sometimes result in impaired cognition, amnesia, and ataxia.[55] Based on preclinical studies,[105-107] such impairments would not be expected in patients treated with hypocretin-orexin receptor antagonists.

Unlike older hypnotics such as the barbiturates, the benzodiazepines are relatively benign in overdose when taken alone by a medically healthy individual. They may be very toxic, however, when taken in combination with other CNS depressants such as alcohol. Because a significant portion of overdoses involve a combination of drugs, it is wise to treat them as potentially toxic or even lethal agents. In practice, this translates into being very conscious of the possibility that a patient seeking help for sleep disturbance may be suffering from unrecognized depressive illness, and if it is present, initiation of antidepressant therapy may be more appropriate.

EFFECTS ON SLEEP

Polysomnographic studies of benzodiazepines indicate that, consistent with their clinical effects, sleep latency and wake time after sleep onset are generally reduced, and total sleep time is increased. As with barbiturates and ethanol, spindle activity may be increased. REM sleep time may be reduced mildly, in contrast to the very potent REM suppression induced by barbiturates. In the early years after the introduction of benzodiazepines into the U.S. market, much was made of this observation, which was interpreted to mean that they somehow produced a more natural sleep. In hindsight, many investigators recognize that the psychological effects of REM deprivation are much less clear than originally thought and that, in the case of depressed patients, REM deprivation may be therapeutic, so whether having only mild REM suppressant properties translates into a clinical advantage seems uncertain.

In contrast to the barbiturates, however, the benzodiazepines are potent suppressors of slow wave sleep. The same dilemma that arises in terms of REM sleep recurs: Because the function of slow wave sleep has not been clearly determined, the clinical significance of pharmacologically suppressing this stage remains uncertain. The nonbenzodiazepine zolpidem shares with the benzodiazepines the very mild effects on REM sleep but, in contrast, does not alter slow wave sleep, which may even increase toward more expected values in patients with insomnia and low baseline levels. The development of neuroimaging techniques is also beginning to give insights into how the newer nonbenzodiazepines may affect the brain. A positron emission tomography study indicates that following eszopiclone administration, there is a more rapid decline in metabolic activity in the midbrain and pontine reticular formation during the transition from waking to NREM sleep.[108] One interpretation would be that the drug is reducing the hyperarousal often seen in insomnia.

Clinical efficacy studies indicate that virtually all these agents improve polygraphic measures of sleep and result in better subjective ratings of sleep quality during short-term use. One of the few distinctions among the benzodiazepines is that the longer acting agents such as flurazepam may not have much effectiveness on sleep latency until the second night of administration. A concern that was initially raised regarding the short-acting agents such as triazolam was the possibility that the relatively rapid metabolism might lead to sleep disturbance after several hours; later studies analyzing awakenings in the latter part of the night have generally found no evidence that this is the case.[109,110]

Another pharmacologic property to consider is the potential for dependence. A review of this topic by a panel from academia, industry, and the government concluded that the dependence potential of currently available hypnotics in patients without a history of substance abuse is minimal.[111] Ramelteon has been found to have no dependence-producing properties in standard measures. Griffiths and Johnson[112] present an excellent review of abuse liability and toxicity of 19 hypnotics from the days of the barbiturates through ramelteon. A variety of other issues at a clinical level include determination of whether tolerance develops during long-term nightly administration, and the effectiveness of nonnightly use.

FUTURE HYPNOTICS

Since the previous edition of this volume zolpidem became generic in the United States, and the landscape for hypnotic development has changed markedly. The previous edition listed a number of compounds that were under development at the time, including antagonists to specific serotonin receptor subtypes, alpha-2-delta calcium-channel blockers, and orexin antagonists. The serotonin subtype 2 receptor has been of interest for some time in relation to the action of some antidepressants and most atypical antipsychotics. Serotonin subtype 2A antagonists, including eplivanserin, pimavanserin, and others, were previously under development as hypnotics, with studies emphasizing improvements in sleep continuity and minimal daytime sedation. Some compounds combine antagonism at serotonin 2C receptors with melatonin receptor agonism. Compounds that bind to alpha-2-delta subunits of voltage-gated calcium channels, such as the anticonvulsants pregabalin and gabapentin, increase calcium ion entry into neurons and alter GABA concentrations. Both appear to increase slow wave sleep and decrease awakenings, and were also under development as hypnotics. An antagonist to orexin receptors, GW649868, was under development as a hypnotic as well. To the best of our knowledge, the above programs have ceased. On the other hand, several new drugs have been approved. Low-dose administration of the tricyclic antidepressant doxepin, which has significant antihistaminergic properties as well as antagonism at $alpha_1$ adrenoceptors and other properties, was approved by the FDA in 2010. The first orexin antagonist, suvorexant, was approved in mid-2014, and others are still in development. Advances in medicinal chemistry have enabled the creation of hypnotics that may come in

multiple preparations varying in duration of action. There has been interest in novel modes of administration including inhalation, oral sprays, and oral dissolving tablets. A sublingual form of zolpidem was approved by the FDA in 2011. However, the existence of a safe and relatively effective hypnotic in generic form raises the bar for the development of new clinical entities for insomnia. Only compounds directed at novel mechanisms such as the orexin receptors are likely to be able to address shortcomings of current nonbenzodiazepine hypnotics to overcome this economic barrier.

CLINICAL PEARL

Most currently available hypnotics induce sleep by acting on various moieties of the $GABA_A$-benzodiazepine receptor complex; knock-in studies of the alpha-1 subunit suggest that it is unlikely that hypnotics that act by this mechanism can be developed that will be able to have sleep-inducing, but not amnestic, properties. Hypnotics with novel mechanisms of action such as the orexin receptor antagonisms have recently become available for clinical use.

SUMMARY

Most currently available hypnotics induce sleep by acting at various moieties of the $GABA_A$-benzodiazepine receptor complex. This receptor is a member of the ligand-gated receptor family and is composed of several subunits, each of which appears in multiple isoforms. Microinjection studies indicate that the neuroanatomic sites of action of hypnotics from a variety of pharmacologic classes include the preoptic area and hypothalamus. These brain areas express GABA receptors and also contain GABAergic neurons. Connections between the preoptic area and the forebrain and brainstem, and the integrative nature of the anterior hypothalamus, help explain the mechanism by which sleep—and drugs that affect sleep—interact with a variety of physiologic systems. Novel hypnotic agents directed toward the hypocretin-orexin receptors have recently become available for clinical use.

Selected Readings

Brown RE, Basheer R, McKenna JT, et al. Control of sleep and wakefulness. *Physiol Rev* 2012;**92**:1087–187.

Buysse DJ. Insomnia. *JAMA* 2013;**309**:706–16.

de Lecea L. Hypocretins and the neurobiology of sleep-wake mechanisms. *Prog Brain Res* 2012;**198**:15–24.

Liu J, Clough SJ, Hutchinson AJ, et al. MT_1 and MT_2 melatonin receptors: a therapeutic perspective. *Annu Rev Pharmacol Toxicol* 2016;**56**.

Mieda M, Tsujino N, Sakurai T. Differential roles of orexin receptors in the regulation of sleep/wakefulness. *Front Endocrinol* 2013;**4**:57.

Okun ML, Ebert R, Saini B. A review of sleep-promoting medications used in pregnancy. *Am J Obstet Gynecol* 2015;**212**(4):428–41.

Palmaro A, Dupouy J, Lapeyre-Mestre M. Benzodiazepines and risk of death: results from two large cohort studies in France and UK. *Eur Neuropsychopharmacol* 2015;**25**(10):1566–77.

Scalo J, Desai P, Rascati K. Insomnia, hypnotic use, and health-related quality of life in a nationally representative sample. *Qual Life Res* 2015;**24**(5):1223–33.

Sigel E, Steinmann ME. Structure, function, and modulation of GABA(A) receptors. *J Biol Chem* 2012;**287**:40224–31.

Tannenbaum C, Diaby V, Singh D, et al. Sedative-hypnotic medicines and falls in community-dwelling older adults: a cost-effectiveness (decision-tree) analysis from a US Medicare perspective. *Drugs Aging* 2015;**32**(4):305–14.

Winkelman JW. Clinical practice. Insomnia disorder. *N Engl J Med* 2015;**373**(15):1437–44.

A complete reference list can be found online at ExpertConsult.com.

Chapter
42

Clinical Pharmacology of Other Drugs Used as Hypnotics

Daniel J. Buysse; Shachi Tyagi

Chapter Highlights

- Insomnia is the most prevalent sleep disorder and affects large proportions of the population on a situational, recurrent, or persistent basis. There is a growing awareness in the population regarding insomnia, which has led to an increase in diagnosis and treatment of insomnia in primary care.
- Pharmacologic treatment of insomnia is managed by hypnotic drugs from several classes. While benzodiazepine receptor agonists remain the most widely used hypnotics approved by the U.S. Food and Drug Administration (FDA), melatonin receptor agonists are also FDA approved for the treatment of insomnia. In addition, various other drugs originally developed as antidepressants, anticonvulsants, and antipsychotics, as well as hormones and other "natural" substances, have been used as hypnotics.
- Safe use of these drugs in clinical practice depends on the knowledge of pharmacokinetics, pharmacodynamics, sleep effects, and side effects of these medications. The purpose of this chapter is to provide a comprehensive review of the pharmacotherapeutic agents available for insomnia treatment to guide students and practitioners in treating the condition and to stimulate additional research to refine our understanding of this condition.

Benzodiazepine receptor agonists (BzRAs) are the most widely used drugs for the treatment of insomnia. Although BzRAs are popular among prescribers, they can be associated with rebound insomnia on discontinuation and a variety of well-described side effects such as tolerance, dependence, and abuse. Increasing concern has also been raised about their potential for more serious long-term adverse effects, including increased risk for dementia[1] and mortality.[2] Off-label use of sedating antidepressants is considered to be relatively safe and less likely to induce dependence; however, there are few clinical data establishing their efficacy and safety in insomnia, and side effects can be limiting.

Pharmacoepidemiology data indicate that prescribers often use non-BzRA drugs to treat insomnia. Data from National Health and Nutrition Examination Survey (NHANES) between 1999 and 2010 show that trazodone is one of the most commonly prescribed medications for insomnia, second only to zolpidem.[3] Data from a variety of other sources also demonstrate that trazodone and other antidepressants continue to be either the first- or second-choice hypnotic agents prescribed in the United States.[4-6] Use of drugs from other classes as hypnotics, although widely accepted clinically, raises several concerns. Most serious is the lack of evidence regarding safety and efficacy. Because most of these drugs have not been rigorously evaluated with regard to their sleep effects, vital information regarding appropriate dose, efficacy, and side effects is largely unknown.

Recent development of hypnotic drugs has arguably become more focused. Research is focused not only on novel targets (e.g., suvorexant targeting the orexin system) but also on developing new formulations of existing drugs specifically for insomnia treatment (e.g., low-dose doxepin).

A MODEL OF SLEEP-WAKE REGULATION RELEVANT TO SLEEP-PROMOTING DRUGS

Recent findings from the basic and clinical neuroscience of sleep-wake regulation permit specific models of sleep-wake regulation and insomnia to be advanced and tested (see also Chapters 7, 12, and 26). The model in Figure 42-1 describes non–rapid eye movement (NREM) sleep and wakefulness in humans as the products of dynamic interactions among several neural systems.[7-13] Wakefulness and arousal states are generated by the ascending activity of monoaminergic brainstem nuclei, histaminergic nuclei in the posterior hypothalamus, and cholinergic nuclei of the pontine tegmentum and basal forebrain (a to c in Fig. 42-1). Activity of these arousal systems is promoted by input from orexin (hypocretin) neurons in the perifornical lateral hypothalamus (LHA) during wakefulness and inhibited by gamma-aminobutyric acid (GABA) and galaninergic neurons in the ventrolateral preoptic area (VLPO) and median preoptic area (MnPO) of the hypothalamus at the onset of sleep (c). The LHA and VLPO/MnPO are in dynamic equilibrium with each other to ensure stable sleep-wake states (d). The timing and overall level of activity in the LHA and VLPO are tightly regulated by homeostatic sleep drive and the circadian timing system (e), which also indirectly affect output of brainstem arousal centers (f). Descending inputs from cortical-diencephalic cognitive and affective centers further modulate sleep-wake balance (g) and are in turn modulated by

ascending input from arousal centers, both directly and through thalamic pathways (a, b). Finally, corticothalamic oscillations underlie the characteristic electrophysiologic events of NREM sleep, sleep spindles, and delta oscillations.

Findings from human functional neuroimaging studies are consistent with this model. Compared with wakefulness, NREM sleep is associated with decreased global blood flow and glucose metabolism as well as large decreases in subcortical (brainstem, thalamus, basal ganglia, basal forebrain) and cortical (prefrontal and orbitofrontal cortex, anterior cingulate cortex, and precuneus) regions.[14-17] The influence of the circadian timing system is suggested by diurnal variation in regional metabolism during wakefulness in humans, with increased brainstem-hypothalamic activity in the morning compared with evening.[18] Finally, increasing homeostatic sleep drive with total sleep deprivation (TSD) leads to large reductions of metabolism in cortical, diencephalic, and brainstem structures during wakefulness and sleep.[19-21]

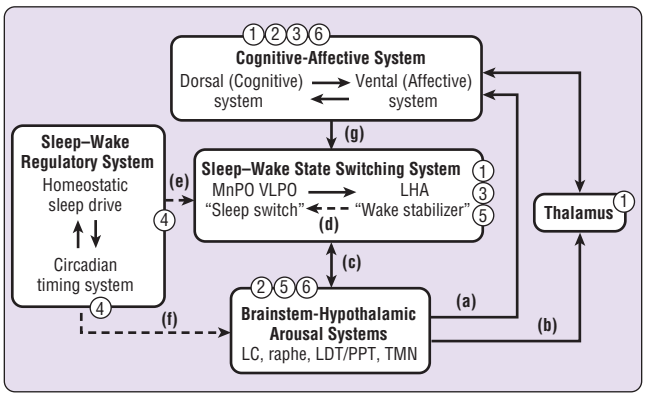


Figure 42-1 Direct (*solid arrows*) and indirect (*dotted arrows*) anatomic or physiologic pathways. LC, Locus coeruleus; LDT, laterodorsal pontine tegmentum; LHA, lateral hypothalamus perifornical area; MnPO, median preoptic area; PPT, pedunculopontine tegmentum; TMN, tuberomamillary nucleus of the posterior hypothalamus; VLPO, ventrolateral preoptic area. See text for explanation of letters and numbers.

Sleep-promoting medications may affect sleep-wake regulatory systems at several levels, as indicated in Figure 42-1. Benzodiazepine receptor agonists (1 in Fig. 42-1) may directly affect the sleep-wake state-switching system but also have direct cortical, thalamic, and brainstem effect due to the widespread distribution of GABA-A receptors. Sedating antidepressant and antipsychotic medications (2), through their activity on monoaminergic systems, affect corticolimbic systems and brainstem-hypothalamic arousal systems. Antihistamines (3) antagonize histamine-1 (H_1) receptors in the hypothalamus and cortex that receive projections from the tuberomamillary nucleus. Melatonin and melatonin receptor agonists (4), through their effects on melatonin-1 (MT_1) and MT_2 receptors, influence the "wake signal" from the suprachiasmatic nucleus and circadian timing system. Orexin antagonists[22] (5) inhibit the effect of orexin-hypocretin on brainstem and hypothalamic arousal centers, and 5-HT_2 antagonists (6) are most likely to have corticolimbic and brainstem sites of action. Thus different types of sleep-promoting drugs achieve their effects through very different actions on very different components of the sleep-wake regulatory system.

SEDATING ANTIDEPRESSANTS

Overview

More than two dozen drugs are approved by the U.S. Food and Drug Administration (FDA) as antidepressants, but several of these are used, mainly in low doses, for the treatment of insomnia. These include the tricyclic antidepressants (TCAs) doxepin, trimipramine, and amitriptyline, as well as the heterocyclic drugs trazodone, and mirtazapine. Of these, low-dose doxepin is the only FDA-approved antidepressant for insomnia treatment. Pharmacokinetic properties of these drugs are summarized in Table 42-1. The complex receptor effects of sedatives antidepressants are summarized in Table 42-2.

Most data regarding the effects of antidepressant drugs on human sleep come from studies in patients with depression,

Table 42-1 Pharmacokinetic Properties of Sedating Antidepressant Drugs*

Drug	Drug Class	Time to Maximal Concentration (h)	Metabolism (CYP Enzymes)	Elimination Half-Life (h) (Range)	Usual Dose (mg)	
					Antidepressant	Hypnotic†
Doxepin	Tricyclic	2–8	Major: 2D6, 2C19 Minor: 1A2, 3A4	20 (10–30)	100–300	3, 6
Amitriptyline	Tricyclic	2–8	Major: 2D6, 2C19 Minor: 1A2, 3A4	30 (5–45)	100–300	10–150
Trimipramine	Tricyclic	2–8	Major: 2D6, 2C19 Minor: 1A2, 3A4	25 (15–40)	100–300	25–150
Trazodone	Phenylpiperazine	1–2	3A4, 2D6	9 (3–14)	200–600	25–150
Nefazodone	Phenylpiperazine	1	3A4, 2D6, 2C19	2–4 (6–18 for active metabolites)	150–450	50–150
Mirtazapine	Noradrenergic and specific serotonergic antidepressant	1–3	3A4, 2D6, 1A2	25 (13–40)	15–45	7.5–30

CYP, Cytochrome P-450 (individual letters and numbers in table represent specific CYP enzymes).
*Data from references 23, 25, 53, 54, 74, 161, and 213.
†Except doxepin, hypnotic doses are based on published studies and common clinical practice without formal dose-ranging studies for this indication.

Table 42-2 Receptor Pharmacology of Sedating Antidepressant Drugs*

Drug	Receptor Effects†						
	NE Reuptake	5-HT Reuptake	5-HT_2 Receptor Antagonism	α_1 Antagonism	M_1 Antagonism (Anticholinergic)	H_1 Antagonism (Antihistamine)	Other Effects
Doxepin	+	0/+	+	+++	++	+++	
Amitriptyline	+	++	+	+++	+++	++	
Trimipramine	0	0	+	+++	++	+++	
Trazodone	0	+	++	++	0	0/+	5-HT_{1A}, 5-HT_{1C}, and α_2 antagonism
Nefazodone	0	++	++	++	0	+	
Mirtazapine	0/+	0	++	+	+	+++	α_2 and 5-HT_1 antagonism

5-HT, Serotonin; α, α-adrenergic receptor; H, histamine receptor; M, muscarinic cholinergic receptor; NE, norepinephrine. Numbers after receptor abbreviations indicate specific receptor subtype.
*Data from references 23, 26, 53, 161, and 214. Receptor actions are noted for usual antidepressant doses of each medication. Receptor actions at hypnotic doses may differ. Please refer to text.
†"+" Indicates strength of effect relative to other antidepressant drugs; "0" indicates no significant effect.

Table 42-3 Polysomnographic Effects of Sedating Antidepressant Drugs on Sleep*

Drug	Sleep Latency	Sleep Continuity†	Stage 3/4 NREM Sleep Amount (%)	REM Sleep	Other
Doxepin	↓	↑	↔	↓ amount, % of REM ↑ phasic eye movements (REM density)	↓ sleep apnea (minor effect); ↔ or ↑ periodic limb movements; ↑ restless legs symptoms; may induce eye movements during NREM sleep
Amitriptyline	↓	↑	↔	↓ amount, % of REM ↑ phasic eye movements (REM density)	
Trimipramine	↓	↑	↔	↔ amount, %	
Trazodone	↓	↔ to ↑	↑	↔ amount, % (↓ to ↑ in individual studies)	
Nefazodone	↔	↑	↔	↔	
Mirtazapine	↓	↑	↔	↔	

*Reported effects are based on a preponderance of evidence from published studies (see text for details). Many effects are inconsistent between individual studies. ↑ indicates increase from pretreatment baseline; ↓ indicates decrease from pretreatment baseline; ↔ indicates no change from pretreatment baseline. Effects noted are for antidepressant doses of each drug. Effects of hypnotic doses may differ.
†Sleep continuity refers to the proportion of sleep relative to wakefulness after sleep onset, as reflected by measures such as sleep efficiency. Other indicators of sleep continuity, such as wakefulness after sleep onset or number of awakenings, would have opposite signs. Thus ↑ indicates improvement in overall sleep continuity.

using doses higher than those typically used for insomnia. Formal dose ranging studies have not been conducted to determine optimal hypnotic doses, except for doxepin. Sleep effects of sedating antidepressant drugs have been discussed elsewhere[23,24] and are summarized in Table 42-3.

Tricyclic Antidepressant Drugs

TCAs share a core cyclic structure and differ from one another in their specific side chains (see Fig. 42-1). TCAs can be classified according to side chain structures as tertiary or secondary amines; tertiary amine TCAs undergo hepatic metabolism to form secondary amine TCAs. Tertiary TCAs are generally more sedating than secondary TCAs, and their metabolites have measurable blood levels and pharmacologic activity.

Pharmacokinetics

TCAs are absorbed moderately quickly from the gastrointestinal tract, with maximum concentrations occurring within 2 to 8 hours. They undergo extensive (30% to 90%) first-pass metabolism and are extensively protein bound (e.g., 90% or greater). Consequently, bioavailability is low following oral administration.[23,25,26] TCAs are very lipophilic, ensuring large volumes of distribution and high concentrations in the brain. TCAs have half-lives of approximately 15 to 30 hours.

TCAs undergo extensive hepatic metabolism, including oxidative demethylation of the side chains and hydroxylation of the ring structure[23-25] followed by conjugation and renal excretion. Hepatic metabolism of TCAs primarily

involves CYP2C19 and CYP2D6, as well as CYP1A2 and CYP3A4.[24-26] At usual doses, metabolism of TCAs follows linear kinetics, that is, drug blood levels (and metabolism) are proportional to dose, but at higher doses, metabolic enzymes can become saturated, leading to nonlinear pharmacokinetics, that is, higher blood levels than predicted by dose alone.[24] Wide interindividual variability in metabolism of TCAs is due to population variability in CYP2D6 polymorphisms and activity.[27] About 7% of patients in the population metabolize TCAs slowly due to a variant CYP2D6 isoenzyme, causing up to a 30-fold difference in plasma concentrations among different patients given the same TCA dose.[26] Age is associated with decreased metabolic clearance by CYP3A4 and decreased renal clearance.[23,25]

CYP3A4 can be induced by drugs such as barbiturates and tamoxifen, leading to reduced TCA levels, whereas CYP2D6 can be inhibited by drugs such as antipsychotic drugs, methylphenidate, fluoxetine, and paroxetine, leading to increased TCA levels.[24] Grapefruit juice is a common, if unsuspected, inhibitor of CYP3A4.

Pharmacodynamics and Receptor Pharmacology

TCAs interact with receptors of a variety of neurotransmitter receptors, including serotonin, norepinephrine, acetylcholine, and histamine. Effects on sleep likely represent the combined effects of many of these actions. Amitriptyline and doxepin inhibit both serotonin and norepinephrine reuptake transporters, whereas trimipramine has minimal reuptake effects. Tertiary TCAs, including amitriptyline and doxepin, have relatively more pronounced effects on serotonin than norepinephrine reuptake, whereas their secondary amine metabolites have more pronounced effects on norepinephrine reuptake.[24,28] After chronic dosing, additional effects on serotonergic and noradrenergic neurotransmission are also observed: desensitization of presynaptic 5-HT_{1A} autoreceptors; upregulation (sensitization) of postsynaptic 5-HT_{1A} receptors; and both downregulation and antagonism of postsynaptic 5-HT_2 receptors.[29] At 5-HT_2 receptors, antagonism of the 5-HT_{2C} subtype is more strongly associated with increasing slow wave sleep compared with 5-HT_{2A} receptors.[30] The net effect of TCA receptor effects is an enhancement of 5-HT effects in the central nervous system (CNS). Noradrenergic effects of TCAs include desensitization of presynaptic α_2 autoreceptors and a compensatory downregulation of postsynaptic β receptors,[23] with a net effect of increasing noradrenergic neurotransmission. TCAs also antagonize peripheral α_1 and α_2 adrenergic receptors, which accounts for their cardiovascular effects.

Sedating TCAs also antagonize M_1 muscarinic cholinergic and H_1 receptors. Amitriptyline is the most anticholinergic of all antidepressants, and doxepin is a more potent antihistamine than many drugs marketed as antihistamines, including diphenhydramine. Doxepin at low doses is also highly selective for this effect, with more than seven times greater affinity for the H_1 receptor relative to any other receptor.[31,32] As a result, low doses of doxepin (3 and 6 mg) can achieve selective H_1 blockade, with no serotonergic, adrenergic, and cholinergic effects.[33-36]

Effects on Human Sleep

Subjectively, tertiary tricyclic drugs are perceived as sedating, associated with reports of decreased sleep latency and wakefulness during the sleep period and improved sleep quality among depressed patients. Secondary TCAs such as desipramine are less sedating and may even be subjectively alerting. The polysomnography (PSG) effects of sedating TCAs in depression have been studied extensively. Reduced sleep latency, reduced wakefulness during sleep, and increased sleep efficiency have been reported in depressed patients treated with amitriptyline,[37-39] doxepin,[40] and trimipramine[41,42] and in primary insomnia patients treated with doxepin[43,44] and trimipramine.[45] By contrast, secondary TCAs such as desipramine and nortriptyline have little or no effect on sleep onset and continuity measures in depressed patients.[38,46]

Doxepin and amitriptyline have consistent effects on sleep stage architecture, including reductions in rapid eye movement (REM) sleep percentage and increases in phasic REM activity and REM sleep latency.[37,39,40,47,48] Trimipramine differs from most TCAs in its effects on REM sleep, which has been reported to increase, decrease, or remain unchanged during treatment.[39,41,42,45] Doxepin, amitriptyline, and trimipramine have inconsistent effects on stage 3/4 REM sleep, which has been reported to increase in some studies,[48] but not in most others.[39,41,42,44] The effects of a TCA on PSG sleep are illustrated in Figure 42-2.

In early studies of insomnia patients, moderate doses of doxepin (25 to 50 mg) and trimipramine (50 to 200 mg) have been associated with improved overall subjective sleep quality

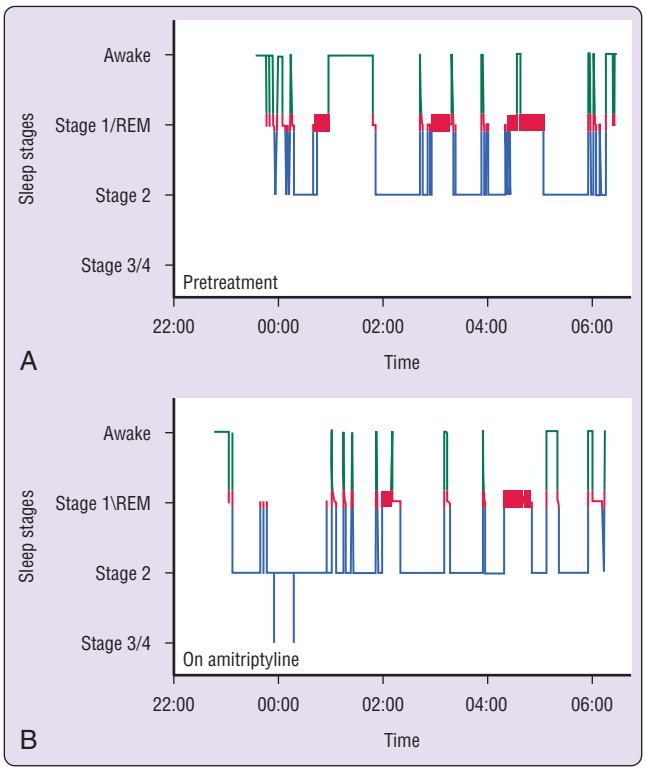


Figure 42-2 Effects of amitriptyline on electroencephalogram sleep in a 54-year-old depressed woman. **A,** Baseline sleep histogram during acute depressive episode showing prolonged sleep latency, reduced sleep efficiency, and reduced stage 3/4 sleep. **B,** Sleep histogram after treatment with amitriptyline for 24 days (final dose, 200 mg) and remission of symptoms. Compared with baseline, the histogram during treatment shows reduced sleep latency, improved sleep efficiency, reduced REM sleep, and prolonged REM latency.

and daytime well-being compared with placebo.[44,45] PSG effects of doxepin include reduced wakefulness and sleep latency and increased sleep time and sleep efficiency.[49] Studies using low doses of doxepin (1, 3, and 6 mg) demonstrated increased total sleep time, reduced wakefulness after sleep onset, and improved sleep efficiency.[33-36] Sleep efficiency was increased in the last quarter of the night, but with no demonstrable effect on next-day alertness or psychomotor performance. Trimipramine has been reported to reduce wakefulness and improved sleep efficiency without affecting sleep time or sleep latency.[50]

Other effects of TCAs on PSG sleep can include increases in periodic limb movements during sleep and the appearance of eye movements during NREM sleep; such effects are particularly noted with strongly serotonergic TCAs such as clomipramine. TCAs do not worsen sleep apnea and may have a small beneficial effect.[47] The acute effects of TCAs on sleep in depression are maintained over 1 to 3 years of maintenance treatment.[47,51] Rebound insomnia, indicated by decreased sleep time and sleep efficiency, may occur on discontinuation of sedating TCAs.[52]

Side Effects

Anticholinergic side effects of high- and moderate-dose doxepin and amitriptyline include dry mouth, increased perspiration, constipation, and urinary retention. More serious effects include precipitation of ocular crises in patients with narrow angle glaucoma, seizures, and anticholinergic delirium, which are dose related and typically occur at blood levels higher than 300 ng/mL.[23] Side effects related to antihistaminic properties include sedation and weight gain. Studies of very low-dose doxepin (1 to 6 mg) show a low incidence of side effects, with somnolence and headache being the most common.[33-36] Side effects related to α_1 antagonism include orthostatic hypotension with attendant risks for lightheadedness, syncope, and falls. TCAs typically increase heart rate. They also slow cardiac electrical conduction due to type I antiarrhythmic (quinidine-like) effects, which can lead to prolongation of the QRS duration and PR and QT intervals and heart block. The overdose lethality of TCAs is largely due to their cardiovascular toxicity, which can occur at doses as low as 10 times the therapeutic antidepressant daily dose. On the other hand, TCAs have no effect on cardiac contractility, and they can suppress atrial and ventricular ectopy.[23]

Trazodone

Trazodone is a tetracyclic antidepressant drug, and nefazodone is structurally similar. Nefazodone is not recommended for the treatment of insomnia because of potential hepatotoxicity and is not further discussed here.

Pharmacokinetics

Trazodone is rapidly absorbed, with peak plasma concentrations occurring 1 to 2 hours after oral doses. Like TCAs, it is highly (85% to 95%) protein bound. Trazodone has a relatively short half-life, approximately 5 to 9 hours.

The major metabolic pathway for trazodone is *N*-dealkylation to produce meta-chlorophenylpiperazine (m-CPP), an active metabolite that possesses serotonergic activity.[26,53] Trazodone also undergoes oxidation. Trazodone and m-CPP are substrates for CYP2D6, and trazodone is also metabolized to a lesser extent by CYP3A4. Drugs that inhibit CYP3A4, such as ketoconazole, inhibit trazodone metabolism and decrease m-CPP formation.[53]

Pharmacodynamics and Receptor Pharmacology

Trazodone is a relatively weak but specific inhibitor of the serotonin reuptake transporter with minimal affinity for norepinephrine or dopamine reuptake. Trazodone also inhibits serotonin 5-HT_{1A}, 5-HT_{1c}, and 5-HT_2 receptors. It has essentially no affinity for M_1 receptors, but it does have moderate H_1 receptor antagonism. Finally, trazodone is a relatively weak antagonist of α_2 adrenergic receptors and a somewhat more potent antagonist of α_1 receptors.[26,53]

The potent 5-HT agonist properties of the active metabolite m-CPP may play a role in the mechanism of action of trazodone. m-CPP inhibits serotonin reuptake and is a partial agonist at postsynaptic 5-HT_{2C} receptors. These actions can lead to increased side effects from trazodone among individuals who are "poor metabolizers" through CYP2D6 or who are taking inhibitors of CYP2D6, such as fluoxetine.[54]

Effects on Human Sleep

Studies of trazodone effects on human sleep have been limited by small sample sizes, particularly in PSG studies, and by study designs that have typically included sleep only as a secondary end point. In addition, many studies specifically addressing trazodone's sleep effects have not included double-blind, placebo-controlled, randomized study designs. Sedation is a commonly reported effect of trazodone in the treatment of depression, reported by more than 40% of patients. When specifically assessed, subjective sleep quality is improved by trazodone in healthy controls[55] and in patients with depression,[56-59] fibromyalgia,[60,61] and alcohol dependence.[62] However, some negative studies with respect to sleep quality in depressed subjects have also been reported.[63]

PSG studies have not consistently reported trazodone's effects on sleep continuity. The available evidence is inconsistent with regard to effects on sleep latency, total sleep time, and sleep efficiency; about half of the published studies show improvements in these measures.[64] Trazodone improved sleep efficiency, total sleep time, and wakefulness during sleep in older controls[55]; patients with depression,[57,65,66] including depressed patients concurrently treated with selective serotonin reuptake inhibitors[59]; and abstinent alcohol-dependent patients.[67] Some negative studies with regard to sleep continuity have also been reported in younger healthy controls,[68] in which ceiling effects may have been an issue, and in some studies of depressed patients with small numbers of subjects.[56,63]

Unlike most TCAs, trazodone has little effect on the amount of REM sleep, with most studies showing no significant change[57,65,66,68] or a small decrease.[55,63] Trazodone is also associated with increased stage 3/4 NREM sleep in most studies.[56,57,59,68] In one study, a small reduction in apnea-hypopnea index and no change in periodic limb movements were noted.[57] Although no information has been published regarding its long-term effects on sleep, rebound insomnia has been noted on discontinuation after several weeks of use.[55]

Only two studies have been published on the effects of trazodone in primary insomnia, compared with placebo[69] and both placebo and zolpidem.[70] Compared with placebo, trazodone 50 mg was associated with fewer night-time awakenings, minutes of stage 1 sleep,[69] reduced subjective sleep

latency, awakenings and wake time during sleep, and increased subjective sleep time, ease of falling asleep, and sleep quality.[70] The magnitude of effects with trazodone was similar to that for zolpidem 10 mg.[70]

Side Effects

Trazodone can produce side effects including orthostatic hypotension, lightheadedness, and weakness.[53] Unlike TCAs, trazodone does not have anticholinergic side effects, but it can have antihistaminic effects such as weight gain. Case reports suggest a potential for ventricular tachyarrhythmias.[71] A potentially serious effect of trazodone is priapism, sustained painful erections in men.[72] Although uncommon, priapism can be serious, requiring prompt surgical treatment. The risk appears to be greatest early in the course of treatment and can occur even at low doses. The incidence of abnormal erections during trazodone treatment is about 1 per 6000 male patients treated. Fatalities have been reported with overdoses of trazodone, although most of these occurred in conjunction with other drug ingestion. Unlike zolpidem and triazolam, trazodone is not associated with subjective effects associated with abuse potential.[73]

m-CPP, the metabolite of trazodone, can contribute to the development of the "serotonin syndrome" when this drug is used in combination with other serotonergic drugs such as selective serotonin reuptake inhibitors. The serotonin syndrome includes symptoms of confusion or delirium, restlessness similar to akathisia, muscular irritability, hyperreflexia, and autonomic instability, including hypotension.

Mirtazapine

Pharmacokinetics

Mirtazapine is rapidly absorbed, undergoes extensive first-pass metabolism, and is about 85% protein bound, yielding bioavailability of about 50%.[74,75] Mirtazapine has an elimination half-life of approximately 20 to 40 hours.

Mirtazapine undergoes *N*-demethylation (producing an active metabolite) and *N*-oxidation, followed by conjugation and excretion. CYP2D6 and, to a lesser extent, CYP3A4 and CYP1A2 are the major enzymes involved in mirtazapine metabolism. Mirtazapine follows linear pharmacokinetics, but its metabolism is affected by both age and sex; metabolic clearance is reduced in women and in older adults, as well as those with liver disease.[74] Although mirtazapine itself does not strongly inhibit or induce hepatic enzymes, mirtazapine blood levels are decreased by medications such as carbamazepine that induce CYP enzymes and increased by medications such as fluoxetine that inhibit CYP enzymes.

Pharmacodynamics and Receptor Pharmacology

Mirtazapine is a very weak inhibitor of noradrenergic reuptake and has no effect on serotonin reuptake.[76] However, similar to TCAs, it increases serotonergic and noradrenergic neurotransmission through blockade of α_2 autoreceptors and heteroreceptors.[54] It also has prominent antagonist activity at $5HT_2$, $5HT_3$, H_1, and α_1 adrenergic receptors,[76] which may contribute to its hypnotic effects.

Effects on Human Sleep

Mirtazapine is reported to be subjectively sedating in clinical studies of depression. In a PSG study in healthy adults, mirtazapine decreased sleep latency, awakenings, and stage 1 NREM sleep and increased stage 3/4 sleep.[77] In a study of 10 depressed patients, mirtazapine had similar effects, increased sleep continuity, with significant increases in slow wave and low delta sleep.[78] Thus, although potentially promising, mirtazapine has not yet been adequately evaluated as a hypnotic.

Side Effects

In addition to causing sedation, mirtazapine is associated with increased appetite, weight gain, and dry mouth, probably related to its antihistaminic properties. Clinical observations suggest that mirtazapine may be less sedating at doses greater than 30 mg per day than at lower doses. This is hypothesized to be related to greater noradrenergic effects relative to antihistaminic and serotonergic effects at lower doses.[79] Mirtazapine has not been associated with serious toxicity or death in overdose.

Serotonin Receptor ($5\text{-}HT_{2a/2c}$) Antagonists

As discussed previously, several antidepressants are antagonists of serotonin $5\text{-}HT_2$ receptors, which is thought to underlie some of their sedative effect. Other compounds that antagonize $5\text{-}HT_{2A}$ and $5\text{-}HT_{2C}$ receptors have been investigated more specifically for their sleep effects, in addition to possible antidepressant or antianxiety effects. $5\text{-}HT_{2A/2C}$ receptors are G-protein–coupled receptors that are widely distributed in the CNS, including the cortex, hippocampus, amygdala, thalamus, hypothalamus, and brainstem. They functionally interact with and affect the activity of GABAergic, dopaminergic, and cholinergic neurons.[80]

Early evidence in healthy humans showed that ritanserin, a $5\text{-}HT_{2A/2C}$ antagonist, increased slow wave sleep and delta electroencephalogram (EEG) activity in a dose-dependent fashion, with inconsistent and smaller magnitude effects on subjective sleep quality and PSG sleep latency and awakenings.[81,82] Ritanserin was found to have few effects on sleepiness or psychomotor function, even with daytime administration. Among poor sleepers, ritanserin had a small effect on the number of awakenings, in addition to the previously described SWS effect and improved subjective sleep quality.[83,84] Another $5\text{-}HT_2$ antagonist, SR 46349B, was observed to have very similar effects to ritanserin, increasing visually scored slow wave sleep and suppressing REM sleep.[85] Ultimately, ritanserin was not further developed as a hypnotic medication, and several newer $5\text{-}HT_2$ antagonists have also been taken out of development for lack of efficacy or adverse effects.

MELATONIN AND MELATONIN RECEPTOR AGONISTS

Biosynthesis, Physiologic Regulation, and Specific Compounds

Melatonin is a hormone endogenously synthesized from serotonin and produced in the pineal gland, retina, and intestinal tract. The noradrenergic sympathetic nervous system, acting through the superior sympathetic ganglion, stimulates pineal melatonin production. Specifically, β_1 stimulation with α_1 amplification leads to increased availability of *N*-acetyltransferase, the rate-limiting enzyme in melatonin biosynthesis.[86] Hence, β- and α-adrenergic antagonists may affect melatonin synthesis. Environmental light, acting through the hypothalamic tract and superior cervical ganglion,

suppresses endogenous pineal melatonin secretion. A number of synthetic melatonin receptor agonists have been developed and tested in clinical studies. These include two melatonin receptor agonists, ramelteon and tasimelteon, and agomelatine (S-20098), a serotonergic-melatonergic antidepressant.[87,88]

Pharmacokinetics

Exogenous melatonin is rapidly absorbed, with peak levels occurring in about 20 to 30 minutes. It has a 40- to 60-minute elimination half-life.[89] When formulated with absorption-retarding binders, systemic availability can be prolonged to mimic the normal period of nocturnal secretion.[90,91] About 85% of an oral dose is removed by hepatic first-pass metabolism. Metabolism includes oxidation by CYP1A2 and CYP2C19, followed by glucuronate and sulfate conjugation and renal excretion.[92] Melatonin is secreted in breast milk. Ramelteon is also rapidly absorbed, with time to maximal concentration of 0.75 to 1 hour and elimination half-life of 0.8 to 2.5 hours.[93] Tasimelteon's time to reach maximum plasma concentration ranges from 1.9 hours with low dose to 3 hours with higher doses.[94] Agomelatine is rapidly absorbed, irrespective of food intake, reaching peak plasma concentration in 1 to 2 hours.[95]

Pharmacodynamics and Receptor Effects

Three subtypes of melatonin receptors have been identified, with the MT_1 and MT_2 subtypes being most prevalent in the suprachiasmatic nucleus and retina and mediating the phase-shifting effects of melatonin. Melatonin receptors are also found in reproductive organs, immune cells, and vasculature, where they can mediate both vasoconstriction and vasodilation. Melatonin has effects in regulating retinal light sensitivity, seasonal breeding patterns, systemic immunity, and cancer cell growth. Tasimelteon, like ramelteon, has a high affinity for MT_1 (pKi = 9.45 ± 0.04 [0.35 nM]) and MT_2 (pKi = 9.8 ± 0.07 [0.17 nM]), which is similar to the affinity of melatonin for both receptors.[96] Agomelatine, on the other hand, is not only an MT_1 and MT_2 receptor agonist but is also a 5-HT_{2C} receptor antagonist that has chronobiotic, antidepressant, and anxiolytic effects.[97]

Effects on Human Sleep

In humans, melatonin has been studied both as a chronobiotic (where the intended effect is to phase-shift circadian rhythms) and as a hypnotic (where the intended effect is to induce or maintain sleep). Its use as a chronobiotic is discussed elsewhere in this volume. The potential hypnotic actions of melatonin have been studied in healthy adults,[98,99] adults and children with insomnia,[100,101] older adults with insomnia,[90] dementia patients,[91,102] medically ill adults,[103] and patients with affective disorders.[104] These studies have used a wide range of doses from less than 1 mg to greater than 80 mg and a variety of administration schedules.[105,106] A dose-response effect has not been demonstrated.

A number of studies have shown that melatonin is associated with reduced sleep latency and improved quality of sleep by self-report, with less consistent effects on sleep maintenance and duration.[107] Studies using actigraphy and PSG outcomes in pediatric, older adult, and nursing home populations also suggest modest effects on sleep latency, with less consistent increases in total sleep time.[108,109] Melatonin has been reported to decrease sleep latency and awakenings and increase sleep time in children with neurodevelopmental disabilities and sleep or circadian rhythms disturbances.[110,111] Paradoxically, melatonin may act preferentially as a hypnotic in situations of low homeostatic drive for sleep, such as during the daytime period of low endogenous secretion.[112-114] A meta-analysis of 17 studies, predominantly in healthy subjects, showed that melatonin was associated with a mean decrease of 4 minutes in sleep latency, mean increase of 2.2% in sleep efficiency, and mean increase in sleep duration of 12.8 minutes.[115] Heterogeneity in melatonin doses and subject characteristics may have affected these findings, as well as the fact that most studies were not conducted in subjects with insomnia.

Ramelteon is associated with reduced sleep latency and increased sleep time, but not with consistent changes in wake after sleep onset (WASO) or other sleep continuity measures. In several studies, changes on PSG measures were more robust than self-report measures. When administered to healthy adults in the novel sleep laboratory environment, ramelteon was associated with improved subjective sleep quality, reduced subjective and PSG sleep latency, and increased PSG sleep time, but not changes in sleep continuity.[116] In short-term (two nights) administration to adult chronic insomnia patients, ramelteon was associated with reduced subjective and PSG sleep latency, increased PSG sleep time, reduced slow wave sleep, and no effect on WASO[117]; a similar study in older adults yielded similar findings but with a significant effect on sleep efficiency and no effect on self-report measures. Five-week studies in adults with chronic insomnia show a similar pattern of reduced sleep latency and increased sleep time by self-report and PSG, but no change in sleep efficiency.[118,119] Effect on sleep latency was persistent after 6 months of treatment, without any change in total sleep time, sleep efficiency, or subjective sleep measures.[120] Ramelteon had no effect on sleep-disordered breathing or other measures of sleep in patients with sleep apnea.[121,122]

Agomelatine, an antidepressant with both melatonin MT_1/MT_2 agonist and serotonin 5-HT_{2C} antagonist properties, may also have effects on sleep and circadian rhythms. A study of agomelatine administration in the evening hours to healthy elderly men showed no effects on PSG measures of sleep, but phase advances were observed in circadian rhythms of body temperature and cortisol.[123] In 6-week studies of depressed patients, agomelatine was associated with improved subjective measures of sleep latency and sleep quality,[124] decreased PSG WASO, and increased sleep efficiency and slow wave sleep.[125]

Tasimelteon is a newer melatonin receptor agonist. It has recently been approved by the FDA for treatment of non–24-hour sleep-wake disorder,[126] but there have been no studies assessing its efficacy as a hypnotic. In a phase 2 study, tasimelteon showed a dose-dependent shift in plasma melatonin rhythm to an earlier hour.[94] Tasimelteon 100 mg advanced the melatonin rhythm by an average of 2 to 3 hours within hours of administration, which led to the postulation that this drug might improve sleep, at least partly, by shifting the circadian pacemaker.

Side Effects

Melatonin has a wide therapeutic index in humans. Despite widespread use of melatonin at low doses, no obvious public health risk has yet emerged. Nonetheless, the actual risks of protracted melatonin consumption remain unknown. The most common side effect reported from taking melatonin is

headache. Acutely, increased sleepiness and fatigue might contribute to the loss of vigilance in critical work situations. Melatonin is nonaddicting. Ramelteon is generally well tolerated, with the most common side effects including headache, dizziness, somnolence, fatigue, and nausea. Rebound insomnia has not been observed, and there is no evidence for abuse potential.

Headache was the most commonly reported side effect from tasimelteon (more than 10%). Fewer (1% to 10%) patients experienced increased alanine transaminase, nightmares or abnormal dreams, upper respiratory infections, and urinary tract infections. Tasimelteon is a pregnancy category C medication, so benefits must outweigh any potential risks.

ANTIHISTAMINES

Specific Agents

Antihistamine drugs used in the treatment of insomnia are reversible antagonists of H_1 receptors. Antihistamines are a diverse group of drugs broadly divided clinically into two groups based on their sedative potential.[127] First-generation agents include doxepin (discussed earlier under antidepressants), diphenhydramine, doxylamine, chlorpheniramine, hydroxyzine, meclizine, promethazine, and cyproheptadine. Essentially all of the over-the-counter antihistamine drugs marketed for insomnia treatment include diphenhydramine, the prototype of this class, or doxylamine. In addition, most analgesic-hypnotic combination medications ("p.m." preparations) contain diphenhydramine. Second-generation nonsedating antihistamine drugs are used primarily for treatment of seasonal, environmental, and other allergic reactions and are not used for the treatment of insomnia.

Pharmacokinetics

Antihistamines are well absorbed from the gastrointestinal tract and widely distributed throughout the body, including the CNS. Most first-generation antihistamines, including diphenhydramine, achieve peak plasma concentrations in 2 to 3 hours, and effects usually last 4 to 6 hours; however, specific agents such as chlorpheniramine, hydroxyzine, and meclizine may last up to 24 hours. Diphenhydramine is extensively metabolized by CYP enzymes and has an elimination half-life of 4 to 8 hours.[128]

Pharmacodynamics and Receptor Pharmacology

Histamine is widely present throughout the CNS and the body. Three types of histamine receptors have been described, labeled H_1, H_2, and H_3. Histamine is synthesized from histidine by the action of L-histidine decarboxylase and is metabolized by methylation and oxidative deamination.[127] In most tissues, histamine is stored in mast cells, which mediate allergic responses. In the CNS, histamine serves as a neurotransmitter localized to the tuberomamillary nucleus of the posterior hypothalamus. Histaminergic neurons project widely to the brainstem and cerebral cortex and promote wakefulness.[129] Histaminergic neurons fire actively during wakefulness, reinforced by excitatory input from hypocretin-containing neurons of the lateral hypothalamus, and are inhibited during sleep by the activity of GABAergic projections from the ventrolateral preoptic area.[130]

Histamine has a variety of effects in the periphery, including H_1 receptor–mediated vasodilation, increased capillary permeability, bronchoconstriction, contraction of the gut, and H_2-mediated gastric acid secretion. H_3 receptors serve as presynaptic autoreceptors that mediate feedback inhibition of histamine synthesis and release.

H_1 receptor antagonists have a variety of CNS and systemic effects. Well-described CNS effects include sedation, decreased alertness, decreased reaction times, and sleepiness.[131] Paradoxically, a minority of patients respond with CNS activation, including restlessness, anxiety, and increased alertness. Systemic effects of antihistamines include inhibition of immediate hypersensitivity reactions mediated by mast cell release of histamine. In addition, H_1 antagonists decrease capillary permeability, thus inhibiting edema, wheal reactions, and the itch response. H_1 antihistamines also have minor effects on antagonizing respiratory smooth muscle, relaxing bronchospasm, and promoting vasodilation.[132]

Many of the early sedating antihistamines, including diphenhydramine, have muscarinic anticholinergic effects similar to atropine. In addition, diphenhydramine increases serotonergic neurotransmission and antagonizes α-adrenergic receptors. The "nonsedating" H_1 antagonists have very little CNS effect because of their inability to penetrate the blood-brain barrier, thus limiting their actions to the periphery.

Effects on Human Sleep

Antihistamines such as diphenhydramine and doxylamine are associated with subjective drowsiness and sleepiness, leading to their widespread use as over-the-counter hypnotic agents. However, their efficacy has not been well studied. A study of 10 healthy adults showed increased motor activity with diphenhydramine compared with placebo and no significant effect on self-reported sleep outcomes.[133] Clinical trials using doses of 12.5 to 50 mg in hospitalized and outpatient samples of poor sleepers have shown subjective improvements in sleep latency, nocturnal awakenings, sleep duration, and sleep quality.[134-136] A comparative trial of temazepam 15 mg and diphenhydramine 50 mg in older adults with insomnia showed a reduction in awakenings, but not in sleep latency, total sleep time, or sleep quality with diphenhydramine.[137] Approximately 50% of subjects experienced rebound symptoms with temazepam and diphenhydramine, and the rate of other side effects was similar. The largest published trial compared diphenhydramine and valerian-hops combination to placebo in a total of 184 adults.[138] Diphenhydramine was associated with significantly greater change in self-reported sleep efficiency, but not in sleep latency or total sleep time, relative to placebo; a global measure of self-report outcome, the Insomnia Severity Index, showed significantly greater reductions in the diphenhydramine groups than placebo groups. PSG-measured sleep latency, sleep efficiency, and total sleep time also showed no significant group differences. Other PSG studies have examined daytime sleepiness with sedating and nonsedating antihistamines. These studies confirm the hypnotic effect of drugs such as diphenhydramine, while showing no significant sedation with nonsedating drugs such as astemizole, loratadine, or cetirizine,[139-141] and also show tolerance to the daytime sedative effect of diphenhydramine.[142]

Side Effects

Impairment of psychomotor performance with diphenhydramine has been well documented.[131] Epidemiologic studies have also suggested cognitive impairment associated with

diphenhydramine use in the older adults.[143] Other side effects related to CNS activity include dizziness, fatigue, and tinnitus. Peripheral side effects can include decreased appetite, nausea, vomiting, diarrhea, and constipation as well as weight gain. A number of case reports have also documented potentially serious side effects of doxylamine, including coma, rhabdomyolysis, and resultant kidney failure.[144]

VALERIAN

Pharmacokinetics

Valerian preparations include more than 400 extracts, mainly derived from the roots of the plant species *Valeriana officinalis*. These extracts contain a number of chemicals with CNS activity, including sesquiterpenes, valepotriates, valerinic acid, and various other alkaloids, in unknown proportions.[145] The constituents of a specific valerian preparation also depend on the actual valerian species used, the method of extraction (aqueous versus ethanolic), and the combination with other herbs or agents such as hops.[146] Because of the multiple constituents of valerian preparations, their pharmacokinetics have not been well described. Doses in clinical studies have typically ranged from 400 to 900 mg per day.

Pharmacodynamics and Receptor Pharmacology

The exact mechanism of action of valerian preparations is also unknown. The GABA-like activity of valerian extracts is suggested by their sedative, anxiolytic, myorelaxant, and possible anticonvulsant effects.[147] Valerian extracts contain a small amount of GABA, but GABA is not transported across the blood–brain barrier, and valerian extracts do not act on benzodiazepine receptors. Valerenic acid may actually inhibit brain GABA metabolism. Other potential mechanisms of action include serotonin receptor activity and adenosine receptor antagonism. Finally, some components of valerian, including valepotriates, may act as prodrugs transformed into homobaldrinal, a compound that inhibits motor activity in mice.[145]

Effects on Human Sleep

The effects of valerian extracts on sleep in humans have been investigated in healthy young adults and middle-aged and older adults with insomnia. Duration of treatment has ranged from 1 to 14 days, and outcome measures have included self-reports, actigraphy, and PSG. Subjective effects of valerian preparations include decreased sleep latency, improved sleep quality, and decreased awakenings.[148,149] Effects on PSG sleep include increased stage 3/4 and reduced stage 1 NREM sleep.[150,151] Although improved sleep latency and sleep efficiency have been observed in some PSG studies, sleep continuity effects of valerian are inconsistent.[150-152] A systematic review of valerian preparations found no significant effects of ethanolic extracts on subjective or objective sleep measures; inconsistent effects of aqueous valerian extracts; limited and variable evidence for valepotriate preparations; and mixed but predominantly negative evidence regarding valerian combinations.[146] Overall, this review concluded that research findings to date do not support the efficacy of valerian preparations for improving subjective sleep outcomes.

Side Effects

Side effects associated with valerian have been reported to be few and mild and include headache and weakness. Morning sleepiness is an infrequent side effect.[145]

GABAPENTIN AND PREGABALIN

Gabapentin and pregabalin were initially developed as anticonvulsant drugs, but they have subsequently found widespread use in treating neuropathic and fibromyalgia pain, periodic limb movement disorder, bipolar mood disorder, and insomnia. Some evidence suggests that GABA levels in the brain (occipital cortex, anterior cingulate gyrus, and thalamus) are lower in patients with primary insomnia compared with control participants.[153,154] Therefore medications that indirectly increase GABA might be therapeutic in patients with insomnia.

Typical daily doses of gabapentin range from 300 to 2100 mg, taken in divided doses with larger doses in the evening hours. Gabapentin absorption is moderately rapid (2 to 3 hours to maximal absorption), and it has a bioavailability of 35% to 60%, with lower availability at higher doses (nonlinear pharmacokinetics). It is distributed extensively and has an elimination half-life of 5 to 9 hours. Doses of pregabalin are typically 150 to 600 mg in divided doses. Absorption of pregabalin is rapid (1 hour to maximal absorption) and is not affected by dose (linear pharmacokinetics). It has a half-life of 4.5 to 7 hours. Gabapentin and pregabalin are not metabolized in humans, are not bound to plasma proteins, and are excreted unchanged, mainly in the urine.[155] Therefore doses of both drugs may need to be adjusted in patients with reduced renal clearance.[156,157]

Gabapentin and pregabalin may have several mechanisms of action for their analgesic and sedative effects. Both drugs are structural analogues of GABA but do not appear to interact with GABA-A or GABA-B receptors or influence GABA reuptake, although they may promote formation of GABA in the CNS.[155,158] The drugs selectively bind with high affinity to the alpha-2-delta subunit of N-type voltage-gated calcium channels, which both increases GABA concentration in synapses and decreases stimulus-evoked release of excitatory neurotransmitters such as glutamate and norepinephrine.[159,160] Other potential mechanisms of action include antagonism of NMDA receptors and interaction with the L-amino acid transporter receptor.[161] Side effects associated with gabapentin and pregabalin include sedation and fatigue, dizziness, headache, ataxia, and a less common risk for leukopenia. Gabapentin and pregabalin are consistently associated with improvement on self-reported sleep measures in patients with various pain conditions (e.g., fibromyalgia, neuropathic pain, postherpetic neuralgia, postsurgical pain).[157,159,162]

Although these drugs are sometimes used clinically for primary insomnia or other forms of comorbid insomnia, their sleep effects have not been systematically evaluated. A small study of gabapentin administered for 16 days to healthy adults showed no change in subjective sleep quality for most subjects, trends for reduced awakenings and periodic limb movement arousal index, and a significant increase in slow wave sleep.[163] Pregabalin administered to healthy adults was likewise associated with improved subjective sleep quality, reduced sleep latency and REM sleep, and increased sleep efficiency, total sleep time, and slow wave sleep.[164] Similar effects on self-report and PSG sleep continuity, but not slow wave sleep, were seen when pregabalin was added to monotherapy with other drugs in epilepsy patients.[165] In patients with restless legs syndrome, gabapentin has been associated with improved sleep continuity, increased stage 3/4 sleep, and reduced periodic limb movements.[166] An open pilot study of gabapentin

in alcohol-dependent patients with comorbid insomnia showed improvements in insomnia symptoms and feelings of tiredness in the morning.[62] Another international, multicenter, double-blind, placebo-controlled trial of pregabalin in patients with fibromyalgia demonstrated improved sleep as assessed by the Medical Outcomes Study Sleep Disturbance subscale and the Sleep Quality diary.[167] In patients with generalized anxiety presenting with concurrent insomnia, pregabalin was shown to improve anxiety symptoms while specifically improving insomnia.[168]

One small, uncontrolled study with 18 primary insomniacs receiving gabapentin treatment for at least 4 weeks showed improved Pittsburg Sleep Quality Index score, as well as increased PSG-measured sleep efficiency and slow wave sleep, decreased wake after sleep onset, and spontaneous arousal index.[169]

GABOXADOL

Gaboxadol, also known as THIP, is a direct GABA-A receptor agonist with effects distinct from those of BzRA.[170] BzRAs are active at a wide range of GABA receptor subtypes, but especially those with α_1 subunits, which are primarily synaptic in location. Synaptic GABA receptors are activated during fast synaptic transmission, transiently hyperpolarize and inhibit postsynaptic neurons, and desensitize rapidly. Gaboxadol is a potent agonist of GABA receptors that contain α_4, α_6, and δ, subunits, which have more restricted anatomic distribution in the thalamus, hippocampus, and cerebellum and are mainly extrasynaptic in location. Extrasynaptic GABA receptors are sensitive to low concentrations of GABA, they desensitize slowly, and their activation induces sustained neuronal effects.[171,172] Gaboxadol is rapidly absorbed, reaching peak concentration within 30 minutes, with a half-life of approximately 1.5 to 2 hours.[171]

In experimental models using healthy subjects, gaboxadol 5 to 15 mg reduced self-reported and PSG sleep latency and WASO and increased total sleep time and delta and theta EEG activity.[173-175] Short-term studies in patients with primary insomnia show similar findings, comparable to those seen with zolpidem as an active comparator.[176,177] Gaboxadol also increased slow wave sleep. Gaboxadol is not associated with typical indicators of abuse potential in animal or human studies.[171] Common side effects include headache, nausea and vomiting, somnolence, and fatigue. Gaboxadol has also been associated with dissociative reactions including hallucinations and disorientation, which have led to suspension of clinical studies in the United States.

OREXIN ANTAGONISTS

Orexin (hypocretin) is a peptide neurotransmitter localized to the perifornical neurons of the lateral hypothalamus. Orexinergic neurons have widespread excitatory projections to the brainstem and posterior hypothalamic arousal centers and are key neuropeptides responsible for generating and maintaining wakefulness.[178-180] Two receptors respond to orexin signaling: Orexin-1 receptor (OX1R) and orexin-2 receptor (OX2R) both have partially overlapping nervous system distributions. Suvorexant is a potent and selective antagonist of OX1R and OX2R and was the first drug in this class approved by FDA for the treatment of insomnia. It is orally bioavailable, has good brain penetration, and has high receptor occupancy.[181] Suvorexant's peak concentrations occur at a median of 2 hours (30 minutes to 6 hours) under fasted conditions. The mean absolute bioavailability of 10 mg is 82%. It is extensively bound (>99%) to plasma proteins and has a terminal half-life of 9 to 12 hours,[22] with some next morning residual effects.[182] CYP3A4 and, to a lesser extent, CYP2C19 are the major enzymes involved in suvorexant metabolism. In healthy subjects, suvorexant 50 mg and 100 mg decreased latency to onset of persistent sleep and wake after sleep onset and increased sleep efficiency and total sleep time, whereas suvorexant 10 mg decreased wake after sleep onset. In patients with primary insomnia, 4 weeks of suvorexant treatment improved sleep efficiency and wake after sleep onset. Total sleep time also improved with greater time spent in REM and stage 2 sleep.[183] During 1-year treatment with suvorexant, insomniac patients reported subjective improvements in total sleep time, time to sleep onset, WASO, and sleep quality.[184]

The most common adverse events reported with suvorexant are somnolence, fatigue, and dry mouth. The key safety concerns are residual sedation, rapid onset of somnolence if administered during the daytime, motor impairment, driving impairment, and hypnogogic hallucinations.[182,184,185] At higher dosages (50 and 100 mg), the medication significantly decreased reaction time and reduced subjective alertness tested the morning after drug administration.[182] Although effects resembling cataplexy are a theoretical concern, given the role of deficient orexin neurotransmission in narcolepsy-cataplexy,[186] these effects were not found in clinical trials. Adverse events appeared to be dose- and plasma-exposure-dependent. It is possible that these residual effects are related not only to half-life but also to a combination of pharmacokinetic (slow elimination or metabolism) and pharmacodynamic effects (slow equilibration and off rates).[22] Moreover, endogenous orexin production appears to follow a circadian pattern with a peak in the late waking period,[187] which could lead to more potent effects of orexin receptor antagonists during daytime than nighttime hours.

Given the stricter FDA policy for hypnotics to use the lowest effective dose to minimize safety risk, 10 and 20 mg strengths of suvorexant have been approved,[185] as opposed to 30 and 40 mg strengths, which were tested in a phase 3 trial.[184]

CHLORAL HYDRATE

Chloral hydrate has been used as a hypnotic and as a sedative in children undergoing clinical procedures. Chloral hydrate is a prodrug, rapidly converted by alcohol dehydrogenase in the liver to the active compound trichloroethanol, which acts at the barbiturate recognition site on GABA-A receptors. Metabolism of trichloroethanol occurs through hepatic conjugation, with a half-life of approximately 5 to 10 hours. Sleep effects include subjective and objective reduction in sleep latency and improvement in sleep continuity, with little effect on stage 3/4 or REM sleep.[131] Because chloral hydrate is a skin and mucous membrane irritant, it can have side effects of unpleasant taste, gastrointestinal distress, nausea, and vomiting. Other potential side effects include lightheadedness, nightmares, and ataxia. More serious potential side effects include hepatic injury. Fatal overdoses are possible, and chronic use can result in severe withdrawal. Chloral hydrate is not recommended for treatment of insomnia in adults or children,

given its low therapeutic index and the availability of safer alternative drugs.

SEDATIVE ANTIPSYCHOTIC DRUGS

Second-generation antipsychotic drugs have significantly higher rates of somnolence than placebo in clinical trials.[188] This effect may be clinically useful in the treatment of insomnia, particularly among patients with severe depression, bipolar disorder, and psychotic disorders. Although many different antipsychotic drugs have sedative effects (see Monti and Monti[189] for review), olanzapine and quetiapine are the drugs most commonly used in nonpsychotic and nonbipolar patients for this purpose. Typically olanzapine is administered in doses 2.5 to 20 mg and quetiapine in doses of 25 to 200 mg at bedtime.

Unlike older antipsychotic drugs that antagonize primarily dopamine receptors, olanzapine has a variety of receptor effects, including antagonism of serotonin 5-HT_{2A}, muscarinic cholinergic, H_1, and α_1-adrenergic receptors, as well as activity at serotonin 5-HT_{2C}, 5-HT_3, and 5-HT_6 receptors.[161,190,191] Olanzapine is structurally similar to benzodiazepines. Quetiapine, like olanzapine, is an antagonist of serotonin 5-HT_{2A}, H_1, and α_1 receptors. It has somewhat more potent dopamine D_2 receptor antagonism than olanzapine, but its dopamine binding is rapidly reversible.

Olanzapine is rapidly absorbed, but a significant portion of the drug is metabolized in first-pass circulation. Its peak concentration occurs at about 6 hours, and it has a terminal elimination half-life of 20 to 54 hours. It is metabolized through the activity of CYP1A2 and CYP2D6. Quetiapine is also rapidly absorbed but reaches peak concentration in about 1.5 hours and has a terminal elimination half-life of approximately 6 hours. Quetiapine is also metabolized in the liver, primarily through CYP3A4.

Both of these antipsychotic drugs have a lower incidence of extrapyramidal side effects than traditional antipsychotic drugs such as haloperidol. However, both can cause hypotension. In addition, olanzapine has been associated with weight gain and glucose intolerance, as well as neurocognitive impairment at higher doses. Quetiapine has been associated with prolongation of the QT_c interval on electrocardiogram.

Both olanzapine and quetiapine are subjectively sedating. Uncontrolled and placebo-controlled treatment studies using these medications as primary or adjunctive treatments demonstrate improved subjective sleep quality and reduced sleepiness in patients with schizophrenia,[192,193] unipolar depression,[194,195] and bipolar depression.[196] In PSG studies with small numbers of healthy control subjects, olanzapine is associated with decreased sleep latency, wakefulness, and stage 1 NREM sleep; increased sleep efficiency, stage 2, and stage 3/4 NREM sleep; no consistent effect on REM; and improved subjective sleep quality.[197-201] Similar self-report and PSG effects have been demonstrated in small clinical studies of patients with depression,[202] mania,[203] and schizophrenia.[198] Quetiapine administered acutely to healthy subjects has been reported to decrease sleep latency; increase sleep time, sleep efficiency, and subjective sleep quality; and reduce REM sleep.[204] One double-blind randomized controlled trial evaluating efficacy of quetiapine 25 mg in primary insomnia showed no significant improvement of self-reported total sleep time and sleep onset latency.[205]

Small uncontrolled and controlled studies have also examined the effects of other second-generation antipsychotics, including risperidone and clozapine, in patients with schizophrenia.[189] Both drugs are associated with improved sleep continuity, and risperidone is associated with increased slow wave sleep.

Given their potentially significant neurologic and metabolic side effects, antipsychotic drugs are best reserved for treatment of individuals who have insomnia comorbid with major psychiatric disorders, particularly psychotic and bipolar disorders.

SODIUM OXYBATE (GAMMA-HYDROXYBUTYRATE)

Sodium oxybate, the sodium salt of gamma-hydroxybutyrate (GHB), is FDA approved for the treatment of cataplexy in patients with narcolepsy and is also recommended for the treatment of excessive sleepiness.[206] GHB is an endogenous short-chain fatty acid that is synthesized from GABA. GHB acts as a neuromodulator and neurotransmitter, has two specific neuronal recognition sites, and is also a ligand for $GABA_B$ receptors. It is widely distributed in the brain, including the hippocampus, nucleus accumbens, basal ganglia, cortex, and brainstem. GHB acts primarily to inhibit the release of colocalized neurotransmitters, but its net effect may be to increase or decrease neuronal activity, depending on which other neurotransmitter (e.g., dopamine, GABA, serotonin, glutamate) is affected. Pharmacologic concentrations of GHB act primarily to decrease neuronal activity through $GABA_B$ modulation, but this brief period of inhibition may be followed by increased neuronal activity; this effect may explain the initial sedative effect of GHB when administered at night, followed by increased alertness the following day.[207] GHB effects on the CNS include sedation and, in higher doses, coma. GHB has few effects on cardiovascular or respiratory systems.

GHB is absorbed rapidly after oral administration, particularly because it is administered as a liquid, with peak concentrations approximately 30 to 60 minutes after administration. GHB is not bound to plasma protein. It is metabolized to a limited extent to GABA. GHB is also decomposed to water and carbon dioxide and exhaled. The mean half-life is quite short, ranging from 20 to 70 minutes (mean, 53 minutes).[208]

GHB is subjectively sedating. In healthy subjects, GHB increases stage 3/4 sleep, decreases stage 1 sleep, and reduces REM sleep latency.[209,210] When administered to patients with narcolepsy, its PSG effects include reduced REM latency and awakenings and increased stage 3/4 sleep, sleep efficiency, and sleep duration.[207,211] A study in fibromyalgia patients showed similar results, with reduced sleep latency and REM sleep, increased stage 3/4 sleep, and a reduction in alpha EEG activity intrusion during NREM sleep.[212] GHB has not been formally assessed for its hypnotic properties in patients with other types of insomnia.

The rapid sedative effects of GHB, particularly when combined with alcohol, have led to abuse. Other side effects of GHB include excess salivation, increased dreaming, sleepwalking, and gastrointestinal effects such as vomiting. It is also associated with amnesia, similar to other BzRA hypnotic agents. In overdoses, GHB can be associated with acute

delirium.[208] High-dose recreational users of GHB have been described to have a withdrawal symptom characterized by insomnia, tremor, and anxiety. Concerns regarding safety and abuse, as well as its strict regulation by the U.S. Drug Enforcement Agency, make sodium oxybate an impractical choice for the treatment of insomnia.

CLINICAL PEARL

Antidepressants, antihistamines, and other drugs (Tables 42-4 and 42-5) are often considered safer alternatives to BzRAs for treatment of insomnia. However, in most cases, their efficacy has not been well demonstrated, and they can have clinically important side effects. Understanding the clinical pharmacology and known sleep effects of these medications is critical to their rational use in clinical practice. These drugs may be useful when BzRAs are contraindicated or ineffective or when comorbidities such as severe psychiatric illness are present.

Table 42-4 Summary of Other Drugs Used to Treat Insomnia*

Drug	Drug Type	Time to Maximal Concentration	Metabolism	Elimination Half-Life	Mechanism of Action
Melatonin	Hormone	20–60 min	Conjugation; oxidation by CYP enzymes	40–60 min	Agonist at melatonin type 1 and type 2 receptors
Diphenhydramine	Ethanolamine antihistamine	2–2.5 h	Hepatic demethylation, oxidation	4–8 h	Antagonize H_1 receptors
Doxylamine	Ethanolamine antihistamine	2–3 h	Most excreted unchanged in urine; some hepatic metabolism	10	Antagonize H_1 receptors
Valerian	Plant extract	Uncertain because of multiple constituents	Uncertain because of multiple constituents	Uncertain because of multiple constituents	Uncertain; may increase GABA formation, interact with L-amino acid transporter receptor, or act as adenosine receptor agonist
Gabapentin	Anticonvulsant (structural analog of GABA)	3–3.5 h	Renal excretion (unchanged)	5–9 h	Uncertain; may affect GABA release or interact with L-amino acid transporter protein
Tiagabine	Anticonvulsant	1–1.5 h	CYP3A4	8 h	Inhibits GABA transporter GAT-1
Suvorexant	Orexin receptor antagonists	30 min to 6 h	CYP3A4, CYP2C19	9–12 h	Blocks the binding of wake-promoting neuropeptides orexin A and orexin B to receptors OX1R and OX2R
Choral hydrate	Two-carbon molecule	Short	Converted to trichloroethanol, which undergoes conjugation	5–10 h (for trichloroethanol)	Barbiturate-like effect at GABA-A receptors
Olanzapine	Thienobenzodiazepine antipsychotic	4–6 h	CYP1A2, CYP2D6	20–54 h	Antagonizes H_1, α_1, α_2, M_1, 5-HT_2, D_2 receptors
Quetiapine	Dibenzothiazepine antipsychotic	1–2 h	CYP3A4	6 h	Antagonizes H_1, α_1, M_1, 5-HT_2, D_2 receptors
Gamma–hydroxybutyrate (GHB)	Endogenous four-carbon molecule	30–45 min	Metabolized to GABA, succinic semialdehyde, H_2O and CO_2	20–70 min	May act directly as neurotransmitter, increase brain dopamine levels

5-HT, 5-Hydroxytryptamine (serotonin); α, α-adrenergic receptor; CYP, cytochrome P-450 system (individual letters and numbers represent CYP families); D_2, dopamine type 2 receptor; GABA, gamma-aminobutyric acid; H_1, histamine type 1 receptor; M_1, muscarinic cholinergic type 1 receptor.
*Data compiled from sources indicated in text.

Table 42-5 Polysomnographic Effects of Other Drugs Used to Treat Insomnia*

Drug	Sleep Latency	Sleep Continuity†	Stage 3/4 NREM Sleep Amount (%)	REM sleep	Other
Melatonin	↓	↔ to ↑	↔	↔	
Diphenhydramine	↓	↔ to ↑	↔ to ↑	↓	
Valerian	↓	↔ to ↑	↔ to ↑	↔ to ↑	Inconsistent effects on sleep continuity, stage 3/4 across studies
Gabapentin	↔	↔ to ↑	↑	↔	Reduced periodic limb movements
Tiagabine	↔	↑	↑	↔	Results based on single study
Suvorexant	↔	↑	↔	↑	
Chloral hydrate	↓	↑	↔	↔ to ↓	Rapid tolerance may develop
Olanzapine	↔ to ↓	↑	↑	↔ to ↓	Reports of increased periodic limb movements, sleep-eating
Gamma-hydroxybutyrate (GHB)	↔ to ↓	↑	↑	↔ to ↓	↓ Alpha NREM intrusions in fibromyalgia patients

*Reported effects are based on preponderance of evidence from published studies (see Buysse[131] and text for details). Many effects are inconsistent between individual studies. ↑ Indicates increase from pretreatment baseline; ↓ indicates decrease from pretreatment baseline; ↔ indicates no change from pretreatment baseline.

†Sleep continuity refers to the proportion of sleep relative to wakefulness after sleep onset, as reflected by measures such as sleep efficiency. Other indicators of sleep continuity, such as wakefulness after sleep onset or number of awakenings, would have opposite signs. Thus ↑ indicates improvement in overall sleep continuity.

SUMMARY

Pharmacologic treatment of insomnia is managed by hypnotic drugs from several classes. Although BzRAs remain the most widely used FDA-approved hypnotics, melatonin receptor agonists are also FDA approved for the treatment of insomnia. In addition, various other drugs originally developed as antidepressants, anticonvulsants, and antipsychotics, as well as hormones and other "natural" substances, have been used as hypnotics. Safe use of these drugs in clinical practice depends on the knowledge of pharmacokinetics, pharmacodynamics, sleep effects, and side effects. Sedating tricyclic and other antidepressant drugs primarily act through serotonin, norepinephrine, and histamine receptor effects but show considerable heterogeneity in terms of biologic half-lives, receptor pharmacology, and sleep effects. Efficacy data for most of these drugs on sleep continuity have been derived from studies of depressed patients. Some antidepressants increase slow wave sleep and reduce REM sleep. Low-dose doxepin, one of the tricyclic compounds, is FDA approved for treatment of insomnia. Melatonin and melatonin receptor agonists reduce sleep latency by acting on melatonin receptors in the suprachiasmatic nucleus and cortical regions. Two melatonin receptor agonists are FDA approved, one of which, ramelteon, is approved for insomnia. Antihistamines antagonize the effects of histamine, a wake-promoting neurotransmitter synthesized in posterior hypothalamus with widespread cortical projections. They are widely used because of their subjective sedation. A limited amount of evidence exists regarding their effects on nocturnal sleep, and like antidepressants, they can have clinically important side effects. Valerian extracts have uncertain pharmacokinetics and mechanisms of action. They appear to affect primarily sleep latency, although some studies also show increased slow wave sleep. Small numbers of studies suggest that sedating antipsychotic drugs, tiagabine, gabapentin, and sodium oxybate (GHB) may all increase slow wave sleep, with variable effects on sleep continuity. These drugs have a wide variety of receptor effects, and the mechanisms of their effects on human sleep are less well understood. Few clinical studies have been conducted with any of these agents in patients with insomnia. A newly approved dual orexin receptor antagonist, suvorexant, has been shown to improve sleep continuity in individuals with insomnia. Future studies will be needed to understand the appropriate role of these drugs in the treatment of sleep disorders and specifically how they will fit into emerging treatment algorithms for insomnia.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Paula Schweitzer, PhD, and Douglas E. Moul, MD, MPH, to earlier versions of this chapter, including the chapter in *Principles and Practice of Sleep Medicine*, 4th edition, 2005. Supported by National Institutes of Health grants MH024652, AG015138, AG020677, and AG024827.

Selected Readings

Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep* 2014;**37**:343–9.

Callander GE, Olorunda M, Monna D, et al. Kinetic properties of "dual" orexin receptor antagonists at OX1R and OX2R orexin receptors. *Front Neurosci* 2013;**7**:230.

Farkas R. *Suvorexant safety and efficacy*. U.S. Food and Drug Administration; 2013. p. 1–58.
Gnjidic D, Hilmer SN, Hartikainen S, et al. Impact of high risk drug use on hospitalization and mortality in older people with and without Alzheimer's disease: a national population cohort study. *PLoS ONE* 2014;**9**:e83224.
Hickie IB, Rogers NL. Novel melatonin-based therapies: potential advances in the treatment of major depression. *Lancet* 2011;**378**:621–31.
Krystal AD, Durrence HH, Scharf M, et al. Efficacy and safety of doxepin 1 mg and 3 mg in a 12-week sleep laboratory and outpatient trial of elderly subjects with chronic primary insomnia. *Sleep* 2010;**33**:1553–61.
McCleery J, Cohen DA, Sharpley AL. Pharmacotherapies for sleep disturbances in Alzheimer's disease. *Cochrane Database Syst Rev* 2014;(3): CD009178.
Mignot E, Taheri S, Nishino S. Sleeping with the hypothalamus: emerging therapeutic targets for sleep disorders. *Nat Neurosci* 2002;**5**(Suppl.): 1071–5.
Plante DT, Jensen JE, Schoerning L, Winkelman JW. Reduced gamma-aminobutyric acid in occipital and anterior cingulate cortices in primary insomnia: a link to major depressive disorder? *Neuropsychopharmacology* 2012;**37**:1548–57.
Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* 2005;**437**:1257–63.
Vermeeren A, Sun H, Vuurman EF, et al. On-the-road driving performance the morning after bedtime use of suvorexant 20 and 40 mg: a study in non-elderly healthy volunteers. *Sleep* 2015 [Epub ahead of print].
Wang YQ, Takata Y, Li R, et al. Doxepin and diphenhydramine increased non-rapid eye movement sleep through blockade of histamine H1 receptors. *Pharmacol Biochem Behav* 2015;**129**:56–64.

A complete reference list can be found online at ExpertConsult.com.

Chapter
43

Wake-Promoting Medications: Basic Mechanisms and Pharmacology

Seiji Nishino; Emmanuel Mignot

Chapter Highlights

- Central nervous system stimulants currently used in sleep medicine include amphetamine-like compounds (L- and D-amphetamine and methamphetamine, L- and D-methylphenidate, pemoline), mazindol, modafinil-armodafinil, some antidepressants with stimulant properties (e.g., bupropion), and caffeine.
- The effects of most of these drugs on wakefulness are primarily mediated by an inhibition of dopamine reuptake and transport and in some cases by increased dopamine release. Inhibition of adrenergic uptake also likely has some stimulant effects.
- Biogenic amine transporters (for dopamine, norepinephrine, and serotonin) are located at nerve terminals and are important in terminating transmitter action and maintaining transmitter homeostasis. The results of pharmacologic studies using animals suggest the importance of the dopamine transporter for the mode of action of amphetamines and amphetamine-like compounds on wakefulness.
- The mode of action of modafinil, a more recent compound that rapidly became a first-line treatment for excessive daytime sleepiness in narcolepsy, is controversial but is increasingly suggested to primarily involve dopamine reuptake inhibition.
- Other agents with mechanisms of action involved in wake promotion include adenosine receptor antagonists, such as those found in caffeine. More recently, novel classes of wake-promoting therapeutics are being developed, including glutamatergic and histaminergic modulators, and preclinical and clinical evaluations are in progress.

CENTRAL NERVOUS STIMULANTS: DEFINITIONS

Although widely used, the term central nervous system (CNS) stimulant is a loosely defined scientific term. In *Drugs and the Brain* by S. Snyder, stimulants are "drugs that have an alerting effect; they improve the mood and quicken the intellect." In *Handbook of Sleep Disorders* by J. D. Parkes, CNS stimulation implies "an increase in neuronal activity due to enhanced excitability, with a change in the normal balance between excitatory and inhibitory influences. This may result from blockage of inhibition, enhancement of excitation, or both." In *A Primer of Drug Action* by R. M. Julien, the term "psychomotor stimulants (psychostimulants)" is used, and "psychostimulants" are said to induce excitement, alertness, euphoria, a reduced sense of fatigue, and increased motor activity. Psychostimulants include dopamine (DA) uptake blockers, DA-releasing agents, adenosine receptor blockers, and acetylcholine receptor stimulants. In *The Pharmacological Basis of Therapeutics* by Goodman and Gilman, the term "indirect sympathomimetic amines" refers to amphetamines as the "most potent compounds with respect to stimulation of the CNS."

In this chapter, the generic term *CNS stimulants* will be used for all wake-promoting compounds of potential use in the treatment of excessive daytime sleepiness (EDS) (see Chapters 4 and 44 for the classification of EDS disorders and the indication of CNS stimulants for patients affected with sleep disorders). EDS is a common symptom in patients with sleep disorders and in the general population at large. CNS stimulants are generally effective in patients with EDS independently of its underlying cause; however, they sh ould be used cautiously because of their potential for misuse and abuse. This chapter reviews the neurochemical, neurophysiologic, and neuropharmacologic properties of the CNS stimulants most commonly used in sleep medicine. This will be followed by a perspective on future stimulant treatments.

AMPHETAMINES AND AMPHETAMINE-LIKE COMPOUNDS

Historical Perspective

Amphetamine was first synthesized by Alles in 1897, but its stimulant effects were not recognized until 1929. Alles wanted to find a synthetic substitute for the recently banned ephedrine, a compound isolated from the *Ephedra vulgaris* plant in 1925. Amphetamine increases energy, elevates mood, prevents fatigue, increases vigilance and prevents sleep, stimulates respiration, and causes electrical and behavioral arousal from natural or drug-induced sleep. It was rapidly shown to be a safer and cheaper alternative to ephedrine as a stimulant. In World War II, amphetamine was supplied to paratroopers

and commandos. British troops alone were issued 72 million tablets. In Japan, methamphetamine, initially used for munitions factory workers, flooded the civilian market at the end of the war; 5% of the Japanese population between the ages of 16 and 25 years became dependent on the drug. More than 50 "amphetamine" preparations containing amphetamine or derivatives, alone or in combination with other drugs (most notably barbiturates), were on the market after World War II.

Narcolepsy was probably the first condition for which amphetamine was used clinically. It revolutionized therapy for the condition, although it was not curative. The piperazine derivative of amphetamine, methylphenidate, was introduced in 1959 by Yoss and Daly.[1] The use of amphetamine in treating parkinsonism dates back to 1937, when it was first used to alleviate the muscular rigidity of postencephalitis parkinsonism. By 1968, its use in the treatment of this condition was largely suspended owing to the use of more effective dopaminergic agents. Until the dangers of amphetamine dependence and abuse became recognized, amphetamine was widely used in the treatment of obesity. It was also prescribed in the treatment of sedative abuse and alcoholism to offset sleepiness and lethargy.

Bradley and Bowen (1941) were the first to report on the use of amphetamine to modify antisocial behavior in children[2]: "When children are withdrawn or lethargic, the amphetamine tended to make them more alert, more accessible to persons and the environment." A paradoxical calming effect was also noted in some children and aggressive adults. Most notably, a selected group of children who were "hyperactive" tended to move less, to be calmer, and to be less quarrelsome after being treated with amphetamine. In 1958, methylphenidate was introduced to treat hyperactivity in children.[3] These observations preceded reports on the effects of amphetamine and methylphenidate in children who are hyperkinetic, a disorder now referred to as attention-deficit/hyperactivity disorder (ADHD).

Although no controlled trials have investigated the use of stimulants in depression, many case series suggest the effectiveness in some treatment-resistant cases. The use of stimulants with monoamine oxidase (MAO) inhibitors is generally not advised but has not been reported to induce significant hypertension or hyperthermia. Amphetamines are often prescribed in combination with low (anticataplectic) doses of tricyclic agent in narcolepsy-cataplexy without any problem, and combining these substances in depression has been shown to be effective, although not recommended because of the risk for dependence and abuse. Part of the beneficial effects of amphetamine on depression may be due a reduction of fatigue and apathy rather than a genuine antidepressant effect.

From a historical perspective, the number of indications for amphetamine stimulants has narrowed considerably over the years to primarily include narcolepsy, ADHD, and treatment-resistant depression. The rationale for this change has been the realization of the risk for abuse and dependence with these compounds. The introduction of other effective therapies for these conditions (e.g., modafinil for narcolepsy, atomoxetine for ADHD) has also led to narrower indications, although many new formulations and isomer-specific preparations have been recently developed and are increasingly used, mostly for the treatment of ADHD.

Structure-Activity Relationships and Major Chemical Entities

Distinguishing *potency* and *efficacy* is helpful to the understanding of the pharmacology of stimulant drugs; these terms are too often used incorrectly when using colloquial language. Efficacy refers to the therapeutic effects that can be achieved by a drug, whereas potency describes the amount of the drug needed to achieve therapeutics effects. In general, potency correlates with the affinity of the drug for its target, whereas efficacy reflects how much maximal effect can be achieved when the targets are fully occupied. These two characteristics are uncorrelated.

Phenylisopropylamine (amphetamine) has a simple chemical structure resembling endogenous catecholamines (Figure 43-1). This scaffold forms the template for a wide variety of pharmacologically active substances. Although amphetamine possesses strong central stimulant effects, minor modifications can result in a broad spectrum of effects, including nasal decongestion, anorexia, vasoconstriction, antidepressant effects (for the MAO inhibitor tranylcypromine), or hallucinogenic properties (methylenedioxymethamphetamine [MDMA] and methylenedioxyamphetamine [MDA]).

The phenylisopropylamine molecule can be divided into three structural components: an aromatic nucleus, a terminal amine, and an isopropyl side chain. Substitution on the aromatic nucleus generally produces less potent, if not entirely inactive, stimulants.[4] The substitution of two or more methoxy groups and the addition of ethyl, methyl, or bromine groups on the aromatic nucleus creates hallucinogens of various potencies. MDMA ("Ecstasy") is built on a methamphetamine backbone, with a dimethoxy ring extending from the aromatic group. If a similar compound is synthesized with a primary amine (without the methyl group), then it creates "Love" (MDA). Substitution at the amine group is the most common alteration. Methamphetamine, which is characterized by an additional methyl group attached to the amine (a secondary substituted amine), is more potent than amphetamine, probably because of increased CNS penetration. An intact isopropyl side chain appears to be needed to maintain stimulant efficacy. Changing the propyl to an ethyl side chain, for example, creates phenylethylamine and an endogenous neuroamine, which has mood- and energy-enhancing properties but is less potent and has a much shorter half-life than amphetamine.

The pharmacologic effects of most amphetamine derivatives are isomer specific. These differential effects occur both at the pharmacokinetic level (absorption, brain penetration, metabolism, distribution volume, elimination) and in terms of pharmacodynamic profile (actual pharmacologic effects). D-Amphetamine, for example, is a far more potent stimulant than L-amphetamine. In electroencephalographic (EEG) studies, D-amphetamine is four times more potent in inducing wakefulness than L-amphetamine.[5] The relative effects of the D- and L-isomers of amphetamine on norepinephrine (NE) and DA transmission explains some of these pharmacodynamic differences (for details, see the pharmacology discussion for each compound). Not all effects are stereospecific, however. For example, both enantiomers are equipotent in suppressing rapid eye movement (REM) sleep in humans and rats and in producing amphetamine psychosis.

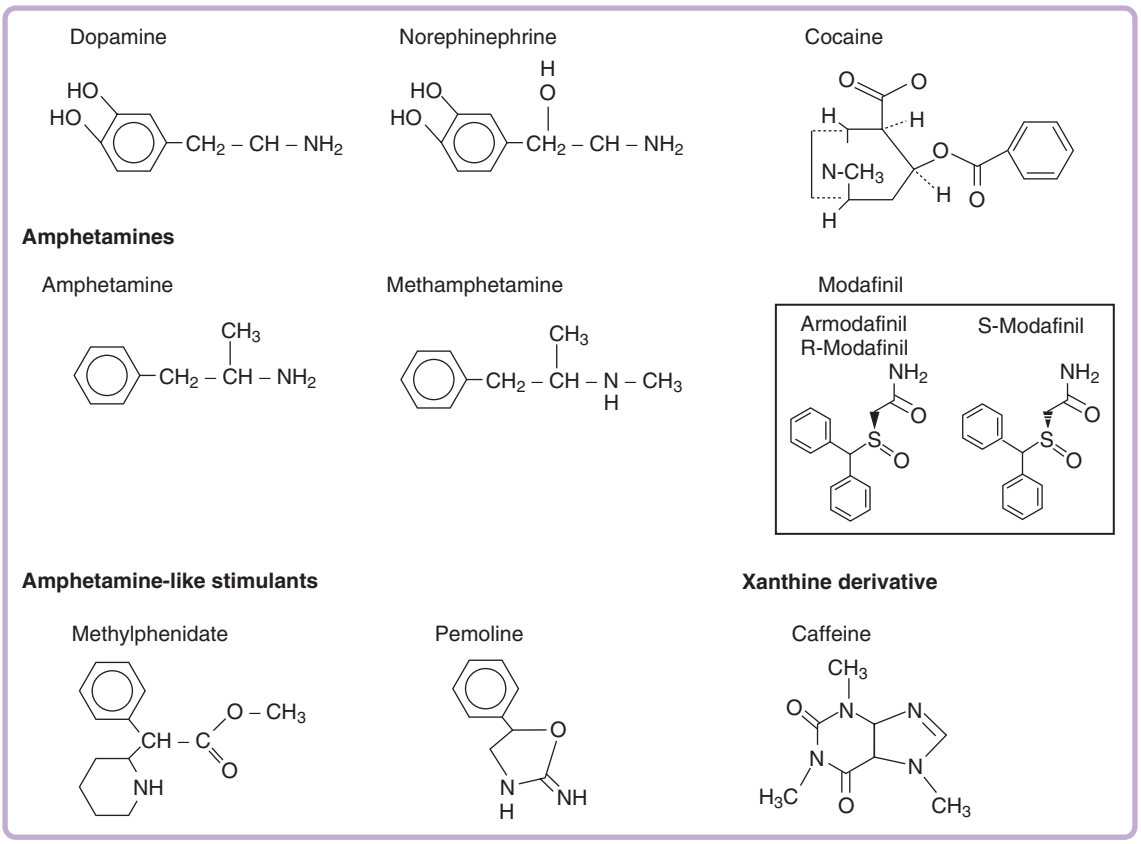


Figure 43-1 Chemical structures of amphetamine-like stimulants, modafinil, armodafinil, and xanthine derivatives compared with catecholamine.

Amphetamine-like compounds, such as methylphenidate, pemoline, and fencamfamin, are structurally similar to amphetamines; all compounds include a benzene core with an ethylamine group side chain (see Figure 43-1). Both methylphenidate and pemoline were commonly used for the treatment of EDS in narcolepsy, but pemoline has been withdrawn from the market in several countries because of liver toxicity (Table 43-1). The most commonly used commercially available form of methylphenidate is a racemic mixture of both a D- and L-enantiomer. In this preparation, the D-methylphenidate mainly contributes to its clinical effects, especially after oral administration. This is because L-methylphenidate, but not D-methylphenidate, undergoes a significant first-pass metabolism (by deesterification to L-ritalinic acid). A single isomer form of D-methylphenidate is also marketed under the brand name of Focalin.

Cocaine also mediates its psychostimulant effects by blocking catecholamine reuptake (mainly DA), but its structure is different from amphetamine-like compounds (see Figure 43-1). The fact that cocaine and some DAT inhibitors are drugs of abuse is responsible for schedule labeling of such drugs by the U.S. Food and Drug Administration (FDA).

Amphetamines are highly lipid soluble molecules that are well absorbed by the gastrointestinal tract. Peak levels are achieved approximately 2 hours after oral administration, with rapid tissue distribution and brain penetration. Protein binding is highly variable, with an average volume of distribution of 5 L/kg.

Both hepatic catabolism and renal excretion are involved in the inactivation of amphetamine. Amphetamine can be metabolized in the liver through either aromatic or aliphatic hydroxylation, yielding parahydroxyamphetamine or norephedrine, respectively, both of which are biologically active. The metabolism of amphetamine and amphetamine-like compounds is pH dependent. Amphetamine is metabolized into benzoic acid (23%), which is subsequently converted to hippuric acid or to parahydroxyamphetamine (2%). This in turn is converted to parahydroxynorephedrine (0.4%). Thirty-three percent of the oral dose is excreted unchanged in the urine. Urinary excretion of amphetamine and many amphetamine-like stimulants is greatly influenced by urinary pH. At urinary pH of 5 the elimination half-life of amphetamine is short, about 5 hours, but at pH of 7.3 it increases to 21 hours. Sodium bicarbonate will delay excretion of amphetamine and prolong its clinical effects, whereas ammonium chloride will shorten amphetamine action (and can possibly induce toxicity).

Methylphenidate is almost totally and rapidly absorbed after oral administration. Methylphenidate has low protein binding (15%) and is short acting; effects last approximately 4 hours, with a half-life of 3 hours. The primary means of clearance is through the urine, in which 90% is excreted.

Molecular Targets of Amphetamine Action

The molecular targets mediating amphetamine-like stimulant effects are complex and vary depending on the specific

Table 43-1 Commonly Used Pharmacologic Compounds for Excessive Daytime Sleepiness

Stimulant Compound	Usual Daily Doses*	Half-Life (hr)	Side Effects, Notes
Amphetamines and Amphetamine-like CNS Stimulants			
D-Amphetamine sulfate (schedule II)	5–60 mg (15, 100 mg)	16–30	Irritability, mood changes, headaches, palpitations, tremors, excessive sweating, insomnia
Methamphetamine HCl[†] (schedule II)	5–60 mg (15, 80 mg)	9–15	Same as D-amphetamine; may have a greater central over peripheral effects than D-amphetamine[‡]
Methylphenidate HCl (schedule II)	10–60 mg (30, 100 mg)	~3	Same as amphetamines; better therapeutic index than D-amphetamine with less reduction of appetite or increase in blood pressure; short duration of action
Pemoline (schedule IV)	20–115 mg (37.5, 150 mg)	11–13	Less sympathomimetic effect, milder stimulant, slower onset of action; occasionally produces liver toxicity; had been withdrawn from the U.S. market
Dopamine and Norepinephrine Uptake Inhibitor			
Mazindol (schedule IV)	2–6 mg (NA)	10–13	Weaker CNS stimulant effects; anorexia, dry mouth, irritability, headaches, gastrointestinal symptoms; reported to have less potential for abuse
Other Agents for Treatment of EDS			
Modafinil[§] (schedule IV)	100–400 mg (NA)	9–14	No peripheral sympathomimetic action; headaches, nausea; reported to have less potential for abuse
Armodafinil (schedule IV)	100–300 mg (NA)	10–15	Similar to those of modafinil
MAO Inhibitors with Alerting Effect			
Selegiline	5–40 mg (NA)	2	Low abuse potential; partial (10%-40%) interconversion to amphetamine
Xanthine Derivative			
Caffeine[¶]	100–200 mg (NA)	3–7	Weak stimulant effect; 100 mg of caffeine roughly equivalent to one cup of coffee; palpitations, hypertension

*Dosages recommended by the American Sleep Disorders Association are listed in parentheses (usual starting dose and maximal dose recommended).
[†]Methamphetamine is reported to have more central effects and may predispose more to amphetamine psychosis. The widespread misuse of methamphetamine has led to severe legal restriction on its manufacture, sale, and prescription in many countries.
[‡]L-Amphetamine (dose range, 20-60 mg) is not available in the United States but probably has no advantage over D-amphetamine in the treatment of narcolepsy (slightly weaker stimulant).
[§]The half-life of the s-enantiomer of modafinil is short (3-4 hr) and thus the half-life of racemic modafinil mostly reflects the half-life of armodafinil (r-enantiomer).
[¶]Caffeine can be brought without prescription in the form of tablets (No Doz, 100 mg; Vivarin, 200 mg caffeine) and is used by many patients with narcolepsy before diagnosis.
CNS, Central nervous system; EDS, excessive daytime sleepiness; MAO, monoamine oxidase; NA, not applicable.

analogue or isomer used and the dose administered. Amphetamine increases catecholamine (DA and NE) release and inhibits reuptake from presynaptic terminals. This results in an increase in catecholamine concentrations in the synaptic cleft and enhances postsynaptic stimulation. The presynaptic modulations by amphetamines are mediated by specific catecholamine transporters[6] (Figure 43-2). Axelrod and colleagues first demonstrated that epinephrine could be rapidly and selectively taken up by the heart, spleen, and glandular organs, each of which has significant sympathetic innervation. It was subsequently discovered that NE-containing neurons bind and take up NE against a concentration gradient, suggesting the existence of selective norepinephrine transporters (NETs). Further experiments also found that these transporters can not only carry catecholamine back into nerve terminals but can release catecholamines through reverse efflux.

The molecules responsible, the dopamine transporter (DAT) and the NET, have now been cloned and characterized. The DAT and NET proteins are about 620 amino acid proteins with 12 putative membrane-spanning regions. Amphetamine derivatives are known to inhibit the uptake and enhance the release of DA, NE, or both by interacting with the DAT and the NET. These transporters normally move DA and NE from the outside to the inside of the cell. This process is sodium dependent; sodium and chloride bind to the

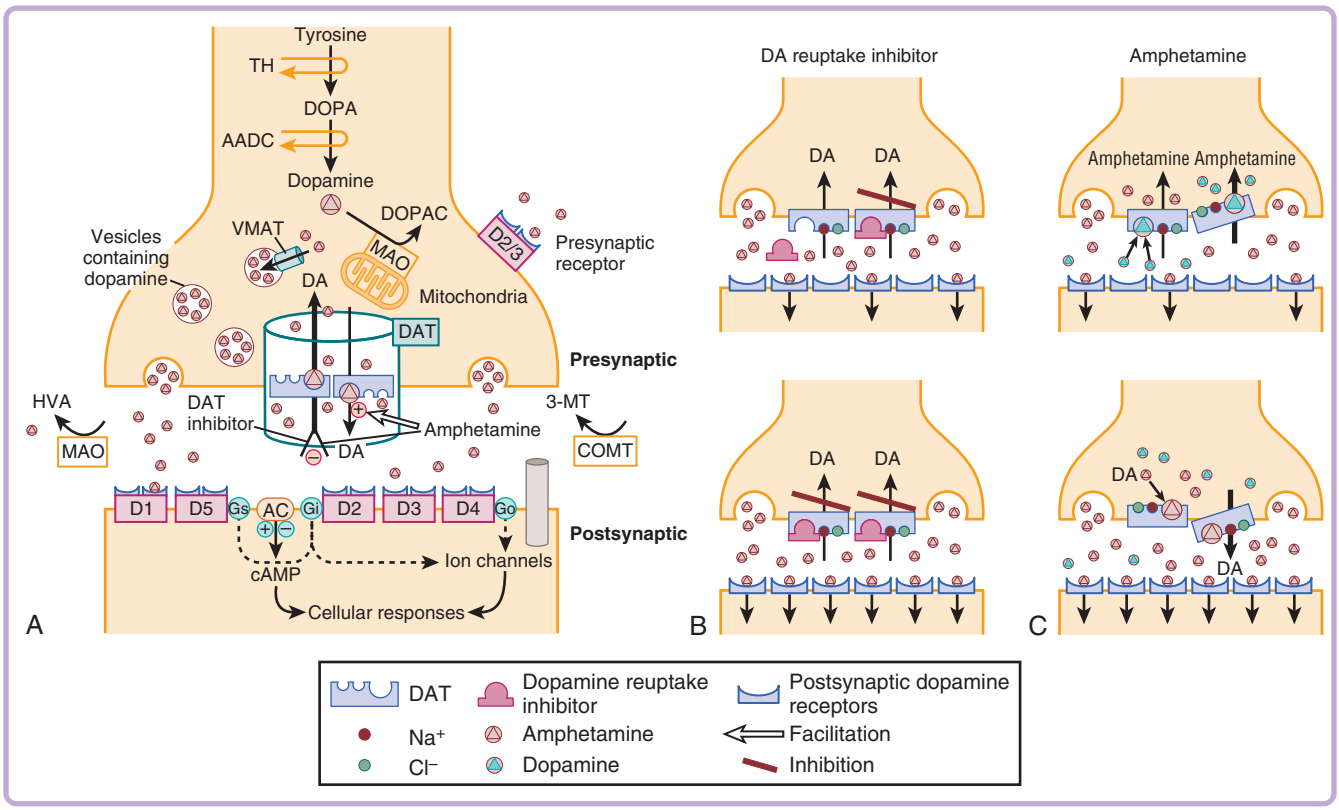


Figure 43-2 A, Schematic representations of dopaminergic terminal neurotransmission in relation to mode of action of dopamine (DA) reuptake inhibitors and amphetamine and effects of DA reuptake inhibitors and amphetamines at the dopaminergic nerve terminal. Dopamine transporter (DAT) is one of the most important molecules located at the dopaminergic nerve terminals and regulates dopaminergic neurotransmission. Amphetamine interacts with the DAT carrier to facilitate DA release from the cytoplasm through an exchange diffusion mechanism. At higher intracellular concentrations, amphetamine also disrupts vesicular storage of DA and inhibits monoamine oxidase (MAO). Both these actions increase cytoplasmic DA concentrations. Amphetamine also inhibits DA uptake by virtue of its binding to and transport by the DAT. These mechanisms all lead to an increase in DA synaptic concentrations, and these are independent of the phasic activity of the neurons. Increased synaptic concentration of DA stimulates postsynaptic DA receptors (D1 type [1, 5] and D2 type [2, 3, 5] receptors). **B,** Sodium and chloride bind to the DAT to immobilize it at the extracellular surface. This alters the conformation of the DA binding site on the DAT to facilitate substrate (i.e., DA) binding. DAT reuptake inhibitors bind to DAT competitively and inhibit DA-DAT bindings, resulting in increasing DA concentrations in the synaptic cleft. **C,** Amphetamine, in competition with extracellular DA, binds to the transporter. Substrate binding allows the movement of the carrier to the intracellular surface of the neuronal membrane, driven by the sodium and amphetamine concentration gradients, resulting in a reversal of the flow of DA uptake. Amphetamine dissociates from the transporter, making the binding site available to cytoplasmic DA. DA binding to the transporter enables the movement of the transporter to the extracellular surface of the neuronal membrane, as driven by the favorable DA concentration gradient. DA dissociates from the transporter, making the transporter available for amphetamine and thus another cycle. AADC, Aromatic acid decarboxylase; AC, adenylyl cyclase; cAMP, cyclic adenosine monophosphate; COMT, catechol-Omethyltransferase; D1 to D5, dopamine receptors 1 through 5; DOPA, 3,4-dihydroxyphenylalanine; DOPAC, dihydroxyphenylacetic acid; Gi, Go, and Gs, protein subunits; HVA, homovanillic acid; TH, tyrosine hydroxylase; VMAT, vesicular monoamine transporter.

DAT or NET to immobilize it at the extracellular surface and to alter the conformation of the DA or NE binding site so that it facilitates substrate binding. Substrate binding allows movement of the carrier to the intracellular surface of the neuronal membrane, driven by sodium concentration gradients. Interestingly, in the presence of some drugs such as amphetamine, the direction of transport appears to be reversed (see Figure 43-2). DA and NE are thus moved from the *inside* of the cell to the *outside* through a mechanism called exchange diffusion, which occurs at low doses (1 to 5 mg/kg) of amphetamine. This mechanism, rather than a simple inhibition of monoamine reuptake, is involved in the enhancement of extracellular catecholamine release by amphetamine. It explains why amphetamine in particular is more potent than expected based on its relatively low binding affinity for DAT and NET.[7,8] A recent in vitro experiment has shown that amphetamine transport causes an inward sodium current.[6] As intracellular sodium ions become more available, a DAT-mediated reverse transport of DA occurs, producing DA release through the DAT transporter.

At higher dose other effects are involved. Increased serotonin (5-HT) release is also observed. Moderate to high doses of amphetamine (>5 mg/kg) also interact with the vascular monoamine transporter 2 (VMAT2).[6] The vesicularization of the monoamines (DA, NE, 5-HT, and histamine) in the CNS is dependent on VMAT2; VMAT2 regulates the size of the vesicular and cytosolic DA pools. Amphetamine is highly lipophilic and easily enters nerve terminals by diffusing across plasma membranes. Once inside, amphetamine depletes vesicular monoamine stores by several mechanisms. First, it binds directly, albeit with low affinity, to VMAT2, thereby inhibiting vesicular uptake. Second, amphetamine, a weak base, diffuses across the vesicular membrane in its uncharged (lipophilic) form and accumulates in the granules in its charged form (because of the lower pH of the synaptic vesicle interior). As vesicular amphetamine concentration increases, the buffering capacity of the catecholamine-containing vesicle is lost. The vesicular pH gradient diminishes, a loss of the free energy necessary for monoamine sequestration occurs, and vesicular monoamine uptake decreases. In addition, the collapse of the gradient purportedly results in a competition for protons between the native monoamines and amphetamine, thereby increasing uncharged vesicular neurotransmitter concentrations. All these mechanisms lead to a diffusion of the native monoamines out of the vesicles into the cytoplasm along a concentration gradient. Amphetamine can therefore be viewed as a physiologic VMAT2 antagonist that releases the vascular DA and NE loaded by VMAT2 into the cytoplasm. The high doses of amphetamine also inhibit MAO activity. These mechanisms, as well as the reverse transport and the blocking of reuptake of DA and NE by amphetamine, all lead to an increase in NE and DA synaptic concentrations,[6] and these are independent on the phasic activity of the neurons.

Various amphetamine derivatives have slightly different effects on all these systems. For example, methylphenidate also binds to the NET and DAT and enhances catecholamine release. It has less effect, however, on the VMAT granular storage site than native amphetamine. Similarly, D-amphetamine has proportionally more releasing effect on the DA versus the NE system when compared with L-amphetamine. MDMA has more effect on 5-HT release than on catecholamine release. Of note, other medications acting on monoaminergic systems, including DA, NE, and 5-HT (e.g., bupropion or mazindol, see later), tend to exert their actions by simply blocking the reuptake mechanism.

Some amphetamines have neurotoxic effects on monoaminergic systems. This is well established for MDMA and serotoninergic systems in both humans and animals. Similarly, amphetamine derivatives with strong effects on monoamine release (typically methamphetamine and less so derivatives with simple monoamine reuptake inhibition effects, e.g., methylphenidate) have neurotoxic effects on DA systems at high dose in animal studies, especially in the context of repeated administration mimicking binges of stimulant abuse administration.

Presynaptic Modulation of the Dopaminergic System Primarily Mediates the Electroencephalographic Arousal Effects

Although amphetamine-like compounds are well known to stimulate catecholaminergic transmission, the exact mechanism by which they promote EEG arousal is still uncertain. A canine model of the sleep disorder narcolepsy has been used to explore its mechanism. Canine narcolepsy is a naturally occurring animal model of the human disorder.[7] Similar to human patients, narcoleptic dogs are excessively sleepy (i.e., short sleep latency), have fragmented sleep patterns, and display cataplexy.[7]

Using narcoleptic and control Doberman dogs, the effects of ligands specific for the DA (GBR-12909, bupropion, and amineptine), NE (nisoxetine and desipramine), or both the DAT and NET (mazindol and nomifensine), as well as amphetamine and a nonamphetamine stimulant, modafinil, were studied to dissect wake-promoting mechanisms.[8] The results indicate that prototypical DA uptake inhibitors such as GBR-12909 and bupropion, dose-dependently increased EEG arousal in narcoleptic dogs, whereas nisoxetine and desipramine, two potent NE uptake inhibitors, had no effect on EEG arousal at doses that almost completely suppressed REM sleep and cataplexy (Figure 43-3).[8] Furthermore, the EEG arousal potency of various DA uptake inhibitors correlated tightly with in vitro DAT-binding affinities (see Figure 43-3), whereas a reduction in REM sleep correlated with in vitro NET-binding affinities,[8] suggesting that DA uptake inhibition is critical for the EEG arousal effects of these compounds.

D-Amphetamine has a relatively low DAT-binding affinity but potently (i.e., need for a low mg/kg dose) promotes alertness (see Figure 43-3). It is also generally considered more efficacious (i.e., can produce more alertness with high dose) than pure DAT reuptake inhibitors in promoting wakefulness. As described in the pharmacology discussion, D-amphetamine not only inhibits DA reuptake but also enhances DA release (at lower dose by exchange diffusion and at higher dose by antagonistic action against VMAT2) and inhibits monoamine oxidation to prevent DA metabolism. The DA-releasing effects of amphetamine are likely to explain the unusually high potency and efficacy of amphetamine in promoting EEG arousal.

In vitro studies have demonstrated that the potency and selectivity for enhancing release or inhibiting uptake of DA and NE vary between amphetamine analogs and isomers.[9] Amphetamine derivatives thus offer a unique opportunity to study the pharmacologic control of alertness in vivo. To dissect wake-promoting effects of amphetamine, the effects of various amphetamine analogs (D-amphetamine, L-amphetamine, and L-methamphetamine) on EEG arousal and in vivo effects on brain extracellular DA levels were compared using narcoleptic dogs.[10] In canine narcolepsy, D-amphetamine is 3 times more potent than L-amphetamine and 12 times more potent than L-methamphetamine in increasing wakefulness and reducing slow wave sleep (see Figure 43-3, *A*).[10]

Microdialysis experiments in the same narcoleptic dogs suggest that wake-promoting effects of amphetamine derivatives correlate well with their effects on dopamine efflux (i.e., intracellular concentration, a net effect of dopamine release and dopamine uptake block). The local perfusion of D-amphetamine raised DA levels nine times above baseline (Figure 43-4, *B*).[10] D-Amphetamine also increased DA levels by up to seven times, but peak DA release was only obtained at the end of the 60-minute perfusion period. L-Methamphetamine did not change DA levels under these conditions. NE was also measured in the frontal cortex

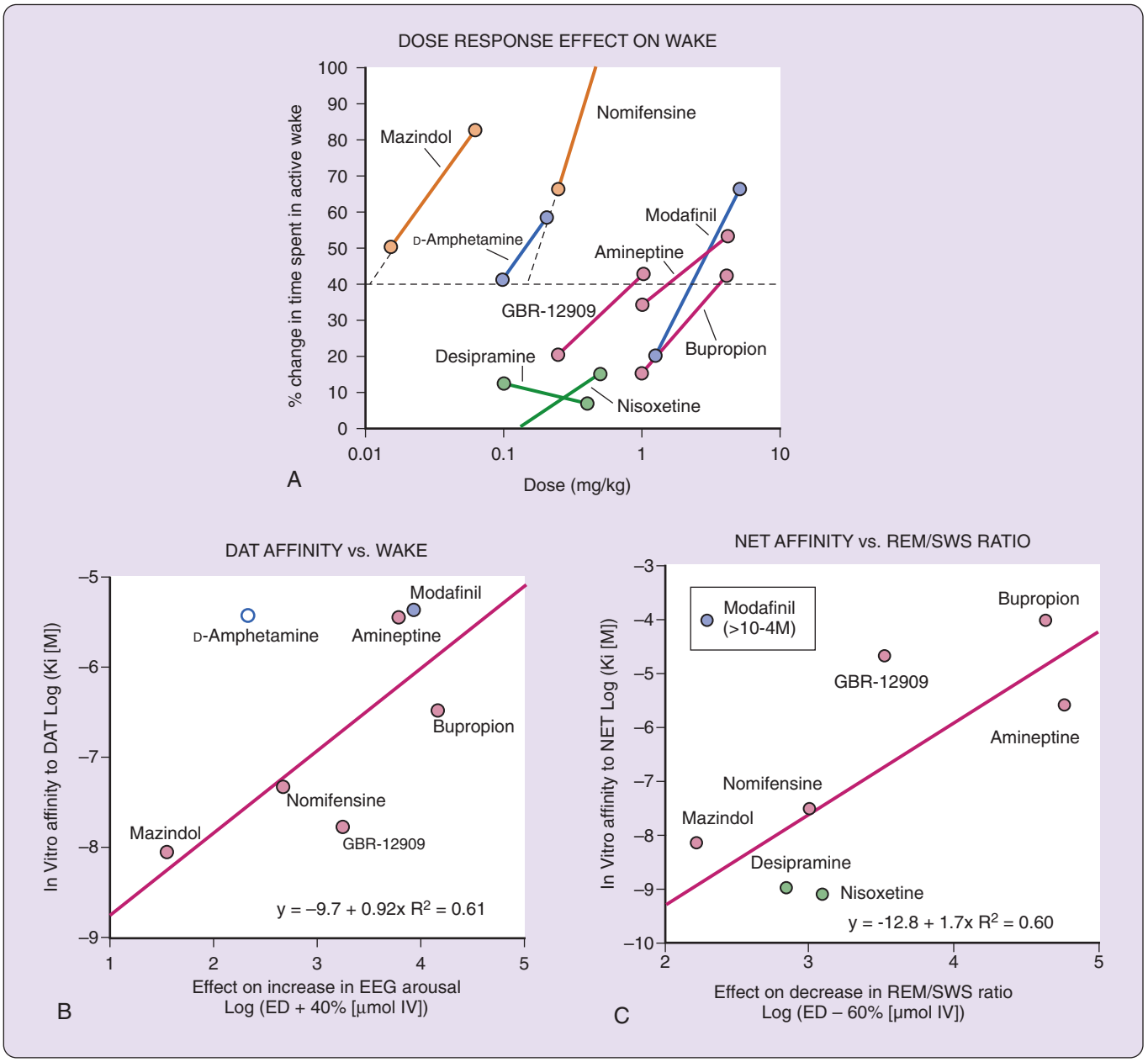


Figure 43-3 Effects of various dopamine (DA) and norepinephrine (NE) uptake inhibitors and amphetamine-like stimulants on the electroencephalographic (EEG) arousal of narcoleptic dogs and correlation between in vivo EEG arousal effects or REM sleep and in vitro DA or NE transporter binding affinities. **A,** The effects of various compounds on daytime sleepiness were studied using 4-hour daytime polygraphic recordings (10:00 to 14:00) in four to five narcoleptic animals. Two doses were studied for each compound. All DA uptake inhibitors and central nervous system (CNS) stimulants dose-dependently increased EEG arousal and reduced slow wave sleep (SWS) in comparison to vehicle treatment. In contrast, nisoxetine and desipramine, two potent NE uptake inhibitors, had no significant effect on EEG arousal at doses that completely suppressed cataplexy. Compounds with both adrenergic and dopaminergic effects (nomifensine, mazindol, D-amphetamine) were active on both EEG arousal and cataplexy. The effects of the two doses performed for each stimulant were used to approximate a dose-response curve; the drug dose that increased the time spent in wakefulness by 40% above baseline (vehicle session) was estimated for each compound. The order of potency of the compounds obtained was: mazindol > (amphetamine) > nomifensine > GBR-12909 > amineptine> (modafinil) > bupropion. **B,** In vitro DAT binding was performed using [^{3}H]-WIN 35,428 onto canine caudate membranes. Affinity for the various DA uptake inhibitors tested varied widely between 6.5 nM and 3.3 mM. In addition, it was found that both amphetamine and modafinil have low but significant affinity (same range as amineptine) for the DAT. A significant correlation between in vivo and in vitro effects was observed for all five DA uptake inhibitors and modafinil. Amphetamine, which had potent EEG arousal effects, has a relatively low DAT binding affinity, suggesting that other mechanisms, most probably monoamine-releasing effects or monoamine oxidase inhibition, are also involved. In contrast, there was no significant correlation between in vivo EEG arousal effects and in vitro NE transporter binding affinities for DA and NE uptake inhibitors. These results suggest that presynaptic enhancement of DA transmission is the key pharmacologic property mediating the EEG arousal effects of most wake-promoting CNS stimulants. **C,** In vitro NE transporter binding was performed using [^{3}H]-nisoxetine. A significant correlation between in vivo potencies on the REM/SWS and in vitro affinity to the NE transporter suggests that presynaptic modulation of NE transmission is important for the pharmacologic control of REM sleep. This may explain why most monoamine uptake inhibitors and monoamine oxidase inhibitors strongly reduce REM sleep in humans and experimental animals.

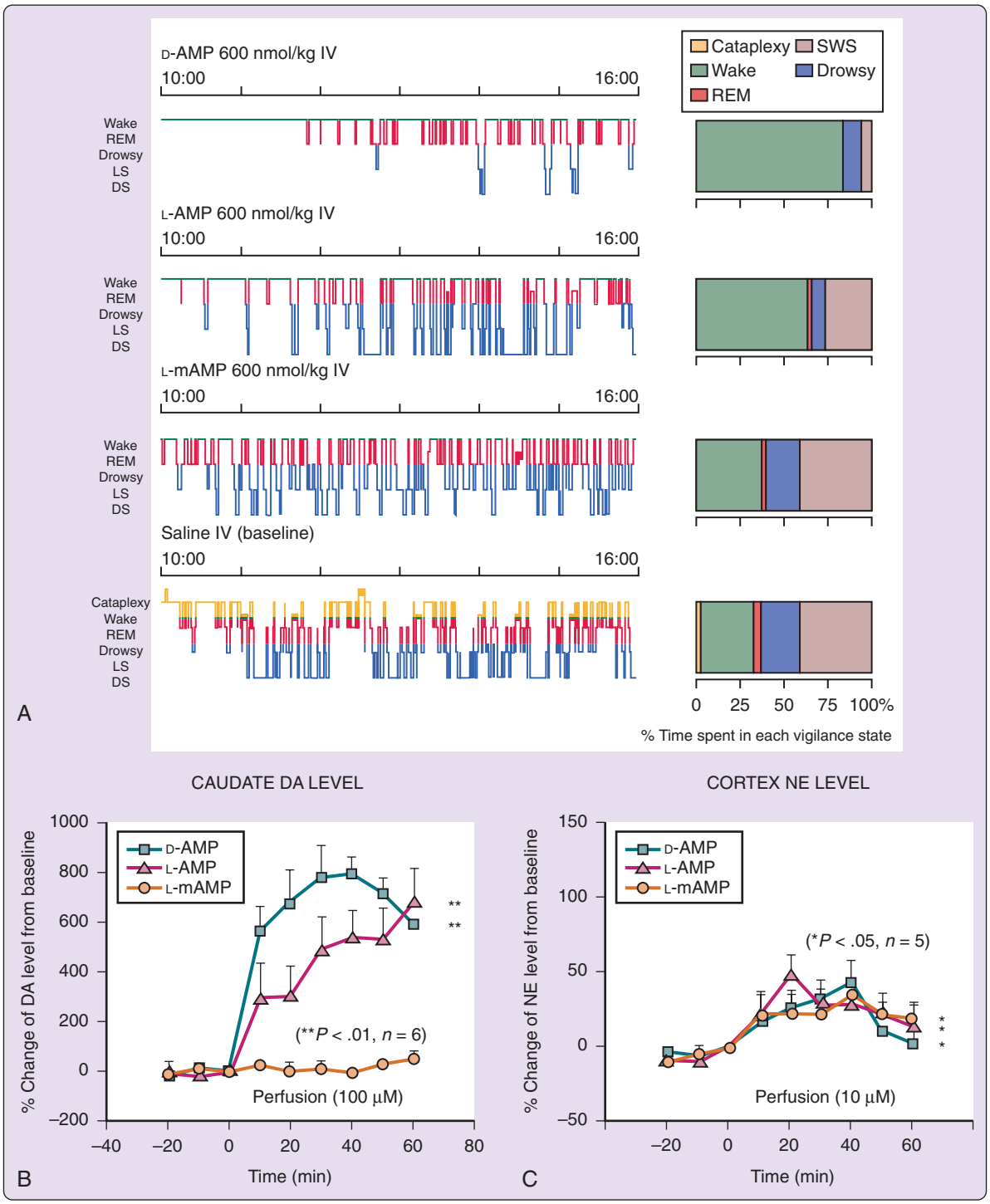


Figure 43-4 A, Effect of amphetamine derivatives on sleep parameters during 6-hour electroencephalogram (EEG) recordings in a narcoleptic dog (600 nmol/kg IV). Representative hypnograms with and without drug treatment are shown. Recordings lasted 6 hours, beginning at approximately 10:00 AM. Vigilance states are shown in the following order from top to bottom: cataplexy, wake, REM sleep, drowsy, light sleep (LS), and deep sleep (DS). The amount of time spent in each vigilance stage (expressed as % of recording time) is shown on the right side of each hypnogram. D-Amphetamine (D-AMP) was found to be more potent than L-amphetamine (L-AMP), and L-methamphetamine (L-mAMP) was found to be the least potent, whereas all isomers equipotently reduced REM sleep. **B,** Local perfusion of D-AMP (100 µM) raised dopamine (DA) levels eight times above baseline. L-AMP also increased DA levels up to seven times above baseline, but this level was obtained only at the end of the 60-minute perfusion period. L-mAMP did not change DA levels under these conditions. **C,** In contrast, all three amphetamine isomers had equipotent enhancements on norepinephrine (NE) release. These results suggest that the potency of these derivatives on EEG arousal correlated well with measurements of DA efflux in the caudate of narcoleptic dogs, whereas effects on NE release may be related to REM suppressant effects.

during perfusion of D-amphetamine, L-amphetamine, and L-methamphetamine. Although all compounds increased NE efflux, no significant difference in potency was detected among the three analogs.

The fact that the potency of amphetamine derivatives on EEG arousal correlates with effects on DA efflux in the caudate of narcoleptic dogs further confirms that the enhancement of DA transmission by presynaptic modulation mediates the wake-promoting effects of amphetamine analogs. This result is also consistent with data obtained with DAT blockers (see Figure 43-3). Considering the fact that other amphetamine-like stimulants, such as methylphenidate and pemoline, also inhibit DA uptake and enhance release of DA, presynaptic enhancement of DA transmission is likely to be the key pharmacologic property mediating wake promotion for all amphetamines and amphetamine-like stimulants. In contrast, there is little evidence that enhancing adrenergic transmission is wake promoting in animal studies.

The role of the DA system in sleep regulation was further assessed using mice, which genetically lacked the DAT gene. Consistent with a role of DA in the regulation of wakefulness, these animals have reduced non–rapid eye movement (NREM) sleep time and increased wakefulness consolidation (independently from locomotor effects).[11] The most striking finding was that DAT knockout mice were completely unresponsive to the wake-promoting effects of methamphetamine, GBR-12909, and modafinil. These results further confirm the critical role of DAT in mediating the wake-promoting effects of amphetamines and modafinil (see Figures 43-3 and 43-4)[11] (see Modafinil and Armodafinil section). Interestingly, DAT knockout animals were also found to be more sensitive to caffeine,[11] suggesting functional interactions between adenosinergic and DA systems in the control of sleep and wakefulness (see Caffeine section).

Anatomic Substrates Mediating Dopaminergic Effects on Wakefulness

Anatomic studies have demonstrated two major subdivisions of the ascending DA projections from mesencephalic DA nuclei (ventral tegmental area [VTA], substantia nigra [SN], and retrorubral field [A8]): (1) The mesostriatal system originates in the SN and retrorubral nucleus and terminates in the dorsal striatum (principally the caudate and putamen)[12]; and (2) The mesolimbocortical DA system consists of the mesocortical and mesolimbic DA systems. The mesocortical system originates in the VTA and the medial SN and terminates in the limbic cortex (medial prefrontal, anterior cingulated, and entorhinal cortices). Interestingly, DA reuptake is of physiologic importance for the elimination of DA in cortical hemispheres, limbic forebrain, and striatum, but not midbrain DA neurons.[13] It is thus possible that amphetamine, modafinil, and DA uptake inhibitors have greater effect on DA terminals of the cortical hemispheres, limbic forebrain, and striatum and that it is this effect that induces wakefulness. Local perfusion experiments of DA compounds in rats and canine narcolepsy have suggested that the VTA, but not the SN, is critically involved in EEG arousal regulation.[14] DA terminals of the mesolimbocortical DA system may thus be important in mediating wakefulness after DA-related CNS stimulant administration. The involvement of other, less studied dopaminergic cell groups, such as those located in the hypothalamus or in the ventral periaqueductal gray (recently suggested to be wake active),[15] is also possible and would be worth exploring further.

Dopamine agonists and L-DOPA (dopamine precursor) drugs typically used in the therapy of Parkinson disease are generally not strongly wake promoting in clinical practice but instead are mildly sedative. This has been explained by the primary presynaptic effect of these compounds at low dose, an effect that may in fact reduce DA transmission in some projection areas.[16]

Indications

Amphetamine and methylphenidate are primarily indicated for narcolepsy, idiopathic hypersomnia, and ADHD. Other therapeutic uses are controversial because of their abuse potential. This potential also imparts them a schedule II classification under the Controlled Substances Act of 1970. Moreover, certain states (e.g., Wisconsin) have passed even more restrictive legislation limiting the access and the use of these substances to specific indications.[17] The use of these compounds is highly regulated by federal policy and in some states requires triplicate prescription and monthly renewal.

Side Effects and Toxicology

Amphetamine releases not only DA but also NE. NE indirectly stimulates α- and β-adrenergic receptors, a profile common to all indirectly acting sympathomimetic compounds. This results in significant cardiovascular effects. α-Adrenergic stimulation produces vasoconstriction, thereby increasing both systolic and diastolic blood pressure. Heart rate may slightly slow down in reflex (this effect is more pronounced that indirect β-adrenergic stimulation on heart rate at low dose), but with large doses, tachycardia and cardiac arrhythmia may occur. Cardiac output is not modulated by therapeutic doses, and cerebral blood flow is unchanged. In general, smooth muscles respond to amphetamine as they do to other sympathomimetic drugs. There is a contractile effect on the urinary bladder sphincter. Pain and difficulty in micturition may occur.

Other acute side effects include mild gastrointestinal disturbance, anorexia, dryness of the mouth, tachycardia, cardiac arrhythmias, insomnia, restlessness, headaches, palpitations, dizziness, and vasomotor disturbances. Agitation, confusion, dysphoria, apprehension, and delirium may also occur. Other documented side effects include flushing, pallor, excessive sweating, and muscular pains. Tiredness and sleepiness, as well as lethargy and listlessness, may occur when the effects wear off, together with a mild depression of mood. For common side effects of CNS stimulant drugs in narcoleptics, refer to Table 43-1.

Common side effects occurring during long-term treatment in narcolepsy include irritability, headache, bad temper, and profuse sweating (reported by more than one third of subjects). Less common side effects are anorexia, gastric discomfort, nausea, talkativeness, insomnia, orofacial dyskinesia, nervousness, palpitations, muscle jerking, chorea, and tremor. Psychiatric symptoms, such as delusions or hallucinations, may also occur but are rather rare in narcoleptic patients who take amphetamine.

Methamphetamine (and to a lesser extent, amphetamine) can be neurotoxic at high dose. This effect is mediated by

a free radical increase, causing mitochondrial damage and decreasing adenosine triphosphate synthesis. In dopaminergic neurons, the neurotoxicity is mediated by formation of peroxynitrite, which can be reduced by antioxidants or L-carnitine. L-Carnitine is needed to transport long-chain fatty acids to mitochondria for fatty acid oxidation, preventing the generation of free radicals and peroxynitrite. MDMA, another amphetamine derivative with a preferential effect (and toxicity) on serotoninergic neurons, appears to also decrease glutathione and vitamin E in the brain. Mice deficient in vitamin E were found to have greater susceptibility to both MDMA neurotoxicity and hepatic necrosis, a finding further supporting a free radical mechanism for amphetamine toxicity.

The side-effect profile of methylphenidate is similar to that of amphetamine and includes nervousness, insomnia, and anorexia as well as dose-related systemic effects such as increased heart rate and blood pressure. Methylphenidate overdose may lead to seizures, dysrhythmias, or hyperthermia.

Abuse and Misuse of Amphetamine Stimulants

Methamphetamine, amphetamine, and methylphenidate all have clear street value for abusers. Whereas reinforcement occurs in the early stages of drug use, tolerance is common during long-term administration. Appetite-suppressing effects are also common. Interestingly, anecdotal data suggest that psychostimulant abuse in narcoleptic subjects is extremely rare,[18,19] a finding also supported by some animal data.[20] Nevertheless, there is a negative stigma associated with the administration of amphetamine-like compounds in patients with narcolepsy.

The mechanisms underlying abuse of amphetamine-like stimulants are complex but have been shown to primarily involve stimulation of the VTA-DA systems.[21] Downstream changes in adrenergic and serotoninergic systems, particularly through α_{1b}-adrenergic receptors and 5-HT_{2A}, may also be important.[22,23]

Drug-Drug Interactions

Drug-drug interactions with amphetamine and methylphenidate are generally pharmacodynamic or neurochemical in nature.[24] Small percentages of the metabolism of amphetamine and methylphenidate occurs through cytochrome P-450 2D6, and drugs that inhibit 2D6 metabolism can theoretically increase plasma levels of amphetamine. This is rarely, however, a significant problem with therapeutic doses. Tricyclic drugs inhibit the metabolism of amphetamine and amphetamine-like stimulants and enhance their behavioral effects. The combination of amphetamine with tricyclics could theoretically further blood pressure increases (because of the combined effects of NE reuptake and release), but in practice amphetamine 10 to 16 mg, methylphenidate 10 to 60 mg, and mazindol 2 to 12 mg have been given safely with imipramine and clomipramine, 10 to 100 mg, to treat narcolepsy-cataplexy. The dosage of amphetamine required to control narcolepsy may be reduced by one third with the simultaneous use of tricyclic drugs. MAO-A inhibitors (e.g., nialamide, pargyline, and tranylcypromine) inhibit the removal of amphetamine by the liver and greatly potentiate the behavioral effects of amphetamine.[25] Coadministration of MAO inhibitors and amphetamine derivatives is generally contraindicated. In contrast to tricyclics and MAO-A inhibitors, haloperidol, reserpine, and atropine have no effect on amphetamine hydroxylation in the animal liver, although they may reduce the central effects of amphetamine.[26] Chlorpromazine, trifluoperazine, perphenazine, and thioproperazine increase the half-life of amphetamine in the brain but inhibit central behavioral effects, such as stereotyped behavior in animals and euphoria in humans.[26] Hypnotic drugs will prevent many behavioral effects of amphetamines, although chlordiazepoxide and diazepam increase amphetamine tissue levels.[26]

MODAFINIL AND ARMODAFINIL

Racemic modafinil (2-[(diphenylmethyl)sulfinyl]acetamide; see Figure 43-1) was first developed in France and has been available in Europe since 1986. Modafinil was first approved in 1998 in the United States for the treatment of narcolepsy. More recently, it has been approved for shift work sleep disorder and for the treatment of residual sleepiness in treated with obstructive sleep apnea syndrome. Modafinil is a primary metabolite of adrafinil, a vigilance-promoting compound developed in France in the 1970s. Modafinil lacks adrafinil's terminal amide hydroxy group (see Figure 43-1) and is better tolerated.

Pharmacokinetics

Modafinil is rapidly absorbed but slowly cleared. It is approximately 60% bound to plasma proteins and a volume of distribution of 0.8 L/kg, suggesting that the compound is readily able to penetrate into tissues. Its half-life is 9 to 14 hours. Up to 60% of modafinil is converted into modafinil acid and modafinil sulfone, both of which are inactive metabolites. Metabolism primarily occurs through cytochrome P-450 3A4/5, but the compound has also been reported to induce P-450 2C19 in vitro.[27] Modafinil is currently available as a racemic mixture of two active isomers and as an r-isomer-only preparation (armodafinil). Importantly, the r-enantiomer of modafinil has a half-life of 10 to 15 hours, which is longer than that of the s-enantiomer (3 to 4 hours).[28] The dual pharmacokinetic properties of the racemic mixture may explain why modafinil is often more potent when taken twice per day at the beginning of therapy, during the period of drug accumulation.

Indications

Modafinil is one of the few compounds that have been specifically developed for the treatment of narcolepsy. Early clinical trials in France and Canada showed that modafinil 100 to 300 mg is effective in improving EDS in narcolepsy and hypersomnia without interfering with nocturnal sleep, but that it has limited efficacy in cataplexy and other symptoms of abnormal REM sleep.[29-31] Pharmacologic experiments in canine narcolepsy also demonstrated that modafinil has no effects on cataplexy, but it significantly increases time spent in wakefulness.[32] A double-blind trial of 283 narcoleptic subjects in 18 centers in the United States revealed that 200 mg and 400 mg of modafinil significantly reduced EDS and improved patients' overall clinical condition.

Armodafinil was approved by the FDA in 2007 for the treatment of sleepiness in association with narcolepsy, treated obstructive sleep apnea syndrome, and shift work sleep

disorder (i.e., for the same indications as those of racemic modafinil).[28] Armodafinil has been shown to be potent for a longer period of time after administration. In patients in whom once-daily modafinil does not cover the entire day, armodafinil may be useful. Further, lower doses of armodafinil, 150 mg and 250 mg, were used in a phase III trial, whereas earlier modafinil trials used 200 mg and 400 mg. Armodafinil is available at lower doses than modafinil, suggesting an improved safety profile. Although armodafinil may not be a revolutionary improvement compared with modafinil, it may have its place in the therapeutic arsenal.[28]

In addition to the FDA-approved indications for modafinil and armodafinil, several reports have suggested that modafinil is also effective for the treatment of ADHD, fatigue in multiple sclerosis, and EDS in myotonic dystrophy or Prader-Willi syndrome.[33] Modafinil is also being used in the treatment of periodic hypersomnia (Klein-Levin syndrome), for which treatment immediately after initiation of the episode may be critical.[34]

Side Effects

Modafinil is well tolerated. The most frequent reported side effects are headache and nausea.[35] In addition, however, modafinil, because of its dual hepatic and renal elimination profile, should be used at lower dose in hepatic and renal insufficiency cases, although dosage recommendations in such patients cannot be made.[33] Modafinil has a number of potential drug interactions. In vitro, modafinil produces a reversible inhibition of CYP2C19 in human liver microsomes. It also causes a small but concentration-dependent induction of CYP1A2, CYP2B6, and CYP3A4 activities and suppression of CYP2C9 activity in primary cultures of human hepatocytes. Clinical studies have been conducted to examine the potential for interactions with methylphenidate, dexamfetamine, warfarin, ethinylestradiol, and triazolam. The most substantive interactions observed were with ethinylestradiol and triazolam, apparently through induction of CYP3A4, primarily in the gastrointestinal system. For this reason, it is suggested that women taking low-estrogen contraception be informed of alternative or concomitant methods of contraception. Interestingly, modafinil has been shown to be safe and to have additive effects on alertness when administered with sodium oxybate in narcolepsy.

Several factors make modafinil an attractive alternative to amphetamine-like stimulants. First, animal studies suggest that the compound does not affect blood pressure as much as amphetamines do; only high doses (800 mg) have been found to be associated with higher rates of tachycardia and hypertension. Recent clinical studies showed only small average increases in mean systolic and diastolic blood pressure in patients receiving armodafinil compared with placebo. Increased monitoring of blood pressure may be appropriate in patients taking modafinil. Second, data obtained to date suggest that dependence is limited in humans with this compound,[29,36] although a recent animal study suggests that cocaine-like discriminative stimulus and reinforcing effects of modafinil in rats and monkeys. Most notably, modafinil is almost certainly used as a convenience drug by some to fight sleepiness resulting from sleep deprivation or jet lag. Modafinil is not attractive to cocaine or stimulant abusers and does not have a high street value. Third, modafinil has minimal effects on the neuroendocrine system. In a study of healthy volunteers who were sleep deprived for 36 hours, those who received modafinil did not differ from those who did not with respect to cortisol, melatonin, and growth hormone levels.[37] Fourth, clinical experience suggests that the pharmacologic profiles of modafinil might be qualitatively different from those observed with amphetamine.[29] In general, patients feel less irritable or agitated with modafinil than with amphetamines[29] and do not experience severe rebound hypersomnolence (seen in patients with amphetamine) after modafinil is eliminated. This differential profile is substantiated by animal experiments. In rats and dogs, modafinil does not increase locomotion beyond the effect expected in association with increased wakefulness.[32,38] Similarly, modafinil acutely decreases both REM and NREM sleep in rats for up to 5 to 6 hours, but the effect is not followed by a rebound hypersomnolence. This profile contrasts with the intense recovery sleep seen after amphetamine-induced wakefulness.[39] The safety profile of modafinil is likely the basis for the fact that it has replaced amphetamine-like stimulants as a first-line treatment for EDS in narcolepsy.[40]

Mechanism of Action

The mechanism of action of modafinil-armodafinil is the subject of controversy, although in our opinion, it is, as in the case of other stimulants, most likely related to DAT inhibition. Because there are a limited number of studies addressing the mode of action of armodafinil, this section mostly discusses the actions of the racemic modafinil mixture. Modafinil-armodafinil has not been shown to bind to or inhibit receptors or enzymes for most known neurotransmitters, with the exception of the DAT protein.[41,42] In vitro, modafinil-armodafinil binds to the DAT and inhibits dopamine reuptake.[28,41,42] These binding inhibitory effects have been shown to be associated with increased extracellular DA levels in the striatum of rats and dogs, suggesting functional effects. Finally and most important, modafinil effects on alertness are entirely abolished in mice without the DAT protein[11] and in animals lacking D1 and D2 receptors.[43] A similar abolition of wake promotion in DAT knockout mice is also observed with amphetamine and GBR-12909 (a selective DAT blocker), drugs known to work through the DAT. Modafinil promotes wakefulness in hypocretin-deficient narcolepsy. Modafinil also promote wakefulness in noradrenaline-depleted animals (by DSP-4 administration)[44] and in histamine-deficient animals (histidine decarboxylase knockout mice),[45] suggesting that the wake-promoting effects of modafinil are seen independent from the availability of these wake-promoting neurotransmitters.

Given these similarities in mechanism to other DAT inhibitors, it is puzzling that modafinil has a low potential for abuse, a property that we believe may be due to the insolubility of the compound (inability to use another formulation, e.g., intravenously), its low potency (impossibility to greatly increase the dose), its slow absorption (no rapid brain effects), or its atypical binding interaction with the DAT transporter.

Adrenergic effects have also been suggested to be involved in the wake promotion effects of modafinil, but we believe these to be insignificant in vivo. When first introduced, an involvement of α_1-adrenergic systems was suggested as the

primary mode of action of modafinil on wakefulness, based on the ability of the α_1 antagonist, prazosin, to antagonize modafinil-induced increases in motor activity in mice and wakefulness in cats. Problematically, however, modafinil does not bind α_1 receptors in vivo (Ki $>10^{-3}$ M, obtained from prazosin binding using canine cortex).[32] It also does not produce smooth muscle contraction in vas deferens preparations and is still wake promoting in noradrenaline-depleted animals (see earlier).[44] Further, the hyperlocomotion produced by amphetamine, like that of modafinil, also largely depends on α_{1b} receptors, a finding now explained by remodeling of the DA system in α_1 knockout mice.[46] Finally, previous studies in the canine model of narcolepsy have shown that α_1-adrenergic agonists are potent anticataplectic agents[47] and have significant acute hypertensive effect. Modafinil has neither anticataplectic activity nor hypertensive effects, suggesting that its alerting properties are unrelated to adrenergic α_1 stimulation.

Clinical observations provide even stronger evidence that modafinil is not a primarily adrenergic compound. Amphetamine and adrenergic reuptake blockers cause dilation of the pupils by increasing NE signaling, but modafinil has no effect on pupil size. Some studies have noted slight increases in heart rate or blood pressure with high doses of modafinil. However, these changes were small, and most clinical studies on modafinil, including a meta-analysis of six large clinical trials of modafinil (the most comprehensive study on this issue), have found no changes in heart rate or blood pressure. In contrast, adrenergic reuptake blockers are well known to slightly increase blood pressure and heart rate. These clinical observations suggest that at usual clinical doses, modafinil does not increase adrenergic signaling in humans.

Interestingly, Madras and colleagues[48] recently reported, in a study involving rhesus monkeys undergoing positron emission tomography (PET), that modafinil (given intravenously) occupied striatal DAT sites (5 mg/kg, 35%; 8 mg/kg, 54%). In the thalamus, modafinil occupied NET sites (5 mg/kg, 16%; 8 mg/kg, 44%) (Figure 43-5). The authors also showed that modafinil inhibited [^{3}H]-dopamine (IC_{50} = 6.4 M)

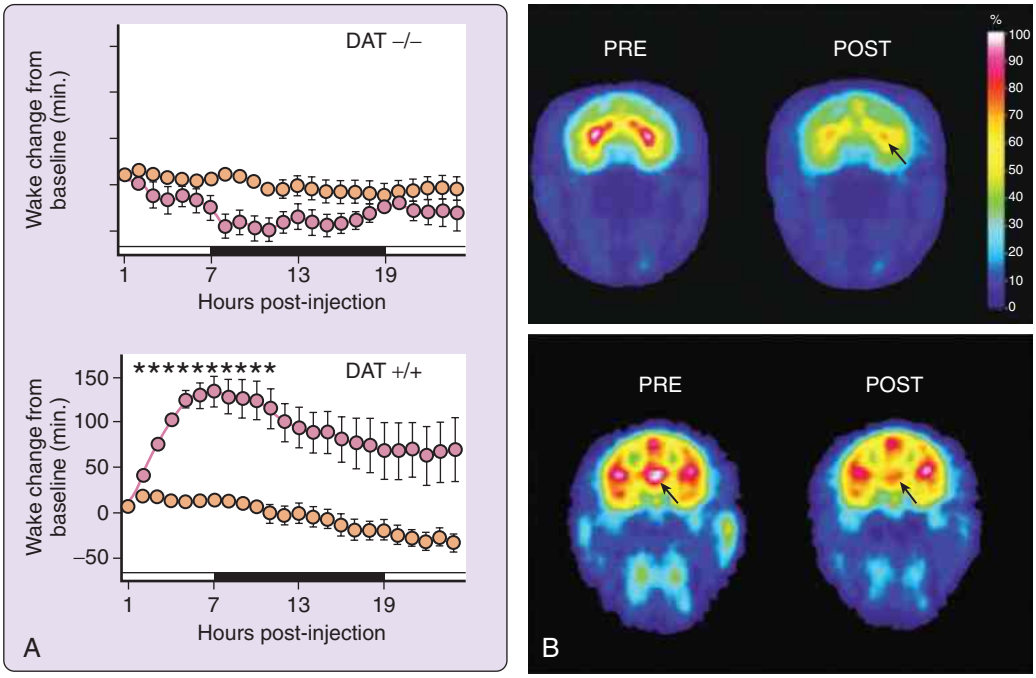


Figure 43-5 **A,** Wake-promoting effects of modafinil were completely abolished in dopamine transporter (DAT) knockout (KO) mice, suggesting that intact DAT function is required for the mediation of wake-promoting effects of modafinil. **B,** Modafinil (8 mg/kg) occupancy by the DAT in caudate putamen is shown as detected by positron emission tomography (PET) of the DAT with [^{11}C]CFT. *Left,* an adult rhesus monkey was injected with [^{11}C]CFT and scanned over 60 minutes to develop baseline measures of DAT-binding potential in the caudate putamen. Images were color-transformed to display occupancy of the DAT with [^{11}C]CFT, with highest levels detected in caudate putamen (*white-red*), as designated by the *arrow*, and lowest levels in *blue-purple*. Regions of interest are drawn over the caudate putamen. *Right,* After decay of [^{11}C]CFT radioactivity, modafinil was injected intravenously, and [^{11}C]CFT was injected again 1 hour later. [^{11}C]CFT accumulation was significantly lower compared with baseline levels of accumulation (*left*). **C,** Modafinil (8 mg/kg) occupancy by the norepinephrine transporter (NET) in the thalamus, as detected by PET imaging of the NET with [^{11}C]MeNER. *Left,* an adult rhesus monkey was injected with [^{11}C]MeNER and scanned over 60 minutes to develop baseline measures of NET binding potential in the thalamus. Images were color-transformed to display occupancy of the NET by [^{11}C]MeNER, with high levels detected in the thalamus (*white-red*), as designated by the *arrow*, and lowest levels in *blue-purple*. Regions of interest are drawn over the thalamus. *Right,* after decay of [^{11}C]MeNER radioactivity, modafinil was injected intravenously, and 1 hour later, [^{11}C]MeNER was injected. [^{11}C]MeNER accumulation was significantly lower compared with baseline levels of accumulation. (Modified from Wisor JP, Nishino S, Sora I, et al. Dopaminergic role in stimulant-induced wakefulness. *J Neurosci* 2001;21:1787–94; and Madras BK, Xie Z, Lin Z, et al. Modafinil occupies dopamine and norepinephrine transporters in vivo and modulates the transporters and trace amine activity in vitro. *J Pharmacol Exp Ther* 2006;319:561–9.).

transport 5 times and 80 times more potently than [^{3}H]-norepinephrine (IC_{50} = 35.6 M) and [^{3}H]-5-HT (IC_{50} = 500 M) transport, respectively, in cell lines that expressed the human DAT, NET, and 5-HT transporter. These data provide compelling evidence that modafinil occupies the DAT in the living brains of rhesus monkeys, consistent with the DAT hypothesis, but suggest that modafinil may also act on NET, depending on drug dose, brain structure, and other physiologic conditions.

Furthermore a recent human PET study in 10 healthy humans with [^{11}C]-cocaine (DAT radioligand) and [^{11}C]-raclopride (D2/D3 radioligand sensitive to changes in endogenous dopamine) also demonstrated that modafinil (200 mg and 400 mg given orally) decreased [^{11}C]-cocaine-binding potential in caudate (53.8%; $P < .001$), putamen (47.2%; $P < .001$), and nucleus accumbens (39.3%; $P = .001$),[49] the results being consistent with the DAT hypothesis. In addition, modafinil also reduced [^{11}C]-raclopride-binding potential in caudate (6.1%, $P = .02$), putamen (6.7%; $P = .002$), and nucleus accumbens (19.4%; $P = .02$) (see Figure 43-5), suggesting that the increases in extracellular dopamine were caused by DAT blockades.[49] These results are highly consistent with the previously mentioned results of the animal studies; the effects of modafinil on alertness are entirely abolished in mice without the DAT protein[11] and in animals lacking D1 and D2 receptors.[43]

MAZINDOL

Mazindol is a schedule IV controlled drug that is rarely used in the United States. At 2 to 8 mg daily, mazindol produces central stimulation, a reduction in appetite, and an increase in alertness but has little or no effect on mood or the cardiovascular system.[50] Mazindol is effective for the treatment of both EDS and cataplexy in humans[51] and in canine narcolepsy, possibly owing to its blocking properties of DA and NE reuptake.[42] This compound has a high affinity for DAT and NET,[42] yet interestingly this compounds has a low abuse potential. Problematically, however, mazindol often causes significant side effects, including anorexia, gastrointestinal discomfort, insomnia, nervousness, dry mouth, nausea, constipation, urinary retention, and occasionally angioneurotic edema, vomiting, and tremor.

BUPROPION

Bupropion is not scheduled by the U.S. Drug Enforcement Administration. Although the selectivity for the dopamine transporter is not absolute, bupropion blocks DA uptake. Bupropion shows a weak inhibition of NE reuptake and very limited serotoninergic effects. Although not indicated for these uses, bupropion may be useful for the treatment of EDS associated with narcolepsy at 100 mg three times daily.[42,52] It may be especially useful in cases associated with atypical depression.[52] Risk for convulsion increases dose dependently (0.1% at 100 to 300 mg; 0.4% at 400 mg).

SELEGILINE (L-DESPRENYL)

Selegiline is a methamphetamine derivative and a potent, irreversible, MAO-B selective inhibitor primarily used for the treatment of Parkinson disease.[53,54] Because it is often considered a simple MAO-B inhibitor, it is worth mentioning that selegiline is an amphetamine precursor. This compound is metabolized into L-amphetamine (20% to 60% in urine) and L-methamphetamine (9% to 30% in urine).[53]

In the canine model of narcolepsy, selegiline (2mg/kg given orally) was demonstrated to be an effective anticataplectic agent, but this effect was found to be mediated by its amphetamine metabolites rather than MAO-B inhibition.[55] Several trials in human narcolepsy have demonstrated a good therapeutic efficacy of selegiline in both sleepiness and cataplexy with relatively few side effects.[56,57] Selegiline 10 mg daily has no effect on the symptoms of narcolepsy, but 20 to 30 mg improves alertness and mood and reduces cataplexy, showing an effect comparable to D-amphetamine at the same dose. Selegiline may be an interesting alternative to the use of more classic stimulants because its potential for abuse has been reported to be very low.

ATOMOXETINE AND REBOXETINE

Atomoxetine and reboxetine (in Europe) are selective adrenergic reuptake inhibitors. Both compounds were developed as antidepressants, but atomoxetine is now mainly used in the therapy of ADHD.[58] Although these compounds are not stimulants per se, they are slightly wake promoting[59,60] and reduce REM sleep. These compounds can be helpful in some cases of narcolepsy and idiopathic hypersomnia. Atomoxetine needs twice-daily administration owing to its short half-life. Reboxetine was shown to reduce MSLT mean sleep latency in narcoleptic patients.[59] These compounds, however, increase heart rate and blood pressure. Sexual side effects are also common, but there is no risk for abuse.

CAFFEINE

Caffeine, a xanthine derivative isolated from plants, may be the most popular and widely consumed CNS stimulant in the world. An average cup of coffee contains 50 to 150 mg of caffeine. Tea, cola drinks, chocolate, and cocoa all contain significant amounts of caffeine. Caffeine can also be bought over the counter (No Doz, 100 mg caffeine; Vivarin, 200 mg caffeine) and is commonly used by narcoleptic patients before diagnosis.

Taken orally, caffeine is rapidly absorbed. The half-life of caffeine is 3.5 to 5 hours. The behavioral effects of caffeine include increased mental alertness, a faster and clearer flow of thought, wakefulness, and restlessness.[61] Fatigue is reduced and sleep-onset delayed.[61] The physical effects of caffeine include palpitations, hypertension, increased gastric acid secretion, and increased urine output.[61] Heavy consumption (12 or more cups/day, or 1.5 g of caffeine) causes agitation, anxiety, tremors, rapid breading, and insomnia.[61]

Adenosine has been proposed to be a sleep-promoting substance that accumulates in the brain during prolonged wakefulness[62] and possesses neuronal inhibitory effects. In animals, sleep can be induced after administration of adenosine A1 receptor (A1R) or A2A receptor (A2AR) agonists, such as N6-L-(phenylisopropyl)adenosine, adenosine-5′-*N*-ethylcarboxamide, and cyclohexyladenosine. Adenosine content is increased in the basal forebrain after sleep deprivation. Adenosine has thus been proposed to be a sleep-inducing substance accumulating in the brain during

prolonged wakefulness.[62] The mechanism of action of caffeine on wakefulness involves nonspecific adenosine receptor antagonism. In particular, Huang and colleagues[63] recently reported that wake-promoting effects of caffeine are abolished in A2AR knockout mice, whereas the effects were not altered in A1R knockout mice, suggesting a primary effect of caffeine through the A2AR, at least in this species. Interestingly, the A2AR interacts strongly with dopaminergic transmission. A2AR forms a heterodimer with dopamine D2 receptors, and 2AR knockout mice have been shown to have reduced amphetamine-induced locomotor simulation and reward.[64-66] Recently, Lazarus and colleagues demonstrated the specific neurons on which caffeine acts to produce arousal using selective gene deletion strategies for A2ARs in animals.[67] The authors reported that the A2ARs in the shell region of the nucleus accumbens (NAc) are responsible for the effect of caffeine on wakefulness. Caffeine-induced arousal was not affected in rats when A2ARs were focally removed from the NAc core or other A2AR-positive areas of the basal ganglia. The authors claim that caffeine promotes arousal by activating pathways that traditionally have been associated with motivational and motor responses in the brain.

Caffeine is metabolized into three active metabolites: paraxanthine, theobromine, and theophylline. We recently demonstrated that paraxanthine significantly promoted wakefulness and proportionally reduced NREM and REM sleep in both control and narcoleptic mice.[68] The wake-promoting potency of paraxanthine (100 mg/kg given orally) is greater than that of the parent compound, caffeine (92.8 mg/kg given orally), and comparable to that of modafinil (200 mg/kg given orally). High dose of caffeine and modafinil induced hypothermia and reduced locomotor activity, whereas paraxanthine did not. In addition, behavioral testing revealed that the compound possessed lesser anxiogenic effects than caffeine. Although further evaluation in humans should be needed, paraxanthine may be a better wake-promoting agent for normal individuals as well as patients who have hypersomnia associated with neurodegenerative diseases.

FUTURE STIMULANT TREATMENTS

Hypocretin-Based Therapies

Hypocretin deficiency is a main cause of human narcolepsy. Intracerebroventricular injections of hypocretin strongly promote wakefulness in dogs, mice, and rats. Animal experiments using ligand-deficient narcoleptic dogs show that very high systemic doses are required for hypocretin to penetrate the CNS and that only a short-lasting therapeutic effect is observed after intravenous administration of hypocretin. Stable and centrally active hypocretin analogs (possibly nonpeptidic synthetic hypocretin ligands) after peripheral administration will need to be developed.[69,70] Studies have also noted a normalization of the sleep-wake patterns and behavioral arrest episodes (equivalent to cataplexy and REM sleep onset) in hypocretin-deficient mice following the central administration of hypocretin-1.[71] Hypocretin may, therefore, one day prove to be effective in the treatment of both EDS (i.e., fragmented sleep-wake pattern) and cataplexy. Such studies also open the door to the possibility of cell transplantation–based and gene-based therapies.

To address whether hypocretin receptor function is intact after long-term hypocretin deficiency, Mishima and colleagues[72] recently studied hypocretin receptor gene expressions of ligand deficient narcolepsy in mice, dogs, and humans. A substantial decline (by 50% to 71%) in the expression of hypocretin receptor genes was observed in both ligand-deficient humans and dogs. Similar murine studies suggested that this decline is progressive over age. Importantly, however, about 50% of baseline expression was still observed in old ligand-deficient narcoleptic human subjects. Furthermore, because narcoleptic Doberman dogs heterozygous for the hypocretin receptor-2 mutation (with 50% receptor levels and normal levels of hypocretin) are asymptomatic, it is likely that an adequate ligand supplementation will prevent narcolepsy in hypocretin-deficient patients even if receptors are partially nonfunctional.

Histamine-3 Antagonists

Histamine has long been implicated in the control of vigilance because histamine-1 (H_1) antagonists are strongly sedative. The excitatory effects of hypocretins on the histaminergic system through hypocretin receptor-2 are likely to be important in mediating the wake-promoting properties of hypocretin.[73] In fact, brain histamine levels are reduced in narcoleptic dogs.[74] Reduction of histamine levels is also observed in human narcolepsy and other hypersomnias of central origin.[75,76] Although centrally injected histamine or histaminergic H_1 agonists promote wakefulness, the systemic administration of these compounds induces various unacceptable side effects through peripheral H_1 receptor stimulation. In contrast, the histaminergic H_3 receptors are regarded as inhibitory autoreceptors and are enriched in the CNS. H_3 antagonists enhance wakefulness in normal rats and cats[77] and in narcoleptic mice models.[78] Histaminergic H_3 antagonists might be useful as wake-promoting compounds for the treatment of EDS or as cognitive enhancers and are being studied by several pharmaceutical companies.[45]

Thyrotropin-Releasing Hormone

Another possible avenue of treatment, although one that currently enjoys less interest by pharmaceutical companies, is the use of thyrotropin-releasing hormone (TRH) direct or indirect agonists. TRH itself is a small peptide that penetrates the blood-brain barrier at very high doses. Small molecules with agonistic properties and increased blood-brain barrier penetration have been developed (i.e., CG3703, CG3509, or TA0910), thanks, in part, to the small nature of the starting peptide.[79] TRH (at the high dose of several mg/kg) and TRH agonists increase alertness, have been shown to be wake promoting and anticataplectic in the narcoleptic canine model,[80,81] and have excitatory effects on motoneurons. Initial studies demonstrated that TRH enhances DA and NE neurotransmission and that these properties may partially contribute to the wake-promoting and anticataplectic effects of TRH. Interestingly, recent studies have suggested that TRH may promote wakefulness by directly interacting with the thalamocortical network; TRH itself and TRH type 2 receptors are abundant in the reticular thalamic nucleus. Local application of TRH in the thalamus abolishes spindle wave activity,[82] and in the slice preparations, TRH depolarized thalamocortical and reticular-perigenuculate neurons by inhibition of leak K^+ conductance.[82] TRH injected in the lateral

hypothalamus induced locomotor activation in mice, but this effect was attenuated in hypocretin knockout mice, suggesting that the stimulant effects of TRH are partially mediated by stimulation of hypocretin neurons.[83] TRH also excites the histaminergic tuberomammillary nucleus.[84] Considering that TRH provokes arousal from hibernation,[85] TRH may be a potentially important wake-promoting system, although further studies are needed to disclose the roles of TRH in sleep-wake regulation.

Glutamatergic Compounds

Glutamatergic transmission is the major excitatory transmission of the mammalian brain and is increasingly believed to play a role in the generation of sleep homeostasis through changes in cortical synaptic plasticity.[86] Not surprisingly, therefore, compounds that are allosteric modulators of glutamatergic transmission, the ampakines, are being developed as wake-promoting compounds and may have counteracting effects on sleep deprivation.[87] Similarly, GluR subtype–specific compounds are likely to regulate sleep based on available knockout data and pharmacologic experiments.[88,89]

Among GluR subtypes, accumulating data support the therapeutic potential of glutamate metabotropic (mGluR2) receptors for treatment of psychiatric disorders such as depression, anxiety, and schizophrenia. The mGluR2 receptors are localized predominantly in presynaptic terminals of glutamate neurons, where they are inhibitory receptors and control glutamate release and glutamatergic neurotransmission on target networks.[90] Ahnaou and colleagues recently demonstrated that blockade of mGluR2, such as with the specific mGluR2 antagonist LY341495 or negative allosteric modulator Ro-4491533, in animals induced an immediate and endured desynchronized cortical activity associated with enhanced theta and gamma oscillations.[91] The wake-promoting effects are associated with marked lengthening of sleep-onset latency, an increased number of state transitions from light sleep to waking. The arousal response to mGluR2 blockade was not accompanied by sharp sleep rebound as found with the classic psychostimulant amphetamine, and further studies are needed to disclose the roles of mGluR2 receptors in sleep-wake regulation and their therapeutic use as new wake-promoting compounds.

CLINICAL PEARLS

- Almost all the currently available stimulants used to treat excessive daytime sleepiness in clinical practice (amphetamines, amphetamine-like stimulants, and modafinil-armodafinil) act presynaptically to increase dopaminergic transmission, either by stimulating dopamine release or by blocking dopamine reuptake. These effects are believed to be critically involved in the mediation of the wake-promoting effects of these compounds.
- Some (e.g., amphetamine) stimulants also increase adrenergic neurotransmission. Selective adrenergic uptake inhibitors have limited wake-promoting effects but potently reduce REM sleep or cataplexy. Increased adrenergic neurotransmission may play a minor role in stimulant-induced wake-promoting effects.
- Caffeine (as an over-the-counter supplement, coffee, tea, cola drinks, chocolate, and cocoa) is a nonselective adenosine receptor blocker. The potency and efficacy of caffeine are too low to provide substantial relief in the treatment of EDS associated with narcolepsy.
- Agents stimulating the hypocretinergic and histaminergic pathways may be promising future wake-promoting compounds but are not yet available.
- When using stimulants for the treatment of sleepiness, it is suggested to start with compounds that inhibit dopamine reuptake first (modafinil > methylphenidate) and then move on to the use of dopamine-releasing agents (e.g., amphetamines) only if the other compounds are not effective enough.

SUMMARY

Amphetamine-like stimulants have been used in the treatment of narcolepsy and various other conditions for decades, yet only recently has the mode of action of these drugs on vigilance been characterized. In almost all cases, the effects on vigilance were found to be mediated by effects on the DAT, leading to the widely accepted notion that the wake-promoting effects of these agents cannot be disentangled from their abuse potential. Importantly, however, the various medications available have differential effects and potency on the DAT and on monoamine storage and release. The various available stimulants are more or less selective for dopamine versus other amines. Although much work remains to be done in this area, it appears more and more likely that other properties, for example, the ability to release DA rather than simply block reuptake, plus the combined effects on other monoamines (such as serotonin) may be important to explain abuse potential. Differential binding properties on the DAT itself may also be involved, together with drug potency and compound solubility. The lack of solubility of some low-potency compounds may, for example, result in an inability to administer the drug by snorting or intravenously. Finally, lower abuse potential of these compounds has long been suspected in narcolepsy-cataplexy patients either because of the biochemical hypocretin abnormality or because of the social aspects of treating narcolepsy as a disease.

The mode of action of modafinil remains controversial and probably involves dopaminergic rather than nondopaminergic effects. Whatever its mode of action, the compound is generally found to be safer and to have a lower abuse potential than amphetamine stimulants. Its favorable side-effect profile has led to an increasing use outside the narcolepsy indication, most recently in the context of shift work sleep disorder and residual sleepiness in treated sleep apnea patients. This recent success exemplifies the need to develop novel wake-promoting compounds with low abuse potential. Other mechanisms of action involved in wake promotion include adenosine receptor antagonists, such as those found in caffeine. Novel classes of wake-promoting therapeutics are being developed, including glutamatergic and histaminergic modulators, and preclinical and clinical evaluations are in progress. A need for treating daytime sleepiness extends well beyond the relatively rare indication of narcolepsy-cataplexy.

Selected Readings

Battleday RM, Brem AK. Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: a systematic review. *Eur Neuropsychopharmacol* 2015 Aug 20. pii: S0924-977X(15)00249-7.

Kanbayashi T, Nishino S, Honda K, et al. Differential effects of D- and L-amphetamine isomers on dopaminergic trasmission: implication for the control of alertness in canine narcolepsy. *Sleep Res* 1997;**26**: 383.

Kuczenski R, Segal DS, Cho A, Melega W. Hippocampus norepinephrine, caudate dopamine and serotonin and behavioral responses to the stereoisomers of amphetamine and methamphetamine. *J Neurosci* 1995;**15**: 1308–17.

Lazarus M, Shen HY, Cherasse Y, et al. Arousal effect of caffeine depends on adenosine A2A receptors in the shell of the nucleus accumbens. *J Neurosci* 2011;**31**:10067–75.

Lu J, Jhou TC, Saper CB. Identification of wake-active dopaminergic neurons in the ventral periaqueductal gray matter. *J Neurosci* 2006;**26**: 193–202.

Mücke M, Mochamat, Cuhls H, et al. Pharmacological treatments for fatigue associated with palliative care. *Cochrane Database Syst Rev* 2015;**5**: CD006788.

Nishino S, Mao J, Sampathkumaran R, et al. Increased dopaminergic transmission mediates the wake-promoting effects of CNS stimulants. *Sleep Res Online* 1998;**1**:49–61.

Nishino S, Okuro M. Armodafinil for excessive daytime sleepiness. *Drugs Today (Barc)* 2008;**44**:395–414.

Okuro M, Fujiki N, Sokoloff P, Nishino S. Evaluations of wake promoting effects of paraxanthine in orexin/ataxin-3 narcoleptic mice. *Sleep* 2009; **32**(Abst. Suppl.):A35.

A complete reference list can be found online at ExpertConsult.com.

Chapter
44

Wake-Promoting Medications: Efficacy and Adverse Effects

Mihaela Bazalakova; Ruth M. Benca

Chapter Highlights

- A variety of wake-promoting medications are used to treat excessive sleepiness, ranging from over-the-counter caffeine to Schedule II amphetamine-like compounds. Each has a potential role in the clinical treatment of excessive sleepiness due to sleep disorders. Because their pharmacologic profiles are diverse, the clinician may guide selection of the agent based on a variety of factors: time of onset, length of activity, degree of tolerance in chronic use, expected side effects, and abuse liability. It is important to recognize that wake-promoting medications provide symptomatic treatment but do not modify the underlying etiology of sleepiness.[1]
- Although the traditional stimulants have been prescribed most widely for disorders such as narcolepsy, nonsympathomimetic compounds such as modafinil and its R-enantiomer armodafinil are now considered first-line wake-promoting agents for this disorder. These compounds have also been approved by the U.S. Food and Drug Administration (FDA) for treatment of excessive sleepiness due to shift work sleep disorder and in patients with obstructive sleep apnea whose sleepiness fails to remit despite optimal treatment with nasal continuous positive airway pressure.
- The rapid-acting hypnotic medication sodium oxybate also improves daytime alertness in people with narcolepsy and has received FDA approval for use in this patient population. Understanding the underlying pharmacology of the range of alerting agents available may clarify the qualitative aspects of wakefulness that they affect.

Wake-promoting medications fall into two categories: those that support wakefulness directly and are taken during the daytime and the hypnotic sodium oxybate, which gradually improves wakefulness over months of regular use at night. Of the daytime medications, there are three chemical classes: (1) direct-acting sympathomimetics, such as the α_1-adrenergic agonist phenylephrine; (2) indirect-acting sympathomimetics (frequently referred to as "stimulants" in clinical practice), such as methylphenidate and amphetamine; and (3) nonsympathomimetics (frequently referred to as "wake-promoting agents" in clinical practice), such as modafinil and caffeine. This chapter focuses on the clinical use of alerting medications.

THE HISTORY OF WAKE-PROMOTING MEDICATIONS

The known history of wake-promoting substances dates back to the early epochs of human civilization. Psychostimulants have been used for centuries in tonics and other preparations to allay fatigue and treat a variety of ailments (for reviews, see Haddad[2] and Angrist and Sudilovsky[3]).

Caffeine

Caffeine is the most widely consumed psychoactive substance in the world today—a testament to the apparently universal need for and widespread perceived benefit of an alertness-promoting agent.[4] Caffeine can be extracted from plants such as coffee and tea or synthetically produced. Caffeine is also an important central nervous system (CNS) active constituent of chocolate and "energy drinks." The most popular drinks in the world—coffee, tea, and many carbonated soft drinks—contain caffeine (Table 44-1), with carbonated beverages constituting the primary source of caffeine for children.[5]

Coffee's stimulant effects were likely first discovered in East Africa many centuries ago. Legends describe Ethiopian goat herders noticing the energizing effects of coffee beans on their herds, with the coffee plant eventually making its way to Yemen, where it has been cultivated since the 6th century, via the port city of Mocha or Mokha. Reports of coffee bean roasting date back to the 1400s, with writings by Abd al-Qadir al-Jaziri describing Sheikh Jamal-al-Din al-Dhabhani using coffee to "[drive] away fatigue and lethargy."[6] By the mid-1600s, coffee became popular in Europe, where it substituted alcohol-based staples, such as beer soup, at breakfast,[7] thus likely transforming European health and habits.[8] Today, 83% of U.S. adults report drinking coffee, with 63% consuming coffee daily and 75% reporting coffee intake at least once per week.[9]

Historical records suggest tea was first discovered as early as 2737 BCE by the Chinese Emperor Shen-Nung, who boiled the first pot of tea using bush leaves.[10] Like coffee, tea became popular in Europe in the 1600s. The tradition of the afternoon tea is ascribed to Anna, Duchess of Bedford, who introduced afternoon tea to Queen Victoria's court to "ward off that sinking feeling." Close to 80% of U.S. households reported tea consumption in 2012, totaling 3.6 billion gallons per year.[11]

Table 44-1 Caffeine per Serving and Product

Product	Serving Size	Caffeine Content (mg)
Coffees*		
Coffee, brewed	8 oz	110 (range, 100–200)
Coffee, decaf	8 oz	5 (range, 3–12)
Starbucks coffee, grande	16 oz	330 (range 260–560)
Starbucks coffee, tall	12 oz	260
Starbucks coffee, short	8 oz	180
Espresso	1 oz	64 (range 30–90)
Espresso, decaf	1 oz	10
Instant coffee	8 oz	75 (range 27–173)
Caffé latte	8 oz	120 (range 63–175)
Arizona Blue Luna iced coffees	8 oz	40–50
Arizona iced coffees	8 oz	40–50
Coffee ice cream	8 oz	58
Teas*		
Yerba Mate	8 oz	85
Arizona iced tea, black tea	8 oz	16
Arizona iced tea, green tea	8 oz	7.5
Arizona iced tea, Rx Power and Energy	8 oz	30
Brewed, imported brands	8 oz	60
Brewed, major U.S. brands	8 oz	40 (range 40–120)
Lipton Brisk iced tea	8 oz	6
Mistic teas	8 oz	17 (average)
Snapple iced tea, all kinds	8 oz	21
Soft Drinks		
Josta	12 oz	58
Mountain Dew	12 oz	55.5
Surge	12 oz	52.5
Diet Coke	12 oz	46.5
Coca-Cola	12 oz	34.5
Dr. Pepper, regular or diet	12 oz	42
Sunkist orange soda	12 oz	42
Pepsi-Cola	12 oz	37.5
Diet Pepsi	12 oz	36
Diet RC	12 oz	54
Barqs Root Beer	12 oz	22.5
Barqs Diet Root Beer	12 oz	0
7-Up or Diet 7-Up	12 oz	0
Sprite or Diet Sprite	12 oz	0
Mug Root Beer	12 oz	0
Caffeine-Free Coke or Diet Coke	12 oz	0
Caffeine-Free Pepsi or Diet Pepsi	12 oz	0
Minute Maid orange soda	12 oz	0

Continued

Table 44-1 Caffeine per Serving and Product—cont'd

Product	Serving Size	Caffeine Content (mg)
Caffeinated Waters and Energy Drinks		
Wired X344	16 oz	344
Spike Shooter	8.4 oz	300
5-Hour Energy	1.9 oz	200
Monster	16 oz	160
Full Throttle	16 oz	144
Java Water	500 mL	125
Krank 20	500 mL	100
Aqua Blast	500 mL	90
Red Bull	8.3 oz	80
Water Joe	500 mL	60–70
Aqua Java	500 mL	50–60
Chocolate		
Hershey's Chocolate Bar	1.55 oz	9
Hershey's Dark Chocolate Bar	1.45 oz	31
Baker's chocolate	1 oz	26
Chocolate milk beverage	8 oz	5
Chocolate-flavored syrup	1 oz	4
Cocoa beverage	8 oz	6
Dark chocolate, semi-sweet	1 oz	20
Milk chocolate	1 oz	6
Medications		
Anacin	2 tablets	26
Aqua Ban	1 tablet	100
Cafergot	1 tablet	100
Caffedrine	2 capsules	200
Coryban-D	1 tablet	30
Darvon Compound	1 tablet	32
Dexatrim	1 tablet	200
Dristan	1 tablet	30
Excedrin, max strength	2 tablets	130
Fiorinal	1 tablet	40
Midol	1 tablet	32
Migralam	1 tablet	100
Neo-Synephrine	1 tablet	15
NoDoz, maximum strength; Vivarin	1 tablet	200
NoDoz, regular strength	1 tablet	100
Percodan	1 tablet	32
Permathene Water Off	1 tablet	100
Pre-Mens Forte	1 tablet	50
Prolamine	1 tablet	140
Triaminicin	1 tablet	30
Vanquish	1 tablet	33

*The listed caffeine content is average for a standard brewed cup of coffee or tea; certain brewing methods may increase or decrease the average caffeine content per cup.

Together, coffee, tea, and energy drink consumption in 2011 totaled 30 gallons per capita per year in the United States (18.5, 10.3, and 1.2 gallons per person per year, respectively), exceeding bottled water consumption (28.3 gallons per capita per year) and only exceeded by yet another source of caffeine: carbonated soft drinks (44.7 gallons per capita per year).[12]

In a 7-day, diary-based population study of 42,851 consumers 2 years and older performed between October 2010 and September 2011, Mitchell and colleagues confirmed that 85% of the U.S. population consumes at least one caffeinated beverage daily, with mean and 90th percentile caffeine intake of 165 and 380 mg/day, or 2.2 and 5 mg/kg/day, respectively, for all ages.[5] Caffeine consumption increased with age, with highest levels found in adults 50 to 64 years old (226 mg/day), and women reported higher caffeine consumption than men when adjusted for body weight. A notable finding included increased caffeine consumption in all age groups, including children, compared with a similar survey from 1999. Although consumption of caffeinated carbonated soda drinks decreased over the same time period, there was a concomitant increase in coffee consumption, the main course of caffeine in adults. Interestingly, the yearly per capita consumption of coffee varies significantly around the globe, ranging from a high of 12 kg per person in Finland to less than 0.8 kg per person in Southeast Asia. Annual consumption in Canada (6.5 kg per person) and Brazil (5.6 kg per person) outpaces U.S. use (4.2 kg per person).[13]

Sympathomimetics

The native peoples of Peru and Bolivia used cocaine, a crystalline alkaloid derived from the leaves of the coca plant, for pleasure and to increase stamina. From 1886 to 1905, cocaine was an ingredient in Coca-Cola. The medicinal use of cocaine was advocated by Freud.[14] However, cocaine's profound potential for abuse and addiction soon limited the role of this stimulant in modern medicine. In 1931, Doyle and Daniels described the use of ephedrine to treat the sleepiness of narcolepsy.[15] Despite its clinically noteworthy efficacy, it was soon apparent that side effects, incomplete patient acceptance, rapid development of tolerance, and cost limited its usefulness. In 1935, Prinzmetal and Bloomberg suggested that amphetamine sulfate would be appropriate treatment for narcolepsy because of its close relationship to ephedrine and epinephrine, its low toxicity and low cost, its prolonged action, and its lack of pronounced sympathomimetic side effects.[16] By 1949, amphetamine (racemic B-phenylisopropylamine), in one or another of several oral preparations as a phosphate or sulfate, had become the treatment of choice for excessive sleepiness due to narcolepsy. Methylphenidate, a piperidine derivative, was introduced in 1959 by Yoss and Daly.[17] Pemoline, an oxazolidine compound, was later introduced as a mild CNS stimulant, whereas the mild stimulant mazindol, an imidazoline derivative, was marketed as an appetite suppressant. Neither pemoline nor mazindol are currently available as wake-promoting medications because of their adverse effects.

Nonsympathomimetics

Modafinil (2-iphenylmethylsulfinyl acetamide) is a racemic compound unrelated to the amphetamines or other CNS stimulants. Of all the alerting agents, modafinil has the most specific and selective wake-promoting properties and usually has minimal side effects. Modafinil appeared on the world market for the indication of narcolepsy and CNS hypersomnia in the early 1990s and is now considered a first-line agent for the treatment of these conditions.[18] Its R-enantiomer, armodafinil, was introduced in 2007. Additional U.S. Food and Drug Administration (FDA)–approved indications for modafinil and armodafinil took effect in the mid-2000s, including treatment of patients with excessive sleepiness due to shift work sleep disorder (SWSD) and treatment of patients with obstructive sleep apnea (OSA) to augment nasal continuous positive airway pressure (CPAP). These additional indications have helped fuel discussion of the more general need for assessment and treatment of pathologic sleepiness in clinical practice.

Hypnotics (Sodium Oxybate)

The most recent addition to the armamentarium of wake-promoting treatments is, paradoxically, a hypnotic, sodium oxybate, the sodium salt of γ-hydroxybutyrate (GHB). GHB, a naturally occurring inhibitory neurotransmitter that binds to γ-aminobutyric acid B (GABA-B) and GHB receptors, was first used as an anesthetic and neuroprotective agent in the 1960s. In 1979, Broughton and Mamelak described improvements in nighttime sleep, daytime alertness, and cataplexy symptoms in a group of 16 narcolepsy patients who took GHB at night, with sustained treatment effects over 20 months.[19] Subsequent research in the 1980s and 1990s confirmed GHB as an effective anticataplectic agent that also evoked improvements in daytime alertness,[20] although often patients required additional daytime use of traditional stimulant medication. GHB had previously been described as a "cataplexy antagonist and mild stimulant" but has more recently been recognized as a wake-promoting agent.[21] In 2002, sodium oxybate was granted an FDA indication for the treatment of cataplexy in narcolepsy, and an additional indication for the treatment of excessive sleepiness in narcolepsy was added in 2005.

WAKE-PROMOTING AGENTS: CAFFEINE

Mechanism of Action

Caffeine's main mechanism of action on the CNS is antagonism of adenosine receptors. Adenosine-releasing neurons are found in the hypothalamus and project to cells in the cortex, basal forebrain, and reticular activating system. It is known that endogenous adenosine levels rise with continued wakefulness and may be a fundamental part of the homeostatic sleep mechanism.[22] Exogenous adenosine promotes slow wave sleep, whereas xanthines, including caffeine, block the A1 adenosine receptors, thereby inhibiting sleep onset and maintenance. Caffeine inhibits sleep in other mammals and insects through similar mechanisms.[23]

Pharmacokinetics and Dynamics

Following oral ingestion caffeine reaches peak plasma levels within 30 to 120 minutes. Caffeine then undergoes hepatic metabolism, with metabolites excreted in the urine. The half-life of caffeine varies, ranging between 4 and 6 hours.[24] In smokers, clearance rate is increased by more than 50%.[25] In contrast, in women taking oral contraceptives and during pregnancy, caffeine's half-life may be prolonged twofold to threefold, possibly through CYP1A interactions.[26]

Alerting Effects and Clinical Efficacy

Caffeine improves alertness, mood, and cognitive performance. The usual dose in tablet form is 50 to 200 mg, and beverages contain amounts within this range as well (see Table 44-1). In standard daily practice, 85% of Americans use caffeine, many to foster wakefulness when arising from sleep.[27] This culturally accepted truism has been empirically examined, and it is clear that caffeine effectively eliminates the cognitive fog of sleep inertia on psychomotor tasks.[28] The combination of caffeine and naps, with caffeine intake immediately preceding a short 20-minute nap, appears to be especially effective in reducing subjective sleepiness and postnap sleep inertia, and improving objective performance in working memory tasks during the "midafternoon dip."[29]

Use in Sleep Deprivation

The sleepiness caused by sleep deprivation in young, healthy, non–caffeine-dependent volunteers can clearly be attenuated both subjectively and objectively using caffeine supplements. Using doses of 600 mg of a sustained-release preparation, caffeine reduced slow wave activity on the electroencephalogram and improved psychomotor performance tasks after up to 36 hours of sleep deprivation.[30] Similarly, two 300-mg doses of sustained-release caffeine significantly improved both vigilance and performance during 64 hours of continued wakefulness.[31] In a study of U.S. Navy SEALs randomly assigned to receive caffeine doses of 100, 200, or 300 mg or placebo after 72 hours of sleep deprivation and with continuous exposure to other stressors, caffeine at doses of 200 mg or above clearly improved tests of vigilance, alertness, and reaction time. However, it did not improve marksmanship, a task that requires fine motor control that tends to be worsened by caffeine.[32]

A more recent study demonstrated that caffeine (5 mg/kg) administered after 36 hours of sleep deprivation in normal subjects significantly improved reaction times as well as physical performance compared with placebo.[33] One of the few head-to-head studies comparing the effects of caffeine (600 mg), dextroamphetamine (20 mg), and modafinil (400 mg) on psychomotor vigilance after 44 hours of wakefulness found similar improvements in performance with all three stimulants, although caffeine had a shorter duration of action.[34]

Although caffeine can promote wakefulness, it should not be assumed that it will reverse all the effects of sleep loss on cognition and emotional regulation. Studies have begun to address this question and assess the effects of caffeine on restoring higher order executive function during sleep deprivation. In one study of extended sleep deprivation for 3 nights (77 hours), administration of 200 mg of caffeine every 2 hours from 1 AM to 7 AM each night (800 mg total dose per night) improved planning speed, response time, and throughput compared with placebo on a visuospatial planning and sequencing task known to be mediated by the dorsolateral prefrontal cortex.[35] Administration of caffeine was also reported to reduce the increases in a risk-taking behavior task produced by 75 hours of sleep deprivation.[36]

Use in Shift Work

Because it is such a widely available alerting agent, caffeine stands in a unique position to help improve the safety of shift workers and drivers, but not without caveats. Caffeine has been shown to substantially improve alertness in a simulated night shift.[37] In general, night shift workers tend to consume more caffeine than day workers, yet they continue to be at risk for accidents both on the road and at the workplace.[38] If used in sufficient doses (usually at least 200 mg), caffeine may significantly improve the alertness and cognitive skills that become impaired by sleepiness, especially in individuals who are not already moderately caffeine dependent. Caffeine's effectiveness is greater with the sustained-release form of a 600-mg daily dose, which has been demonstrated to extend the benefit of short naps following partial sleep deprivation in a driving simulation task.[39] Compared with subjects who had received a placebo, normally rested subjects who took 200 mg caffeine or a 30-minute nap 1 hour before driving 200 km at night (between 2 and 3:30 AM) in a driving simulator showed significantly lower incidence of impaired driving as indicated by inappropriate line crossings and subjective sleepiness.[40] Clearly, there are multiple factors, including tolerance or habituation and the timing of driving home in relation to the circadian nadir, that compromise the ability of caffeine (or any alerting agent) to mitigate severe sleepiness.

Potency

Compared with the potency of other alerting medications, caffeine is a moderately effective alerting agent when taken on an intermittent basis. Parkes and Dahlitz estimated that a dose of six cups of strong coffee has about the same alerting effect as 5 mg dextroamphetamine.[41] The duration of caffeine's effect on alertness appears to be dose dependent, with 75 to 150 mg of caffeine (1 cup of coffee) lasting up to 90 minutes after administration, 200 mg (approximately 2 cups of coffee) improving performance up to 4 hours after administration, and 300 to 400 mg (3 to 4 cups of coffee) sustaining alertness for up to 5.5 to 7.5 hours.[24] High doses of caffeine (200 to 600 mg) may approximate the efficacy of standard doses of modafinil (200 to 400 mg) in maintaining alertness and performance during long-term sleep deprivation.[42] Importantly, conditions of prolonged sustained sleep deprivation (24 to 44 hours) may dissociate the effect of caffeine on alertness from its effect on cognitive performance because decision making may remain impaired despite improved vigilance.[43]

Despite its documented ability to promote wakefulness, the potential benefits of caffeine to counteract sleep loss or shift work for alertness may be suboptimal because (1) it may not be consumed in adequate doses; (2) acute benefits are relatively short lived, and so it must be taken at the right time; and (3) development of tolerance leads to reduced efficacy overall. Caffeine may also be insufficiently potent in situations of new or worsening hypersomnia, and it is ineffective as monotherapy for the severe sleepiness of sleep disorders such as narcolepsy and idiopathic CNS hypersomnia.

Side Effects and Morbidity

The most common side effect of caffeine use is disrupted nighttime sleep. If taken before sleep, caffeine postpones sleep onset and reduces the amount of slow wave sleep.[44] The disruptive effects of caffeine on sleep maintenance are also well known; typically, if caffeine is consumed within a few hours of bedtime, sleep efficiency and total sleep time are both decreased. A recent study reported that 400 mg of caffeine

administered even 6 hours before bedtime in normal sleepers led to a reduction in total sleep time of more than 1 hour.[45] Individual sensitivity to caffeine's effects varies, likely based on multiple factors. Genetic studies in humans have demonstrated differential sensitivity to both wake-promoting and anxiety-eliciting effects of caffeine in relation to polymorphisms in the adenosine A2A receptor gene.[46] The decline in metabolic rate with age, leading to an increased half-life in older adults, is another factor that makes caffeine an increasingly likely contributor to sleep fragmentation in some, especially older, adults.

At high doses (above 4 mg/kg body weight), caffeine stimulates the medullary vagal, vasomotor, and respiratory centers,[47] as well as skeletal muscle,[48] giving rise to a variety of common side effects: nausea and diarrhea, flushing, sweating, increased heart and respiratory rates, muscle twitches and cramps, tremor, and nervousness. The lethal dose of caffeine is quite high—more than 10 g for an adult, or the equivalent of 100 cups of coffee.

Although coffee may exacerbate several disorders, such as osteoporosis, fibrocystic breast disease, irritable bowel syndrome, and peptic ulcer disease,[49] caffeine use in moderation appears to be generally safe. In fact, in 2012 the FDA stated that doses up to 400 mg/day do not appear to be associated with adverse health effects in healthy adults.[5]

Additional Health Benefits and Uses

Caffeine or coffee consumption has been associated with, among others, weight loss and insulin sensitization; lower risk for type 2 diabetes, hypertension, depression, symptomatic gallstones, and hepatocellular and colorectal malignancies; and possible neuroprotection, with lower incidence of Parkinson and Alzheimer disease.[50] Contrary to long-standing clinical suspicion, coffee consumption not only appears nonharmful but in fact also is possibly beneficial for cardiovascular morbidity and mortality. Although acute coffee intake does increase systolic blood pressure, habitual use of up to 6 cups of coffee per day was not associated with development of hypertension in the Nurses' Health Study. Caffeine doses as high as 500 mg/day did not precipitate or worsen ventricular arrhythmias, and increased coffee consumption (3 to 4 cups per day) is protective against atrial fibrillation and is associated with reduced incidence of stroke, heart failure, and coronary artery disease.[51]

Withdrawal

Caffeine in even moderate daily doses has been shown to produce a withdrawal syndrome after abrupt cessation. In one double-blind, placebo-controlled study, an average of 235 mg/day was consumed. On discontinuation, subjects reported headache, increased sleepiness and fatigue, fogginess and difficulty concentrating, and depressed mood, with symptoms emerging within 12 to 24 hours and peaking between 20 and 51 hours after caffeine cessation.[52] Although there are clearly both physical and mental changes associated with withdrawal, an expectation of symptoms may also increase the likelihood that they emerge.[53]

Tolerance

Although regular caffeine consumers frequently report decreased effectiveness of caffeine in the maintenance of subjective alertness, physiologic tolerance to caffeine may manifest in some (e.g., mood, response time) but not other (e.g., working memory) aspects of cognitive function.[24] Nevertheless, objectively measured sleep latencies on the Multiple Sleep Latency Test (MSLT) increase most notably on the first day of caffeine supplementation and subsequently decline, although values remain significantly higher than placebo, suggesting persistent benefit despite some possible development of tolerance.[54]

Dependence and Abuse Potential

Although caffeine discontinuation leads to withdrawal symptoms and cravings, lack of significant decrements in social, emotional, or physical well-being generally prevents substance abuse experts from considering caffeine dependence as a serious addiction.[51]

WAKE-PROMOTING AGENTS: SYMPATHOMIMETICS

Mechanism of Action

As discussed in detail elsewhere (see Chapter 44), the sympathomimetics directly or indirectly increase the activity in dopaminergic and noradrenergic pathways by blocking dopamine (DA) and norepinephrine (NE) reuptake and inducing DA/NE release through the dopamine (DAT) and norepinephrine (NET) transporters. The primary effect on alertness is mediated through the dopaminergic ventral tegmental area and the noradrenergic locus coeruleus, which both project widely throughout the brain. The additional activation of subcortical target areas (e.g., striatum, nucleus accumbens) accounts for the side effects typical of the sympathomimetics (e.g., tics) and abuse liability. It is important to recognize that wake-promoting medications provide symptomatic treatment but do not modify the underlying pathophysiologic processes leading to sleepiness, which are frequently not understood.[1]

Pharmacokinetics and Dynamics

Currently, there is a long list of immediate- and delayed-release sympathomimetic stimulants, whose development and use have been driven primarily by the attention deficit-hyperactivity disorder (ADHD) field. The available preparations offer a range of half-lives and therefore dosing strategies to treat sleepiness. Immediate-release amphetamines are absorbed rapidly and, on average, reach peak plasma levels within 2 hours of oral ingestion and have half-lives in the range of 4 to 6 hours. They undergo hepatic metabolism and renal excretion, the latter significantly increased at low urinary pH. Therefore urine acidification (e.g., with orange juice or ascorbic acid consumption) significantly reduces the elimination half-life and, thus, efficacy of amphetamines and methylphenidate, whereas urine alkalinization (e.g., with sodium bicarbonate or acetazolamide) prolongs their elimination half-life and may cause toxicity.[55]

Several preparations of amphetamine have been developed as oral compounds that vary in terms of the concentration of the dextro-isomer and whether a phosphate or sulfate salt is used. Although most methylphenidate preparations include racemic mixtures, Focalin consists of D-methylphenidate alone. Immediate-release methylphenidate has a rapid onset and shorter half-life on average (3 to 4 hours) compared with the amphetamines and thus can be administered two to four times daily. Sustained-release formulations of

methylphenidate and amphetamine have longer half-lives (8 to 16 hours). Even the pharmacokinetics of different formulations of the same stimulant vary and may affect a patient's level of alertness throughout the day. For instance, two commercial preparations of sustained-release methylphenidate (Ritalin LA and Concerta ER)[56] compared across several subjects exhibit similarly timed bimodal peaks in plasma levels after dosing but significantly different blood levels between formulations.

Combination of sympathomimetics with monoamine oxidase inhibitors (MAOIs, such as tranylcypromine, pargyline, phenelzine, and high-dose selegiline) is contraindicated because MAOIs inhibit hepatic metabolism of amphetamine and may result in hypertension or hyperthermia. However, coadministration of sympathomimetics and low-dose tricyclic antidepressants (TCAs) used as anticataplectic agents (e.g., imipramine, protriptyline, and clomipramine at 10 to 100 mg) appears generally safe, although TCAs also inhibit amphetamine and methylphenidate metabolism and thus may lead to reduced dosing requirements of the sympathomimetics.[57] Synergistic effects between methylphenidate and selective serotonin reuptake inhibitors (SSRIs) have been reported, possibly owing to increased monoaminergic tone at the synapse or decreased SSRI metabolism, with at least one case report of serotonin syndrome in a patient taking sertraline and methylphenidate.[58]

Alerting Effects and Clinical Efficacy

The clinical treatment of excessive sleepiness due to narcolepsy originated with the traditional stimulants, and the dosing guidelines have changed little since their development in the first part of the last century. Clinical practice parameters are thus based on a few small trials, without effective assessment of risk-benefit ratios, long-term efficacy, and side-effect profiles.[1]

In one double-blind, randomized protocol comparing 8 narcoleptics with cataplexy to matched controls, MSLT sleep latencies increased from 4.3 minutes (placebo) to 9.3 minutes (methamphetamine 60 mg) in narcoleptics and from 10.3 minutes (placebo) to 17.1 minutes (methamphetamine 10 mg) in controls.[59] Thus although mean sleep latencies increased with high-dose methamphetamine in narcoleptics, they did not completely normalize compared with controls, remaining pathologically low (less than 10 minutes). Importantly, functional improvement accompanied the reduction in sleepiness, with significantly fewer objects hit on a driving stimulator test following amphetamine administration both in the narcoleptics (0.3 versus 3) as well as in the controls (0.16 versus 0.8).

In another randomized controlled trial, Mitler and colleagues compared 13 narcoleptic patients given methylphenidate at 10, 30, or 60 mg total daily dose (taken in divided doses three times per day) to 5 narcoleptic patients taking dextroamphetamine 10, 30, or 60 mg maximum total daily dose (also taken divided into three doses over the day) and 9 control subjects taking placebo.[60] After 7 days of drug or placebo administration, participants underwent objective (Maintenance of Wakefulness Test [MWT]) and subjective evaluation of sleepiness as well as cognitive testing (Wilkinson Addition and Digit-Symbol Substitution Tests). Both methylphenidate and dextroamphetamine showed dose-dependent improvements in mean sleep latency on MWT. However, dextroamphetamine 60 mg showed significantly greater relative improvement, with mean latency improving from 35% to 70% of control values (18.9 minutes), compared with an increase from 55% to 80% of control with methylphenidate 60 mg. The difference in baseline sleep latency values between dextroamphetamine and methylphenidate treatment groups may partially be related to the small sample size and additional factors such as mean age (50 years old in the methylphenidate group versus 39 years old in the dextroamphetamine group). Subjective improvement in sleepiness and cognitive testing was only seen at the maximum daily dose of 60 mg methylphenidate but was seen with all doses of dextroamphetamine, including the lowest 10-mg dose.

Current commonly used sympathomimetics for the treatment of hypersomnia include immediate-release amphetamines, D-amphetamine (Dexedrine), racemic D/L amphetamine (Benzedrine), and methamphetamine (Methadrine); delayed-release amphetamine formulations Adderall and Vyvanse; short-acting racemic methylphenidate (Ritalin); and the bioactive d-methylphenidate formulation (Focalin), supplemented by the osmotic release oral system (Concerta) (Table 44-2). In general, the specific pharmacokinetic profile must be considered when prescribing stimulants because it is often the most important element in shaping a patient's wakefulness throughout the day and ability to tolerate one formulation of a medication better than others.

Potency

A useful distinction is one between the amphetamine derivatives and the piperazine derivative methylphenidate. Although both amphetamines and methylphenidate block DA and NE reuptake and induce catecholamine release through interactions with DAT and NET, amphetamine also binds the intracellular vesicular monoamine transporter, thus additionally potentiating catecholamine release compared with methylphenidate, which does not bind vesicular monoamine transporter.

Additionally, the dextro-isomer d-amphetamine appears to increase DA release preferentially to NE release compared with L-amphetamine and induces wakefulness more potently. Methamphetamine, which has an additional methyl group attached to the amine, has increased CNS penetration and is thus more potent than amphetamine. These subtle molecular differences may account for the differing clinical efficacy and side-effect profiles of the sympathomimetics such that, for example, methylphenidate may be less efficacious but also easier to tolerate.

Side Effects and Morbidity

Common stimulant side effects include irritability, nervousness or tremulousness, insomnia, orofacial dyskinesias, and headache. Sympathomimetic activation may cause palpitations, tachycardia and hypertension, diaphoresis, anorexia, and vomiting.[17,61] The reported frequency of side effects of stimulants in clinical practice and in clinical trials varies from 0% to 73%; the extreme variation reflects, at least in part, differences in methods of determining side effects and the definitions of side effects. Studies show that at high doses, most patients experience side effects, including disturbed nocturnal sleep.[59]

Cardiac and vascular complications due to prescribed sympathomimetics have been reported only rarely in people with narcolepsy. These drugs do not appear to cause clinically significant increases in blood pressure at commonly used doses

Table 44-2 Medication and Dosage

Medication	Dose	Onset to Peak Concentration	Half-Life	Common Side Effects	Serious Side Effects	Important Drug Interactions	Contraindications and Precautions	Comments
Amphetamine/Dextroamphetamine								
Amphetamine/dextroamphetamine IR (Adderall)	5–60 mg	2–3 hr	7–34 hr (average 10 hr)	Weight loss, headache, insomnia, tremor, abdominal pain, anorexia, xerostomia, dysphoria, euphoria, anxiety, restlessness	Cardiomyopathy, chest pain, sudden death, MI, irregular heart rate, immune hypersensitivity reaction, CVA, Tourette syndrome, seizure, hypertension, palpitations, psychotic disorder with prolonged use	MAO inhibitors: hypertensive crisis SSRIs, SNRIs: increased risk for serotonin syndrome Sodium bicarbonate: amphetamine toxicity by decreasing urinary excretion/increasing half-life Ascorbic acid: increased urinary excretion/decreased half-life	Advanced atherosclerosis, cardiovascular disease, concomitant use of MAOIs or within 14 days of MAOI use, drug dependence, structural cardiac abnormalities, hyperthyroidism, moderate to severe hypertension Can lower seizure threshold	Black box warning: high potential for abuse
Dextroamphetamine IR (Dexedrine)	5–60 mg	2–3 hr	10 hr					
Lisdexamphetamine (Vyvanse)	30–70 mg	1 hr	4 hr					
Methamphetamine (Desoxyn)	5–60 mg	30–60 min	4–5 hr					
Amphetamine/dextroamphetamine XR (Adderall XR)	10–60 mg	7 hr	12 hr					
Dextroamphetamine SR	5–60 mg	8 hr	12 hr					
Methylphenidate								
Methylphenidate hydrochloride (Ritalin, Concerta)	10–80 mg	1–2 hr (food slows absorption)	3 hr	Loss of appetite, irritability, anxiety, restlessness	Hypertension (frequent), tachyarrhythmia (frequent), thrombocytopenia, hallucinations	Warfarin: increased plasma concentrations and an increased risk for bleeding	Caution in patients with a history of drug dependence or alcoholism	Black box warning: high potential for abuse
Methylphenidate hydrochloride ER (Ritalin ER, Concerta ER, Metadate CD, Methylin ER)	10–60 mg	1.3–4 hr (food slows absorption)	3.5 hr (6–12 hr)			Warfarin: increased plasma concentrations and increased risk for bleeding MAO inhibitors: hypertensive crisis Phenytoin, phenobarbital increased serum levels	Caution in patients with a history of drug dependence or alcoholism Contraindicated in patients taking MAOIs and patients with glaucoma, motor tics, Tourette syndrome	
Other Drugs								
Modafinil (Provigil)	100–800 mg	2–4 hr	9–14 hr (R isomer 15 hr, S isomer 4–5 hr	Headache, nausea, anxiety, insomnia, dizziness	Drug hypersensitivity syndrome, Stevens-Johnson syndrome, toxic epidermal necrolysis due to drug hypersensitivity reaction, hypertension	OCPs: decreased bioavailability and reduced effectiveness Other interactions: diazepam, propranolol, phenytoin, cyclosporine, carbamazepine, clomipramine	Angioedema, hypersensitivity reaction, anaphylactoid (rare)	Low abuse potential, Schedule IV
Armodafinil (Nuvigil)	50–300 mg	2–4 hr	10–15 hr, plasma levels remain elevated significantly longer compared with modafinil					

Continued

Table 44-2 Medication and Dosage—cont'd

Medication	Dose	Onset to Peak Concentration	Half-Life	Common Side Effects	Serious Side Effects	Important Drug Interactions	Contraindications and Precautions	Comments
Ritanserin	10–30 mg	140 min	40 hr	Constipation	Prolongation of the QTc interval	Droperidol: increased risk for cardiotoxicity (QT prolongation, torsades de pointes, cardiac arrest)	Patients concurrently receiving class III antiarrhythmic agents or drugs known to cause hypokalemia, prolongation of QT interval arrhythmias	Some improvement in daytime sleepiness in patients with narcolepsy
Selegiline	5–10 mg	40–90 min	10 hr	Decreased systolic arterial pressure, orthostatic hypotension, weight loss, diarrhea, indigestion, headache, insomnia, xerostomia	Hypertensive crisis, suicidal thoughts	Meperidine, methadone, propoxyphene, tramadol: severe hypertension or hypotension, hyperpyrexia, coma, death SSRIs: increased risk for serotonin syndrome Albuterol: increased risk for tachycardia, agitation, or hypomania TCAs: hyperpyrexia, convulsions, death Fentanyl: severe and unpredictable potentiation of opioid analgesic effects	Meperidine, methadone, propoxyphene, tramadol, carbamazepine, oxcarbazepine, cyclobenzaprine, bupropion, mirtazapine, St. John's wort, SSRIs, albuterol TCAs, fentanyl: contraindicated Concomitant use of dextromethorphan: can cause psychosis or unusual behavior Pheochromocytoma Caution with tyramine-rich foods, increased risk of hypertensive crisis	Black box warning: increased the risk for suicidal thinking and behavior in children, adolescents, and young adults with major psychiatric disorders Very few data to support use for daytime sleepiness
Sodium oxybate (γ-hydroxybutyrate [GHB]) (Xyrem)	2.25–9 g in divided doses	15–30 min to peak concentration Fatty food delays absorption	30–60 min	Nausea, vomiting, enuresis, dyspepsia, abdominal pain, confusion, dizziness, somnolence, headache, incontinence	Respiratory suppression, sleepwalking, depression	Benzodiazepines Opiates may have additive CNS and respiratory depressant effects	Succinic semialdehyde dehydrogenase deficiency, concurrent treatment with sedative-hypnotics, alone or combined with alcohol, has a high propensity to induce a comatose state	Can be used as drug of abuse Need to be registered to prescribe in the United States

CNS, Central nervous system; CVA, cerebrovascular accident; MAO, monoamine oxidase; MAOI, monoamine oxidase inhibitor; MI, myocardial infarction; OCP, oral contraceptive pill; SNRI, serotonin-norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

in normotensive individuals.[17,61] Isolated cases of severe disease such as stroke, cardiomyopathy, and ischemic vascular complications have been reported in the context of chronic use of sympathomimetics, especially at high doses. Although advanced cardiovascular disease is a reasonable contraindication to sympathomimetic therapy, there are no systematic studies indicating that well-controlled hypertension is exacerbated by moderate doses of stimulants. Again, methylphenidate appears to result in less hypertension (as well as appetite reduction, another common side effect) compared with the amphetamines.[62]

Psychiatric complications with the use of sympathomimetics, including delusions, paranoia, and mania, are dose dependent and more likely to occur in patients with coexisting or preexisting psychiatric conditions.[63] Psychosis and hallucinations are rare in narcoleptic patients treated with stimulants. There is no evidence that different agents confer a greater or lesser risk for psychotic symptoms, although the use of short-acting forms is associated with mood swings and irritability. Methylphenidate 20 to 60 mg does not appear to worsen clinical measures of impulsivity or addictive behaviors in narcoleptic patients.[64]

A variety of complications can occur with intravenous, intranasal, or oral amphetamine or methamphetamine abuse. In healthy volunteers, repetitive oral administration of 5 to 10 mg of dextroamphetamine produces paranoid delusions, often with blunted effects, after a cumulative dose of 55 to 75 mg.[65] Other symptoms of amphetamine abuse are motor tics, stereotypic movements, and perseveration: repetitive thoughts or organized, goal-directed, but meaningless activity, such as repetitive cleaning or elaborate sorting of small objects.[66] In young adults, the relative risk for stroke is estimated to be 6.5 times greater for drug abusers compared with nonabusers, with amphetamines implicated in a substantial proportion of young drug abusers with strokes.[67]

Additional Health Benefits and Uses

Sympathomimetic stimulants appear effective in treatment-resistant depression, although no controlled trials have been performed to confirm this effect and to investigate true antidepressant qualities rather than fatigue reduction or increased motivation as a result of amphetamine intake.[55] Bronchodilation and weight loss are known side effects of the sympathomimetics, which may be beneficial in certain clinical scenarios.

Withdrawal

Abrupt discontinuation of amphetamines can result in prolonged bouts of recovery sleep, disrupted sleep including vivid or unpleasant dreams, depressive mood, and worsening of daytime sleepiness.[68]

Tolerance

In people with narcolepsy, tolerance to alerting effects appears to occur with variable frequency. In one review, 10 of 100 patients had discontinued stimulants owing to failure to respond, tolerance, or side effects, and 31 others had required doubling of dosage over a 1-year period for the same control of symptoms.[61] Other studies have found a similar or higher amount of tolerance evident clinically in patients using sympathomimetic agents.[69] Tolerance to stimulants appears to be more likely, or at least more evident, in patients taking high doses. There is little evidence that the incidence of tolerance and side effects is less in people with narcolepsy than in others taking sympathomimetics. Furthermore, it does not appear that tolerance reported by some patients is an effect of inadequate nocturnal sleep rather than true tolerance, nor does tolerance appear less likely to occur with methylphenidate than with dextroamphetamine.[70]

Dependence and Abuse Potential

Amphetamines and related compounds have a high abuse potential and can produce dependence. Although most users do not become addicted, controlled use may become compulsive use, especially when high doses or rapid route of administration are used.[71] A sequence of euphoria, dysphoria, paranoia, and psychosis can occur after a single exposure to a high dose or with chronic exposure to low doses. Because of its increased lipophilicity and thus rapid CNS penetration and onset of action, methamphetamine has the greatest abuse potential.

WAKE-PROMOTING AGENTS: MODAFINIL AND ARMODAFINIL

Mechanism of Action

Modafinil (the racemic mixture of R- and S-enantiomers) and armodafinil (the R-enantiomer preparation) are chemically unrelated to the sympathomimetics agents and are sometimes referred to as somnolytics rather than stimulants.[72] The precise mechanism through which modafinil enhances wakefulness remains unclear. A comprehensive discussion of modafinil's mechanism of action is included elsewhere (see Chapter 43). However, modafinil likely blocks dopamine reuptake predominantly, through differential involvement of the dopamine rather than norepinephrine transporters,[55] possibly accounting for its more benign cardiovascular and tolerance and abuse side-effect profiles compared with the sympathomimetics.[73] It has been postulated that modafinil exerts its effects by modulating the homeostatic sleep drive (e.g., by decreasing recovery sleep duration following prolonged sleep deprivation). However, studies have not demonstrated clear effects on sleep homeostasis, such as increased homeostatic sleep pressure inducing rebound sleepiness following discontinuation of the drug. Furthermore, beyond the initial acclimatization period, modafinil's alerting effects do not appear to disrupt the evolution of normal sleep architecture. Thus modafinil likely exerts its alerting effects through activation of dopaminergic wake-promoting mesocortical pathways.[74]

Pharmacokinetics and Dynamics

Modafinil is absorbed quickly and reaches peak plasma levels within 2 to 4 hours, with a half-life of 9 to 14 hours. The onset of action and half-life of armodafinil are similar to those of modafinil, but the pharmacokinetics of the two drugs are quite different, partially owing to the much shorter half-life of the S-enantiomer (3 to 4 hours), which is present in modafinil but not armodafinil.[75] Thus armodafinil plasma levels remain elevated significantly later in the day compared with modafinil, allowing once-daily armodafinil dosing, whereas modafinil is frequently used in divided doses. Modafinil is primarily metabolized by CYP3A4 and is renally excreted; lower doses should be used in patients with renal and hepatic dysfunction. Modafinil is a CYP3A4 inducer and thus may increase metabolism, thereby decreasing efficacy of oral contraceptives as well as triazolam, diazepam, and phenytoin. Alternative contraceptive methods should be used by women

of childbearing age. Armodafinil is also a moderate inhibitor of CYP2C19, which metabolizes coumadin, and thus potential dose reductions may be necessary.

Alerting Effects and Clinical Efficacy

Use in Narcolepsy

Although in some individuals 100 mg of modafinil is sufficient to sustain alertness for several hours, most patients with excessive sleepiness require doses of 200 mg per day or higher. In two large populations of narcoleptic patients taking 200 to 400 mg per day, alertness measures (MWT, Epworth Sleepiness Scale [ESS], Clinical Global Impression of Change) gradually increased over 9 weeks of double-blind treatment.[76,77] In one study, mean latencies on the MSLT improved by 1.8 and 19.9 minutes with modafinil 200 and 400 mg, respectively (from baseline of 2.9 and 3.3 minutes), and ESS score declined by 3.5 and 4.1 points from 17.9 and 17.1, respectively.[76] Some individuals with severe sleepiness may require modafinil at 600 to 800 mg per day in divided doses (morning and no later than early afternoon to avoid insomnia) for effective control of their symptoms.[78,79] Although these doses are significantly above the FDA-indicated guidelines, if lower doses are well tolerated but ineffective, then it is reasonable to titrate up to higher doses. Armodafinil's potency is estimated at approximately twice that of modafinil; thus initial dosing may start as low as 50 mg, increasing to as much as 250 mg in the morning. Armodafinil was found to increase mean sleep latencies on MWT of narcoleptics by 1.3 minutes (from baseline of 12.1 minutes) and 2.6 minutes (from baseline of 9.5 minutes) at 150 mg versus 250 mg respectively.[80]

Use in Idiopathic Hypersomnia

Both sympathomimetics (methylphenidate) and modafinil have been used in the treatment of idiopathic hypersomnia (IH), although no large randomized and controlled trials have been performed. A recent cohort study compared modafinil (50 to 600 mg/day) in patients with diagnoses of IH with ($n = 59$) and without ($n = 45$) long sleep time and in patients with diagnoses of narcolepsy with cataplexy ($n = 126$). This study found similar improvements in subjective sleepiness in IH (ESS, −2.6) and narcolepsy and cataplexy (ESS, −3) patients. As a group, IH patients without long sleep time appeared more impaired at baseline (ESS, 18) and showed greater benefit with modafinil (ESS, 12) compared with the IH patients with long sleep time (ESS, 15 at baseline and 13.7 following modafinil treatment). The side-effect profile was similar, with more frequent side effects reported in the IH groups (nervousness 14%, palpitations 13%, headache 11%) compared with the narcolepsy and cataplexy group.[81] A small randomized crossover double-blind placebo-controlled trial showed objective improvement with higher MWT mean sleep latencies in 13 patients with narcolepsy and 14 patients with IH who took modafinil 400 mg/day for 5 days (30.8 minutes) compared with narcolepsy and IH patients taking placebo (19.7 minutes; controls = 39.6 minutes); this improvement was also correlated with better performance on an open highway driving test.[82]

Use in Sleep-Disordered Breathing

Modafinil and armodafinil are approved by the FDA for patients with OSA who have disabling sleepiness despite OSA-specific treatments such as nasal CPAP. Nasal CPAP treatment has been clearly demonstrated to improve alertness in patients with OSA,[83] but even with optimal mechanical therapy, chronic sleepiness remains a problem for some patients with sleep-disordered breathing. Indeed, a recent study of patients with OSA demonstrated a clear dose-response relationship between hours of CPAP use during sleep and both subjective and objective daytime sleepiness. However, about 20% of those study subjects with an average of 8 hours of use of CPAP per night remained excessively sleepy by self-report.[84] It has been hypothesized that this residual sleepiness in OSA patients is a long-term effect of the intermittent hypoxic episodes that their wake-promoting brain areas were exposed to before therapy.[85]

Whatever the underlying cause, it is clear that a subset of patients with OSA experience chronic residual sleepiness despite their compliance with mechanical treatments during sleep. For these patients, the adjunctive use of modafinil appears to be a reasonable and safe measure to improve their safety and quality of life. In a large, double-blind placebo-controlled study of patients with OSA reporting residual excessive sleepiness while on CPAP, modafinil at doses of 400 mg improved alertness by 2.6 points on the ESS above placebo-treated patients, and more than half the modafinil-treated patients reported normal ESS values (score of less than10) by the study end point.[86] Further, in a 12-week follow-up open-label study, adjunct modafinil treatment improved objective measures of alertness on the MSLT (8.6 minutes compared with 7.4 minutes at baseline). However, a small drop in CPAP use was also noted (5.9 hours/night in modafinil group compared to 6.3 hours/night during double-blind baseline). Subsequent studies have demonstrated the efficacy of modafinil in daily doses of 200 to 400 mg for improving alertness in CPAP-treated patients with OSA and residual sleepiness and confirmed a relative absence of adverse consequences in this patient population.[87]

Use in Shift Work

Millions of adults keep nonstandard work hours, with many experiencing chronic, problematic sleepiness as a result. Although many shift workers adapt adequately to the constraints of their schedules, there are many more who suffer at least transiently from the effects of both sleep deprivation and circadian misalignment. Furthermore, it is estimated that approximately 10% of the adults working nonstandard hours have persistent complaints of excessive sleepiness or insomnia consistent with the diagnosis of SWSD.[88] A double-blind placebo-controlled study of more than 200 night-shift workers demonstrated this group to be pathologically sleepy at baseline (MSLT average sleep latencies approximately 2 minutes), with significant cognitive impairment on a psychomotor vigilance task, as well as numerous mistakes, near misses, or accidents at work or while driving home after work. All of these measures improved substantially after treatment with 200 mg modafinil taken at the beginning of their night shift; for example, MSLT mean sleep latencies improved by +1.7 with modafinil versus 0.3 minutes with placebo. Furthermore, this treatment did not interfere with their ability to sleep during time off duty.[89] Armodafinil 150 mg increased MSLT-measured mean sleep latencies by 3 minutes from 2.3 minutes at baseline in shift work disorder.[90] On the basis of this and other evidence, the FDA approved modafinil for the treatment of excessive sleepiness due to SWSD in 2004. Together with a program

of nonpharmacologic measures to protect sleep time and sleep ability in this patient population, modafinil or armodafinil is a potentially life-saving treatment for these adults.

Potency

In a paradigm comparing modafinil 100 to 400 mg to caffeine 300 mg ingested at 10 PM during an overnight work period spanning 7 PM to 8:45 AM, healthy non–sleep-deprived subjects reported less subjective sleepiness and performed better in vigilance, attention, and recall tasks at all doses of modafinil and caffeine compared with placebo, but modafinil at 300 and 400 mg outperformed caffeine 300 mg.[74]

Side Effects

Side effects are fewer with modafinil than with sympathomimetics. The most common adverse events in the initial modafinil and armodafinil trials were headache, nausea, and anxiety, which increased in frequency if the dose was high or increased too quickly; side effects were usually transient, resolving with acclimatization.[76] Insomnia has not been reported widely and again appears to be dose-related and transient. In one study, modafinil 300 to 400 mg was shown to disrupt recovery daytime sleep following acute overnight sleep deprivation when ingested 11 hours before recovery sleep, with increased sleep latency, reduced sleep efficiency, and greater wake time after sleep onset compared with placebo, caffeine 300 mg, or modafinil 100 to 200 mg.[74] Modafinil 100 to 600 mg does not appear to worsen clinical measures of impulsivity or addictive behaviors in narcoleptic patients.[64]

There have been no clinically significant cardiovascular adverse effects from modafinil or armodafinil treatment in the clinical trials to date, including among patients with OSA.[86] However, at least one small study in 12 healthy volunteers showed increases in resting heart rate (+9 beats/minute on average) and systolic blood pressure (+7.3 mm Hg on average) following ingestion of 400 mg of modafinil on 3 consecutive days. Interestingly, this was not reflected in measures of peripheral sympathoexcitation, namely peroneal microneurographic activation, and 33% of the participants were presyncopal with tilt table testing with either placebo or modafinil ingestion, thus raising a question of underlying pathophysiologic confounders.[91] So far, it appears that only patients with a history of sensitivity to activating medications (e.g., those with mitral valve prolapse) experienced cardiovascular side effects from modafinil (e.g., palpitations, chest pain), and these symptoms reversed when the medication was discontinued. Psychotic symptoms have developed rarely and only at high doses of modafinil.[92] Cases of modafinil-induced hypersensitivity reactions, including rare cases of life-threatening Stevens-Johnson syndrome, have also been reported.

Additional Uses

In randomized placebo controlled trials, modafinil (200 to 300 mg) showed equal or superior efficacy compared to methylphenidate (20 to 30 mg/day) in improving ADHD symptoms in children.[71] However FDA approval for ADD treatment was not granted because of rare cases of Stevens-Johnson syndrome.[93] Modafinil has also been used in Parkinson disease, myotonic dystrophy, multiple sclerosis, traumatic brain injury, depression, and chronic fatigue syndrome. Studies in Parkinson disease have been somewhat contradictory, with smaller studies showing reduction in subjective (ESS), but not objective (MWT), measures of sleepiness at 100 to 200 mg/day,[94] with no significant improvement in ESS or MSLT (–0.16 vs. –0.7 in placebo vs. modafinil 400 mg) in a larger study of 37 patients.[95] Modafinil appears to reduce sleepiness in MD.[71] It was not found to be efficacious in MS patients or patients with chronic fatigue, whereas subjective improvement in sleepiness and fatigue was reported by patients with major depression, but the effect did not extend beyond the first 2 weeks of treatment. Two trials have shown benefit with modafinil 200 to 400 mg for fatigue but not sleepiness in patients with traumatic brain injury.[96]

Withdrawal

In large clinical trials of patients taking stable doses of modafinil for sleepiness, abrupt discontinuation did not elicit specific symptoms of withdrawal with rebound hypersomnia; rather, patients simply returned to their initial level of sleepiness. Patients who discontinue modafinil will typically experience a full return of sleepiness symptoms within 2 to 3 days of cessation. There are generally no obvious recovery changes to nighttime sleep because modafinil does not appear to significantly alter nighttime sleep architecture during treatment. This lack of rebound following discontinuation is a significant advantage for patients already on modafinil requiring diagnostic polysomnography and daytime sleep testing because it significantly reduces the time patients need off medication before testing. Patients withdrawn from modafinil can be expected to be fully back to baseline within 5 days, whereas patients withdrawn from chronic stimulants will generally need at least several weeks before sleep testing to allow normalization of sleep architecture patterns and recovery from rebound hypersomnolence.

Tolerance

Modafinil appears to have a very low, or idiosyncratic, occurrence of tolerance. Clinical and subjective self-assessments of efficacy remained stable for most of those patients who enrolled in open-label studies taking the same dose of modafinil for 3 years.[97,98]

Dependence and Abuse Potential

Modafinil is a Schedule IV medication, with limited potential for abuse and dependence. In abuse liability studies conducted with seasoned substance abusers, modafinil was similar to caffeine in its rating as producing some "good effects" on a subjective rating scale, and it did not elicit any desire to procure more (i.e., "amount willing to pay" was $0).[99] These and other studies demonstrate that the effects of modafinil are clearly different from predictably dose-dependent euphoria and the desire to have more drug that is seen with traditional stimulants like amphetamine. Moreover, modafinil has a slower onset of action, and its water-insoluble properties make it impossible to snort or inject, so it is not pharmacokinetically amenable to abuse. Postmarketing surveys and medical literature to date have identified only idiosyncratic cases of people developing addictions or cravings for modafinil. However, there are some reports that at doses as high as 800 mg, polysubstance abusers described a "high" similar to methylphenidate, while healthy users reported "liking" similar to d-amphetamine.[71] Some feared that modafinil would be abused to extend the wake period by college students or others

in similar situations, but this abuse has so far not been reported to be a widespread phenomenon. This form of abuse of wake-promoting medication is likely limited by the ultimate need to sleep—no medication is really an effective substitute for sleep—and more likely to occur with the more robustly arousing, traditional stimulants.

ADDITIONAL DAYTIME WAKE-PROMOTING AGENTS

Atomoxetine (Strattera) is a nonstimulant NET-specific inhibitor, originally developed as an antidepressant but currently used primarily in ADHD. It has modest wake-promoting as well as anticataplectic effects, but tachycardia, hypertension, and sexual dysfunction are limiting side effects.

Bupropion, a low-potency nonspecific monoamine reuptake inhibitor that also has DAT- inhibitory properties, is sometimes used to combat excessive sleepiness.[100] Bupropion may be especially effective when depression is a major comorbidity. Dose-dependent risk for seizures has been reported.

Finally, selegiline, a methamphetamine derivative, is an MAO-B inhibitor with wake- promoting and anticataplectic properties, the former effect likely owing to its metabolites L-amphetamine and L-methamphetamine. Doses of 20 to 30 mg appear clinically effective in comparison to similar doses of D-amphetamine, with lower potential for abuse.[55]

FUTURE WAKE-PROMOTING CANDIDATE AGENTS

Histamine-3 (H_3) receptor antagonists or inverse agonists, both enhancing histamine release, have been proposed as wake-promoting agents. At least one randomized controlled crossover trial compared a one-time dose of an H_3 receptor inverse agonist to modafinil (200 mg) or placebo in 56 subjects with diagnosis of OSA (apnea-hypoxia index greater than 15), compliant with positive airway pressure and with self-reported regular bedtimes and nightly sleep opportunity of 6.5 to 8 hours. The study showed efficacy of the new agent but no improvement over modafinil (MWT sleep latencies were 8.1 and 10.2 minutes longer than placebo values for H_3 receptor inverse agonist and modafinil), with higher incidence of insomnia (29% for H_3 receptor inverse agonist versus 9% for modafinil and 6% for placebo).[101]

The report of a GABA-A receptor activating compound identified in the cerebrospinal fluid of 32 hypersomnolent patients points to GABA-A receptor antagonists, such as flumazenil, as possible wake-promoting agents.[102] Unfortunately, flumazenil is not currently available in an oral formulation, but a case report of continuous subcutaneous administration of flumazenil for 26 days in one subject with idiopathic hypersomnia described a decrease in self-reported sleep time from 13.5 hours/day at baseline to 9.5 hours/day. ESS score decreased from 21 to 11 as well.[103] Sixty-four percent of the hypersomnolent patients whose cerebrospinal fluid increased GABA-A receptor activity reported subjective reduction in sleepiness and showed improved psychomotor vigilance with clarithromycin treatment (mean dose 1098 mg), presumably through clarithromycin's GABA-A receptor antagonist function, which has been demonstrated in vitro.[104] However, case reports of clarithromycin-induced hypersomnia in children exist as well.[105]

Finally, thyrotropin-releasing hormone agonists and hypocretin replacement, either through delivery of a synthetic ligand or gene therapy, have been proposed as wake-promoting treatments, but no human clinical trials have been done at this time.[55]

SODIUM OXYBATE

Mechanism of Action

Commercially available sodium oxybate, Xyrem, is the sodium salt of GHB. It is a rapidly acting sedative-hypnotic medication used for the treatment of daytime sleepiness and cataplexy in narcolepsy patients. Although the precise mechanism of action is unknown, the effects may be mediated in part through interaction with GABA-B and GHB receptors[106] because GHB is a GABA derivative present endogenously in the mammalian brain.[107] Its highest concentration is in the dopaminergic regions such as the substantia nigra and ventral tegmental area, suggesting that endogenous GHB may also modulate the activity of dopamine neurons.[108] What is interesting, and unknown, is how GHB reduces sleepiness—or elevates wakefulness—in patients, although it could be related to increases in delta power during sleep; sodium oxybate leads to dramatic increases in the density and duration of slow wave sleep each night.[109] Patients with narcolepsy frequently experience side effects as soon as GHB treatment is initiated, but much like antidepressant therapy, improvements in daytime functioning become robustly evident only after sustained use for weeks to months. Indeed, most patients must initially remain on daytime wake-promoting and anticataplectic medications. Although it is unclear how sodium oxybate effects improvements in alertness and reduces cataplexy, it seems reasonable to suppose the mechanism may be related to (or a consequence of) the changes to sleep it evokes each night.

Pharmacokinetics and Dynamics

GHB is rapidly absorbed after oral administration with nonlinear pharmacokinetics such that increases in dosing result in disproportionately higher plasma levels and thus a narrow safety margin.[107] The drug is metabolized rapidly to succinic semialdehyde, then oxidized to succinic acid and ultimately metabolized to carbon dioxide in the Krebs cycle. Onset of action is as fast as 15 minutes and the half-life is as short as 30 to 60 minutes, potentially resulting in sleep maintenance insomnia and necessitating a second dose within 2.5 to 4 hours from the first nightly dose. Oral bioavailability is affected by food, especially high-fat food; therefore administration should remain as consistent as possible and meals should ideally be consumed several hours before bedtime. Sodium oxybate is contraindicated in patients with succinic semialdehyde dehydrogenase deficiency, owing to inability of these patients to metabolize the drug.

Alerting Effects and Clinical Efficacy

The first report that GHB could be an effective treatment for excessive sleepiness in narcolepsy was published in 1979 by Broughton and Mamelak.[19] This study, along with follow-up reports demonstrating use of GHB for treatment of cataplexy in narcolepsy, led to larger research protocols to confirm its effects. FDA approval of sodium oxybate for cataplexy was based on two randomized, double-blind, placebo-controlled trials in patients with narcolepsy who were also being treated

with traditional stimulants.[110,111] Subsequent large trials demonstrated the efficacy of sodium oxybate for the treatment of sleepiness associated with narcolepsy, allowing an expanded indication for the use of this medication in narcolepsy.

In one placebo-controlled, randomized study involving 136 narcoleptic patients with cataplexy, sodium oxybate improved subjective sleepiness as determined by the ESS in a dose-related manner, but the effect was only statistically significant at a dose of 9 g per night.[110] At this dose, the median ESS dropped from 17 to 12, with some patients falling in the normal range (ESS lower than 10). There was also a significant reduction in the number of unintended naps or sleep attacks seen at doses of 6 and 9 g. In another multicenter, randomized, double-blind, placebo-controlled, parallel-arm trial study of 228 patients with narcolepsy with moderate to severe excessive daytime sleepiness and cataplexy symptoms, sodium oxybate demonstrated a significant median increase of more than 10 minutes in the MWT, significant reduction in median ESS, and reduction in weekly unintended naps.[112] In both these studies, most patients remained on wake-promoting medications at stable doses during the study.

Sodium oxybate is taken nightly, in divided doses on an empty stomach (no food within 2 to 4 hours of bedtime). As discussed earlier, food will mitigate absorption of sodium oxybate, so a variable eating schedule is a common source of adverse effects and inconsistent efficacy of similar dosing in patients. The usual effective dose range is 4.5 to 9 g per night, with half the total dose taken immediately before lying down to go to sleep and the second half 2.5 to 4 hours later. Although some patients will respond to this medication at lower doses, often patients will require the recommended dose of 6 to 9 g per night. To minimize side effects, for some patients it may be necessary to begin at much lower doses (e.g., 1 to 2 g per night) and increase sodium oxybate by 0.5- to 1.0-g increments once every several nights. At the 6-g/day dose sleep paralysis appears to be decreased more consistently than hypnagogic hallucinations.

Potency

In a multicenter, double-blind, placebo-controlled study of subjects with narcolepsy who had been taking modafinil, the effects on sleepiness of switching them to sodium oxybate, modafinil, the combination of sodium oxybate with modafinil, or placebo for 8 weeks were assessed.[111] Patients treated with sodium oxybate alone (6 g for 4 weeks and then 9 g for the subsequent 4 weeks) were compared with patients treated with modafinil (200 to 600 mg daily); both of these wake-promoting medications caused similar improvements in the patients' alertness on the MWT compared with the placebo treatment group. The greatest improvement in the MWT was seen in the group of patients taking both sodium oxybate and modafinil, suggesting an additive effect of each medication. ESS scores and weekly inadvertent naps and sleep attacks were also significantly reduced in the sodium oxybate and sodium oxybate–modafinil groups but not in the modafinil group. A limitation of the study was that the modafinil-treated patients remained on doses established before the study and were not further titrated to a maximally effective dose during the study. Additionally, to date there have been no studies that directly compare sodium oxybate to traditional stimulant medications. Furthermore, there have been no randomized, placebo-controlled trials using sodium oxybate for the treatment of excessive daytime sleepiness in patients other than those with narcolepsy.

Side Effects and Morbidity

The most commonly reported adverse events associated with the use of sodium oxybate in placebo-controlled trials (n = 655) and post-marketing use in 26,000 patients between 2002 and 2008 included nausea (2.2%), insomnia (1.4%), headache (1.4%), dizziness (1.3%), vomiting (1%), and somnolence (0.9%).[113] Enuresis and sleep walking have been reported as well, presumably related to increased slow wave sleep. Paradoxical sleep initiation insomnia (0.8%) has been reported with initial administration of GHB. Side effects appear to be dose dependent.

Psychiatric side effects of sodium oxybate are increasingly recognized and include emergent depression (0.6%), suicidal ideation, confusion, and psychosis.[113] Symptoms or pre-existing or new onset major depression and suicidality thus need to be monitored carefully and addressed immediately prior to initiation of and during use of sodium oxybate. Given the potent CNS depressant effects, care must be taken to prevent accidental access to the medication by young children or other household members, as accidental or intentional overdose can result in death.

Sodium oxybate has the potential to impair respiratory drive and thus should be used with great caution or not at all in patients being treated with sedative hypnotic agents or other CNS depressants, and should not be combined with alcohol. Sleep disordered breathing should be ruled out or adequately treated with CPAP or oral appliances prior to initiation of sodium oxybate,[114] and patients should be monitored for worsening of OSA or emergent central sleep apnea. Because impaired motor or cognitive function may occur when taking sodium oxybate, the elderly may be at higher risk of falls and injury. Patients should not drive or operate machinery for at least 6 hours after taking sodium oxybate.

Additional Health Benefits and Uses

Sodium oxybate has been investigated for treatment of alcohol withdrawal, fibromyalgia, and rapid eye movement (REM) behavior disorder, among other conditions.[113] Sleep-deprived normal subjects who took sodium oxybate before a 3-hour nap following a night of sleep deprivation slept a similar amount of time as the placebo group, but they also had a higher percentage of slow wave sleep during and longer MSLTs following the nap, reported less subjective sleepiness, and had faster reaction times on the psychomotor vigilance test.[115] Findings such as these suggest that sodium oxybate may exert its effects by increasing slow wave sleep.

Withdrawal

The discontinuation effects of sodium oxybate have not been systematically evaluated in controlled clinical trials, but an abstinence syndrome has not been reported in clinical investigations. After cessation of treatment, patients can expect a gradual return to baseline levels of sleepiness and recurrence of cataplexy symptoms over days to weeks.[112]

Tolerance

Tolerance development to sodium oxybate has not been reported.

Dependence and Abuse Potential

Sodium oxybate was available over the counter as a food supplement for many years, and it became popular with weightlifters who discovered that it accelerated muscle growth and recovery (no doubt secondary to its effect on growth hormone release during sleep). After reports of overdosing by weightlifters, increasing recreational abuse, and reported use as a "date-rape" drug given its sedative and anterograde amnesia side effects, GHB supplements were banned in 1990. Popular pressure to make sodium oxybate illegal was countered by lobbying efforts on the part of narcolepsy patients who testified to the drug's benefits when used appropriately for sleepiness and cataplexy. The end result was a unique dual-schedule mechanism such that sodium oxybate may be prescribed as a Schedule III drug through a centralized pharmacy, and abuse or diversion of sodium oxybate is prosecuted under Schedule I felony charges.

Sodium oxybate continues to have high street value because of its ability to induce euphoria and craving in users. There have been case reports of dependence after illicit use of sodium oxybate at frequent repeated doses in excess of the therapeutic dose range (18 to 250 g/day). Careful monitoring of patients for dependence and abuse is necessary, although as with sympathomimetic stimulants, addiction has not been described in narcoleptics.

IS ALL WAKEFULNESS THE SAME?

An additional aspect of efficacy is the subjective experience of wakefulness that each medication produces. That is, sympathomimetic-induced wakefulness may not feel the same, or in fact be the same, as the wakefulness produced by caffeine, modafinil, or sodium oxybate. It has been suggested that although several neurotransmitter systems facilitate alertness through their extensive projections throughout the cortex, these systems may not be simply redundant, but rather support different aspects of wakefulness.[116] In particular, the monoaminergic projections of the ascending reticular activating system may mediate a sort of "guard duty"—an externally directed vigilance or awareness of one's surroundings—whereas the hypothalamic arousal regions (tuberomammillary nucleus and orexin systems) may perhaps support a form of internally directed vigilance—attention, motivation, insight, and planning. Normally, a healthy balance of activity from these systems should allow a person to focus on a task while being aware of the surrounding milieu. This hypothesis stems in part from the observation that the excess dopamine and norepinephrine release after administration of high-dose amphetamines provokes a state of exaggerated hypervigilance, or paranoia, and impairs executive functions, including judgment, insight, and planning.[65]

The comparison of relative efficacy on cognitive benefits may be more difficult, however, because sleepy patients frequently judge their level of wakefulness not by the degree of mental alertness present but rather by the autonomic arousal that sympathomimetics generate. Conversely, it is important to distinguish wakefulness from cognitive enhancement as the desired clinical end point guiding titration, especially of sympathomimetics. Thus although subjective or objectively measured sleepiness may appear to be well controlled, patients may request increases in stimulant dosing based on perceived cognitive benefits in concentration, attention, and memory abilities. The coexistence in narcolepsy of sleepiness and deficits of attention and concentration is increasingly appreciated. However, no guidelines are currently available to inform dosing of psychostimulants or GHB for cognitive rather than wakefulness enhancement in narcolepsy or hypersomnia patients. Nevertheless, consideration of the varied mechanisms of alerting medications may be useful in understanding clinical outcomes.

SPECIFIC USE OF WAKE-PROMOTING MEDICATIONS

In Children

Side effects of sympathomimetics in children with narcolepsy have not been studied in detail; much of the available data concern the use of these agents for children with ADHD. The potential side effect of greatest concern is growth restriction.[117] For example, deficits in weight gain and height increase may occur after treatment of ADHD with dextroamphetamine or methylphenidate.[118,119] The growth restriction effects of the sympathomimetic agents are due to drug-induced anorexia and the reduction of slow wave sleep and attendant suppression of growth hormone release. The growth deficits may be reversed during summers off medication,[119,120] and with these drug holidays, there is little or no evidence of long-term effects on growth. Obviously, the need for drug holidays to circumvent the effects on growth means that, during treatment interruptions, the child may suffer disabling symptoms that hamper functioning socially and at home. Motor tics can also occur in children taking sympathomimetics,[121] and these may limit dosing. Typical initial doses of these agents for treatment of ADHD in children are methylphenidate 0.3 mg/kg or dextroamphetamine 0.15 mg/kg, followed by dose titration to achieve optimal effects. The safety of higher doses (e.g., methylphenidate 60 mg/day) for children with narcolepsy, compared with doses currently recommended for ADHD, is unknown.

Neither modafinil nor sodium oxybate is currently indicated for patients younger than 16 years. Both medications have potential advantages for use in children because neither medication degrades nighttime sleep or interferes with appetite, so growth restriction may be less likely to occur compared with the sympathomimetic agents. Indeed, phase III studies of modafinil in children with ADHD revealed equal or superior efficacy of modafinil 400 to 600 mg/day compared with sympathomimetics. However, there was increased risk for rash, including one case reported as Stevens-Johnson syndrome.[93] However, in clinical practice modafinil is commonly used off-label for treatment of sleepiness due to narcolepsy in children either as a first-line agent or, more commonly, following failed trials of sympathomimetics. Further safety and efficacy studies are also needed for sodium oxybate, although this medication is frequently used off-label for narcolepsy in children.

In Sustained Military Operations

During the 1991 Persian Gulf War and during the American-led occupation of Iraq, armed forces were issued modafinil and dextroamphetamines for vigilance during sustained operations (S. Lubin, personal communication). Although there are few controlled studies, in a study of U.S. Army helicopter pilots

engaged in flight simulation after prolonged periods of wakefulness, 10 mg of dextroamphetamine, compared with placebo, improved aviator simulator control on descents and turns. Performance was facilitated most noticeably after 22, 26, and 34 hours of continuous wakefulness. Alertness was sustained significantly by dextroamphetamine—there was reduced slow wave electroencephalographic activity and improved rating of vigor and fatigue. No adverse behavioral or physiologic effects were observed.[122,123] Comparable results on performance have also been demonstrated with modafinil during 64 hours of sustained mental work.[124] Interestingly, the recovery sleep after extended periods of modafinil treatment shows a lack of the rebound hypersomnolence characteristic of recovery sleep following amphetamine treatment.[72,125] This difference suggests that modafinil may exert its alerting effects in a novel way, without invoking a rise in the homeostatic sleep drive.

In Sleepiness Due to Insufficient Sleep

Insufficient sleep, beyond the military situation, arises in many circumstances. Common among these circumstances are jet lag and shift work. Modafinil has recently been approved for the treatment of sleepiness in SWSD. The prospective use of alerting agents to enhance the alertness and performance among resident physicians has become a focus of discussion.[126,127] This use of alerting medications is problematic for many physicians, and active debate continues. The key points of this debate center on the relative importance of the potential benefit for safety and performance, especially when a high degree of vigilance is required, and the potential for abuse and dependency associated with these agents. The demand for alerting medications is likely to increase as our society continues to depend on 24-hour operations in the manufacturing, transportation, and service industries, underlining the importance of a careful risk-benefit analysis, currently limited by absence of large controlled clinical studies.

In Circadian Misalignment

SWSD presents a special case of a frequently shortened sleep zone, in which circadian mismatch results in wakefulness required at the natural circadian nadir and insomnia naturally ensuing at the peak of circadian wake, further exacerbating insufficient sleep. The contribution of circadian mismatch, rather than a contracted sleep zone that is purely behavioral in source, cannot be overemphasized in cases of SWSD. Whereas modafinil has shown limited benefit in studies of prolonged partial sleep deprivation, it may play a particularly important role in counteracting the sleepiness associated with the circadian nadir. However, the importance of obtaining sufficient sleep should be emphasized to patients, and the addition of a hypnotic to counteract sleep initiation and maintenance insomnia attributed to circadian mismatch may be indicated as well. The role of sympathomimetics may be more limited by their extensive side-effect profile, including cardiovascular side effects.

Wake-Promoting Agents as "Smart Pills"

Both sympathomimetics and modafinil have been explored as possible "neuroenhancement" agents in healthy people, either in the rested state or following sleep deprivation. Surveys in the mid to late 2000s have reported pervasive "academic doping" using nonmedical stimulants (typically amphetamine and methylphenidate) ,with the goal of cognitive enhancement or "staying awake," that ranges between 8% and 34% in college student populations.[71] The 2008 National Survey on Drug Use and Health reported a prevalence of 12.3% nonmedical stimulant use in 21- to 25-year-olds in the United States.[128] Surprisingly, controlled studies have not always substantiated robust and sustained cognitive benefits, especially for sympathomimetics. Smith and Farah reviewed placebo-controlled studies of sympathomimetic effects on cognition in healthy nonelderly adults and found inconclusive results, suggestive of enhanced long-term declarative memory consolidation and varied effects on executive function. Not only was improvement of working memory and cognitive control not always seen, some subjects were impaired, most notably high performers at baseline as well as those homozygous for the met allele of the catechol-*O*-methyl transferase gene.[128] Another recent meta-analysis suggested that a single dose of methylphenidate may improve motivation and memory, whereas repeated doses bolster subjective energy and attention during a partial-sleep-deprivation protocol (4 hours of sleep) but do not reduce sleepiness or improve cognitive measures during sustained sleep deprivation (64 hours) (for review, see Repantis and colleagues[129]).

A single dose of modafinil improved wakefulness and attention in both rested and sleep-deprived individuals, whereas repeated administration during 4 days of sleep deprivation improved wakefulness, but not cognitive measures.[129] Two important limitations were seen with modafinil administration in healthy individuals. First, repeated daily administration of 400 mg of modafinil increased scores on both positive (elevated mood) and negative (anxiety) affect scales. Second, after 64 (but not 24 or 40) hours of sustained sleep deprivation, there was an "overconfidence" mismatch between subjects' retrospective, self-reported cognitive performance and objective cognitive measures. In addition, there appears to be a narrow therapeutic range, with smaller (100 mg) but not larger (200 mg) doses showing cognitive enhancement properties following single administration in healthy rested people.[130]

In summary it appears that, although potentially helpful during limited bouts of partial sleep deprivation, wake-promoting agents are not adequate for counteracting the effects of sustained, complete sleep loss. Caution should be exercised with modafinil in particular, given reported potential for subjective overestimation of performance following sleep deprivation. Additionally, enhanced alertness and cognitive processes appear beneficial in a dose-dependent manner but only up to a point, beyond which wake-promoting agent efficacy is limited by side effects. Indeed, Lyon and Robbins have described the efficacy of sympathomimetics as an inverted U-shaped curve, with optimal "psychostimulant activation" at intermediate doses, whereas high doses remain limited by undesirable side effects, including stereotypic behaviors, cognitive inflexibility, psychosis, and addiction.[71,131] In addition, an emerging literature suggests that some of the perceived benefits of sympathomimetics may ensue from their mood-modifying[132] or motivation-modifying properties, rather than strictly cognitive (learning and memory or executive function) effects as previously expected.[128]

RECOMMENDATIONS AND TREATMENT PLANNING

Current practices in the use of alerting medications vary considerably. On the whole, patients who are medicated for

excessive sleepiness are still monitored primarily by clinical assessment of their ability to remain alert during sedentary activities, with medication selection and dosing decisions adjusted accordingly with consideration of medication side-effect profiles. Few studies directly compare the relative efficacies of wake-promoting medications, although an earlier comparison of the studies assessing the effects of various medications on MSLT and MWT measures suggested that classic stimulants may be most potent for the majority of sleepy narcoleptics.[133] All prescription wake-promoting medications produce clinically significant improvements in alertness in narcolepsy, but based on the available published data, only a small proportion of very sleepy patients will achieve normal levels of alertness with medication.[134] Clinicians should treat individual patients based on their profile of sleepiness throughout the day and their ability to tolerate side effects. Although many authorities recommend temporary withdrawal of sympathomimetic medications or reduction of dose if tolerance occurs (i.e., drug holidays[135,136]), there are no published studies demonstrating the efficacy of drug holidays. The effect of drug holidays on patient safety and quality of life must also be considered. Another factor that probably influences clinical practice is whether an alerting medication has been placed on Schedule II by the U.S. Drug Enforcement Administration. Because of the extra paperwork required in some states to prescribe Schedule II agents, Schedule IV drugs such as modafinil may be preferentially prescribed. Similarly, the risks for abuse or diversion may deter some clinicians from prescribing sodium oxybate.

Alerting medications, however prescribed, represent only part of a comprehensive therapeutic approach to excessive somnolence. Sound sleep hygiene, attention to other substances and drugs that may disrupt the sleep-wake cycle, and periodic reassessment of symptom severity and the need for and adequacy of treatment modalities are other important aspects of management. The physician should consider the following points in establishing the proper dose of an alerting drug and structuring a management plan:

1. *Diagnosis*. It is important to define as carefully as possible the factors that contribute to a patient's excessive sleepiness. Differentiating an insidious and lifelong condition such as narcolepsy from sleepiness due to sleep-related breathing disorders, for example, is essential for both the patient and the clinician.
2. *Education*. Clarify the goals of treatment, side effects, risks, and benefits. This process involves discussions with the patient and, perhaps, the patient's family members or companions. Normal alertness throughout the day may not be attainable in many patients because of the disease process, drug side effects, work schedules, or other idiosyncratic circumstances.

   In cases of SWSD, advocating for a work schedule change (such as a switch to daytime work hours) may be the ideal alternative to fully restore alertness. Unfortunately, many people do not have the ability to control their schedules directly and must continue to cope with their current situation. In this case, the clinician must support the patient's need for alertness without imposing judgment about the need for lifestyle changes.

   The importance of obtaining sufficient sleep whenever possible should always be emphasized, however, because stimulants promote wakefulness but cannot substitute for sleep.
3. *Dosing*. Begin with a low to moderate dose of a wake-promoting agent and match the drug and dosage to the patient. For most patients, aim to provide even alertness throughout the wake period. Modafinil or long-acting sympathomimetics provide advantages in this respect. Short-acting sympathomimetics may be useful especially for someone who needs rapid alertness on arising from sleep (e.g., in order to drive). Short-acting medications also provide opportunities for napping between doses but can produce unprotected sleepiness. Recommendations for starting and maximal doses of commonly used wake-promoting medications are summarized in Table 44-2.
4. *Follow-up*. Initially, pharmacologic management should be guided by regular (e.g., weekly) contact with the patient. If prescribing sympathomimetics, it is wise to periodically measure growth (height and weight) in children and weight, pulse, and blood pressure in adults. Patients should be monitored frequently to determine the effective dose and preparation for their symptoms and for side effects. After the dose is stable, patients should be seen every 6 to 12 months. Under circumstances in which the patient's safety or the safety of others depends on adequate control of excessive somnolence, laboratory confirmation of therapeutic efficacy with the MSLT or MWT is helpful.
5. *Emphasize sleep hygiene*. Consider short (30-minute), prophylactic naps. The effect of sleep inertia must be factored in if naps are used in the work setting or are prolonged, and may be counteracted by caffeine administration (100 to 200 mg) immediately preceding a 30-minute nap. Consider the use of light therapy, melatonin, or other modalities if circadian factors affect the ability to sleep (see Chapter 40).
6. *Adjust medication dosages* based on clinical information. Narcolepsy and idiopathic hypersomnia are usually stable conditions that do not progressively worsen. For a patient who has been on a stable dose for some time (years) and now appears to require more medication, consider other possible causes of increased sleepiness, such as: (1) interval development of sleep apnea or other primary sleep disorder that can contribute to sleepiness; (2) tolerance to medication; (3) change in schedule (e.g., a change in job shift, causing less sleep at night); (4) change in life situation (e.g., a new baby causing sleep disruption or a new job that requires greater vigilance); (5) stress, anxiety, or depression; and (6) unrealistic expectations. Evaluation should include a detailed history covering the not only these possibilities but also a review of the patient's sleep schedule and napping.
7. *Recommend counseling and long-term support*. A person with pathologic sleepiness who suddenly becomes more alert and energetic during the daytime may evoke strong feelings from family members not used to their active participation. Patients may become depressed or grieve the time "lost to sleep" before treatment after the degree of their prior impairment becomes clear to them. Available evidence suggests that over time, patients tend to take less, not more, of their prescribed stimulant.[137] Although the reasons for this are undoubtedly complex and incompletely understood, it is important that the patient understand the long-term nature of his or her condition and the

benefits that can be obtained with regular use of alerting medications.

CHANGING OR COMBINING MEDICATIONS

For most patients, replacement of one alerting agent with another should present few problems. However, for patients taking high doses of sympathomimetic medications, a gradual weaning period may be prudent. Furthermore, if the patient is switching from a sympathomimetic stimulant to modafinil, the qualitative difference in their alerting effects—and the difference in peripheral side effects—usually necessitates a 3- to 4-week adjustment period during which the stimulant withdrawal effects dissipate and the patient begins to experience what he or she feels like on modafinil alone. Titration toward optimal control of alertness can then be done more clearly. By combining stimulants with different durations of action, it may be possible to maintain wakefulness during the day, allow for periods for napping, and promote long periods of sleep at night without producing medication-induced insomnia.

Except for studies with sodium oxybate and modafinil, there are no systematic studies of chronic treatment with more than one alerting agent at a time. Some patients report satisfactory results on combinations such as modafinil or extended-release sympathomimetics (for long-lasting effects) combined with small doses of short-acting sympathomimetics such as methylphenidate IR, taken on an as-needed basis. There are no known drug interactions that would preclude this practice, which is indeed common. However, in some patients, hypertension may develop or be exacerbated by the coadministration of multiple wake-promoting medications, so appropriate blood pressure monitoring is indicated during treatment. All of the available wake-promoting medications can be safely combined with nonsedating antidepressants used as anticataplectic agents, including TCAs, SSRIs, and serotonin-norepinephrine reuptake inhibitors.

CLINICAL PEARL

The main goal of the treatment of pathologic sleepiness is to address and correct the underlying sleep disorder. When sleepiness remains an issue despite nonpharmacologic treatment—such as in patients with narcolepsy, other central nervous system hypersomnias, or SWSD, and in some patients with OSA using CPAP—prescription alerting medications should be considered for the patient's safety and quality of life, with the recognition that these medications offer symptomatic rather than disease-modifying treatment. Modafinil and armodafinil are first-line agents in patients with excessive sleepiness due to these disorders because it prompts wakefulness without many side effects or rebound hypersomnia. A broad array of sympathomimetic compounds are also available to treat sleepiness and may be required in patients who do not respond adequately to modafinil, but the risks for abuse, tolerance, and side effects makes them second-line agents in treating narcolepsy or for off-label treatment of other sleep disorders. Caffeine is a useful alerting agent in situations of mild sleepiness, such as with shift work, following mild sleep deprivation, or to overcome sleep inertia, but tolerance can develop when taken daily. Sodium oxybate may be helpful in reducing excessive sleepiness when used in patients with narcolepsy and cataplexy.

SUMMARY

The potentially disabling symptom of sleepiness occurs in many sleep disorders. When this sleepiness does not resolve with nonpharmacologic approaches, the use of alerting medications is appropriate.[138] Caffeine is widely available and is consumed by most of the world's population. As an alerting agent, caffeine is most effective when used intermittently at doses of 200 mg or more; tolerance develops with chronic use, however. Severe or chronic sleepiness is best treated with one of the variety of prescription alerting medications. Treatment with alerting medications of excessive sleepiness associated with narcolepsy or idiopathic hypersomnia is almost always indicated to allow wakefulness when sustained vigilance is necessary, for the safety of both the individual and the public. Pharmacologic treatment with modafinil is now also indicated for severe sleepiness in patients with SWSD and in patients with OSA who remain sleepy despite compliance with nasal CPAP. Sodium oxybate may be indicated in patients with narcolepsy and cataplexy. Because the risk for teratogenicity associated with the use of alerting agents is uncertain, these drugs in general should be avoided in pregnancy unless the benefits associated with their use are likely to outweigh the risks.

Selected Readings

Alameddine Y, Klerman EB, Bianchi MT. Caffeine and naps as countermeasures for sleep loss. In: Bianchi MT, editor. *Sleep deprivation and disease*. New York: Springer; 2014. p. 231–42.

Killgore W. *Caffeine and other alerting agents*. Cambridge: Cambridge University Press; 2011.

Khan Z, Trotti LM. Central disorders of Hhypersomnolence: focus on the Nnarcolepsies and idiopathic hypersomnia. *Chest* 2015;**148**(1):262–73.

Lavault S, Dauvilliers Y, Drouot X, et al. Benefit and risk of modafinil in idiopathic hypersomnia vs. narcolepsy with cataplexy. *Sleep Med* 2011;**12**:550–6.

Majid H, Hirshkowitz M. Therapeutics of narcolepsy. *Sleep Med Clin* 2010;**5**:659–73.

Mayer G. The use of sodium oxybate to treat narcolepsy. *Exp Rev Neurotherapeut* 2012;**12**:519–29.

Mignot EJ. A practical guide to the therapy of narcolepsy and hypersomnia syndromes. *Neurother* 2012;**9**:739–52.

O'Keefe JH, Bhatti SK, Patil HR, et al. Effects of habitual coffee consumption on cardiometabolic disease, cardiovascular health, and all-cause mortality. *J Am Coll Cardiol* 2013;**62**:1043–51.

Repantis D, Schlattmann P, Laisney O, Heuser I. Modafinil and methylphenidate for neuroenhancement in healthy individuals: a systematic review. *Pharmacol Res* 2010;**62**:187–206.

Roth T. Pharmacotherapy of excessive sleepiness. *Sleep Med Clin* 2012;**7**:333–40.

Sakai N, Chikahisa S, Nishino S. Stimulants in excessive daytime sleepiness. *Sleep Med Clin* 2010;**5**:591–607.

Sheng P, Hou L, Wang X, et al. Efficacy of modafinil on fatigue and excessive daytime sleepiness associated with neurological disorders: a systematic review and meta-analysis. *PLoS ONE* 2013;**8**:e81802.

Smith ME, Farah MJ. Are prescription stimulants "smart pills"? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychol Bull* 2011;**137**:717–41.

Urry E, Landolt HP. Adenosine, caffeine, and performance: from cognitive neuroscience of sleep to sleep pharmacogenetics. *Curr Top Behav Neurosci* 2015;**25**:331–66.

Wood S, Sage JR, Shuman T, Anagnostaras SG. Psychostimulants and cognition: a continuum of behavioral and cognitive activation. *Pharm Rev* 2014;**66**:193–221.

A complete reference list can be found online at ExpertConsult.com.

Chapter
45

Drugs that Disturb Sleep and Wakefulness

Paula K. Schweitzer; Angela C. Randazzo

Chapter Highlights

- Disturbed sleep and daytime sedation are common side effects of many medications. Sedating drugs may impair waking function if the sedating action occurs during waking hours, from either prolonged duration of action or administration during waking hours. Drugs that disrupt sleep can lead to impaired waking function and daytime sleepiness, whereas drugs that promote alertness may disrupt sleep.
- Principal pharmacologic mechanisms promoting sedation include antagonism of histamine-1 (H_1) receptors, norepinephrine α_1 receptors, muscarinic cholinergic receptors, serotonin type 2 receptors (5-HT_2), or dopamine receptors.
- Principal pharmacologic mechanisms promoting wakefulness include reuptake inhibition of norepinephrine, serotonin, and dopamine, as well as inhibition of monoamine oxidase.
- A number of drugs have pharmacologic effects at the receptors involved in sleep-wake regulation and thus have the potential to disrupt sleep or impair waking function. Psychotherapeutic drugs are the principal drugs with the potential for these negative effects. However, a variety of other drugs may produce negative effects on sleep-wake function, including antiepileptics as well as drugs used in the treatment of cardiovascular disease, Parkinson disease, and pain.

Research on the neural mechanisms involved in sleep-wake regulation suggests that sleep-wake state is controlled by a complex interaction between wakefulness-promoting and sleep-promoting nuclei in the hypothalamus and brainstem.[1-3] Wake-promoting neurons include orexinergic and histaminergic nuclei in the hypothalamus, cholinergic nuclei in the brainstem, adrenergic nuclei in the locus coeruleus, serotonergic nuclei in the raphe nuclei, and dopaminergic nuclei in the midbrain ventral tegmental area. Sleep is promoted by nuclei in the basal forebrain, ventrolateral preoptic area, and anterior hypothalamus through the inhibitory neurotransmitters gamma-aminobutyric acid (GABA) and galanin. Adenosine, which has been proposed to be involved in homeostatic regulation of sleep, may promote sleep through anticholinergic activity in the basal forebrain and brainstem. Drugs with pharmacologic effects at receptors involved in sleep-wake regulation may therefore have effects on sleep-wake behavior. These effects may be therapeutic (e.g., improve sleep or enhance wakefulness) or impairing (e.g., cause sleep disturbance or daytime sedation).

Table 45-1 summarizes pharmacologic mechanisms of drug effects on sleep and waking behavior.[4] Drugs can cause sedation by multiple mechanisms, either by increasing the activity of the sleep-promoting system through GABA enhancement (e.g., benzodiazepine receptor agonists, ethanol) or by inhibiting the wake-promoting system through antagonism of central histamine-1 (H_1) receptors (e.g., first-generation antihistamines, tricyclic antidepressants), norepinephrine α_1 receptors (e.g., certain antidepressant and antipsychotic medications), muscarinic cholinergic receptors (e.g., some antidepressants), serotonin-2 (5-HT_2) receptors (e.g., trazodone, mirtazapine, olanzapine, quetiapine), or dopamine receptors (e.g., certain antipsychotics). Similarly, drugs can disrupt sleep through effects on either the sleep-promoting system or the wake-promoting system. More specifically, wake promotion may occur through blockade of the reuptake of serotonin (5-HT; e.g., fluoxetine), norepinephrine (e.g., venlafaxine), or dopamine (e.g., bupropion), or by inhibition of monoamine oxidase (MAO; e.g., phenelzine), thereby increasing the available amount of norepinephrine, serotonin, and dopamine.

Drugs may also affect homeostatic and circadian processes involved in sleep-wake regulation. Effects on neurotransmitters and neuronal systems involved in the generation of slow wave sleep (SWS) and rapid eye movement (REM) sleep can affect sleep architecture. REM suppression may occur with blockade of cholinergic receptors and increased 5-HT binding to 5-HT_{1A} receptors. SWS may increase through blockade of 5-HT_2 receptors. Drugs can also impair sleep or wakefulness by causing or exacerbating restless legs syndrome (RLS) and periodic limb movements during sleep (PLMS). The mechanism by which this occurs is not clear but may be associated with increasing availability of 5-HT and blockade of dopamine receptors. Dose, half-life, and time to peak concentration are additional important factors that may determine the effects of drugs on behavior. Pharmacologic effects may vary with drug dose. For example, doxepin at low doses is predominantly a histamine antagonist, whereas at higher doses it also inhibits serotonin transporter (5-HTT) and norepinephrine transporter (NET), in addition to having α-adrenergic and

Table 45-1 Pharmacologic Mechanisms of Drug Effects on Sleep and Wake Behavior

Mechanism	Promotes Sleep	Promotes Wake	Suppresses REM Sleep	Increases SWS	Promotes RLS/PLMS
H_1 antagonism	X				
M antagonism	X		X		
5-HT_2 antagonism	X			X	
α_1 antagonism	X				
D_1/D_2 antagonism	X				X
α_2 agonism	X				
α_2 antagonism		X			
β_2 antagonism		X	X		
5-HT reuptake inhibition		X	X		X
NE reuptake inhibition		X			
DA reuptake inhibition		X			
MAO inhibition		X	X		X
5-HT_{1A} agonism		X	X		

5-HT, Serotonin; DA, dopamine; H1, histamine type 1; M, muscarinic anticholinergic; MAO, monoamine oxidase; NE, norepinephrine; PLMS, periodic limb movements during sleep; REM, rapid eye movement; RLS, restless legs syndrome; SWS, slow wave sleep.
Modified with permission from Krystal A. Antidepressant and antipsychotic drugs. *Sleep Med Clin* 2010;5:571–89.

anticholinergic effects. Half-life, combined with drug dose, determines the duration of clinical effects. Time to peak concentration affects the speed with which clinical effects may occur.

Sedating drugs may impair waking function if the sedating action occurs during waking hours, from either prolonged duration of action or administration during waking hours. Drugs that disrupt sleep can lead to impaired waking function and daytime sleepiness, whereas drugs that promote alertness may disrupt sleep. Thus the desired action of a drug may become an undesirable action when its effect occurs at the "wrong" time of day or night. In addition, drugs often act at multiple neural sites involved in sleep-wake regulation. Thus the desired action of a drug may be produced by effects at specific receptor sites, and undesired actions may occur because of concomitant effects at other receptor sites.

This chapter reviews drugs that are used for common medical and psychiatric conditions and that have unintended effects on sleep or wakefulness. Drugs used as hypnotics, stimulants, and drugs of abuse are reviewed elsewhere in this volume.

PSYCHOTHERAPEUTIC DRUGS

Psychotherapeutic drugs have a variety of pharmacologic effects on sleep-wake function. Table 45-1 summarizes the pharmacologic mechanisms likely responsible for the effects of antidepressant and antipsychotic drugs on sleep-wake behavior. Table 45-2 summarizes the effects of psychotherapeutic drugs on sleep-wake behavior. Figure 45-1 displays receptor binding affinity data for many of these drugs.[6]

Antidepressants

Drugs classified as antidepressants are used in a variety of disorders, including depression, obsessive-compulsive disorder, anxiety disorders, neuropathic pain, and others. Sedating antidepressants used as hypnotics are covered elsewhere in this volume.

Antidepressant drugs can improve or disturb sleep as well as affect waking function. Evaluation of the effects of these drugs on sleep and wakefulness is complicated by the fact that many individuals with depression have disturbed sleep[7] as well as daytime complaints such as fatigue, sleepiness, somatic complaints, and decreased cognitive and psychomotor functioning.[8-10]

Tricyclic Antidepressants

Most tricyclic antidepressants (TCAs) are used in the treatment of depression. However, clomipramine and doxepin have U.S. Food and Drug Administration (FDA) indications for obsessive-compulsive disorder. These drugs differ from one another in their relative effects in blocking reuptake of 5-HT compared with norepinephrine as well as in the degree of antagonism of muscarinic cholinergic receptors and H_1 receptors[11] (see Figure 45-1). The more sedating TCAs tend to be more anticholinergic (amitriptyline) and more antihistaminergic (doxepin, trimipramine) but also exhibit proportionately greater inhibition of 5-HT reuptake than norepinephrine reuptake.

Generally, these drugs decrease sleep latency, increase total sleep time (TST), and decrease REM sleep while increasing phasic eye movements during REM.[12,13] TCAs that are more adrenergic (e.g., desipramine, nortriptyline) may decrease TST and increase awakenings.[14] TCAs may increase PLMS and symptoms of RLS.[13] Multiple Sleep Latency Test (MSLT) latency was significantly decreased following a single evening 75-mg dose of amitriptyline.[15] Cognitive, psychomotor, and driving performance are impaired with acute use in normal subjects, but there is evidence that these effects lessen with time.[16-18]

Text continued on p. 487

Table 45-2 Effects of Psychotherapeutic Drugs on Sleep and Wake Behavior

Drug or Class	U.S. Trade Name	FDA Indication	Primary Mechanism of Action	Subjective Data	PSG Data	MSLT Data	Cognitive and Performance Data
Antidepressants							
Tricyclic Antidepressants (TCAs)							
Amitriptyline Amoxapine Imipramine Trimipramine Clomipramine	Elavil Asendin Tofranil Surmontil Anafranil	Depression OCD	5-HT > NE reuptake inhibition; α_1, M, H_1, 5-HT_2 antagonism	Somnolence +++ Insomnia + Nightmares ?RBD	↑TST, ↓W, ↓↓REM, ↑PLMS	↓SL	Mild to moderate impairment acutely
Doxepin	Anafranil Sinequan	OCD	5-HT = NE reuptake inhibition; α_1, M, H_1, 5-HT_2 antagonism	Somnolence +		No data	
Desipramine Nortriptyline Protriptyline	Norpramin Pamelor Vivactil	Depression	NE > 5-HT reuptake inhibition; M antagonism	Insomnia + Nightmares	↑W, ↓TST, ↓↓REM, ↑PLMS		
Selective Serotonin Reuptake Inhibitors (SSRIs)							
Citalopram	Celexa	Depression	5-HT reuptake inhibition	Insomnia + Somnolence +	↓TST, ↑W, ↓REM, ↑PLMS	No data	No impairment
Escitalopram	Lexapro	Depression, GAD, OCD	5-HT reuptake inhibition	Insomnia + Somnolence + Nightmares, RLS	↓REM ? ↑PLMS	No change in MSLT	Improved
Fluoxetine	Prozac, Sarafem	Depression, bulimia nervosa, OCD, panic disorder, PMDD	5-HT reuptake inhibition	Insomnia ++ Somnolence + Nightmares, RLS	↓TST, ↑W, ↓REM,↓SWS, ↑SEMs, ↑PLMS	↑ SL on modified MSLT	Generally no change or mild improvement
Fluvoxamine	Luvox	OCD	5-HT reuptake inhibition	Insomnia ++ Somnolence +	↓TST, ↑W, ↑SL, ↓REM	No change in MSLT	No impairment
Paroxetine	Paxil	Depression, GAD, OCD, panic disorder, PMDD, PTSD, social anxiety	5-HT reuptake inhibition	Insomnia ++ Somnolence ++ Nightmares	↑W, ↓TST, ↑SL, ↓REM ↑PLMS	No data	Mixed results
Sertraline	Zoloft	Depression, OCD, panic disorder, PMDD, PTSD, social anxiety	5-HT reuptake inhibition	Insomnia ++ Somnolence ++	↑SL, ↓TST, ↓REM, ↑REM latency, ↑PLMS		No change or mild improvement
Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)							
Desvenlafaxine	Pristiq	Depression	5-HT and NE reuptake inhibition	Insomnia ++ Somnolence +	No data	No data	No data
Duloxetine	Cymbalta	Depression, diabetic neuropathy, fibromyalgia, GAD, chronic musculoskeletal pain	5-HT and NE reuptake inhibition; weak D reuptake inhibition	Insomnia ++ Somnolence + Nightmares	↓REM, may ↑SWS, variable effects on sleep efficiency	No data	Improvement
Levomilnacipran	Fetzima	Depression	NE, 5-HT reuptake inhibition		No data	No data	No data

Milnacipran	Savella	Fibromyalgia	NE >5-HT reuptake inhibition	Insomnia +	↑W, ↓SL, ↓NREM, ↓REM	No data	No effects
Venlafaxine	Effexor	Depression, GAD, panic disorder, social anxiety	5-HT reuptake inhibition at low doses; NE reuptake inhibition at high doses; weak D reuptake inhibition	Insomnia +++ Somnolence +++ RLS Nightmares	↓TST, ↑W, ↓↓REM ↑PLMs	No data	↑ Performance in normals
Serotonin Antagonist Reuptake Inhibitors (SARIs)							
Nefazodone (no longer available in the U.S.)	Serzone	Depression	5-HT_2 antagonism; weak 5-HT and NE reuptake inhibition	Somnolence +++ Insomnia +++ Nightmares	↑TST		Mixed effects
Trazodone	Desyrel, Oleptro	Depression	5-HT_{2A} antagonism; 5-HT reuptake inhibition; α_1, H_1 antagonism	Somnolence ++	↑TST, ↓SL, ↓W, ↑SWS, ↓REM		↓ Function
Norepinephrine Dopamine Reuptake Inhibitors (NDRIs)							
Bupropion	Wellbutrin, Zyban, Aplenzin	Depression, smoking cessation, seasonal affective disorder	NE and D reuptake inhibition	Insomnia ++ Somnolence + Vivid dreaming, nightmares	↑REM, ↓SWS, ↑PLMS	No data	No impairment
Norepinephrine and Specific Serotonergic Antidepressants							
Mirtazapine	Remeron	Depression	α_1, α_2, H_1, 5-HT_2, 5-HT_3 antagonism	Somnolence +++ Insomnia + Nightmares, RLS	↑TST, ↓SL, ↑SWS, ↑PLMS	No data	↓ Performance acutely
Selective Norepinephrine Reuptake Inhibitors (NRIs)							
Atomoxetine	Strattera	ADHD	NE reuptake inhibition	Somnolence in children + Insomnia in adults +	↑REM latency, ?↓W in ADHD children	No data	?Improvement
Reboxetine (not available in the U.S.)			NE reuptake inhibition	Insomnia +	↓REM, acute ↑W, ↑REM latency		No impairment, ?improvement
Serotonin Partial Agonist Reuptake Inhibitor							
Vilazodone	Viibryd	Depression	5-HT reuptake inhibition; 5-HT_{1A} partial agonism	Insomnia + Abnormal dreams +	↓↓REM, ↑SWS, ↑W	No data	Improvement
Multimodal Antidepressant							
Vortioxetine	Brintellix	Depression	5-HT reuptake inhibition; 5-HT_3, 5-HT_7, 5-HT_{1D} antagonism; 5-HT_{1B} partial agonism; 5-HT_{1A} agonism	Abnormal dreams +	No data	No data	Improvement

Continued

Table 45-2 Effects of Psychotherapeutic Drugs on Sleep and Wake Behavior—cont'd

Drug or Class	U.S. Trade Name	FDA Indication	Primary Mechanism of Action	Subjective Data	PSG Data	MSLT Data	Cognitive and Performance Data
Tetracyclic Antidepressants							
Maprotiline	Ludiomil	Depression, anxiety	NE reuptake inhibition	Somnolence +	Minimal data	No data	Minimal data, no impairment
Mianserin (not available in the U.S.)		Depression	5-HT_2, α_1, α_2, H_1 antagonism	Somnolence +++ Insomnia + RLS	?↑TST	No data	Impairs performance
Monoamine Oxidase Inhibitors							
Moclobemide (not available in the U.S.)			MAO-A inhibition, reversible	Insomnia +		No data	Improvement
Phenelzine	Nardil	Depression	MAO-A, MAO-B inhibition	Insomnia + Somnolence + Nightmares, RLS	↑W, ↓TST, ↓↓↓REM, ↑PLMS, ?RBD	No data	Limited data
Tranylcypromine	Parnate	Depression	MAO-A, MAO-B inhibition	Insomnia + Somnolence +	↑W, ↓TST, ↓↓↓REM	No data	Limited data
Selegiline, selegiline transdermal	Emsam, Zelpar	Depression, Parkinson disease	Inhibits MAO-B at low doses	Insomnia ++	↓TST	No data	No data
Melatonergic Antidepressants							
Agomelatine (not available in the U.S.)			MT_1, MT_2 agonism; 5-HT_{2C} antagonism		↓SL	No data	No data
Anxiolytics							
Benzodiazepines							
Alprazolam Clorazepate Chlordiazepoxide Clonazepam Diazepam Lorazepam Oxazepam	Xanax Tranxene Librium Klonopin Valium Ativan Serax	Anxiety	$GABA_A$ agonism	Somnolence ++	↓SL, ↓SWS, ↓REM	↓MSLT alprazolam, diazepam; no data on other drugs	
Other							
Buspirone	Buspar	Anxiety	5-HT_{1A}, 5-HT_2 agonism; moderate D2 antagonism	Nonsedating	No effect	No effect	No impairment

Antipsychotics and Mood Stabilizers

Traditional Antipsychotics

Chlorpromazine	Thorazine	Schizophrenia, hiccups, nausea/vomiting, mania, porphyria	D2, 5-HT_2, α_1, H_1, M antagonism	Somnolence ++++ RLS	↑TST, ↑SWS ↓REM, ↑PLMS	No data	Limited data
Fluphenazine	Prolixin	Psychoses	D_1, D_2 antagonism	Insomnia ++++		No data	
Haloperidol	Haldol	Schizophrenia, Tourette disorder	D_2, 5-HT_2, α_1 antagonism	Insomnia ++ Somnolence +++ RLS	↓SL, ↑TST, ↓REM, ↑SWS, ↑PLMS	No data	Limited data
Loxapine	Loxitane	Schizophrenia, agitation with schizophrenia or bipolar	D, 5-HT_2 antagonism	Insomnia ++ Somnolence ++++		No data	Limited data
Perphenazine	Trilafon	Schizophrenia, nausea/vomiting	D_1, D_2 antagonism	Insomnia +++ Somnolence ++		No data	Limited data
Pimozide	Orap	Tourette disorder	D_2, 5-HT_7 antagonism	Somnolence ++	No data	No data	No impairment
Thioridazine	Mellaril	Schizophrenia	D_1, D_2, 5-HT_2, α_1, H_1, M antagonism	Insomnia +++ Somnolence ++++ RLS	↑TST, ?SWS ?REM, ↑PLMS	No data	Limited data
Thiothixene	Navane	Schizophrenia	D_1, D_2, D_3, D_4, 5-HT_1, 5-HT_2 antagonism	Somnolence + RLS	↑TST, ↑SWS ↓REM, ↑PLMS	No data	Limited data
Trifluoperazine	Stelazine	Schizophrenia, anxiety	D_1, D_2 antagonism	Insomnia ++++ Somnolence +++		No data	

Atypical Antipsychotics

Aripiprazole	Abilify	Schizophrenia, bipolar mania, agitation with schizophrenia or bipolar, irritability with autistic disorder, adjunctive treatment for depression	D_2, D_3, 5-HT_{1A}, 5-HT_{2A}, antagonism; D_4, 5-HT_{2C}, 5-HT_7, α_1, H1, 5-HTT moderate antagonism	Insomnia +++ Somnolence + RLS	↑PLMS	No data	Possible improvement
Asenapine	Saphris	Schizophrenia, bipolar	5-HT_{1A}, 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C}, 5-HT_5, 5-HT_7, D_1, D_2, D_3, D_4, H_1, α_1, α_2 antagonism	Insomnia + Somnolence ++		No data	Possible improvement
Clozapine	Clozaril	Schizophrenia	5-HT_2, α_1, H_1, M, D_1, D_2 antagonism	Somnolence ++++ RLS	↓SL, ↑TST, ↑SWS, ↑PLMS	↓SL	Possible improvement

Continued

Table 45-2 Effects of Psychotherapeutic Drugs on Sleep and Wake Behavior—cont'd

Drug or Class	U.S. Trade Name	FDA Indication	Primary Mechanism of Action	Subjective Data	PSG Data	MSLT Data	Cognitive and Performance Data
Iloperidone	Fanapt	Schizophrenia, bipolar depression, adjunctive treatment for bipolar	D_3, 5-HT_{2A}, D_4, 5-HT_{2C}, 5-HT_7, α_1, H_1, antagonism; D_2, 5-HT_{1A} agonism	Somnolence ++	No data	No data	Possible improvement
Lurasidone	Latuda	Schizophrenia, bipolar depression	D_2, 5-HT_{2A}, 5-HT_7 antagonism; 5-HT_{1A} partial agonism	Somnolence ++ Insomnia +	No data	No data	Improvement
Olanzapine	Zyprexa	Schizophrenia, bipolar, depression (in combination with fluoxetine)	D_2, 5-HT_{2A}, M, H1, α_1, D_1 antagonism	Somnolence +++ Insomnia ++ RLS	↓SL, ↑TST, ↑↑SWS, ↑REM, ↑PLMS	↓SL	Possible improvement
Paliperidone	Invega	Schizophrenia, schizoaffective disorder, adjunctive with antidepressant or mood stabilizer	α_1, D_2, H1, 5-HT_{2C} antagonism	Somnolence +	No data	No data	No impairment
Risperidone	Risperdal	Schizophrenia	D_2, 5-HT_2, α_1 antagonism	Somnolence +++ Insomnia ++ RLS, nightmares	↓SL, ↑TST, ↓REM, ↑SWS, ↑PLMS	No data	Impairment
Quetiapine	Seroquel	Schizophrenia, bipolar, depression adjunct, OCD	H_1, α_1, α_2, 5-HT_{2A}, D_2 antagonism	Somnolence +++ Insomnia ++ RLS	↓SL, ↑TST, ↓REM, ↑PLMS	No data	Possible improvement
Ziprasidone	Geodon	Schizophrenia, bipolar	D_2, 5-HT_2, α_1, D_1 antagonism	Somnolence ++ Insomnia ++ RLS	↓SL, ↑TST, ↓REM, ↑SWS, ↑REM latency, ↑PLMS	No data	Possible improvement
Other Mood Stabilizers							
Lithium carbonate	Eskalith Lithobid	Manic episodes	Unknown	Somnolence + RLS	No data	No data	Mixed results

5-HT, Serotonin; D, dopamine; GABA, gamma-aminobutyric acid; GAD, generalized anxiety disorder; H, histamine; M, muscarinic anticholinergic; NE, norepinephrine; MAO, monoamine oxidase; MSLT, Multiple Sleep Latency Test; MT, melatonin; MWT, Maintenance of Wakefulness Test; OCD, obsessive compulsive disorder; PLMS, period limb movements during sleep; PMDD, premenstrual dysphoric disorder; PTSD, posttraumatic stress disorder; RBD, REM behavior disorder; REM, rapid eye movement; RLS, restless legs syndrome; SEM, slow eye movement; SL, sleep latency; SWS, slow wave sleep; TST, total sleep time; W, wake.

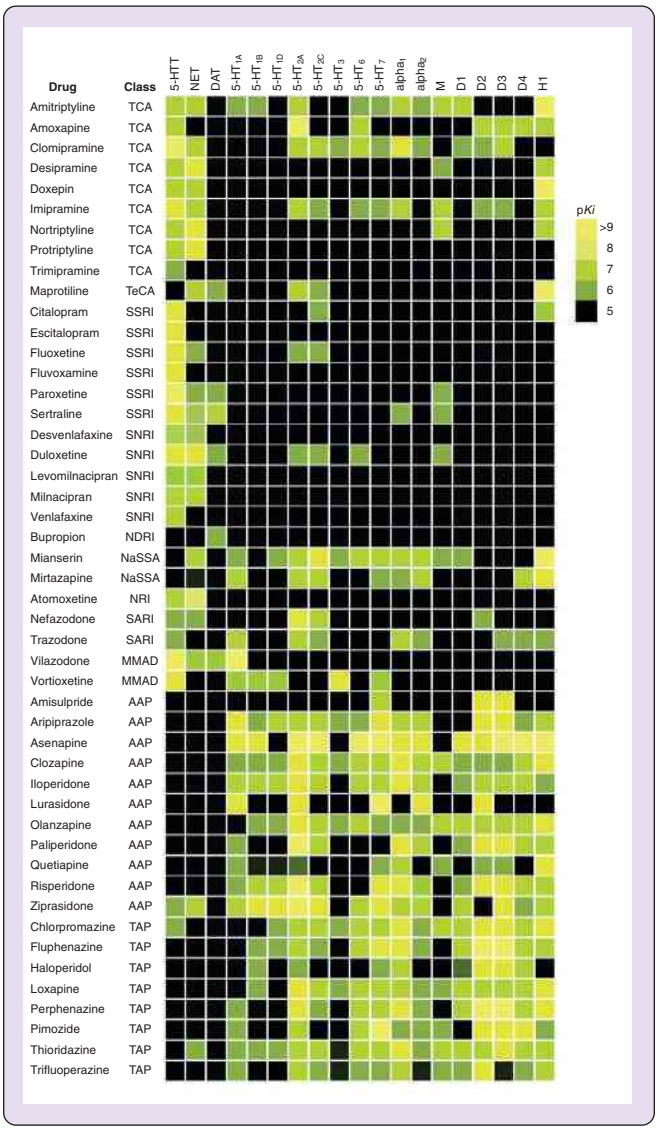


Figure 45-1 Heat map representation of receptor binding affinities for select psychotropic medications. Binding affinities (in p*Ki* values) range from 5 (inactive, *black*) to >9 (highly active, *yellow*). Binding affinity data were retrieved from the Psychoactive Drugs Screening Program (PDSP) database (http://pdsp.med.unc.edu)[5] and from Michl and colleagues.[6] 5-HT, Serotonin; 5-HTT, serotonin transporter; AAP, atypical antipsychotic; alpha, α-adrenergic; D, dopamine; DAT, dopamine transporter; H, histamine; M, muscarinic cholinergic; MMAD, multimodal antidepressant; NaSSA, noradrenergic and specific serotonergic antidepressant; NDRI, norepinephrine-dopamine reuptake inhibitor; NET, norepinephrine transporter; NRI, selective norepinephrine reuptake inhibitor; SARI, serotonin antagonist and reuptake inhibitor; SNRI, serotonin-norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TAP, typical antipsychotic; TCA, tricyclic antidepressant; TeCA, tetracyclic antidepressant. (Modified from Michl J, Scharinger C, Zauner M, et al. A multivariate approach linking reported side effects of clinical antidepressant and antipsychotic trials to in vitro binding affinities. *Eur Neuropsychopharmacol* 2014;24:1463–74.)

Selective Serotonin Reuptake Inhibitors

The primary mechanism of action of selective serotonin reuptake inhibitors (SSRIs) is potent inhibition of 5-HTT. With the exception of escitalopram, however, these drugs are not entirely selective. Citalopram has mild antihistamine properties; fluoxetine blocks the 5-HT_{2C} receptor, likely enhancing both norepinephrine and dopamine release; sertraline weakly inhibits the dopamine transporter (DAT); both sertraline and fluvoxamine are active at the σ_1 receptor, which may account for their anxiolytic effects; and paroxetine weakly inhibits NET. These diverse actions may explain why SSRIs may differentially be associated with insomnia[12,19] and daytime sedation.[8,12,20] Polysomnography (PSG) studies of SSRIs generally indicate disruption of sleep continuity and suppression of REM.[14,15,21] SSRIs are also associated with increased frequency of PLMS, RLS,[22] and REM sleep without atonia.[23]

Fluoxetine, which has been studied most extensively, decreases TST and increases wake time in normal subjects during single-night studies with doses of 20 to 60 mg[24] and in depressed patients with doses of 20 to 80 mg for up to 1 year.[25] Fluoxetine has been associated with the presence of prominent slow eye movements in non–rapid eye movement (NREM) sleep.[26] Paroxetine decreases TST and increases awakenings in normal subjects.[27] Escitalopram showed decreased TST, increased wake time, decreased REM sleep, and increased PLMS following a single 10-mg dose in healthy males.[15] In depressed patients, there is evidence of increased awakenings and sleep fragmentation with paroxetine and fluvoxamine.[28,29] Citalopram produced the typical decrease in REM sleep but no changes in sleep latency or TST in depressed patients during 5 weeks of treatment.[30] Sertraline prolonged both sleep latency and REM latency while decreasing REM sleep, but had no effects on sleep efficiency or wake after 12 weeks of treatment in depressed patients.[31]

SSRIs usually do not negatively affect daytime performance or cognitive functioning and may actually improve functioning in some patients, but data are limited. One placebo-controlled study reported memory impairment with paroxetine but improvement in a verbal task with sertraline.[32] A single nighttime dose of fluvoxamine in healthy subjects showed increased daytime sleep latencies compared with dothiepin but no change compared with placebo in a modified MSLT.[33]

Serotonin-Norepinephrine Reuptake Inhibitors

Serotonin-norepinephrine reuptake inhibitor (SNRI) drugs combine 5-HTT inhibition with varying degrees of NET inhibition, which reportedly results in improved cognitive and physical pain symptoms.[34] These drugs increase 5-HT and norepinephrine throughout the brain as well as increasing dopamine in the prefrontal cortex. Venlafaxine exhibits dose-dependent variability in NET inhibition, such that 5-HT reuptake predominates at low doses and NET inhibition progressively increases as dose increases. Duloxetine, in addition to 5-HTT and NET inhibition, also weakly inhibits DAT and has weak antagonistic activity at 5-HT_{2A}, 5-HT_{2C}, 5-HT_6, and muscarinic receptors.[5] With the exception of milnacipran (approved for treatment of fibromyalgia), all of these drugs have FDA indication for the treatment of depression. Duloxetine has, in addition, been approved for treatment of neuropathic pain, fibromyalgia, generalized anxiety disorder, and chronic musculoskeletal pain. Venlafaxine is additionally approved for use in anxiety and panic disorders and has been studied for use in obsessive-compulsive disorder and posttraumatic stress disorder.

Both insomnia and somnolence are common with venlafaxine (placebo-corrected rates in clinical trials of 14% and 8%, respectively), desvenlafaxine (5% to 8% and 3% to 9%),

and duloxetine (4% and 5%) but rare with milnacipran and levomilnacipran.[35] In normal subjects, 75 to 150 mg of venlafaxine produced increased wake and stage 1 sleep; in addition, in six of the eight subjects, frequent PLMS (more than 25 per hour) were noted.[36] In depressed inpatients treated for 1 month, venlafaxine (maximum dose 225 mg/day) increased PSG-recorded wake after sleep onset.[37] In one study of normal subjects, sleep efficiency was decreased with duloxetine 60 mg twice daily but increased by duloxetine 80 mg daily.[38] In a study of depressed patients, duloxetine 60 to 90 mg daily had no effect on sleep continuity but increased stage 3 sleep.[39] REM was suppressed in both studies. In a non–placebo-controlled study of depressed patients, milnacipran increased TST and decreased REM sleep.[40] There are no studies that objectively evaluate daytime sleepiness and alertness. Neurocognitive function is either unchanged or improved with these drugs.

Serotonin Antagonist and Reuptake Inhibitors

Trazodone weakly inhibits 5-HTT and blocks 5-HT_{2A} receptors at antidepressant doses. However, at the low doses typically used off-label for treatment of insomnia, it does not antagonize 5-HT_{2A} receptors. It shows moderate H_1 antagonism, also inhibits 5-HT_{1A} and 5-HT_{1C} receptors, and demonstrates weak α_1 and α_2 antagonism.[41] Nefazodone (no longer available in the United States) is also a 5-HT_{2A} antagonist and a weak inhibitor of 5-HTT, NET, and DAT but has little affinity for 5-HT_{1A}, α_1, α_2, or H_1 receptors.[41]

Trazodone is currently more likely to be used as a hypnotic than as an antidepressant in doses lower than those used for depression. Drowsiness is the most commonly-reported side effect.[35] PSG studies are limited by small sample sizes, lack of placebo control, and other methodologic factors. These studies indicate that trazodone likely decreases sleep latency, increases SWS, and does not affect REM sleep, but effects on sleep continuity are equivocal.[42] Nefazodone may increase TST but has no consistent effect on sleep latency, SWS, or REM sleep.[12,43] Trazodone impairs performance in healthy individuals,[44] but data on depressed individuals are inconclusive. In a placebo-controlled study, low-dose trazodone (50 mg) administered as a hypnotic before sleep for a week in primary insomniacs resulted in next-day impairment of memory, equilibrium, and muscle endurance but no decrease in MSLT latency.[45]

Norepinephrine and Dopamine Reuptake Inhibitor

Bupropion weakly inhibits DAT and NET.[5] It is approved for the treatment of depression, smoking cessation, and seasonal affective disorder. The low degree of DAT occupancy likely accounts for this drug being activating or stimulating without the abuse potential seen in stimulants that are also NET and DAT inhibitors. Bupropion is associated with insomnia in 5% to 19% of patients in clinical trials.[46] In a PSG study of seven depressed patients, after 4 weeks of treatment bupropion did not affect sleep latency or TST but did decrease REM latency and increase REM sleep percentage.[47] Bupropion is not usually associated with cognitive or psychomotor performance impairment.[44]

Norepinephrine and Specific Serotonin Antagonists

Mirtazapine disinhibits both 5-HT and norepinephrine through blockade of α_2 receptors.[48] It is a potent antagonist of H_1 and 5-HT_2 receptors, which likely accounts for its highly sedating activity and the reason for its use off-label as a hypnotic. PSG studies indicate improvements in sleep latency and sleep continuity.[49] One study suggests RLS symptoms develop or worsen in up to 28% of patients.[50] Mirtazapine impairs driving performance, attention, reaction time, and verbal memory, at least acutely, in normal subjects.[51,52] Bedtime dosing of mirtazapine 15 mg impaired next day road-tracking performance acutely but not after 9 days.[53]

Selective Norepinephrine Reuptake Inhibitors

Atomoxetine, although classified as a norepinephrine inhibitor because of its potent inhibition of NET, also inhibits both 5-HTT and DAT, although much less effectively.[5] This drug was not developed as an antidepressant; it is approved for treatment of attention-deficit/hyperactivity disorder (ADHD). Clinical trials indicate somnolence in ADHD children (2% incidence) but insomnia in ADHD adults (8% incidence).[35] A PSG study in ADHD children showed that atomoxetine increased REM latency but did not adversely affect sleep latency.[54]

Clinical trials of reboxetine, not available in the United States, indicate an incidence of insomnia of approximately 10%.[55] PSG studies in depressed patients showed acute increases in wake and persistent suppression of REM.[56] In healthy normal subjects, reboxetine was subjectively sedating but did not impair cognitive function.[57]

Serotonin Partial Agonist Reuptake Inhibitor

The only drug currently in this category is vilazodone, which is a 5-HTT inhibitor and a partial agonist at 5-HT_{1A} receptors. Insomnia and abnormal dreams have been reported in clinical trials with placebo-corrected incidence rates of 4% and 3%, respectively.[35] A single 20-mg dose of vilazodone in healthy young males resulted in near abolishment of REM sleep, increase in SWS and delta power, and increased wake time, probably a result of 5-HT_{1A} agonism.[58]

Multimodal Antidepressant

Vortioxetine is a multimodal antidepressant that inhibits 5-HTT and is also a 5-HT_3, 5-HT_{1D}, and 5-HT_7 receptor antagonist as well as a 5-HT_{1B} partial agonist and a 5-HT_{1A} agonist. Neither insomnia nor sedation has been reported in clinical trials. However, abnormal dreams occurred at a rate of 1% to 2% (placebo corrected).[35] There are no PSG studies in humans. However, in the rat, vortioxetine increased wakefulness and frontal cortical activity, likely because of 5-HT_7 and 5-HT_3 antagonism and 5-HT_{1A} agonism.[59] Neurocognitive function, including driving performance, does not appear to be impaired and may be improved independent of the improvement in depression.[60,61]

Monoamine Oxidase Inhibitors

Drugs of this type include moclobemide (not available in the United States) and selegeline. MAO inhibitors inhibit the action of MAO enzymes, resulting in increased concentrations of 5-HT, norepinephrine, and dopamine. MAO type A (MAO-A) is primarily involved in the metabolism of 5-HT, norepinephrine, epinephrine, melatonin, and dopamine, whereas MAO type B primarily metabolizes dopamine. The classic MAO inhibitors (e.g., isocarboxazid, phenelzine, tranylcypromine) irreversibly inhibit both MAO-A and MAO-B enzymes. Insomnia and daytime sedation are

commonly-reported side effects (up to 62% and 42% of patients, respectively),[62] but there are no placebo-controlled studies. The most impressive PSG finding is a marked decrease in REM sleep, including almost complete abolishment of REM sleep.[63] TST is also decreased. Although MSLT studies are lacking, actigraphic monitoring in a small group of patients confirmed periods of decreased daytime activity coincident with reported episodes of napping, possibly associated with poor nighttime sleep. Cognitive and psychomotor performance data are limited. Subjective and PSG data suggest that insomnia is more likely with higher doses.[64,65]

Melatonergic Antidepressant

Agomelatine, approved for treatment of depression in Europe but not available in the United States, is a potent agonist at both melatonin-1 (MT_1) and MT_2 receptors as well as an antagonist at 5-HT_{2C} receptors.[5] A single placebo-controlled PSG study showed phase shifts in body temperature and hormonal rhythms in healthy older men but no changes in sleep variables.[66] Studies without placebo control show variable effects on sleep.

Antipsychotic Drugs

Antipsychotic drugs were initially developed for the treatment of schizophrenia. However, many drugs now have clinical approval for the treatment of other disorders such as bipolar disorder, autism, obsessive-compulsive disorder, and major depressive disorder. In addition, these drugs have been used off-label in the treatment of dementia, depression, borderline personality disorder, posttraumatic stress disorder, substance abuse, eating disorders, anxiety, and insomnia.[67] Antipsychotics used as hypnotics (e.g., olanzapine, quetiapine) are covered more fully elsewhere in this volume.

Antipsychotic drugs have complex pharmacologic profiles (see Figure 45-1). Their antipsychotic effects are thought to be mediated primarily by antagonism of dopamine-2 (D_2) receptors in the mesolimbic dopamine pathway.[68] These drugs also block dopamine receptors in other neural pathways, leading to unwanted effects such as anhedonia, extrapyramidal symptoms, and hyperprolactinemia, which are commonly seen with the older or "typical" antipsychotics (e.g., chlorpromazine, haloperidol, thioridazine). These effects are decreased with the second-generation or "atypical" antipsychotic drugs that, in addition to blockade of dopamine receptors, also antagonize serotonin receptors (especially 5-HT_{2A}), have the ability to dissociate from D_2 receptors, may have partial agonist activity at D_2 receptors, or may act as partial agonists at 5-HT_{1A} receptors.[48] These drugs differ in their specificity for D_2 versus D_1 receptors. Typical antipsychotics also have some degree of 5-HT antagonism. Antipsychotics in general, whether typical or atypical, differ in the degree to which they block muscarinic cholinergic, H_1, and α-adrenergic receptors.[5,68] Sedation is more likely in drugs with relatively more potent antagonism of histamine, α-adrenergic, or 5-HT_2 receptors compared with antagonism of dopamine receptors. See Table 45-1 for a summary of the effects of these drugs on sleep and waking behavior.

Patients with schizophrenia commonly have insomnia and circadian rhythm disturbances.[69] Sedation is a common side effect of antipsychotic drugs, but insomnia has also been reported. Among the older drugs, chlorpromazine and thioridazine have very high rates of somnolence (33% to 57%), as does haloperidol (23%).[68] Among the atypical drugs, sedation is very frequent with clozapine (transient sedation 54%, persistent sedation 46%, and sedation requiring drug discontinuation 24%),[70] frequent with risperidone (30%) and olanzapine (29%), and moderately frequent with quetiapine and ziprasidone (12%).[68] Although ziprasidone and quetiapine would be expected to be sedating given their pharmacologic profiles, they appear less sedating than other drugs, possibly because of their short half-lives. Aripiprazole is least sedating (12%) but more likely to cause insomnia.

Insomnia is frequent with the typical antipsychotics haloperidol and thioridazine.[68] Among atypical drugs insomnia is highest with aripiprazole but also frequent with risperidone and olanzapine and moderately frequent with quetiapine and ziprasidone.[68] Although the mechanism is not clear, options include 5-HT_{1A} receptor agonism and RLS symptoms secondary to dopaminergic antagonism.[68] RLS symptoms have been reported in case reports for olanzapine, risperidone, quetiapine, and clozapine.[71-73] Drugs with high potency for blocking dopamine receptors are more likely to trigger RLS and PLMS.[4]

Double blind placebo-controlled PSG studies of healthy normal subjects and schizophrenia patients indicate that olanzapine markedly increases SWS, likely because of 5-HT_2 antagonism. Haloperidol, risperidone, and likely ziprasidone also increase SWS. Olanzapine also increases REM sleep; quetiapine and risperidone decrease REM sleep; and haloperidol and ziprasidone have variable effects on REM sleep. Clozapine, haloperidol, olanzapine, quetiapine, and ziprasidone all increase sleep continuity.[68,71,74-79] The relatively high cholinergic antagonism of chlorpromazine and thioridazine may lead to REM suppression, but PSG data are lacking.

Objective measures of daytime sleepiness are rare. However, both clozapine and olanzapine reduce MSLT latencies in schizophrenic patients.[80] A placebo-controlled study showed increased Epworth Sleepiness Scale scores with quetiapine but not with lurasidone in patients with acute schizophrenia.[81]

Although antipsychotics cause cognitive impairment in healthy subjects,[82] these drugs may have no negative effects in patients, who generally have mild to moderate cognitive impairment when untreated.[83] Although there is evidence that treatment with some atypical antipsychotics may remediate impairment of cognitive function,[84,85] results from large randomized trials have not supported this conclusion.[86] One exception is a study showing superior cognitive performance compared with placebo in patients receiving lurasidone.[87] Risperidone decreased cognitive performance while increasing sleepiness in a study in healthy normal subjects, and paliperidone was no different from placebo in the same study.[88]

Lithium

Lithium, which is used primarily in the treatment of bipolar disorder, is subjectively associated with improved nocturnal sleep and increased daytime sleepiness, at least initially.[89] Sleep disturbance is a prominent feature of mania and similar polysomnographically to that observed in major depression.[90] In healthy volunteers lithium increases SWS, decreases REM sleep,[91] and produces cognitive and psychomotor deficits, including prolonged reaction times, decreased vigilance, and

impairment of semantic reasoning.[90] Similar deficits have been shown in psychiatric patients taking lithium for periods of time ranging from 2 weeks to longer than 3 months,[92] although it is difficult to determine whether the deficits seen in the patient population are caused by the medication or their psychiatric illness. Degree of cognitive deficit increases with age, severity of disease, and lithium concentration.

Anxiolytic Drugs

The primary drugs used in the treatment of anxiety disorders include antidepressants (particularly SSRIs and SNRIs), benzodiazepines, and buspirone, a 5-HT_{1A} partial agonist. Antiepileptics and atypical antipsychotics are also sometimes used off-label, particularly pregabalin, quetiapine, and aripiprazole.[93,94] Prazosin, an antihypertensive with α_1-adrenergic antagonism, has been used to treat nightmares in posttraumatic stress disorder.[95]

Antidepressants, antiepileptics, and antipsychotics are covered elsewhere in this chapter. Benzodiazepines approved for treatment of anxiety disorders (e.g., alprazolam, clonazepam, diazepam, lorazepam) have similar pharmacologic profiles to benzodiazepines used to treat insomnia and thus similar side effects. Given that these drugs enhance GABA at the $GABA_A$ receptor, their most common side effect is sedation.[89] In nonanxious subjects, daytime administration of alprazolam and diazepam produced decreased MSLT latencies on both day 1 and day 7 of treatment, with alprazolam producing greater sleepiness than diazepam on the first day of treatment.[96] Performance impairment, including impairment of actual driving performance,[97] is common with daytime administration of benzodiazepines in studies of normal subjects and patient groups for treatment periods of up to 3 weeks, particularly at higher doses. Well-controlled studies are needed to determine whether longer-term use of benzodiazepine anxiolytics results in tolerance to these performance-impairing effects and whether there are differential effects between younger and older individuals.

Buspirone does not have the hypnotic, anticonvulsant, and muscle relaxant properties of the benzodiazepines. The anxiolytic efficacy of buspirone is similar to that of the benzodiazepines, but its onset of action is much slower, requiring up to 3 to 4 weeks.[98] Buspirone appears to act primarily as a 5-HT_{1A} partial agonist but also has effects on D_2 receptors.[99] It has no affinity for the benzodiazepine receptors and does not affect GABA binding. In clinical studies of anxious patients, buspirone was comparable with placebo in the frequency of subjective reports of sedation.[100] In a study of 12 patients with chronic insomnia, alertness as measured by MSLT was not impaired by 20 mg/day in divided doses over a 3-day period.[101] Compared with benzodiazepine anxiolytics, buspirone appears to have few negative effects on psychomotor, cognitive, or driving performance in healthy volunteers receiving short-term treatment or in patients treated for up to 4 weeks.[101,102]

ANTIEPILEPTIC DRUGS

Antiepileptic drugs (Table 45-3) include compounds with diverse pharmacology and chemistry that share the common property of ability to decrease neuronal excitability. Although these drugs are used to treat epilepsy, a number of them are used in the treatment of neurologic and psychiatric disease, including neuropathic pain, hyperkinetic movement disorders, migraine, RLS, bipolar disorder and schizophrenia.[103] For example, gabapentin is more frequently used for postherpetic neuralgia and RLS than for epilepsy, whereas pregabalin is more frequently used in the treatment of neuropathic pain and fibromyalgia than for seizure control. Both drugs, along with tiagabine, have been evaluated for insomnia. Lamotrigine and carbamazepine are used in the treatment of bipolar disorder, and eslicarbazepine is under study for bipolar treatment. Divalproex sodium and valproic acid are used for migraine prophylaxis.

Primary mechanisms of action of these drugs include (1) blockade of voltage-dependent sodium channels; (2) GABA-mediated enhancement through interaction with specific $GABA_A$ binding sites, inhibition of GABA metabolism, or reduction of neuronal uptake of GABA; and (3) blockade of voltage-gated calcium channels.[104] Many of the newer drugs have more specific mechanisms of action including (1) blockade of glutamate *N*-methyl-D-aspartate (NMDA; e.g., felbamate); (2) blockade of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)/kainite receptors (e.g., topiramate, perampanel); (3) synaptic vesicle 2A binding (levetiracetam); and (4) enhancement of transmembrane potassium currents mediated by KCNQ ion channels (ezogabine).[104] Conventional or "older" antiepileptics are mostly likely to involve GABA neurotransmission, whereas most of the newer compounds are more likely to have other or multiple mechanisms of action; the exceptions are tiagabine and vigabatrin, whose mechanisms involve GABA inhibition.

Sleepiness is one of the most commonly-reported adverse effects of antiepileptic drugs. In general, drugs acting on GABAergic neurotransmission (benzodiazepines, barbiturates, tiagabine, vigabatrin) have the highest incidence of sleepiness or fatigue (15% to 30% or more). Interestingly, although placebo-controlled trials showed no difference between tiagabine and placebo in the incidence of sedation, open-label, long-term studies showed a 25% incidence of sedation with tiagabine.[105] The incidence of sedation with drugs acting primarily via sodium channel blockade (e.g., carbemazepine, phenytoin, eslicarbazepine) is 5% to 10%, while the incidence of sedation with drugs acting via calcium channel blockade or with multiple mechanisms of action varies substantially, being 5% to 15% for gabapentin, lamotrigine, perampanel, pregabalin, zonisamide, and 15% to 27% for levetiracetam and topirimate.[35,104,106-109]

PSG studies of established antiepileptic drugs in general show these drugs produce shorter sleep latency and increase TST.[110] The newer drugs show variable effects on sleep latency and sleep continuity (see Table 45-3). However, gabapentin, pregabalin, and tiagabine increase SWS, whereas lamotrigine decreases SWS.[111-114]

A placebo-controlled study in healthy normal subjects reported no decrement in MSLT latencies with levetiracetam.[114] However, results have been mixed in patient groups. One study of patients with partial epilepsy showed no decrement in MSLT latencies but an increase in Epworth Sleepiness Scale ratings,[115] whereas another reported decreased Maintenance of Wakefulness Test latencies.[116] Neither gabapentin nor topiramate showed increased sleepiness on a 6-minute "awake maintenance test," although gabapentin produced electroencephalogram slowing.[117] In studies without placebo control, patients treated with phenobarbital[118] or

Table 45-3 Effects of Antiepileptic Drugs on Sleep and Wake Behavior

Drug or Class	U.S. Trade Name	FDA Indication Other than Seizure Control	Primary Mechanism of Action	Subjective Data	PSG Data	MSLT / MWT Data	Cognitive / Performance Data
Conventional Antiepileptics							
Benzodiazepines: Clonazepam Diazepam Lorazepam	Klonopin Valium Ativan	Anxiety	$GABA_A$ receptor modulation	Somnolence +++	↓SL, ↓SWS, ↓REM	↓SL	Mild impairment
Carbamazepine	Tegretol	Bipolar disorder, trigeminal or glossopharyngeal neuralgia	Sodium channel blockade	Somnolence ++	?↓SL, ↑TST, ?↓REM, ↑SWS	↓SL	Mild impairment
Ethosuximide	Zarontin		Calcium channel blockade	Somnolence +		No data	No data
Phenobarbital	Phenobarbital		$GABA_A$ receptor modulation	Somnolence +++	↓SL, ↓#W, ↑TST, ↓REM	↓SL	Significant impairment
Phenytoin	Dilantin		Sodium channel blockade	Somnolence ++	↓SL, ↑S1, ↓SWS	No data	Moderate impairment
Primidone	Mysoline		$GABA_A$ receptor modulation	Somnolence +	↓SL	No data	No data
Valproate, valproic acid	Depakote Depakene	Mania Migraine prophylaxis	GABA synthesis, sodium channel blockade	Somnolence ++	↑TST, ↓#W, ↑S1	No change	Mild to moderate impairment
Atypical Antiepileptics							
Brivaracetam	Not approved	Under study for neuropathic pain	Synaptic vesicle SV2A binding, sodium channel blockade	Somnolence +	No data	No data	No impairment
Eslicarbazepine	Aptiom	Under study for bipolar disorder	Sodium channel blockade	Somnolence +	No data	No data	No impairment
Ezogabine	Potiga		Interaction with voltage-gated potassium KCNQ channels and with GABA receptors	Somnolence +			
Felbamate	Felbatol		Blockade of sodium and calcium channels and NMDA	Insomnia + Somnolence +	No data	No data	No data
Fosphenytoin	Cerebyx		Sodium channel blockade	Somnolence ++	No data	No data	No data
Gabapentin	Neurontin Gralise Horizant	Postherpetic neuralgia RLS	GABA turnover and calcium channel blockade	Somnolence ++	↑SWS, ↑REM, ↑TST,	↓MWT	No to moderate impairment
Lacosamide	Vimpat		Sodium channel slow inactivation	Somnolence +	No data	No data	No impairment

Continued

Table 45-3 Effects of Antiepileptic Drugs on Sleep and Wake Behavior—cont'd

Drug or Class	U.S. Trade Name	FDA Indication Other than Seizure Control	Primary Mechanism of Action	Subjective Data	PSG Data	MSLT / MWT Data	Cognitive / Performance Data
Lamotrigine	Lamictal	Bipolar disorder	Sodium channel blockade, possible increase of GABAergic activity	Somnolence + Insomnia +	?↓SWS	No change	No impairment
Levetiracetam	Keppra		Acutely: calcium channel blockade Chronically: synaptic vesicle SV2A binding	Somnolence ++	↑TST ↑SWS	No change in healthy normals, mixed findings in patients	Mild impairment
Oxcarbazepine	Trileptal Oxtellar XR		Sodium and calcium channel blockade, possibly potassium channel effects	Somnolence ++	No data	No data	Driving impairment
Perampanel	Fycompa		Antagonism on AMPA/kainite glutamate receptors	Somnolence ++			
Pregabalin	Lyrica	Fibromyalgia, neuropathic pain, postherpetic neuralgia, under study for anxiety	Selective binding to alpha-2-delta subunits of voltage-gated calcium channels	Somnolence ++	↓SL, ↓W, ↑SWS, ?↓REM	No data	Mild impairment
Rufinamide	Banzel		Probably sodium channel blockade	Somnolence ++	No data	No data	No impairment
Tiagabine	Gabitril		GABA uptake inhibition	Somnolence + to ++++	↑TST, ↑SWS	No data	Mild impairment
Topiramate	Topamax	Migraine prophylaxis	GABA increase, sodium blockade, antagonism of AMPA/kainite glutamate receptors, inhibition of brain carbonic anhydrase activity	Somnolence +++ RLS	No data	No impairment on MWT	Dose-dependent moderate to significant Impairment
Vigabatrin	Sabril	Infantile spasms	GABA transaminase inhibition	Somnolence ++	No data	No change?	No impairment
Zonisamide	Zonegran		Sodium and calcium channel blockade	Somnolence ++	No change	No change	Dose-dependent impairment

#W, Number of awakenings; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; GABA, gamma-aminobutyric acid; MAO, monoamine oxidase; MSLT, Multiple Sleep Latency Test; MWT, Maintenance of Wakefulness Test; NMDA, *N*-methyl-D-aspartate; REM, rapid eye movement; TST, total sleep time.

carbamazepine[119] had lower mean MSLT latencies, and patients on carbamazepine, phenytoin, phenobarbital, or valproate demonstrated increased drowsiness on the "awake maintenance test" compared with healthy control subjects and untreated patients with epilepsy.[106] Drug-naive patients with partial epilepsy on topiramate showed no change in MSLT latencies compared with healthy control subjects either at baseline or 2 months later.[120] Neither lamotrigine nor zonisamide showed change in MSLT when used as add-on therapy in patients with focal epilepsy.[121,122]

Cognitive impairment appears to be more common with phenobarbital and possibly with primidone than with other drugs, more common when multiple medications are used, and more common in children than in adults. Phenobarbital and phenytoin have most frequently been associated with significantly impaired neuropsychological function, particularly in the areas of short-term memory, concentration, and attention.[123] Carbamazepine is moderately impairing.[124] Among the newer antiepileptic drugs, topiramate demonstrates dose-dependent persistent impairment on multiple cognitive domains.[123,125,126] Most of the other newer drugs appear to have fewer and less severe negative cognitive effects. In particular, gabapentin, lamotrigine, vigabatrin, and levetiracetam appear to have few negative effects, although placebo-controlled studies are rare.[127-129] Driving impairment was less with oxcarbazepine than with carbamazepine.[130] A placebo-controlled study of pregabalin showed no driving impairment in healthy individuals but, unlike placebo, there was failure to improve skills with training.[131] It should be noted that well-controlled studies evaluating cognitive function are rare, and methodologic problems include subject composition, choice of neuropsychological test, lack of placebo control, and sample size. Because a number of these drugs are being increasingly used in nonepilepsy disorders, the need for carefully-controlled studies evaluating the effects of these drugs on daytime function is increased.

DRUGS USED IN THE TREATMENT OF PARKINSON DISEASE

Sleep disorders are extremely common in Parkinson disease, the most frequent being insomnia, sleep apnea, REM behavior disorder, RLS, and disorders of daytime alertness.[132] Sleep-related complaints, which tend to worsen with disease progression, may be the result of abnormalities in sleep-wake regulation caused by the disease, accompanying symptoms such as nocturnal motor disturbance, other sleep disorders such as sleep apnea or PLMS, concurrent medical or psychiatric illness, or the medications used for treatment. PSG studies of patients with Parkinson disease show increased sleep fragmentation[133] and decreased REM sleep and SWS,[132] along with increased incidence of REM behavior disorder.[134] PLMS and sleep apnea are also common. MSLT studies indicate a high prevalence of excessive daytime sleepiness.[135] Dopamine replacement is the primary treatment for Parkinson disease. Because sleep is significantly disrupted in Parkinson disease, it is difficult to determine whether changes in sleep and waking behavior following drug administration are due to direct effects of drug or effects of drug on disease.

The principle drugs used to treat Parkinson disease are levodopa/carbidopa and dopamine agonists, which include ergot agonists (apomorphine, bromocriptine, cabergoline, lisuride, piribedil, and pergolide) and nonergot agonists (pramipexole, ropinirole, and rotigotine). Amantadine is less commonly used. Adjunctive treatments include catechol-*O*-methyltransferase inhibitors (e.g., entacapone, tolcapone), which prolong the duration of the effect of levodopa,[136] anticholinergics (e.g., hyoscyamine, benztropine), and selective irreversible MAO-B inhibitors (selegiline, rasagiline) which presumably result in increased dopamine through blocking dopamine catabolism. Selegiline is metabolized to methamphetamine and amphetamine.[137]

Dopamine agonists differ somewhat in their selectivity for dopamine receptor subtypes. There are five subtypes of dopamine receptors divided into two major classes, D_1 (D_1 and D_5 subtypes) and D_2 (D_2, D_3, D_4 subtypes).[138] The nonergot drugs have higher selectivity than the ergot agonists.[139] Pergolide and apomorphine have both D_1 and D_2 agonist activity, whereas pramipexole, ropinirole, and rotigotine are D_2 selective. All three have higher specificity for the D_3 receptor subtype than for the D_2 subtype; however, pramipexole exhibits the highest specificity. Ergot-derived drugs also demonstrate 5-HT_{2A} and 5-HT_{2B} agonism. Long-term use of the ergot agonists cabergoline and pergolide have been associated with valvular heart disease, possibly associated with 5-HT_{2B} (and perhaps 5-HT_{2A}) agonism.[139]

Some data suggest that low doses of dopaminergic medications tend to improve sleep, whereas higher doses are likely to disrupt sleep.[132,139] PSG studies have shown mixed results, including both increased and decreased REM sleep and decreased SWS. Nightmares and visual hallucinations may be increased. Improvement in parkinsonian symptoms may independently improve sleep.

Although sleepiness is a common symptom in Parkinson disease, with prevalence rates of 15.5% to 74%,[140] there is controversy as to whether the sleepiness is related to the pathologic process of the disease itself (due to loss of neurons involved in the control of the sleep-wake cycle), use of specific drugs, or other factors, including comorbid disorders such as sleep apnea. Parkinson disease itself is associated with loss of orexin neurons,[141] and objective sleepiness in Parkinson disease correlates with decrease of cerebrospinal fluid hypocretin levels.[142] Subjective daytime sleepiness measured by Epworth Sleepiness Scale has been associated with nigrostriatal dopaminergic degeneration.[143]

Dopaminergic agonists have been associated with increases in daytime sleepiness, including sudden "sleep attacks."[144] Studies in rats[145] and humans[146] suggest that sleepiness may be a drug effect in that nonergot dopamine D_2 receptor agonists cause selective loss of orexin-immunoreactive neurons, possibly through suppression of glutamatergic inputs to orexin neurons. Although initial reports suggested that unintentional sleep episodes were related specifically to the nonergot dopamine agonists, subsequent data have indicated no difference in sleepiness or unintentional sleep episodes between ergot and nonergot agonists.[147,148] MSLT studies of Parkinson disease patients indicate no differences in sleepiness among various dopamine agonists, including pramipexole, ropinirole, bromocriptine, or pergolide, taken alone or in combination with levodopa,[149,150] or compared with levodopa alone.[151] On the other hand, a study in healthy individuals showed reduced MSLT latency with pramipexole but no differences among L-dopa, bromocriptine, and placebo.[152] (See Ataide and

Table 45-4 β-Adrenergic Antagonists: Selected Pharmacologic Characteristics and Risk for Insomnia

Drug	Lipid Solubility	β Selectivity*	Relative Affinity for 5-HT Receptors	Other Effects	Risk for Insomnia†
Bisoprolol	Moderate	β_1	Low		Low
Atenolol	Low	β_1	Low		Low
Betaxolol	Low	β_1			Low
Acebutolol	Moderate	β_1	Low	ISA‡	Low
Nebivolol	High	β_1	Low	Nitric oxide–mediated vasodilating activity	Low
Nadolol	Low	Nonselective	High		Low?
Sotalol	Low	Nonselective	High		Moderate?
Timolol	High	Nonselective	High		Moderate?
Pindolol	Moderate	Nonselective	High	ISA‡	Moderate
Carvedilol	Moderate	Nonselective	High	α_1 Antagonism	Moderate
Labetalol	Moderate	Nonselective	Low	α_1 Antagonism	High
Metoprolol	Moderate	β_1	Low		High
Propranolol	High	Nonselective	High		High

*Nonselective β antagonists appear to have more central nervous system–related side effects. They also have higher affinity for 5-HT receptors.
†High 5-HT receptor occupancy and high β_2 affinity may be more important factors in disrupting sleep than degree of lipophilicity although lipophilic compounds appear to be more disruptive of sleep than hydrophilic compounds.
‡β antagonists with intrinsic sympathomimetic activity (ISA) act as partial agonists at β_2 receptors.

colleagues[153] for a review of MSLT studies in Parkinson disease.) The association of sleepiness with dopaminergic drugs may be related to dose, with lower doses more likely to improve sleep through D_2 autoreceptor activation, whereas higher doses may impair wake through differential activation of D_1 receptors. Indeed, several studies suggest that higher doses of dopaminergic drugs are more likely to be associated with increased sleepiness.[154-156]

Cognitive and motor deficits are common in Parkinson disease. Anticholinergic drugs have been demonstrated to produce worsening of cognitive function, primarily in the areas of memory function.[157] In healthy subjects, pramipexole impaired cognitive performance while increasing subjective sedation,[158] pergolide impaired delayed response tasks but not memory or executive function,[159] while ropinirole resulted in improved fine motor activity and reaction time.[160] In patients with mild Parkinson disease, pramipexole worsened verbal fluency, but pergolide and L-dopa did not.[161] Evaluation of the effects of these drugs on cognitive function is complicated by the frequent presence of behavioral symptoms that may affect performance. Moreover, the interaction of disease-related severity of dopamine depletion with distinct dopamine replacement therapies may produce different cognitive profiles at various stages of disease.[162]

Dopaminergic medications are also used in the treatment of RLS (covered more fully elsewhere in this volume). Somnolence and fatigue have been reported with pramipexole and ropinirole.[163] PSG data on pramipexole, ropinirole, and cabergoline generally show improvements in sleep in this population.[160,164,165] However, these improvements are likely the result of treatment of symptoms rather than the sedating effect of the drug. In a single MSLT study in healthy normal subjects, ropinirole decreased mean MSLT latency.[166]

CARDIOVASCULAR DRUGS

β-Adrenergic Antagonists

Information on pharmacologic characteristics of β antagonists relevant to central nervous system (CNS) sleep-wake function is given in Table 45-4. CNS side effects reported with β blockers include tiredness, fatigue, insomnia, nightmares and vivid dreams, depression, mental confusion, and psychomotor impairment.[167] There have also been case reports of REM behavior disorder with lipophilic β blockers.[168] Although sleep disturbance appears to be more common with the more lipophilic drugs, high β_2 or 5-HT receptor occupancy may be a more important factor in causing sleep disruption.[169] Plasma concentration (degree of β antagonism) may also be a factor.[170] Drugs that are more selective for β_1 receptors have lower affinity for 5-HT receptors. On the other hand, β antagonists decrease melatonin release through inhibition of β_1 receptors, which could affect sleep.[171] Among these drugs, propranolol, which has high lipid solubility and high 5-HT affinity, is most commonly associated with disturbed sleep. A study of new-onset insomnia following treatment for newly diagnosed hypertension in patients 65 years and older suggests that both selectivity and lipophilicity influence the risk for insomnia because atenolol and bisoprolol were associated with low risk for insomnia, whereas carvedilol, metoprolol, labetalol, and propranolol were associated with high risk.[172] Nebivolol, a highly selective and lipophilic compound with endothelial nitric oxide–mediated vasodilation,

appears to be less disruptive of sleep.[173] β blockers with vasodilating properties through blockade of α_1 receptors (e.g., carvedilol, labetalol) have been associated with fatigue and somnolence.[174] One placebo-controlled study[175] in normal subjects demonstrated increased wake with propranolol, metoprolol, and pindolol, but not with atenolol. However, REM sleep was decreased by all four drugs. Reviews of the effects of β blockers on cognitive and psychomotor performance indicate that these drugs produce few consistent neuropsychological deficits.[167,169]

α_2-Adrenergic Agonists

Sedation is the most common side effect of both clonidine and methyldopa, occurring in 30% to 75% of patients, but the severity apparently diminishes with time.[176] There are also reports of insomnia and nightmares with these drugs. In a double-blind placebo-controlled crossover study, hypertensive men aged 31 to 59 years who were given 0.1 to 0.3 mg of clonidine twice daily showed significantly decreased TST compared with placebo after 3 months of use.[177] Healthy subjects, however, given clonidine acutely showed increases in TST.[178] No MSLT studies exist to quantify daytime sedation objectively. However, one study of a single morning dose of clonidine in young healthy subjects demonstrated microsleeps in six of eight subjects.[179] Few well-controlled studies exist that evaluate the effects of these drugs on performance. Verbal memory impairment and poorer workplace performance[180] have been reported in patients receiving methyldopa.

Hypolipidemic Drugs

Data mining studies indicate that statins, particularly the more lipophilic compounds, are associated with a higher reporting rate of insomnia in comparison with other drugs.[181,182] One PSG study in normal subjects showed increased wake after sleep onset with lovastatin, whereas pravastatin did not differ from placebo.[183] Subsequent placebo-controlled clinical trials and PSG studies of lovastatin, simvastatin, and pravastatin have in general failed to show increased sleep disturbance even with the more lipophilic compounds.[184,185] However, Roth and colleagues[186] reported performance decrements with lovastatin even though nocturnal sleep and daytime sleep tendency (measured by MSLT) were not affected. There have been case reports of short-term memory loss associated with statin use, but randomized studies using neuropsychological testing and a meta-analysis of observational studies suggest that these drugs may actually lower the odds for development of cognitive impairment.[187,188]

Other Cardiovascular Drugs

The α_1 antagonists (e.g., prazosin, terazosin) are sometimes associated with transient sedation. Prazosin has been used in the treatment of nightmares and sleep disturbance in combat-related posttraumatic stress disorder.[95] In placebo-controlled studies prazosin increased TST, REM, and subjective sleep quality and reduced nightmares.[189,190] There are no reports of sleep disturbance or wake dysfunction with the calcium channel blockers (e.g., verapamil, nifedipine); however, these drugs decrease the effectiveness of hypnotics and potentiate the effects of stimulants, at least in studies in animals.[191] Angiotensin-converting enzyme inhibitors (e.g., captopril, cilazapril) reportedly have a low incidence of central side effects. However, a dry, irritating cough is a common side effect,[192] which may contribute to sleep apnea, possibly related to rhinopharyngeal inflammation.[193]

HISTAMINE ANTAGONISTS

Histamine-1 Receptor Antagonists

The first-generation H_1 antihistamines (e.g., diphenhydramine, hydroxyzine) are lipophilic and easily cross the blood–brain barrier, demonstrating H_1 receptor occupancy of up to 60% or more.[194,195] In addition to H_1 antagonism, these drugs demonstrate muscarinic anticholinergic antagonism and may also have α-adrenergic and serotonergic effects. They cause decrements in alertness and performance.[196,197] Because of their sedating effects, these drugs (in particular, diphenhydramine) are widely used as over-the-counter hypnotics. Subjectively, these drugs decrease sleep latency and increase sleep continuity, but PSG data are mixed. Diphenhydramine increases physiologic sleep tendency as measured by MSLT and decreases performance acutely, but tolerance may develop within 3 to 4 days,[198,199] although driving may continue to be impaired.[200] A single presleep dose of diphenhydramine 50 mg resulted in decreased REM and next-day impairment in psychomotor performance as well as increased subjective sleepiness and a trend toward decreased latency on MSLT.[201]

The second-generation H_1 antihistamines (cetirizine, desloratadine, fexofenadine, levocetirizine, loratadine) are hydrophilic molecules that do not easily penetrate the CNS. Although they are much more selective than the first-generation antihistamines, H_1 receptor occupancy varies from almost negligible (e.g., fexofenadine) to 30% (cetirizine).[188,200] Most studies in normal subjects and atopic individuals generally confirm that these drugs are not sedating and do not impair performance when used in recommended doses.[197] Although MSLT studies indicate that cetirizine is nonsedating,[198] it has been classified as sedating by the FDA, and a number of studies suggest it is more subjectively sedating and more likely to impair performance than other second-generation H_1 antagonists.[202,203] Meta-analysis of 18 studies concluded that although second-generation antihistamines caused less performance impairment than the first-generation antihistamine diphenhydramine, mild impairment was still present.[204] There is some evidence that fexofenadine and levocetirzine may be less impairing than other second-generation drugs,[195,205] although sedation may emerge with increase in dose.[206]

Histamine-2 Receptor Antagonists

H_2 antagonists (e.g., cimetidine, ranitidine, famotidine, and nizatidine) are unlikely to impair CNS function because these compounds do not easily cross the blood–brain barrier. However, cimetidine slows the clearance of some benzodiazepine receptor agonists, which may make carryover effects of hypnotics more of a problem.[207] Similarly, cimetidine has been shown to increase levels of theophylline, carbamazepine, and β blockers with resultant increases in the CNS effects of these drugs. Ranitidine has produced some of the same effects, although not to the extent seen with cimetidine.

One crossover study comparing 1-week administration of cimetidine, famotidine, ranitidine, and placebo in normal subjects reported no differences in nocturnal sleep or daytime MSLT latencies although cimetidine produced a slight

increase in subjective estimates of sleepiness.[208] H_2 antagonists do not appear to affect psychomotor performance. However, both cimetidine[209] and ranitidine, administered in conventional doses, have been associated with an increased incidence of lethargy, somnolence, and confusion in patients with renal impairment. In addition, benzodiazepine-produced impairment of psychomotor and cognitive function was prolonged with concomitant administration of cimetidine and ranitidine in healthy volunteers.[210]

DRUGS USED FOR THE TREATMENT OF PAIN

The neurobiology of pain is complex and involves both peripheral and central mechanisms. Simplistically, activation of peripheral nociceptors (by inflammation or tissue damage) is modulated by central mechanisms that affect pain perception. Pain disrupts sleep, whereas sleep loss may increase pain sensitivity.[211] Multiple neurotransmitters modulate pain processing, including substance P, endorphins (through μ-opioid receptors), norepinephrine (α_2 adrenoceptors), and serotonin (5-HT_{1B}, 5-$HT_{1D,}$ 5-HT_3 receptors).[212] Drugs used in the treatment of pain include analgesics (opiates and nonsteroidal antiinflammatory drugs [NSAIDs]), muscle relaxants, antidepressants (particularly TCAs and SNRIs), and antiepileptics.[211,213] The SNRI duloxetine has been approved for treatment of diabetic neuropathy, fibromyalgia, and chronic musculoskeletal pain; the SNRI milnacipran has been approved for fibromyalgia. Among antiepileptics, gabapentin is approved for treatment of postherpetic neuralgia and used frequently off-label for other neuropathic pain. Pregabalin is approved for treatment of neuropathic pain, postherpetic neuralgia, and fibromyalgia. These drugs are discussed in the sections on antidepressants and antiepileptics. In addition, pain and sleep are discussed more fully elsewhere in this volume.

Opioids act at a variety of CNS receptors, including μ, κ, δ, and nociceptin/orphanin FQ.[214] All of these receptor subtypes appear to be involved in the analgesic effect of opioids, whereas the μ subtype plays a prominent role in respiratory control. Sedation is mediated by both μ and κ receptors.[212] The most common clinically-used opioids are relatively selective for μ receptors. Opioid peptides (enkephalins, endorphins, dynorphins) are involved in the regulation of a number of biologic activities, including blood pressure, respiration, mood, pain perception, and possibly sleep.[215] Somnolence is a common side effect of opioid medication.[216] Degree of sedation may depend on the specific drug, dosage, and duration of use as well as severity of the underlying disease.[217] Chronic use in cancer or chronic pain patients is associated with both insomnia and daytime sleepiness and fatigue, but this may be secondary to disease or psychological function.[218] One case-controlled study showed increased subjective sleepiness (measured by Epworth Sleepiness Scale) in patients on stable methadone compared with normal controls, but mean Epworth values for the group were within the normal range; daytime function was impaired (as measured by the Functional Outcomes of Sleep Questionnaire), however.[219] Older adults appear to have increased pharmacodynamic sensitivity to opioids.[220]

The limited PSG data available indicate that opioids used acutely in young healthy volunteers markedly decrease SWS and may decrease stage 2 and REM, particularly at higher doses.[221,222] In current or former addicts, opioids also decrease TST and increase wake after sleep onset.[223] Subjective quality of sleep may be improved, presumably because of improved pain control.[224]

In healthy normal subjects, cognitive and psychomotor function are impaired in a dose-related fashion with parenteral administration of opioids, but findings are mixed with oral dosing.[225] Acute administration in pain patients results in dose-related impairment in cognitive functioning.[217,226] Both acute and chronic opioid use have been associated with deficits in attention, recall, visuospatial skills, and psychomotor speed, whereas long-term use appears to have the greatest effect on executive function.[227] Comorbid medical or psychiatric disease may be more predictive of cognitive impairment than frequency or dose of medication.[228]

The most serious adverse effect of opioids is respiratory depression, particularly during sleep or after surgery. Opioids act directly on the brainstem respiratory centers through μ and δ receptors and at chemoreceptors through μ receptors, resulting in a shift to the right and a change in slope of the carbon dioxide response curve.[229] Opioids depress the pontine and medullary centers involved in the regulation of respiratory rhythmicity, resulting in increased respiratory pauses, irregular breathing, and decreased tidal volume.[230] Respiratory depression increases with increase in opioid dose. Limited data suggest that clinically significant respiratory depression rarely occurs with standard opioid doses used acutely in healthy individuals.[231,232] However, chronic opioid use is a risk factor for the development of central sleep apnea and hypoxemia.[233] Individuals with pulmonary disease or obstructive sleep apnea are at greater risk for sustained hypoxemia during sleep.[234] Concomitant use of other sedatives, including sleep aids, increases the risk for potentially fatal respiratory depression. After surgery, individuals with obstructive sleep apnea receiving intravenous morphine have been shown to have pronounced oxygen desaturation, paradoxical breathing, and slow ventilation.[235]

Centrally-acting α_2 agonists used in the management of pain include clonidine, tizanidine, and dexmedetomidine. Somnolence is common with these drugs.[236] Tizanidine also negatively affects psychomotor performance and neurocognitive function.[237]

Triptans, which are selective 5-$HT_{1B/1D}$ agonists, are currently the primary treatment for acute migraine. Somnolence is a common side effect, but incidence varies among drugs, likely the result of differences in lipophilicity and the presence of active metabolites. Somnolence is highest with eletriptan, zolmitriptan, and rizatriptan, all of which are highly lipophilic and have active metabolites, and lowest with almotriptan, sumatriptan, and naratriptan, which are weakly lipophilic and have no active metabolites.[238] There are no studies which report PSG or MSLT data or which evaluate cognitive function.

NSAIDs may affect sleep because they decrease the synthesis of prostaglandin D_2, suppress the normal nocturnal surge in melatonin synthesis, and attenuate the normal nocturnal decrease in body temperature.[239] Prostaglandin D_2 increases proportionately with increased duration of wake and may be involved in sleep initiation.[240] NSAIDs inhibit cyclooxygenase (COX), blocking the synthesis of inflammatory prostaglandins. The classic NSAIDs inhibit both COX-1 (thereby accounting for their gastrointestinal toxicity) and COX-2 isoenzymes, whereas the newer NSAIDs selectively inhibit COX-2 (found primarily in the CNS, renal cortex, and

vas deferens).[241] Limited PSG data are mixed and have shown both no effect[242] and decreased sleep efficiency with acutely-administered aspirin and ibuprofen in healthy subjects.[239,243] One placebo-controlled study of tenoxicam (not available in the United States) in rheumatoid arthritis patients improved clinical symptoms but did not affect PSG measures.[244] Subjective reports suggest an improvement in sleep quality with use of NSAIDs, presumably because of a reduction in pain. Cognitive deficits are apparently rare with NSAIDs but may be a problem in older adults.[245]

DRUGS USED FOR THE TREATMENT OF ALZHEIMER DISEASE

Drugs used in the treatment of dementia of the Alzheimer type include centrally-acting cholinesterase inhibitors (tacrine, donepezil, rivastigmine, galantamine) and NMDA receptor antagonists (memantine). Insomnia and nightmares have been reported in 1% to 5% (placebo-adjusted rate) of patients in clinical trials of donepezil.[246,247] In contrast, there was no indication of insomnia based on adverse-event reporting in clinical trials of rivastigmine or galantamine.[248,249] Clinical trials of memantine report variable rates of insomnia (placebo-adjusted rate, 0% to 3%) and sleepiness (placebo-adjusted rate, 0% to 6%).[250]

PSG studies in normal subjects and Alzheimer disease patients show an increase in REM percent with donepezil but no effect on TST or other sleep architecture measures.[251,252] However, donepezil in healthy older adults, rivastigmine in healthy older adults and young normal subjects, and galantamine in young normal subjects produced no change in REM percent, although REM latency decreased and REM density increased.[253-256]

OTHER DRUGS

Corticosteroids

Corticosteroids are widely believed to disrupt sleep, but the results of objective studies are inconsistent. Differences in receptor affinities between synthetic and endogenous corticosteroids, dosage, methodologic issues associated with the study of patient populations, and the variety of organ systems affected by corticosteroids, as well as the variety of side effects reported,[257] all contribute to this confusion.

In patient populations, corticosteroids have frequently been associated with sleep disturbance. Approximately 50% of patients treated with prednisone for optic neuritis reported sleep disturbance, compared with 20% on placebo.[258] Patients taking prednisone for oral inflammatory ulcerative disease reported a dose-related incidence of insomnia ranging from 12% to 71%.[259] Parent ratings of sleep disturbance increased when steroids were added to the chemotherapy regimen of children with leukemia or other types of cancer.[260] Insomnia has also been reported more frequently in patients with asthma receiving steroid medications.[261] In addition, numerous anecdotal and case reports implicate systemic corticosteroid use with insomnia.[262] Behavioral observations of 12 healthy subjects given prednisone 80 mg/day for 5 days showed decreased sleep in 25% and mild hypomania in 67%.[263] Inhaled glucocorticoids do not appear to have the same negative effects, but there have been case reports of hyperactivity, insomnia, and psychosis with these drugs as well.

The most consistent effect of corticosteroids on PSG-recorded sleep in normal subjects is a marked decrease in REM sleep.[264] Although less consistent, there is good evidence for increased waking during the night with cortisol, dexamethasone, and prednisone. Dexamethasone, administered before bedtime, resulted in increased daytime alertness the next day as measured with MSLT.[265]

In a single study evaluating performance in healthy subjects, prednisone 80 mg/day given for 5 days produced increased frequency of errors of commission on a verbal memory task.[266] Prednisone was associated with decreased cognitive functioning in a study of patients with systemic lupus erythematosus.[267]

Pseudoephedrine and Phenylpropanolamine

Pseudoephedrine and phenylpropanolamine share the pharmacologic properties of ephedrine but have less potent CNS-stimulating effects. These drugs are used extensively as nasal decongestants and are available in a variety of over-the-counter cold preparations; phenylpropanolamine is also available in over-the-counter diet aids. Although similar in chemical structure to amphetamine, phenylpropanolamine is much less lipophilic and thus has much less potent CNS effects. Phenylpropanolamine, however, has been reported to increase plasma caffeine levels,[268] possibly adding to the stimulant effect of caffeine. These drugs have been reported to cause insomnia. In one study, 27% of patients given 120 mg of extended-release pseudoephedrine for 2 weeks for the treatment of allergic or vasomotor rhinitis complained of insomnia. In a PSG study, the administration of pseudoephedrine in the evening (as part of either a 60-mg four-times-daily or 120-mg sustained-release twice-daily dosing regimen) produced increased wake time during sleep compared with the morning administration of a once-daily controlled-release formulation (240 mg).[269] Further objective evaluation of dosage, timing, and duration of treatment of these drugs would be useful.

Stimulants

Stimulant medications are covered more fully elsewhere in this volume. Stimulants are the primary treatment for both narcolepsy and ADHD. Their use in ADHD has increased significantly over the past 20 years. Sleep problems are common in both children and adults with ADHD.[270,271] Although actigraphic, PSG, and MSLT data in children do not yield consistent findings, it appears likely that ADHD children have decreased sleep efficiency, decreased REM, increased daytime sleepiness, and possibly an increased prevalence of PLMS in sleep.[270] Limited actigraphic and PSG data in adults also indicate decreased sleep efficiency.[271,272] Stimulant medications used in the treatment of ADHD include immediate-release compounds (dextroamphetamine, methylphenidate) as well as extended-release formulations (methylphenidate [Concerta], dextroamphetamine-levoamphetamine [Adderall XR], and lisdexamphetamine [Vyvanse]). Modafinil has been studied in both children and adults but is not currently approved by the FDA for ADHD treatment.[273] Clinical trials and subjective reports indicate increased incidence of insomnia with these drugs. Parental reports, in particular, indicate significant sleep disturbance, which may be higher with the modified-release drugs. However, limited PSG and actigraphic data have yielded conflicting results. Cognitive

function may improve as a result of symptom improvement rather than a direct drug effect.

Theophylline

Theophylline, a respiratory stimulant and bronchodilator, is chemically related to caffeine. Peak plasma concentration is usually reached within 2 hours, but the half-life varies by preparation and is typically shorter in children (3.5 hours) and longer in adults (8 to 9 hours). Absorption is lower at night than in the morning[274] and may be greatly affected by food.[275]

Disturbed sleep is a common complaint among patients taking theophylline. In a prospective study, patients with asthma treated with theophylline were more likely to complain of sleep maintenance difficulty (55%) than were patients treated with other asthma medications (31%),[276] and in a retrospective study of treated patients with asthma, 46% of whom complained of insomnia, only theophylline or corticosteroid therapy was associated with the complaint of insomnia.[261] Most of the studies purporting that theophylline does not adversely affect sleep are limited by the lack of a placebo control or other methodologic difficulties. Because theophylline improves asthma-related symptoms, there have also been reports of improved sleep continuity and decreased nocturnal awakenings associated with its use.[277]

Theophylline, administered for up to 3 weeks, has been shown to disturb PSG-recorded sleep in healthy subjects,[278,279] patients with asthma,[280] children with cystic fibrosis,[281] and patients with sleep apnea[282] or chronic obstructive pulmonary disease.[283] Dose-dependent increase in MSLT latency and performance was noted with short-term administration of theophylline in normals.[279] In a double-blinded study, asthmatic children were more likely to exhibit behavioral or attentional problems when receiving sustained-release theophylline for 4 weeks than when on placebo.[284] However, a meta-analysis of 12 studies of theophylline did not indicate impairment in cognition or behavior.[285] Furthermore, academic achievement did not differ between 72 asthmatic patients who were treated with theophylline and siblings without asthma.[286]

Drugs Used for the Treatment of Obesity

Weight loss drugs with CNS activity include the sympathetic amines phentermine and diethylpropion; lorcaserin, a selective 5-HT_{2C} agonist; and the combination drugs phentermine-topiramate (Qsymia) and naltrexone-bupropion (Contrave). Bupropion, as well as the anticonvulsants zonisamide and topiramate, is used off-label for obesity. Insomnia is a common side effect of phentermine and diethylpropion because these drugs have pharmacologic activity similar to amphetamine. Insomnia has also been reported with lorcaserin, phentermine-topiramate, and naltrexone-bupropion. Impairment of attention and memory has been reported with lorcaserin.[287,288]

CLINICAL PEARL

Drugs that block the activity of wake-promoting neurotransmitters are likely to be sedating. Drugs that inhibit the reuptake or metabolism of wake-promoting neurotransmitters are likely to disrupt sleep. Knowledge of receptor pharmacology and pharmacokinetics is helpful in determining the likelihood of negative effects on sleep or waking function.

SUMMARY

Disturbed sleep, daytime sleepiness, and impaired cognitive functioning are common side effects of many medications. Negative effects may be the result of a direct action of a drug (e.g., disturbed sleep from an activating compound or carry-over sedation from a long-acting hypnotic) or an indirect action (e.g., daytime sleepiness as a result of drug-induced sleep disruption). Sleep-wake regulation involves multiple neuronal systems and neurotransmitters. Drugs that increase the activity of the sleep-promoting system or decrease the activity of the wake-promoting system will be sedating, whereas drugs that act in a reciprocal fashion will be alerting. Whether these effects are considered side effects depends on when they occur. Knowledge of receptor mechanisms and pharmacokinetics can help predict which drugs will negatively affect sleep or wake function. Drugs most likely to affect sleep-wake behavior negatively include psychotherapeutic drugs, antiepileptics, dopaminergic medications used in the treatment of Parkinson disease, first-generation antihistamines, and centrally-acting pain medications. A number of other drugs with CNS effects may produce unwanted sedation or sleep disruption.

Selected Readings

Ataide M, Franco CM, Lins OG. Daytime sleepiness and Parkinson's disease: the contribution of the multiple sleep latency test. *Sleep Disord* 2014; **2014**:767181.

Chang CH, Yang YH, Lin SJ, et al. Risk of insomnia attributable to β-blockers in elderly patients with newly diagnosed hypertension. *Drug Metab Pharmacokinet* 2013;**28**:53–8.

Cheatle MD, Webster LR. Opioid therapy and sleep disorders: risks and mitigation strategies. *Pain Med* 2015;**16**(Suppl. 1):S22–6.

Holst SC, Valomon A, Landolt HP. Sleep pharmacogenetics: personalized sleep-wake therapy. *Annu Rev Pharmacol Toxicol* 2016;**56**.

Krystal A. Antidepressant and antipsychotic drugs. *Sleep Med Clin* 2010; **5**:571–89.

Krystal A, Goforth H, Roth T. Effects of antipsychotic medications on sleep in schizophrenia. *Int Clin Psychopharmacol* 2008;**23**:150–60.

Landmark C. Antiepileptic drugs in non-epilepsy disorders: relations between mechanisms of action and clinical efficacy. *CNS Drugs* 2008;**22**:27–47.

Michl J, Scharinger C, Zauner M, et al. A multivariate approach linking reported side effects of clinical antidepressant and antipsychotic trials to in vitro binding affinities. *Eur Neuropsychopharmacol* 2014;**24**:1463–74.

Saper CB, Scammell T, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* 2005;**437**:1257–63.

Schug S, Garrett W, Gillespie G. Opioid and non-opioid analgesics. *Best Pract Res Clin Anaesthesiol* 2003;**17**:91–110.

Siniscalchi A, Gallelli L, Russo W, DeSarro G. A review on antiepileptic drugs-dependent fatigue: Pathophysiological mechanisms and incidence. *Eur J Pharmacol* 2013;**718**:10–16.

Takada M, Fujimoto M, Yamazaki K, et al. Association of statin use with sleep disturbances: data mining of a spontaneous reporting database and a prescription database. *Drug Saf* 2014;**37**:421–31.

Tan L, Tan L, Wang HF, et al. Efficacy and safety of atypical antipsychotic drug treatment for dementia: a systematic review and meta-analysis. *Alzheimers Res Ther* 2015;**7**:20.

Tashiro M, Mochizuki H, Iwabuchi K, et al. Roles of histamine in regulation of arousal and cognition: Functional neuroimaging of histamine H1 receptors in human brain. *Life Sci* 2002;**72**:409–14.

A complete reference list can be found online at ExpertConsult.com.

Effects of Hypnotic Drugs on Driving Performance

Chapter 46

Joris C. Verster; Aurora J.A.E. van de Loo; Thomas Roth

Chapter Highlights

- There are a variety of methodologies to examine whether it is safe to drive a car the day after being treated with hypnotic drugs. This chapter discusses epidemiologic evidence and explains the experimental methodology and results of the standardized on-the-road highway driving test to determine the effects of hypnotic drugs on driving ability.
- Most classic benzodiazepine hypnotics and zopiclone, when administered at bedtime, significantly impair next-morning driving ability. The magnitude of driving impairment depends on variables such as gender and age, drug dosage, half-life, and the time between drug intake and driving. Depending on dose and half-life, impairment of some benzodiazepines may last until the afternoon, that is, 16 to 17 hours after bedtime administration.
- When allowing a full night of sleep, next-morning driving ability was not impaired after bedtime administration of zolpidem (10 mg) and zaleplon (10 mg). Middle-of-the-night administration of zolpidem (10 mg), however, significantly impaired driving performance 4 hours after that time.
- Currently the only drugs that showed no significant driving impairment 4 hours after middle-of-the-night administration are zaleplon (10 and 20 mg) and sublingual zolpidem tartrate (3.5 mg).
- Despite its short half-life, the melatonin receptor agonist hypnotic ramelteon impairs next-day driving performance. Hence the development of safer yet effective hypnotic drugs is needed.

For most people, driving a car is a daily activity (e.g., to commute to and from work). Typically people with insomnia and other sleep disorders are outpatients; thus it is likely that they routinely drive a car. Driver sleepiness (reduced alertness) accounts for 10% to 30% of accidents.[1] Because a number of patients with insomnia report daytime sleepiness, it is important to determine whether sleep disorders or their pharmacologic treatments negatively affect driving.

Data on driving ability in untreated insomniacs is inconsistent as to whether insomnia impairs driving. The fact that only some insomniacs report daytime sleepiness, and objective assays of sleepiness show them to be alert, probably accounts for the negative findings in insomnia. Surprisingly little research has been conducted to examine the effect of insomnia on driving ability under controlled conditions. One on-the-road driving study found no impairment in patients with insomnia, but these patients used hypnotic drugs infrequently, so they may have benefited from treatment.[2]

To be effective, hypnotic drugs need to put patients to sleep and maintain sleep during the night. However, this induction of sedation needs to dissipate across the night because the patient wants to wake up refreshed without residual daytime sleepiness. The challenge is to find the right balance between efficacy into the later portion of the night and safety of hypnotic drugs. This chapter focuses on the effects of hypnotic drugs on next-day driving.

EPIDEMIOLOGIC EVIDENCE

Several epidemiologic studies examined the effect of hypnotic drug use on driving. Neutel selected 78,070 patients using the benzodiazepine hypnotics triazolam or flurazepam from the Saskatchewan Health Database and compared the risk for traffic accident injury with data from 97,862 "healthy control" subjects.[3] The use of benzodiazepine hypnotics was associated with a significantly increased (3.9 times) risk for traffic accident injury. The data further revealed that the risk for accident injury is highest after treatment initiation and then gradually decreases with continued use. Similarly, Barbone and colleagues reported an increased traffic accident risk for users of benzodiazepine hypnotics (odds ratio [OR] = 1.19; 95% confidence interval [CI], 0.83 to 170).[4] In contrast, McGwin and associates did not find a significant increase in traffic accident risk patients using benzodiazepine hypnotics (OR = 5.2; 95% CI, 0.9 to 30).[5] Importantly, researchers have shown that the risk was, in part, dependent on the half-life of the drug. Surprisingly, whereas classic benzodiazepine hypnotics with a long (>24 hours) or intermediate (6 to 24 hours) half-life showed no significant effect on accident risk, users of other benzodiazepine receptor agonist hypnotics with shorter half-lives (i.e., <8 hours) showed a significantly increased traffic accident risk (OR = 4.00; 95% CI, 1.31 to 12.2). Interestingly, in the latter study 14 drivers were all treated with zopiclone.[4]

For a number of reasons it is hard to draw conclusions on fitness to drive of individual patients from these population-based studies. Importantly these studies provide little information on if, when, and to what extent a drug impairs driving.[6] This is because this type of database research only compares medication distribution files with accident records of patients. It is often not possible to verify whether the drug was actually used on the day of the accident, nor are data collected on actual drug dosage, the specific type of hypnotic drug taken, its half-life, the time between intake and driving, and important accident-related issues (e.g., whether the driver was actually at fault of causing the accident or if it was due to other traffic). Also, sometimes it is unclear whether benzodiazepines were being used for hypnotic or anxiolytic purposes.[7-9] This likely has an effect on the risk for traffic accidents because sleep medication is usually taken at bedtime and anxiolytics during the day, causing a large difference between time of drug intake and driving between the two clinical use patterns. Taken together, current epidemiologic data suggest a potentially increased traffic accident risk in patients treated with hypnotic drugs. However, these data provide little information to the user or the clinician about relative risk associated with different drugs, drugs with different half-lives, and dose for a given drug.

THE ON-THE-ROAD DRIVING TEST

Because of possible ethical restrictions and legislation most countries do not allow the investigation of drug effects on driving in normal traffic. An exception is the Netherlands, where a standardized on-the-road driving test has been used over the past 30 years to examine the effects of drugs on driving.[10-12] The 100-km driving test is conducted on a public highway in normal traffic. Subjects are instructed to drive with a steady lateral position of their own choice within the right (slower) traffic lane while maintaining a constant speed of 95 km/hour. This allows them to drive along with the regular traffic flow. A driving instructor with dual controls guards the safety of the subjects, and the investigator in the back seat monitors the recording equipment. Lateral position data and mean speed are recorded two times per second. The mean lateral position (MLP) and mean speed (MS) are control variables, showing whether the subject conducted the test according to the instructions. These data further allow calculating the traditional primary parameter of the driving test, the Standard Deviation of Lateral Position (SDLP). The weaving of the car (i.e., SDLP) has proved to be an excellent measure of vehicle control. Dose-dependent driving impairment (i.e., SDLP increment) has been shown for alcohol and drugs of abuse[13] as well as for pharmacotherapeutic drugs such as antidepressants, hypnotics, anxiolytics, and antihistamines.[14] Alternatively, driving improvement has also been shown, illustrated by a reduction in SDLP, for example, in patients with attention deficit/hyperactivity disorder who were treated with methylphenidate.[15] The effect of driver sleepiness (e.g., that caused by sedative drugs) on vehicle control is illustrated in Figure 46-1.

Drivers are normally capable of keeping the car within the lane boundaries, and loss of vehicle control by drivers after taking a sedative increases SDLP values. This loss of control may also result in having out-of-lane excursions, but these are a poor predictor of vehicle control (SDLP) because they depend in great part on the choice of lateral position within the right traffic lane.[16] Drivers who choose a lateral position close to the lane boundary are much more likely to have excursions out of lane than those driving in the middle of the road,

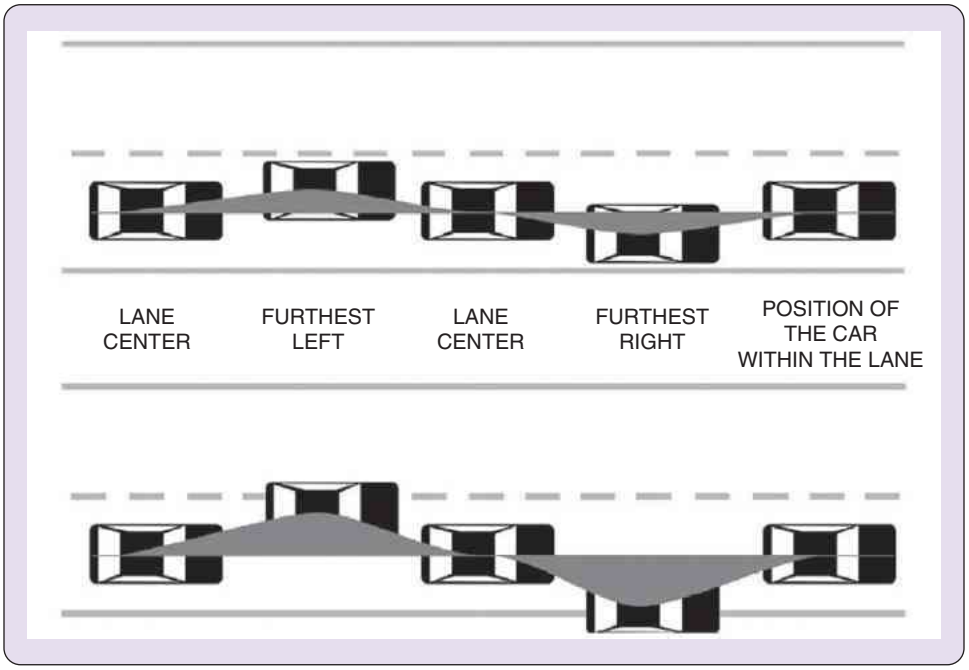


Figure 46-1 Standard Deviation of Lateral Position (SDLP). The *top figure* represents a regular SDLP under placebo condition. If weaving increases (*bottom figure*), the SDLP value becomes higher. (Modified from Verster JC, Veldhuijzen DS, Volkerts ER. Residual effects of sleep medication on driving ability. *Sleep Med Rev* 2004;8:309–25.)

independently from a possible drug effect on driving. A secondary outcome measure of the driving test, the standard deviation of speed, can also provide an indication of loss of vehicle control, but this measure has been shown to be much less sensitive than SDLP.[17]

When interpreting SDLP as an outcome measure it is important to keep in mind that the differences between drug and placebo (ΔSDLP) is the key end point. Absolute SDLP is not as sensitive because values under placebo vary significantly between individual drivers (mean, 18 to 20 cm; range, 10 to 30 cm). However, SDLP values are very stable within individuals (e.g., test-retest reliability above 0.80 have been shown).[12] Aside from demonstrating a significant difference it is also important to determine whether the impairment is clinically relevant as well as how many individuals show a clinically relevant impairment. To determine what should be regarded as clinical relevant impairment, researchers have looked at the ΔSDLP observed after administering alcohol to reach a blood alcohol concentration (BAC) of 0.05%. The driving data indicated that ΔSDLP of greater than 2.4 cm should be regarded as illustrative of clinically relevant driving impairment (Box 46-1).

In many instances in which the driving performance impairment is clinically relevant (i.e., ΔSDLP >2.4 cm), the standard deviation of speed after drug treatment does not differ significantly from placebo.[17] The same is also true for the number of excursions out of lane.[16] Therefore the standard deviation of speed and the number of excursions out of lane are generally not considered as important factors to determine whether driving is safe. Also, the number of collisions and stopped driving tests are poor indicators of a drug's effect on driving (Box 46-2).

An important aspect of the driving test is time on task. The standardized driving test takes about 1 hour to complete. This is necessary to get a good sample of the drug's effect on driving. Research has shown that shortening the driving test makes it less sensitive in showing a true difference between drug and placebo.[21] The latter is caused by the fact that in tests of short duration motivated drivers may successfully counteract impairment by investing more effort to perform the test. Vigilance decrement (i.e., increased performance impairment over distance driven or duration of driving time) is an essential characteristic of the driving test.[21] Hence, on the driving test its gets harder and harder to compensate for drug-induced impairment with increasing time on task. Increased effort sometimes can be effective for short duration (e.g., a 10-minute driving test) but does not last for the full 1 hour.

It has been suggested that on-the-road driving tests can be replaced by psychometric tests measuring driving-related skills and abilities or by driving simulators. At first this seems a safer alternative, and it would be less effortful if on-the-road driving performance and fitness to drive could be predicted by a short test battery that could be conducted at any place of choice (e.g., the physician's office). Unfortunately, comparative research showed that cognitive and psychomotor tests poorly predict on-the-road driving performance.[22,23] The primary reason for this poor correlation is the fact that driving-related skills and abilities are tested in isolation, whereas when on the road these are integrated (e.g., judgment, vision, reaction time). Importantly, overall driving performance is not simply the sum of its components (e.g., tracking, reaction speed, attention) but rather is the integration of these various skills to produce optimal safe driving. Driving simulators attempt to mimic actual driving, and these machines have become more sophisticated over the years. Whereas in the past, driving simulators were often simple computerized divided-attention tasks using a steering wheel instead of a respond box,

Box 46-1 CLINICALLY RELEVANT STANDARD DEVIATION OF LATERAL POSITION INCREMENT

There is only indirect evidence that Standard Deviation of Lateral Position (SDLP) is related to the risk for having car crashes.[18] Hence, to determine whether the magnitude of driving impairment has clinical relevance, often the comparison with impairment seen after different dosages of alcohol is made. Louwerens and colleagues examined driving performance after different dosages of alcohol.[19] The results are depicted in Figure 46-2 and show a clear dose-dependent relationship between SDLP and blood alcohol concentration (BAC).

From these historical data, it was inferred that the cutoff for clinically relevant impairment is an SDLP increment relative to placebo of +2.4 cm, corresponding to a BAC of 0.05%, which is the most commonly reported legal limit for driving a car.

Box 46-2 CRASHES AND STOPPED DRIVING TESTS

At first, the occurrence of crashes may be regarded as the ultimate evidence that a drug negatively affects driving; however, this premise can be debated. First, crashes can be caused by many factors; the effect of hypnotic drugs is only one. For example, a crash may be caused by another driver without any blame to the patient. In the on-the-road driving test, crashes do not occur because the driver is accompanied by a licensed driving instructor. If safety becomes compromised, the driving instructor intervenes and prevents a crash from happening. In driving simulators, crashes are more commonly seen. This likely has to do with increased sleepiness scores and the fact that participants know that having a crash has no real-life consequences in terms of injury or death. Crashes are infrequent and uncommon events during normal driving. Even when driving is significantly impaired, usually crashes do not occur. Therefore counting the number of crashes as an indicator of drug-induced driving impairment is not useful.

Driving tests can be stopped for many reasons. For example, the driver experiences adverse events such as stomach pain or drowsiness and requests to stop the test before completion. Alternatively, the driving instructor may abort the test if he or she feels it is unsafe to continue. In both instances, these are subjective decisions that by no means imply that driving is actually impaired. A comparative analysis of more than 7000 driving tests revealed that stopped driving tests occur both after drug treatment (4.1%) and, although to a lesser extent, in placebo conditions (0.7%).[20] Further analyses revealed that 39.6% of stopped drivers had a lower and 60.4% had a higher SDLP than 35 cm, a cutoff sometimes used to indicate unsafe driving. Because SDLP values of stopped and completed driving tests often do not significantly differ, the number of stopped tests should be regarded as a poor predictor of a drug's effect on driving performance.

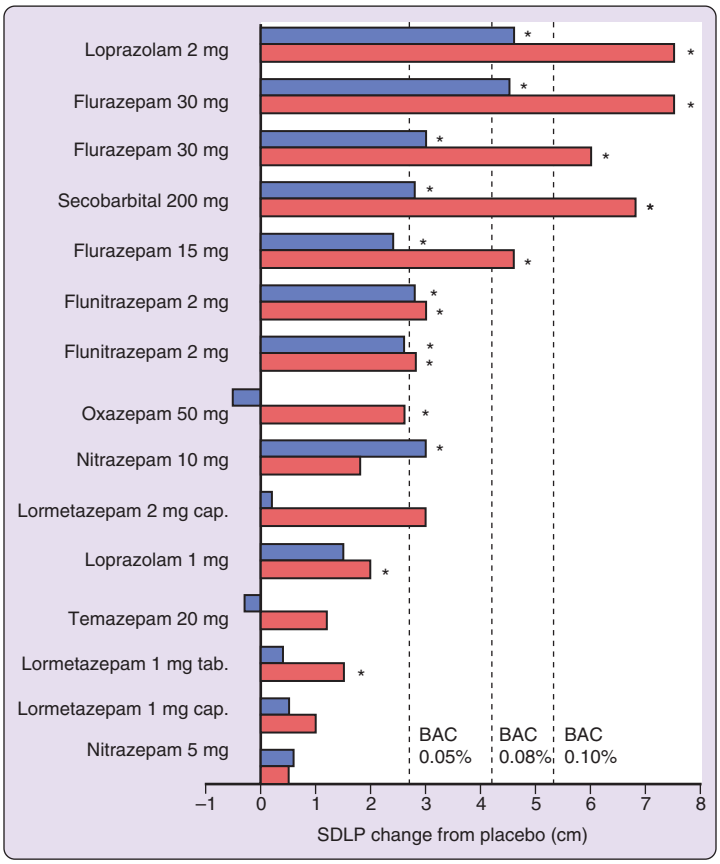


Figure 46-2 Effects of benzodiazepine hypnotics on driving. All treatments were administered at bedtime. Driving tests were performed in the morning, 9 to 10 hours after intake (*red bars*), and the afternoon, 16 to 17 hours after intake (*blue bars*), representing the times many people drive to and from work. Significant differences from placebo ($P < .05$) are indicated by an asterisk (*). (Modified from Verster JC, Veldhuijzen DS, Volkerts ER. Residual effects of sleep medication on driving ability. *Sleep Med Rev* 2004;8:309–25.)

nowadays real cars including car motion and sound are used with realistic scenery projected on a large surrounding screen, including other traffic. Few researchers have directly compared on-the-road and simulated driving. These studies found that SDLP values and sleepiness scores are generally significantly higher in driving simulators.[24] The difference between driving simulator environments and actual driving in terms of risks for having an accident is essential and may account for the observed differences. In many simulators, number of accidents is used as an end point. However, having an accident in a simulator has no consequences in terms of injury or death, whereas these risks are evident during on-the-road driving. Hence a number of people may regard driving in the simulator as a game and thus have a different mindset compared with on-the-road driving. Nevertheless, driving simulators and psychometric tests are useful to determine fitness to drive in general and to examine performance impairment. The decision about whether driving is safe should always be based on the overall available evidence gathered with different research methodologies and thus should include performance on both cognitive and psychomotor tests, driving simulators, and an on-the-road driving test.

The purpose of hypnotic drugs is to make you fall asleep and maintain sleep. It is critical, however, that after waking up 7 to 8 hours later, people who use these drugs are not sedated and can participate safely in activities of daily living such as driving. The next sections of this chapter summarize findings from on-the-road driving studies examining the effects of hypnotic drugs on driving.

EFFECTS OF HYPNOTICS ON DRIVING

The usual design of studies examining the effects of hypnotic drugs on driving performance is a double-blind placebo- and active drug–controlled trial. Treatments are administered at bedtime, and driving tests are typically conducted the following morning and sometimes afternoon (about 9 and 16 hours after intake), occasions that are coincidental with the times people usually drive to and from work. Since the 1980s, the effects on driving of a great number of benzodiazepine and nonbenzodiazepine hypnotic drugs have been examined by applying the standardized on-the-road driving test.[25,26] An overview of the results for benzodiazepine hypnotics is given in Figure 46-3.

It is evident from Figure 46-3 that benzodiazepine hypnotics significantly impair next-morning driving. In the afternoon, impairment is less pronounced, but for several drugs the magnitude of impairment (SDLP increment relative to placebo) still is higher than that seen with a BAC of 0.05%. A recent meta-analysis of these data revealed that driving

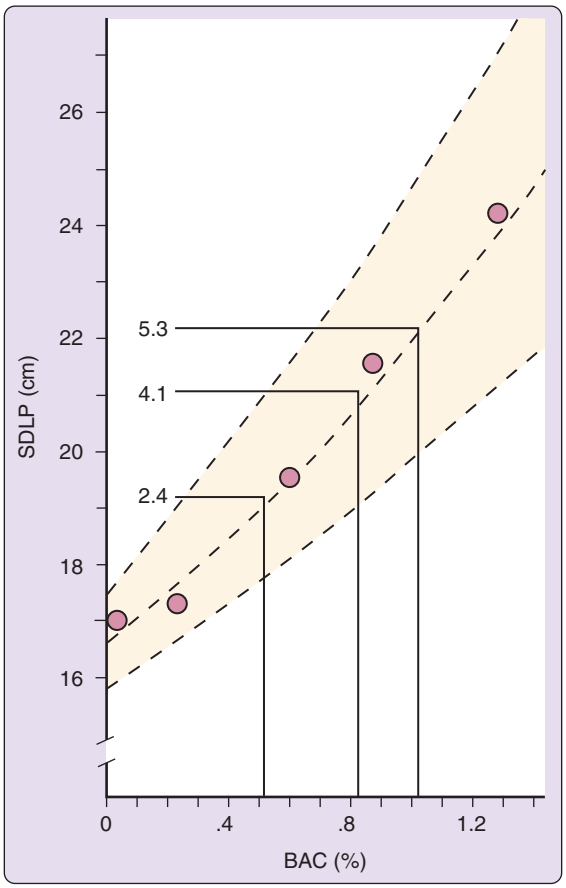


Figure 46-3 Standard Deviation of Lateral Position (SDLP) relative to baseline (no alcohol) at different breath alcohol concentration (BAC) levels.

impairment is dose dependent and is more pronounced with drugs with a longer half-life and when the time between drug intake and driving is shortened.[26] Both intermediate- (6 to 12 hours) and long-acting (>12 hours) drugs cause significant impairment the morning after bedtime administration, whereas short-acting hypnotics (<6 hours) generally do not. Interestingly, interindividual blood drug concentrations at a given time point correlate poorly with an individual's driving impairment.[27] For some benzodiazepine hypnotics, sex differences were seen in the magnitude of driving impairment,[28] emphasizing the importance of including both male and female participants in driving studies. It should be noted that the effects shown in Figure 46-3 were all found after one or two nights of treatment. Up to now, long-term effects of the use of benzodiazepine hypnotics on driving performance have not been extensively examined using the on-the-road driving test. However, given the epidemiologic evidence discussed in this chapter and driving data from studies examining long-term benzodiazepine anxiolytics use (e.g., diazepam for 4 weeks),[29] it can be assumed that at least partial tolerance to the impairing effects can be expected if the drugs are used on a nightly basis over a period of time. However, the rate the development of this tolerance is not currently defined. Importantly, the mechanism that mediates this tolerance has not been investigated. For drugs with a half-life greater than 15 hours, impairment may worsen because of drug accumulation. As a result of the lack of clear data, there is currently no consensus among sleep experts as to whether and when driving is safe after initiating treatment with medications known to impair driving acutely.[30]

The "z-drugs" (i.e., zopiclone, eszopiclone, zolpidem, and zaleplon) also act at the benzodiazepine receptor of the gamma-aminobutyric acid A ($GABA_A$) complex but do so more specifically and have a relatively shorter half-life. Hence, it was anticipated at their introduction that these drugs would be devoid of the next-morning adverse effects seen with benzodiazepine hypnotics. Zopiclone, the first of the z-drugs introduced and commonly prescribed in Europe, is one of the most frequently studied drugs used in hypnotic clinical studies. Bedtime administration of zopiclone (7.5 mg) consistently results in next-morning driving impairment. The magnitude of impairment is about +2 to +3 cm, roughly comparable to that seen with a BAC of 0.05%.[31] For this reason, zopiclone is typically used as positive control in on-the-road hypnotic driving studies. In the afternoon, however, driving after zopiclone (7.5 mg) is not impaired. In contrast, bedtime administration of zolpidem (10 mg) or zaleplon (10 mg) does not impair next-morning driving.[32,33] This is attributable to the short (<3 hours) half-life of these two drugs.

MIDDLE-OF-THE-NIGHT ADMINISTRATION

Because sleep maintenance problems are commonly reported by patients with insomnia, treatments enabling patients to fall asleep more rapidly after middle-of-the-night (MOTN) awakenings have been developed, to be taken in the middle of the night. In addition, many drugs (e.g., zolpidem) are taken off label in the middle of the night. Thus the effects of hypnotics taken in the middle of the night need to be investigated in terms of driving performance the next day. To date, four such on-the-road driving studies have been conducted.[34] In these studies, treatments were administered during the night, 4 to 6 hours before the driving test. Driving performance after MOTN administration of traditional benzodiazepine hypnotics has not been examined, presumably because they already impair next-morning driving after bedtime administration. Zolpidem (10 mg and 20 mg, oral immediate-release tablets) significantly impaired driving in a dose-dependent manner when tested 4 hours after MOTN administration.[35] Also, gaboxadol (15 mg) and zopiclone (7.5 mg) significantly impaired next-morning driving after MOTN administration.[36] In contrast, buffered sublingual zolpidem (3.5 mg) and zaleplon (10 mg and 20 mg) did not significantly affect driving 4 hours after MOTN administration.[36-38]

NON–GAMMA-AMINOBUTYRIC ACID HYPNOTICS

In the search for hypnotic drugs without next-day sedation, development has focused on drugs that do not act at the $GABA_A$ receptor complex. In this context, histamine-1 and orexin receptor antagonists are under investigation. Also, melatonin agonists are considered. For example, the effects on driving of ramelteon (8 mg), a melatonin receptor agonist, have been investigated.[39] Significant driving impairment was observed 8.5 hours after bedtime administration. Significant next-day impairment was also found on reaction time in the Sternberg Memory Scanning Test, reaction speed and

tracking in a divided attention test, and delayed recall in a word learning test. No significant impairment was found on the Digit Symbol Substitution Test and a balance test, which was performed during the night, 2 hours after treatment administration. The magnitude of performance impairment seen with ramelteon (8 mg) was comparable to that of zopiclone (7.5 mg). This is an important finding because ramelteon has a short half-life. Thus the question arises as to the mechanism of this impairment. The two possibilities are that the impairment is due the effect of a long-acting ramelteon metabolite or the effect of ramelteon on shifting circadian phase.

FUTURE DIRECTIONS

Although the effects of drugs on mean SDLP has been the standard measure of impaired driving, the nature of the risk and extent of the risk are not fully defined with this single analytic approach. In the recent past, two modifications to the traditional analysis of on-the-road driving have been investigated. The first is the use of alternate end points to SDLP. Lapses have traditionally been used to assess the impairing effects of sleep deprivation on laboratory-based performance. Recently, lapses have been introduced as an outcome measure of the on-the-road driving test.[40] A lapse in driving is defined as a continuous change in lateral position of greater than 100 cm, lasting for at least 8 seconds. In contrast to weaving (SDLP), a unique feature of lapses is that they occur during short periods of inattention. That is, the presence or absence of lapses may differentiate drivers who are aware of driving impairment from those who are not aware of loss of vehicle control (SDLP increment). If correct, having a lapse may have serious consequences in terms of traffic safety because this period of inattention may increase the risk for having a sleepiness-related accident. Moving forward, research using lapses as an outcome measure is needed to determine the degree of overlap between lapses and SDLP and to what extent lapses and SDLP provide unique information regarding impaired driving and traffic accident risks.

The second innovation involves an alternate method to analyze SDLP data. Although mean SDLP contrasts (drug vs. placebo) provide useful information, they do not address the primary issue of putting individuals at increased risk for having a traffic accident. A large sample size or small variance can lead to statistically significant mean differences that do not correspond to meaningful driver impairment (i.e., nonclinically relevant SDLP increments, less than +2.4 cm). A small sample size or large variance can result in failure to find a difference that in fact corresponds to increased accident risk. Indeed, a large variance may be the result of outlying individuals with impaired driving skills who are the very group of interest. As is the case with all safety data parameters, it is more important to discover whether a treatment produces a large effect in a subset of subjects than whether it produces a relatively clinically meaningless shift across the entire sample. This problem can be addressed by a responder analysis that assesses the proportion of patients on drug versus placebo who exceed a predetermined threshold for clinically meaningful impairment or other thresholds, larger and smaller, that are of interest in understanding the degree of impairment. The statistical test used for such an analysis has been called symmetry analysis because it tests whether the distribution of changes (drug minus placebo) above the threshold and below the threshold is symmetrical around zero.

CLINICAL PEARLS

- Drug type, dosage, time of driving after drug intake, drug half-life, and patient characteristics all have a significant effect on the magnitude of next-day driving impairment.
- Impairment is more pronounced when time between drug intake and driving is shorter, with higher drug dosages, and with drugs with a longer half-life.
- More research is needed to determine effects of chronic drug use on driving and to explore gender, age, and disease differences associated with these effects.

SUMMARY

The World Health Organization has identified traffic accidents as one of the major causes of injury and death around the world. An important factor contributing to traffic accidents is inattention of the driver due to reduced alertness or increased sleepiness. It is therefore important to understand the effects of sedating drugs on driving and their impact on the risk for crashes. A standardized method to examine ability to drive is the on-the-road driving test. Results from 30 years of Dutch on-the-road driving research have demonstrated that some hypnotic drugs are safe whereas others impair driving performance, thereby influencing drug labeling. In addition, differences between drugs in degree of impairment also vary as a function of dosage, half-life, and time since drug ingestion, demonstrating the importance of treatment compliance with directions for use of the drug to ensure driving safety.

DISCLAIMER

Although the information presented in this chapter has been gathered and evaluated with great care, the authors will not accept any liability after use of the information by patients taking the medicines discussed. Patients should always consult their physician concerning whether or not it is safe to drive a car.

DISCLOSURE OF INTEREST

Joris Verster has received grants and research support from The Dutch Ministry of Infrastructure and the Environment, Janssen Research and Development, Takeda, and Red Bull and has acted as a consultant for Canadian Beverage Association, Centraal Bureau Drogisterijbedrijven, Coleman Frost, Deenox, Eisai, Purdue Pharma, Red Bull, Sanofi-Aventis, Sepracor, Takeda, Transcept, and Trimbos Institute.

Thomas Roth has received grants and research support from Aventis, Cephalon, GlaxoSmithKline, Neurocrine, Pfizer, Sanofi, Schering-Plough, Sepracor, Somaxon, Syrex, Takeda, TransOral, Wyeth, and Xenoport. He has consulted for Abbott, Acadia, Acoglix, Actelion, Alchemers, Alza,

Ancil, Arena, AstraZeneca, Aventis, AVER, BMS, BTG, Cephalon, Cypress, Dove, Elan, Eli Lilly, Evotec, Forest, GlaxoSmithKline, Hypnion, Impax, Intec, Intra-Cellular, Jazz, Johnson & Johnson, King, Lundbeck, McNeil, MediciNova, Merck, Neurim, Neurocrine, Neurogen, Novartis, Orexo, Organon, Prestwick, Procter & Gamble, Pfizer, Purdue, Resteva, Roche, Sanofi, Schering-Plough, Sepracor, Servier, Shire, Somaxon, Syrex, Takeda, TransOral, Vanda, Vivometrics, Wyeth, Yamanuchi, and Xenoport.

Selected Readings

Af Wåhlberg AE. Some methodological deficiencies in studies on traffic accident predictors. *Accid Anal Prev* 2002;**859**:1–14.

Barbone F, McMahon AD, Davey PG, et al. Association of road traffic accidents with benzodiazepine use. *Lancet* 1998;**352**:1331–6.

Helland A, Jenssen GD, Lervåg LE, et al. Evaluation of measures of impairment in real and simulated driving: results from a randomized, placebo-controlled study. *Traffic Inj Prev* 2016. [In press].

Kriikku P, Hurme H, Wilhelm L, et al. Sedative-hypnotics are widely abused by drivers apprehended for driving under the influence of drugs. *Ther Drug Monit* 2015;**37**(3):339–46.

Neutel CI. Risk of traffic accident injury after a prescription for a benzodiazepine. *Ann Epidemiol* 1995;**5**:239–44.

Roth T, Eklov SD, Drake CL, Verster JC. Meta-analysis of on the road experimental studies of hypnotics: effects of time after intake, dose, and half-life. *Traffic Inj Prev* 2014;**15**(5):439–45.

Verster JC, Bervoets AC, de Klerk S, Roth T. Lapses of attention as outcome measure of the on-the-road driving test. *Psychopharmacology (Berl)* 2014;**231**:283–92.

Verster JC, Roth T. Standard operation procedures for conducting the on-the-road driving test, and measurement of the Standard Deviation of Lateral Position (SDLP). *Int J Gen Med* 2011;**4**:359–71.

Verster JC, Roth T. Thirty years of Dutch on-the-road driving research: strengths and limitations of the highway driving test and future challenges. *Curr Psychopharmacol* 2012;**1**(1):97–102.

Verster JC, Roth T. Gender differences in highway driving performance after administration of sleep medication: a review of the literature. *Traffic Inj Prev* 2012;**13**(3):286–92.

Verster JC, Roth T. Blood drug concentrations of benzodiazepines correlate poorly with actual driving impairment. *Sleep Med Rev* 2013;**17**(2):153–9.

Verster JC, Roth T. Vigilance decrement during the on-the-road driving tests: the importance of time-on-task in psychopharmacological research. *Accid Anal Prev* 2013;**58**:244–8.

Verster JC, Roth T. Excursions out-of-lane versus Standard Deviation of Lateral Position as outcome measure of the on-the-road driving test. *Hum Psychopharmacol* 2014;**29**(4):322–9.

Verster JC, Veldhuijzen DS, Volkerts ER. Residual effects of sleep medication on driving ability. *Sleep Med Rev* 2004;**8**:309–25.

A complete reference list can be found online at ExpertConsult.com.

Section 7

Psychobiology and Dreaming

Chapter 47

Introduction

Robert Stickgold

The study of dreams and dreaming has made remarkably slow progress over the past century. In defense of this slow progress, it is worth noting that it shares with the study of consciousness several unenviable features: (1) it is only poorly defined, and there is no consensus on a definition; (2) it has no clear physiologic correlates; and, hence (3) its biologic basis is largely unknown; (4) it has no behavioral correlates; and, hence (5) has no known function; and, finally, (6) its existence in species other than humans is a matter of conjecture. In short, we do not know the cause, nature, or consequence of dreaming, a remarkably sad state of affairs! What is, however, absolutely clear is that dreaming is a highly robust human occurrence that is most likely experienced by most humans every night.

DEFINITION

Given our ignorance of underlying biologic mechanisms and functions, it is perhaps not surprising that a clear, agreed-on definition of dreaming has evaded us. Most definitions in science refer either to physical objects to or processes that are defined by their underlying mechanisms or measurable consequences. We have none of these for dreaming. Indeed, a discussion group of the Association of Professional Sleep Societies and the Association for the Study of Dreams concluded that a "single definition for dreaming is most likely impossible given . . . the diversity in currently applied definitions."[1] What all discussants did agree on is that its definition must be phenomenologic; that is, dreaming refers to a state of mind. Although most discussants argued that dreaming should be restricted to mentation occurring during sleep, others argued that daydreaming should be included. Some demanded hallucinatory perceptions, thoughts, and emotions that are altered from normal, but others required one but not another, and some included any reported mental activity during sleep, even something as simple as a report of "I was wondering when you would wake me up."

For these reasons, some authors have eschewed the use of the term altogether, talking instead about *sleep mentation*. Unlike the term *dreaming*, this phrase can be simply defined. Sleep mentation is defined as (1) all mental experiences—perceptions, thoughts, and emotions—occurring during sleep, or (2) the process of experiencing perceptions, thoughts, and emotions during sleep. Thus it combines the sense of the words *dream* and *dreaming*. It is this concept, sleep mentation, that is more properly the subject of this section on dreaming.

METHODOLOGY

It is important to understand that researchers do not study dreams directly. Rather, they study dream reports, written, dictated, or on rare occasions, drawn or acted out, after the subject has awakened from sleep, along with their physiologic and psychological correlates.

In one sense, this is not as big a problem as it is often considered. All scientific measurements are derivative. We measure blood pressure with the aid of a pressure cuff applied externally, listening with a stethoscope for the sounds produced by the movement of blood through the brachial artery, depending on our conscious perception of the moment at

which the intensity of these sounds drops below our threshold of perception. Yet no one questions the legitimacy of the technique. It certainly is not accurate to more than 2 mm Hg at best, but it is good enough.

The difference between dream reports and blood pressure measurements is in the concept of "good enough." Researchers have confirmed the validity of using pressure cuffs by monitoring blood pressure with indwelling catheters in parallel with pressure cuff measurements and confirming the reliability of the pressure cuff technique. There is no way to tell whether a given dream report, or dream reports in general, are good enough, and it is not clear that it will ever be possible to record dreams (or any other conscious experiences) with the same level of confidence and reliability as exists for most other measurements in the biologic sciences.

If we make the presumption that dream reports are a good enough representation of the mental experience during the prior period of sleep, the next question is how our data collection procedures affect the data. Table 47-1 outlines some of the parameters of dream collection that vary from one study to the next.

For each of these parameters, changes in the methodology of report collection are known to affect report content, with differences most notable in report frequency, length, bizarreness, and emotional content. But arguments continue in most cases over whether these differences reflect differences in dreaming or just in accuracy and completeness of the dream reports. Even in the case of laboratory awakenings from rapid eye movement (REM) and non–rapid eye movement (NREM) sleep, arguments continue over how much of the difference seen in reports from the two sleep stages reflects diminished dreaming, as opposed to diminished recall, in NREM. Because a given pair of studies can conceivably differ on all nine of these parameters (allowing more than 3000 distinct protocols), it is not surprising that disputes over report frequencies, lengths, and features are as common as they are.

Table 47-1 Parameters of Dream Collection

Parameter	Common Values
Location	Home (monitored or unmonitored) Laboratory (monitored)
Awakening	Spontaneous Evoked
Timing	Morning Nocturnal
Sleep stage	REM N1 N2 N3
Time in stage	Constant Random Spontaneous Later in stage (esp. REM) later in the night
Time of night	Constant Variable
Nights	One Several (consecutive, fixed schedule, or ad libitum)
Report style	Written Audio recording
Probes	None Fixed Semistructured

DESCRIPTION

A proper description of dreaming (defined now as sleep mentation) should describe the nature of the perceptions, thoughts, and emotions experienced by the dreamer and how these change with condition. Several chapters in this section focus on this question. Unfortunately, dreaming remains incompletely characterized even at the descriptive level. Most studies of dreaming focus on dream perceptions— the highly visual, narratively complex, and delusional hallucinations that we all think of when we think of dreaming. A smaller body of research has looked at the emotions of dreams, and a much smaller body has looked at the thoughts of dreams.

Even in the area of dream perceptions, what is studied varies dramatically: from word counts to counts of characters and objects; to classification of characters, objects, and actions as familiar or novel; to identification of bizarreness (on any of a number of ad hoc scales); to measures of "latent content." Thus it is not uncommon for two studies of "dream content" to lack a single outcome measure reported in both.

Individual and group differences are also poorly understood. Aside from clinical populations and very limited case studies (often from a psychoanalytic perspective), the best comparative data relate to dreaming in children and to the question of whether some individuals dream at all.

Probably the greatest amount of investigation of variability in dreaming relates to the question of REM versus NREM dreaming, and even here huge disagreements exist. At this time, all that can be safely stated is that the two extreme positions, namely that dreaming occurs only during REM or, conversely, that there is no difference between REM and NREM dreaming, no longer appear to have any champions. But between these two extremes, almost every imaginable niche is occupied.

MECHANISMS

Efforts to explain the mechanisms that produce dreams date back to antiquity, when gods and indigestion were major contenders for the origins of dreams. Over the past 150 years, efforts at finding a brain or mind basis for dreaming have progressed slowly. At the moment, one can discern several schools of thought, which generally fall into the fields of neurophysiology, psychology, psychoanalysis, and, more recently, cognitive neuroscience. At their best, each of these takes an inclusive view, acknowledging the contributions of the other fields but focusing on their own. At their worst, they reject the usefulness of each other, with the usefulness of at least neurophysiology and psychoanalysis explicitly rejected by some groups. All of these, except psychoanalysis, are represented in the chapters that follow.

FUNCTIONS

Theories of the function of dreaming often are unclear as to what is actually being discussed, whether it is the

phenomenologic experience of dreaming or the biologic processes that underlie it. This is only important because some researchers state their belief that dreaming per se has no function, but that the biologic processes that underlie it have very important functions. For example, they would argue that the activation of neural networks within association cortex during REM might lead to important, long-term changes in network connectivity and may also fleetingly create a dream, but the dream and the process of dreaming itself are irrelevant to the production of these cortical modifications. Other researchers believe that the phenomenologic dream experience (with or without subsequent waking recall) is the critical element in the functionality of the dream and that the underlying biology is important only insofar as it is necessary to create the dream experience. Still others find the distinction unimportant or even meaningless.

This should not be surprising. It exactly parallels discussions of the functionality of consciousness. Philosophers and neurobiologists alike struggle to understand how to even think and talk about such functionality. Some reject any such functionality, others take it as a given, and others believe the distinction reflects a poorly formed question, that consciousness cannot be separated from its underlying brain basis. For the purpose of this section, we will collapse the two questions, talking about the function of dreaming and its underlying brain mechanisms as if they were interchangeable. While not claiming that they actually are interchangeable, it appears safe to say that we do not know how to measure their effects independently.

Theories of the function of dreaming (or lack thereof) go hand in hand with the approaches taken to studying the mechanisms underlying dreaming. Thus physiologic models tend to see little function in dreaming, whereas psychoanalytic models ascribe highly complex functions to dreaming. Psychological and cognitive neuroscience models tend to lie somewhere in between, ascribing various levels of complexity to dream function.

SUMMARY

In the end, we are left with a field that is very much a work in progress. There is no more than a hint of an even close to complete brain-based model for dream construction, and there is limited experimental evidence that dreaming plays any functional role. With that being said, the past 10 years have shown a resurgence of interest and research on dreaming, and brain imaging studies during sleep have fueled the development of new models of dream construction and function. It is hoped that this section will provide a snapshot of the state of the field at this time, a warning of the questions unanswered, and a guide to how dream research will progress over the next decade.

A complete reference list can be found online at ExpertConsult.com.

Chapter 48

Why We Dream

Robert Stickgold; Erin J. Wamsley

Chapter Highlights

- Sixty years after the discovery of rapid eye movement sleep launched the field of sleep research, we still have surprisingly little insight into the most fundamental question about the sleeping mind: why do we dream?
- In this chapter, we focus on data suggesting that dreaming reflects the activity of brain mechanisms that perform off-line memory consolidation during sleep. We then describe studies suggesting that a multiplicity of memory functions are reflected in the formal properties and content of conscious experience across stages of sleep and wakefulness.
- Together, these findings suggest both a mechanism and function for dreaming.

The search for an understanding of dreaming is thousands of years old. During the past 100 years, three publications have formed the core of most of the scientific discussion of this question: Freud's *Interpretation of Dreams,*[1] published at the end of the 19th century; the report of a correlation between dreaming and the newly discovered rapid eye movement (REM) sleep,[2] in the 1950s; and the proposal of the activation-synthesis model of dreams[3] in the 1970s, which argued that dreaming is initiated by random neural activity in the brainstem during REM sleep. Yet now, in the twenty-first century, there is precious little on which dream researchers agree.

One relatively new approach has been to consider dreaming within a larger neurocognitive framework of off-line memory consolidation during sleep. The rationale of this approach is that dreaming reflects the activity of the brain and that this activity necessarily includes the reactivation of memories and emotions from earlier experiences. Because any neural activity in the brain invariably leaves the activated networks altered, dreaming must modify the networks storing memories and emotions. Dreaming then becomes the conscious experiencing of these activated networks in the process of being modified.

BRAIN ACTIVITY DURING SLEEP

To understand how dreams might be produced and the functions they may serve, it is important to understand how brain activity in REM and non–rapid eye movement (NREM) sleep differs from that in wakefulness. During the past decades, neuroimaging studies have begun to describe how patterns of human brain activity differ across the states of waking, REM sleep, and NREM sleep.[4] In the 1990s, positron emission tomography (PET) studies initially demonstrated that whereas regional cerebral blood flow is generally decreased during slow wave sleep, entry into REM sleep leads to reactivation of some regions along with further deactivation of others.[5-7] The pattern seen suggests a shift in global brain function in REM sleep away from conscious executive control (further decrease in dorsolateral prefrontal cortex activity) and toward hallucinatory (increased activity in sensory association cortices) and emotional (increased amygdala, anterior cingulate, and medial orbitofrontal cortex activity) processing, which might relate to the features of REM dreaming.[6,7] Yet dreams are also commonly recalled from NREM sleep as well. Given this, the fundamental neural substrate for dream experience must be common to all stages of sleep.

More recent imaging studies stressed that activation remains relatively high in certain regions during NREM when controlling for the global decrease in activity, including within several memory-related areas.[8,9] Relative hippocampal activation, for example, has been reported to peak during slow wave sleep, exceeding even that seen during wakefulness.[9] Functional magnetic resonance imaging studies (with vastly superior temporal resolution compared with PET) confirm that NREM sleep is not a homogenous state of "decreased activation"; during NREM sleep, there are in fact brief, transient increases in local brain activation during sleep spindles and slow waves.[4] Fast spindles (>13 Hz) in particular have been associated with increased activity in the hippocampus and medial prefrontal cortex. Both fast spindles[10] and these specific brain regions[11] have been implicated in sleep-dependent memory consolidation.

In both humans and rats, patterns of brain activity observed during waking appear to be "replayed" during NREM sleep (for the most part; one study has reported a similar effect during REM[12]). In rats, recordings of single-unit activity demonstrate that sequences of neuronal firing seen as rats explore an environment are later statistically reiterated during sleep.[13,14] First observed in the hippocampus, this effect also occurs in cortical regions.[15,16] Meanwhile, in humans, PET studies have shown that brain regions activated during the learning of a task are selectively reactivated during the next night's sleep, both in REM and NREM sleep.[9,17] This reactivation of patterned activity in the sleeping brain is thought to support the consolidation of memory during sleep. Indeed, reactivation has recently been reported to promote the

formation of dendritic spines,[18] and disrupting reactivation impairs learning.[19] This discovery, that brain networks involved in prior waking experience are reactivated during later sleep, has been paralleled by a growing body of evidence for sleep-dependent memory consolidation.

SLEEP-DEPENDENT MEMORY CONSOLIDATION

One interpretation of neural changes during sleep is that they reflect a *homeostatic* process working to restore the brain to its state at the start of the previous day. These "rest" or "restorative" models of sleep argue that sleep serves to reverse deleterious changes that inevitably accrue across the day. A current example of such theories is the synaptic homeostasis model of sleep function.[20,21] In contrast, a more powerful, *progressive* model suggests that these changes reflect *off-line processing* of information obtained during the prior day—consolidating, integrating, and sometimes even reversing changes that occurred during waking. Dreaming, as well, might serve either a *homeostatic* or *progressive* function. Freud, for example, proposed a restorative model of dreaming, specifically considering but then rejecting any progressive model.[1] In contrast, others have questioned[22] or rejected[23] the notion of any sort of evolutionary function for dreams.

Although the question of the function of dreaming remains unresolved, there is a growing consensus that sleep serves a function of off-line memory consolidation (for reviews, see Diekelmann and Born[24] and Stickgold[25]). Sleep has been shown to enhance prior learning of perceptual and motor skills,[26,27] paired word associates,[28,29] and emotional memories,[30,31] and even to enhance insight[32] and creativity.[33,34] As with dreaming, memory consolidation appears to vary across sleep stages. In humans, hippocampus-dependent memory has most strongly been associated with slow wave sleep.[9,24,35–37] In contrast, other forms of memory have variously been associated with REM sleep (emotional memory[31,38,39]) or stage 2 sleep (motor learning[27,40]). The particular model of dream function presented here, then, proposes that dreaming relates to this memory function of sleep, participating in, or at least reflecting, the processing of memories for recent daytime experiences.

SLEEP STAGES AND DREAM CONTENT

If memory processing is differentially activated across sleep stages, and dreaming at least parallels and possibly contributes to these memory processes, one would expect to see changes in the content of dream reports collected from different sleep states. In its simplest form, this is exactly what is seen. Reports are more frequent after awakenings from REM sleep,[2,22,41,42] and both REM and NREM sleep reports are more common when awakenings occur later in the night.[42] Reports obtained from REM tend to be longer, more vivid, more storylike, and more bizarre than NREM reports.[22,42] Whereas it has been suggested that some of these differences merely reflect poorer recall after NREM awakenings, little objective evidence supports such a claim, and other REM and NREM differences are not amenable to such an explanation. For example, hallucinations are more prevalent in reports from REM sleep, whereas directed thinking is more common in NREM sleep, a pattern that cannot be explained simply by poorer recall from one stage or another.[43]

Even at this level of analysis, there is the suggestion of homology between dream content and memory function in sleep. In REM sleep, dreams are hallucinatory, emotional, and narrative, with frequent fictive movements (for review, see Hobson and colleagues[22]). Congruently, REM sleep is thought to facilitate consolidation of visual perceptual and emotional memories.[31,38,39] In contrast, NREM sleep (particularly slow wave sleep) has been associated with sleep-dependent improvement on a range of hippocampus-dependent tasks, including the memorization of declarative information[28,29,35] and navigation through spatial environments.[9] Paralleling this mnemonic function, dream reports from NREM sleep tend to be more "realistic" and draw more on material from recent episodic memory.[44] There are notable exceptions (e.g., improvement on a motor task correlates with stage 2 sleep,[27] but REM sleep dreams are decidedly more motoric), but the argument can be made that differences in memory functions across sleep stages are reflected in differences in dream content.

That the mechanisms underlying REM and NREM dreaming may be qualitatively different can be inferred from REM behavior disorder.[45] In this disorder, the inhibition of movement that produces the atonia of REM breaks down, and as a result, patients physically act out their dreams, often with violent consequences. In contrast, no such atonia exists in NREM sleep. Yet there is no equivalent acting out of dream content, suggesting that the neural mechanisms underlying NREM dreaming differ from those of REM dreaming in at least the engagement of motor systems.

DREAMS AND MEMORY SYSTEMS

Most models of dreaming implicitly assume that dreams are constructed from our memories, but they also recognize that this construction need not involve the transparent, direct incorporation of specific memories into the dream scenario. Freud, for example, emphasized that actual memories, events, and their associated emotions underwent "condensation" and "displacement" before appearing in dreams. Whereas he elaborated a complex theory of "dream work" to explain these alterations, modern cognitive neuroscience allows the formation of much simpler, evidence-based explanations. Specifically, dream construction does not include the veridical replay of complete "episodic" memories, whose recall is normally mediated by the hippocampus in waking life. Instead, dreams appear to be constructed from unbound *fragments* of various recent episodes,[46] intermixed with remote memories, semantic memories (facts and general information), and representational memories (sensorimotor images), all stored in and directly accessible in the neocortex.

Within dream content, the distribution of these various types of memory sources appears to differ by sleep stage. Subjects identify episodic memory sources for dream elements more frequently after awakenings from NREM sleep (including sleep onset) than after REM awakenings.[44] This differential rate parallels the decline in directed thinking reported as subjects move from sleep onset to stage 2 NREM sleep and then into REM sleep.[43] At the same time, the frequency of "generic semantic memory sources" is greatest in REM sleep.[44]

Episodic memories are clearly a *source* for the construction of dreams in both REM and NREM sleep, but these episodes are not "replayed" during dreaming in their original form. In

one study, subjects identified waking antecedents for 364 dream elements in 299 dream reports. When the extent of congruence between these dreams and their purported waking sources were analyzed, only 3% of dream elements were judged to have the same location, characters, and actions as the identified waking memory source.[46] For this small percentage of reports, dream content may reflect the same hippocampus-mediated process of episodic recall that characterizes waking recall, allowing identification of particular episodes as the source of their dream elements. What, then, is the basis of this association in the remaining 97% of dreams? The features that showed highest congruence between dream element and putative waking source were theme, emotion, and characters.[46] Whereas the last of these is consistent with the reiteration of a specific episode, the first two are decidedly not. Thematic congruence might instead reflect the activity of nonhippocampal semantic memory systems. As discussed later, evidence from the dreams of amnesiac patients also suggests that the hippocampus is not required for the brain to generate dreams related to recent experience. At the same time, there is increasingly strong evidence that the hippocampal memory system is active during sleep[9,15,15,47,48] and may contribute to dreaming in healthy individuals.[49] Below we turn our attention to the activity of hippocampus-dependent declarative memory systems in sleep.

DREAMING AND DECLARATIVE MEMORY CONSOLIDATION DURING SLEEP

A mounting body of evidence suggests that sleep is critically involved in a wide range of types of memory consolidation, ranging from the consolidation of simple visual and motor skills[26,27,40,50] to the consolidation, integration, and extraction of complex, hippocampus-dependent declarative memories.[28,29,47,48,51-53]

Several studies have employed the "paired associate" learning paradigm to assess the contribution of sleep to declarative memory performance. In this task, subjects are presented with a series of word pairs, and after the entire list has been presented, they are shown the first word of each pair and asked to recall the second. An early study by Plihal and Born[29] suggested that the deeper, slow wave sleep of NREM sleep early in the night is particularly beneficial for the consolidation of such word pair memory compared with the benefits of wake or late-night, REM-rich sleep. Subsequent studies have consistently confirmed that a period of sleep, relative to wakefulness, benefits the retention of this type of memory[11,28,53] and that experimental enhancement of slow wave activity enhances word pair recall.[36,54]

Although most of the literature on sleep and declarative memory focuses on the memorization of verbal or visual stimuli, other research highlights the equally important role of sleep in the *reorganization and transformation* of declarative memories across time. For example, several studies have now reported that sleep aids in the development of transitive inference in a paradigm in which subjects are asked to make judgments extending beyond material learned before sleep.[52,55,56] Payne and colleagues[57] have meanwhile observed that sleep facilitates extraction of the general theme of semantically related word lists. Other studies demonstrate that sleep enhances hippocampus-dependent spatial memory.[47,51,58,59]

In sum, a rapidly accumulating body of research demonstrates that complex, hippocampus-dependent learning is processed during human sleep. Might this reactivation and transformation of declarative memory be expressed within the content of dreams? In partial answer to this question, we later discuss evidence that intensive, engaging learning experiences *are* reliably expressed in the content of dreams during post-training sleep.

INCORPORATION OF WAKING EVENTS INTO DREAMS

Although many of the studies of dreaming described earlier reflect the incorporation of waking events into dreams, few involved experimental manipulations of waking events with the goal of influencing subsequent dream content (for review, see Wamsley and colleagues[60]). A newer line of research investigating this process has focused on hypnagogic dreams, which occur at sleep onset. In the first of these studies,[61] subjects spent 2 to 3 hours per day playing the video game *Tetris* across 2 or 3 days. Three groups of subjects were studied: 12 subjects with no prior *Tetris* experience (novices), 10 with extensive *Tetris* experience (experts), and 5 dense amnesiacs with extensive medial temporal lobe damage resulting from anoxia or encephalitis (amnesiacs). The game involves manipulating game pieces as they "fall" from the top of the computer screen to the bottom in a central "play window." Players can move pieces to the left or right and rotate them as they fall. Their goal is to rotate and position the pieces so that they fill the space at the bottom of the screen without leaving gaps between pieces. On each day of game play, subjects were awakened repeatedly during the first hour of their regular overnight sleep, always within 3 minutes of sleep onset, and asked to report any thoughts, feelings, or images from the prior sleep period. Nine of the novices (75%) and five of the experts (50%) reported visual images of the game in 9.8% and 4.8% of their sleep-onset reports, respectively. Taken together, 64% of the subjects reported instances of game imagery during the hypnagogic period, with images reported in 7.2% of their reports.

Subjects reported remarkably similar imagery, seeing *Tetris* pieces falling in front of their eyes, occasionally rotating and fitting them into empty spaces. Among the 27 reports of imagery, there were no reports of seeing the larger picture surrounding the play window, the scoreboard, or the keyboard or of typing on the keyboard. There were only two reports of seeing a computer screen and none of seeing the desk or room. Thus the imagery was limited to those aspects of the experience that were most salient and to which subjects presumably paid the most attention. As described before, here again we see that dreams are not exact "replays" of waking experience but rather are composed of *fragments* of recent episodic material, often intermingled with other content.

Remarkably, three of five amnesiac patients also reported hypnagogic *Tetris* images, despite being unable to recall playing the game before or after the night's sleep. This observation clearly indicates that the hippocampal memory system, which supports the encoding and recall of episodic memories during wakefulness, is *not* necessary for the construction of sleep-onset imagery related to recent experience. Indeed, the similarity of amnesiacs' reports to those of control subjects was striking. Although unable to recall having played the game (or

even to recognize the experimenter from session to session), patients nonetheless reported, for example, "little squares coming down on a screen," and in one case, the subject explicitly stated that she had no knowledge of the source of the images. Thus she was able to produce dream images of events for which she had no declarative knowledge.

An additional point of note is that two of the five *Tetris* experts reported *Tetris* images explicitly described as being from earlier versions of the game, which they had not played in the last year. One subject reported imagery from a game version that she had not played since high school, 5 years earlier. Thus sleep-onset dream imagery need not be determined only by recent sensory input but additionally can incorporate older, strongly associated memories. Here again, we see that rather than an exact reiteration of waking experience, dreams incorporate salient elements of recent experience into a novel scenario.

This fact is driven home even more forcefully by a study in which subjects were taught three nonsense sentences across a night, one immediately before going to sleep and two more after awakening from REM sleep.[62] The nonsense sentences (translated from the original Italian) are as follows:

> In the bathroom the raven is painting a fish on a radio and spinning a bust on the custard.
>
> In the embers a poster is fining a parcel along the bridge and betting a tooth in the game.
>
> In a liter a cock is tricking a ruble from a palm and nursing a ball in a tub.

Subjects were instructed to memorize the sentences, and after hearing one twice, they repeated it back as accurately as possible. When dream reports were collected after awakening from the next REM period, dream content was frequently judged as related to the previously memorized sentence. For example, after hearing the first sentence, with its reference to "painting a fish," one subject reported "walking with a friend on the seashore."

Of course, such apparent associations can be spurious. Indeed, when judges scored dream reports from a control night, before subjects had heard any of the sentences, apparent associations were again found. However, the experimental design allowed one to use these rates of "pseudoincorporation," obtained from the control night, to correct for such spurious associations, and when this is done, more than one third of all REM reports collected on the experimental night contained actual incorporations of elements from the learned sentence (72% of reports on the experimental night versus 39% on the control night).

Two features of these results are particularly striking. First, the simple act of intentional memorization seems to be sufficient to tag a memory for potential incorporation into subsequent dream content, even when the memory itself has no apparent meaning. Second, as with the previously described study of the incorporation of waking memories into nocturnal dreams,[46] these incorporations are never in the form of an exact replay of the episode, that is, a report of memorizing a sentence. Instead, the brain seems to simply extract specific elements of the memorized sentence (single words or phrases) and to incorporate them into an unrelated scenario. In fact, not only is the scenario changed, but the object is as well, so that the actual word from the memorized sentence (e.g., the fish) is replaced by a semantically related word (e.g., the seashore).

In another study, after training intensively on the downhill skiing arcade game *Alpine Racer II,* 65% of subjects reported images from the game in their subsequent hypnagogic dreams.[63] These sleep-onset dream reports sometimes included places where they frequently crashed or a particularly steep slope, but the game images were again devoid of their original context, always being reported without the arcade game itself being seen or themselves playing it.[63] In agreement with the findings of the *Tetris* study, subjects with prior downhill skiing experience also reported seeing images related not to the arcade game but to actual skiing experiences from their past. These observations thus extend the *Tetris* findings to a second learning paradigm.

In the case of *Alpine Racer II,* it was additionally observed that the nature of game-related sleep-onset imagery became more abstracted from the original experience across the course of the night. A subset of participants were allowed 2 hours of uninterrupted sleep at the start of each night, before being awakened and then, as they fell back asleep, reporting hypnagogic imagery at this later time. In this "delayed awakening" protocol, subjects reported dramatically fewer skiing images. Instead, they reported imagery more indirectly related to the game, for example, of "falling down a hill" or of "moving through some kind of forest" with their "entire upper body incredibly straight."[63] Thus imagery related to recent experience appears to become more abstracted from the original memory source later in the night, a phenomenon also suggested in the transformation of "fish" into "seashore" in the study described previously.[64]

Finally, there is evidence that the incorporation of recent experiences into dream content directly reflects the process of memory consolidation. When newly learned information is incorporated into dream content, even in abstracted form, this is associated with enhanced memory for that information. In fact, such evidence dates back to the 1970s. Fiss and colleagues, for example, found that after reading the text of a short story, participants who reported dreams related to the story exhibited superior memory for the text the following morning.[65] De Koninck and colleagues also examined dreams and verbal learning, exploring dream content as a corollary of language learning in an academic setting.[66] Among students enrolled in a French-immersion class, those with the strongest language acquisition across the 6-week course incorporated French into dream content more often than students who were less successful in the class. Most recently, our laboratory has demonstrated that dreaming of a virtual maze navigation task is associated with enhanced consolidation of spatial memory both across a nap[67] and across a full night of sleep.[68]

EMOTION IN THE SLEEPING BRAIN

REM dreaming is commonly associated with the experience of intense emotion.[22,42] As such, literature suggesting that sleep supports the reactivation and transformation of *emotional* memories, in particular, may be relevant to our understanding of the dreaming process. In one study,[31] the recall of emotionally charged stories was selectively enhanced after REM. In this protocol, subjects learned both neutral and emotional texts before sleep. Relative to wakefulness, periods

of posttraining sleep rich in REM were particularly beneficial for these memories. Furthermore, memory for the *emotional* texts benefited from this REM-rich sleep to a greater degree than did neutral material. Remarkably, when the same research participants were again tested 4 years later,[69] the beneficial influence of REM versus wakefulness on memory for the original emotional (but not neutral) texts was maintained. Similarly, a study from Payne and colleagues[30] demonstrated that sleep selectively enhances memory for emotional objects in the foreground of visual scenes, suggesting that sleep selects the information most relevant to an individual for further processing while allowing memory for less salient aspects of an experience to decay. These observations are also in line with other literature suggesting that sleep modulates emotional responsiveness more generally.[70]

The most well-known example of emotional processing during sleep is perhaps also the least well understood. This is the phenomenon of "sleeping on a problem," involving situations in which a difficult decision must be made. Anecdotal reports indicate that people can go to bed at night with such a problem on their mind and wake up the next morning with a clear solution in their mind. There are several features of this phenomenon that are worth noting. First, it is a remarkably robust effect, with most people casually surveyed believing that, as often as not, it successfully yields results over a single night. Second, the decision normally becomes apparent without an explicit rationale. People report knowing at a "gut level" that they have come to the correct decision but without a clear and rational justification for it. Third, there is usually considerable confidence that the decision reached is the correct one and little sense that further deliberation would be of any added benefit. At the same time, the process does not appear to be useful for the recall of forgotten information, such as a phone number or address. Rather, it serves to analyze available information to come to a decision on the basis of some unknown algorithm that appropriately weights the relevant information. Whereas these features of the process are clear from anecdotal observations, few objective studies of the phenomenon have been made. Still, several studies do suggest that laboratory-induced problems are solved more easily following a period of sleep compared with a period of waking incubation.[34,71,72]

A NEUROCOGNITIVE MODEL OF DREAM CONSTRUCTION AND FUNCTION

When memory networks are activated in the brain, they are inevitably altered. This is one of the most striking findings of cognitive neuroscience in the past decade. Indeed, it might be true that no neural circuits are ever activated without being at least subtly altered. This is true whether the activity of a particular circuit is perceived by the conscious mind or not. Thus every time a young child hears a sentence spoken, neural circuits are activated that over time will extract rules of grammar that will allow her to speak with nearly perfect grammar without explicitly knowing those rules or even that they exist. This is a hallmark of the brain's construction—that it extracts similarities and rules without conscious knowledge that they exist. In addition, studies of mathematical insight[32] and transitive inference[52] have shown that sleep dramatically facilitates this process.

Of course, the activation of memory networks can also be accompanied by conscious experience, as is the case now, as you read and consider the arguments presented here. As there are distinct memory systems in the brain, some accessible to consciousness and some not, so also are there distinct mechanisms for activating and manipulating these memories, some of which are accessible to consciousness and some of which are not. So it should come as no surprise when we suggest, first, that dreaming must inevitably alter the memories accessed in the process of dream construction and, perhaps more important, that dreaming may be a byproduct of mechanisms that evolved to facilitate sleep-dependent memory consolidation and integration.

As in wake, these mechanisms would often activate systems that remain outside of our awareness. This perhaps is the case when a visual texture discrimination skill[23] or a finger-tapping motor sequence[22] is consolidated during sleep. At other times, the sleep-dependent reactivation of memory networks might occur in a manner that brings the patterns of activation in these networks into conscious awareness. Thus, after training on a hippocampus-dependent virtual maze navigation task, participants dream of the maze, observing the flow of images, thoughts, and feelings that occur during this memory reactivation process, which ultimately leads to improved memory performance following sleep. Critically, we do not necessarily *see* the changes in the memory systems that result from this process, and thus we see the content of dreams but neither their underlying purpose nor their ultimate effects. In short, the content of recalled dreams does not in itself reveal any obvious function; as we saw earlier, dreams typically do not appear to be a rehearsal of things that are important to remember.

The question of the function of dreaming may be reducible to a question of the function of the sleep-dependent memory processes that result in the conscious experience of dreaming. Accumulating evidence suggests that sleep has, as one of its most critical functions, the incremental modification of cortical networks. Such a model has been put forward in some detail elsewhere.[24,25] Because sleep has been shown to enhance (1) the experientially controlled modification of visual circuitry during early development[73,74]; (2) visual, auditory, and motor skill learning[26,27,75]; (3) emotional memories[30,31,39,70,76]; (4) declarative and hippocampus-dependent memories[28,29,47,51-53,77]; and (5) creative insight,[32-34] it is clear that sleep's role in memory consolidation spans an impressively wide range of brain circuits and functions. The question of dream function now becomes two questions: For which of these circuits and functions does activity during sleep enter our conscious awareness? and Does this conscious awareness in turn have an effect on these brain circuits? As described earlier, there is evidence that the consolidation of memory in the sleeping brain does, in some cases, directly affect dream content. However, there are still no ways to address the question of whether actual conscious experience can ever alter brain activity, either in waking or in sleep.

CLINICAL PEARL

Emerging data suggest that dream experiences reflect the offline consolidation, integration, and analysis of recent memories during sleep. As such, studying dreaming may facilitate a greater understanding of the sleep-related processes underlying long-term memory formation, as well as the dysfunction of these processes seen in pathologic conditions such as posttraumatic stress disorder.

SUMMARY

In the end, then, we propose that dreaming is simply the conscious perception of the stream of images, thoughts, and feelings evoked in the brain by one or more of the many forms of off-line learning and memory processing that occur during sleep. At the same time, it reflects one of the most sophisticated forms of processing that the brain performs: the analysis and interpretation of the events of our lives in a manner that provides meaning to these events and can guide our future behavior.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by NIMH grants R01-MH48832 and R01-MH65292, as well as a KL2/Catalyst Medical Research Investigator Training award to E. W. (an appointed KL2 award) from Harvard Catalyst (National Institutes of Health Award 8KL2TR000168-05). The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

Selected Readings

Baylor GW, Cavallero C. Memory sources associated with REM and NREM dream reports throughout the night: a new look at the data. *Sleep* 2001;**24**(2):165–70.

Blagrove M, Fouquet NC, Henley-Einion JA, et al. Assessing the dream-lag effect for REM and NREM stage 2 dreams. *PLoS ONE* 2011;**6**(10):e26708.

De Gennaro L, Marzano C, Cipolli C, Ferrara M. How we remember the stuff that dreams are made of: neurobiological approaches to the brain mechanisms of dream recall. *Behav Brain Res* 2012;**226**(2):592–6.

Fosse MJ, Fosse R, Hobson JA, Stickgold RJ. Dreaming and episodic memory: a functional dissociation? *J Cogn Neurosci* 2003;**15**(1):1–9.

Graveline YM, Wamsley EJ. Dreaming and waking cognition. *Trans Issues Psychol Sci* 2015;**1**:97.

Herlin B, Leu-Semenescu S, Chaumereuil C, Arnulf I. Evidence that non-dreamers do dream: a REM sleep behaviour disorder model. *J Sleep Res* 2015;**24**(6):602-9.

Hobson JA, McCarley RW. The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. *Am J Psychiatry* 1977;**134**(12):1335–48.

Ji D, Wilson MA. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat Neurosci* 2006;**10**(1):100–7.

Nielsen TA. A review of mentation in REM and NREM sleep: "covert" REM sleep as a possible reconciliation of two opposing models. *Behav Brain Sci* 2000;**23**(6):851–66.

Nir Y, Tononi G. Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn Sci* 2010;**14**(2):88–100.

Payne JD, Nadel L. Sleep, dreams, and memory consolidation: the role of the stress hormone cortisol. *Learn Mem* 2004;**11**(6):671–8.

Wamsley EJ, Stickgold R. Dreaming and offline memory processing. *Curr Biol* 2010;**20**(23):R1010–13.

Wamsley EJ, Tucker M, Payne JD, et al. Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Curr Biol* 2010;**20**(9):850–5.

A complete reference list can be found online at ExpertConsult.com.

Chapter 49

Dream Content: Quantitative Findings

Antonio Zadra; G. William Domhoff

Chapter Highlights

- Researchers and clinicians have long been fascinated by the content of dreams, and considerable progress has been made in the systematic study of dream content. The most frequently used methods for collecting dream reports, laboratory awakenings, home dream logs, questionnaires, and most recent dreams collected in group settings, all have their uses and inherent advantages and disadvantages. However, reliable, comprehensive, and validated instruments for the actual analysis of dream content reports have been developed, and complementary tools are now available to all researchers on the Internet.
- Quantitative data on dream content from laboratory and nonlaboratory settings generally converge in depicting a reliable picture about the nature of dream content in the general adult population. Both data sets indicate that, for the most part, dreams are a reasonable simulation of waking life characters, social interactions, activities, and settings and that dreams show systematic relationships to various dimensions of the dreamer's waking life but not to day-to-day events.
- Results from a variety of studies show that developmental changes occur in dream content until late adolescence, when dream content becomes generally stable and consistent throughout adulthood and old age. In addition, clinically oriented investigations suggest that affect and social interactions are two key dream content variables that are most strongly related to measures of psychological well-being.
- The findings presented in this chapter have several implications for theories of dreaming and provide convincing evidence that dreams are a unique and meaningful psychological product of the mind.

Researchers and clinicians have long been fascinated by the content of dreams. Although many contemporary dream researchers suggest that dreaming is functionally significant and may subserve a biologically important function, some argue that dreams are a by-product of neurophysiologic activity during rapid eye movement (REM) sleep and have no value in and of themselves even though evidence suggests they have psychological meaning.

There is no consensus on what distinguishes "dreaming" from other cognitive processes, such as thinking or daydreaming, nor on what constitutes "dream content." Interdisciplinary groups from the International Association for the Study of Dreams and the American Academy of Sleep Medicine concluded that "a single definition for dreaming is most likely impossible given the wide spectrum of fields engaged in the study of dreaming, and the diversity in currently applied definitions."[1] Thus, depending on one's perspective, dreaming can be synonymous with the term *sleep mentation*, which refers to the experience of *any mental activity* (e.g., perceptions, bodily feelings, thoughts) during sleep, or can be restricted to more elaborate, vivid, and storylike experiences recalled on awakening. As highlighted by others,[2] using a broadly inclusive versus more restrictive definition of dreaming has a direct and significant effect on the nature and sense of empiric data and theoretical modeling in the field.

In this chapter, the term *dream* is conceptualized as having four interrelated meanings. First, a dream is a form of thinking during sleep that occurs when there is a certain, as yet undetermined, minimal level of brain activation in a context in which external stimuli are typically occluded and the cognitive system that keeps us aware of our surroundings is shut down. Second, a dream is something people experience as a series of actual events (e.g., a sequence of perceptions, thoughts, and emotions) because the thought patterns simulate waking reality in a manner that is now often called *embodied simulation*. Third, a dream is what people remember on awakening, so it is a memory of the dreaming experience. Finally, a dream is the spoken or written report provided to investigators based on the memory of the dreaming experience. The empiric studies discussed in this chapter reveal that the events of a dream always include the dreamer as an observer or participant and that they almost always include at least one other character besides the dreamer (either a person or an animal). In addition, the dreamer or the other characters in the dreams are invariably engaged in one or another activity (e.g., looking, walking, running) or a social interaction. Thus the sense of participation in an event, along with characters, activities, and social interactions, is what distinguishes dreams from the more fleeting, fragmented, and thoughtlike forms of sleep mentation.

METHODS FOR COLLECTING DREAM REPORTS

Researchers never study dream experiences directly. Instead, they collect and have access to descriptions of the experience, the dream report. The nature and content of the verbal or written report obtained can be influenced by a number of factors. These include the setting (e.g., home, laboratory, classroom, psychotherapy), method of awakening (e.g., spontaneous, induced), time of awakening (e.g., early, middle, or late in the sleep period), sleep stage before awakening (e.g., REM, non–rapid eye movement [NREM] sleep), type of collection instrument (e.g., questionnaire, dream journal), reporting method (e.g., written by the subject, written by the experimenter, audio recording), instructions provided (e.g., report anything that was going through your mind before your awakening, not only your dreams), probes on reported content (none, fixed, or semistructured questions), interpersonal situation (e.g., reporting directly to an experimenter, clinician), time delay between when the dream was experienced and when it is reported, study duration, and subject characteristics (e.g., gender, personality, habitual level of dream recall).

The degree to which the content of dream reports is influenced by these various factors either individually or in combination varies as a function of the collection method used. The principal sources of dream reports are the sleep laboratory, home dream journals, questionnaires, psychotherapy sessions, and classroom or other group settings where a most recent dream can be collected from everyone willing to participate. Although there is convincing evidence that working with patients' dreams can be clinically useful,[3] dream reports from the psychotherapy relationship are rarely used in systematic studies and thus this source is not covered here.

Sleep Laboratory

Sleep laboratories are an excellent source of dream reports because they provide the opportunity for collecting a representative sample of a subject's dream life, both within and across nights, under controlled conditions. Awakening subjects from several REM or NREM periods results in the collection of dream reports that may have been otherwise forgotten by the participants on normal awakening in the morning. Awakenings during REM or from stage II NREM sleep late in the sleep period maximize the probability of recall and make it possible to collect as many as four or five dreams in a single night. On the other hand, frequent awakenings can be difficult for participants, and factors such as sleep inertia and one's desire to return to sleep may interfere with the quality of the dream reports. However, a complementary cued morning report of dreams recalled during the night can yield new and reliable information as to the dreams' original contents.[4]

The main problem with the laboratory collection of dream reports is that it is a very costly and time-consuming process, and even though several dreams can be collected each night, it still can take many months to obtain 10 or more dreams from each of a dozen participants. Furthermore, some types of dreams, including nightmares and sexual dreams, rarely occur in the sleep laboratory, presumably because of sociocognitive factors. In addition, approximately 20% of laboratory REM dream reports will reflect direct incorporations of the laboratory environment, even when collected over several consecutive nights. For our purposes, the most important outcome of detailed laboratory studies is that they provide a baseline for assessing the quality of dream reports collected by other methods.

Dream Logs

Prospective daily logs are used by an increasing number of dream researchers even though they require a greater investment of time and resources than do questionnaires. In fields like nightmare research, home journals are considered the gold standard for the measurement of nightmare frequency.[5] Although limitations associated with longer-term retrospective assessments of dream recall and dream content are increasingly recognized, variations in home logs have received little attention. Prospective logs can take two different forms. The first is the checklist format in which participants indicate if there was dream recall and, if so, the number and type of dreams recalled (e.g., nightmare). The second is the narrative log, in which participants are requested to provide a complete written transcript of each dream recalled. Findings from one comparison[6] of these two methods of data collection suggest that narrative log participants, having a more time-consuming task, do not take the required time to provide a complete narrative of all of their recalled dreams, as Strauch[7] found with teenage boys. Instead, they may choose to focus on their more memorable, exciting, or salient dreams, which would typically include bad dreams and nightmares. By comparison, people completing checklist logs would be more likely to record all of their dreams (including relatively banal or poorly recalled ones) because each entry is just as quickly completed regardless of dream type.

Although writing down one's dreams remains the most frequently used method to collect dream content, participants may also use tape recorders to dictate their reports. This approach may be particularly useful with children and younger adolescents. It also proved highly useful in a study of blind participants.[8]

Questionnaires

In questionnaire studies, participants' retrospective self-reported information concerning their dream experiences is viewed as a modest but acceptable way of assessing different aspects of the dream experiences themselves, but research suggests they are of limited value in assessing the frequency or content of dreams. Three types of information are generally collected.

First, subjects can be queried about the frequency with which they experience certain kinds of dreams (e.g., everyday dreams, nightmares) over a determined period of time. There is increasing evidence, however, that data obtained with retrospective estimates differ considerably from daily prospective home logs. For instance, compared with results from daily home logs, retrospective self-reports significantly underestimate current nightmare frequency,[9,10] and this rate of underestimation is not attributable to an increase in recalled dreams caused by keeping a dream log.[10] Similarly, one study[9] found that the magnitude of the association between trait anxiety and nightmare frequency decreased significantly when daily logs were used to measure nightmare frequency instead of retrospective self-reports. This led the authors to suggest that anxious individuals do not necessarily have more nightmares, but rather are more likely to remember and report nightmares retrospectively. Finally, a meta-analysis[11] of studies examining the relationship between dream recall frequency and various

personality dimensions found that scores on personality measures were not related to dream recall frequency per se, but rather to people's tendency to retrospectively underestimate or overestimate their dream recall. Taken together, these findings indicate that correlates of retrospective measures of dream recall should not be assumed to be correlates of log measures of dream recall. Contrary to prospective log measures, retrospective indexes of dream recall are best viewed as measures of peoples' cognitive representations of their dream life.

A second kind of information sometimes elicited by questionnaires focuses on specific dimensions of people's dreams or their beliefs about their general dream life. This approach assumes that there exists a valid relationship between self-reported information on the content of one's everyday dreams and the dream experiences themselves. However, comparisons of self-report measures and log-based data indicate that this assumption may be unwarranted. For instance, one[12] comparison of participants' questionnaires and 2-week logs found no relationship between estimated frequency for the appearance of aggressive, friendly, and sexual elements and their frequency in the dream reports. Similarly, a subsequent study[13] showed that when people's level of dream recall is poor, their beliefs about the level of anxiety in their dreams is not related to the actual affective content of their everyday dreams as recorded prospectively in home logs. These findings suggest that the relation between beliefs people hold about the content of their dreams and their actual dream experiences is mediated by autobiographic memory and that these beliefs are particularly inaccurate when dream recall is low (i.e., when memories of one's dreams are not readily available).

Lastly, questionnaires are used to investigate whether participants ever experienced a specific type of dream and, if so, to report the most recent occurrence as best recalled. This approach allows for the investigation of certain types of dreams that, because of their infrequency, are difficult to capture in laboratory settings or with home dream logs (e.g., recurrent dreams, existential dreams) or dreams that stand out in the person's past (e.g., earliest dream recalled, most terrifying nightmare). Although useful in some research settings, the resulting dream content findings must be treated cautiously because of possible memory distortions and biases.

In sum, although some dream questionnaires have good internal consistency and test-retest reliability,[14] studies of their relationship to dream content and frequency findings obtained from dream journals reveal important discrepancies and raise questions as to their validity.

Classroom and Other Group Settings

Settings such as classrooms provide an objective and structured context for the efficient and inexpensive collection of dream reports. Anonymous participants are instructed to write down the most recent dream that they can recall on a standardized form while revealing only basic background information such as age and gender. The Most Recent Dream method has been used with children as young as ages 10 to 11 years in different countries with surprisingly similar cross-national results.[15] A more recent study[16] using the Most Recent Dream method with Greek children (ages 8 to 12 years) and adolescents (ages 13 to 18 years) also showed consistent age changes along with ongoing gender differences similar to those for children studied in the other countries. However, there is reason to believe that young children up to ages 10 to 11 years are using their waking imaginations to provide a report that fits cultural stereotypes about the nature of dreams. The main drawback with this method is that there is not usually time to collect any personality or cognitive measures on the people providing the reports.

ANALYZING DREAM CONTENT: INSTRUMENTS AND ISSUES

Most past dream research used either rating scales at the ordinal level of measurement ("more" or "less" of a characteristic) or discrete categories at the nominal level of measurement (an element is "present" or "absent"). Rating scales are most useful for those characteristics of dream reports that have degrees of intensity in waking life. Cohen[17] reports that four dimensions of dream salience can be rated by participants in dream studies: emotionality, bizarreness, activity, and vividness. A factor analysis of the ratings of 100 REM dream reports suggests that rating scales boil down to five basic dimensions: (1) degree of vividness and distortion, (2) degree of hostility and anxiety, (3) degree of initiative and striving, (4) level of activity, and (5) amount of sexuality.[18] However, it is often difficult to establish reliability with some scales, and much of the specific information in dream reports is lost or unused with general rating scales.

Of the 150 dream rating and content analysis scales reviewed by Winget and Kramer,[19] the Hall and Van de Castle (hereafter HVDC) coding system[20] is the best validated and remains the most widely used system for analyzing dream content. The HVDC system, which provides many of the findings presented in the rest of this chapter, rests on the nominal level of measurement and uses percentages and ratios as content indicators that can correct for the varying length of dream reports from sample to sample. The dream reports used in the original normative sample, as well as the coding of them, are available to researchers through www.dreambank.net.[21] The normative findings reveal a pattern of gender differences that needs to be taken into account when doing studies of individuals. The coding system employs nonparametric statistics for determining *P* values and effect sizes, which can be obtained instantly after entering codes into the DreamSAT spreadsheet available to all researchers on www.dreamresearch.net.[22] The general HVDC norms can be used with confidence for a variety of purposes because they have been replicated in several studies.[23,24] The coding system and the norms also have been found to be useful in studies of college students in several different countries.[25]

As documented by Winget and Kramer,[19] there exist numerous other coding systems, and many new ones have been created since their comprehensive review. However, unlike the HVDC system, most of these instruments have only been used by the original investigators (limiting potential for comparisons across laboratories), many use weighting systems of questionable validity, and few are based on clearly defined and objective scoring criteria that yield good interrater reliability. Moreover, as detailed elsewhere,[23] many of these scoring systems can be duplicated by combining two or more elements of the HVDC system.

Some research questions (e.g., self-reflectiveness in dreams,[26] contextualizing images in dreams[27]) have necessitated the creation of new instruments. The DreamThreat rating scale[28] was developed to test an evolutionary theory of

dreams that stipulates that the function of dreaming is to simulate threatening events with the intent of improving the subject's capability to recognize and avoid diverse threats in real life. Although this rating scale has been criticized, it is noteworthy in that it has been used by different groups to assess various kinds of dreams[28,29] and that it yields good to excellent interrater agreement. Taken together, findings indicate that a significant proportion of dreams contain a wide range of threats, but few of these dreams present realistic life-threatening events, and the dreamer rarely succeeds in escaping the threat.

Finally, because of the time-consuming nature of traditional coding systems, programs for word and phrase searches have been created to study specific characters (e.g., "my mother) or activities (e.g., "making love"), and lengthy word strings have been developed for coding concepts characterized by a relatively circumscribed set of terms (e.g., specific emotions). The program for single words, phrases, and word strings on www.dreambank.net calculates frequencies, percentages, *P* values, and effect sizes when two sets of dream reports are compared.[21] A comprehensive and carefully constructed set of 40 word strings covers classes of characters, types of activities, natural settings, and much else.[30,31] These 40 word strings have normative findings based on the same dream reports used to create the HVDC norms, and they provide results comparable to the HVDC findings for several categories.

Problems in Studying Emotions and Bizarreness in Dreams

Although both rating scales and the HVDC nominal coding categories have proved useful for most dimensions and elements of dream content, there are methodologic problems relating to the study of both emotions and bizarreness in dreaming. Several different studies using blind coders find that negative emotions outnumber positive ones.[20,32] Further, a laboratory study that compared ratings of emotions by independent judges with similar ratings by participants immediately after each awakening showed no differences.[33] However, different results emerge when the participants themselves make a global rating of each of their dream reports on a pleasant-unpleasant dimension. Such studies regularly find that the dreamers rate the emotions in their dreams as at least equally pleasant and unpleasant, and sometimes as more pleasant.[34,35] Furthermore, some studies[36] show that a greater proportion of laboratory as well as home dream reports are rated as containing emotions when these are scored by the dreamer compared with external raters. More recently, one study[37] of emotional experiences in REM dream reports found that self-ratings of emotions provided by the dreamer on awakening differed from ratings given by external judges using the same rating scales, with self-ratings resulting in greater estimates of emotional dreams, positively valenced dreams, and positive and negative emotions per dream.

Dreamers also tend to attribute many more emotions to their home dreams than do blind judges when they are later asked to recall the emotions that accompanied reports they wrote down at an earlier time. However, it is an open question in need of further study as to why dreamers often say their dreams are more pleasant than might be expected based on judges' ratings and attribute more emotions to their home dream reports than judges do. These differences may result from two extrinsic factors, namely the demand characteristics of such a rating task and the waking-life assumption that certain emotions would logically be present in many of the situations experienced in the dreams. It is also likely that the use of different rating scales and instructions for the scoring of emotions in dreams (e.g., dream's overall emotional tone versus number of emotions reported per dream narrative, frequency versus intensity of emotions, number of discrete positive and negative emotions to be rated, scoring of inferred versus explicitly reported emotions) affects the ratings obtained, whether they are self-reported by the dreamer or scored by external judges.

There is also lack of agreement on how to assess unusual or bizarre elements in dreams, which leads to widely varying prevalence and frequent estimates. In studies that focus on clearly impossible events, the figure is 10% or below for large samples of both REM and home dreams.[38,39] When sudden scene changes, uncertainties, and small distortions are included, the figure rises to between 30% and 60%.[40,41] Using a rating scale based on the degree to which any dimension of the dream differs from waking experience and behavior, it was found that 75% of 500 REM reports from adult men and women had at least one bizarre aspect, as compared with 7% to 8% that were bizarre in three or more ways.[42] In addition, other than for one study[43] showing that scene changes can be similar in REM dream reports and waking mentation, studies of bizarreness in dreams have been handicapped by the fact that there have been no other adequate studies comparing the nature and frequency of bizarre elements in dreams and waking thought samples from the same participants, which seems to be an essential step given the evidence that waking thought often contains unexpected and anomalous elements.[44]

The Importance of Adequate Sample Sizes and Minimum Report Lengths

One important variable that is all too often overlooked when investigating dream content is the sample size required to detect changes in various content variables. The use of an approximate randomization algorithm provides evidence that it takes 100 to 125 dream reports to detect significant content differences for many of the HVDC content indicators because some dream elements appear in half or less of dream reports; in addition, effect sizes are often modest.[45] It should also be noted, as detailed elsewhere,[23] that it is unlikely that repeatable and scientifically useful results can be obtained with dream reports much shorter than 50 words, especially when using HVDC content categories.

Finally, although coefficients of internal consistency for dream diaries indicate that everyday dream recall is relatively stable over time,[6,46] several dream content variables appear infrequently in dream reports and show large intraindividual fluctuations. For this reason, it is suggested that correlational studies involving relatively rare or unstable dream content variables be based on at least 20 dream reports from any given participant.[14]

Quantitative Findings on Dream Content

Dream Reports from Laboratory Awakenings

The best starting point for the systematic study of dream content remains the classic studies completed by dream researchers during the heyday of laboratory dream research in the 1960s and early 1970s. They show that dream content simulates everyday life to a far greater degree than had been

anticipated based on the clinical cases that had been the basis for theorizing before the laboratory era of dream research.[47] They characterize a prototypical REM dream report as a "clear, coherent, and detailed account of a realistic situation involving the dreamer and other people caught up in very ordinary activities and preoccupations, and usually talking about them."[47]

For example, of all the dream settings that were described, only 5% were "exotic," in the sense of highly unusual or out of the ordinary, and less than 1% were "fantastic," in the sense of unrealistic.[47] Using a conservative standard to guard against imputing any emotions to the dreamers, specific emotions were judged to be present in only 30% to 35% of the reports, with unpleasant emotions outnumbering pleasant ones by 2 to 1. Anxiety and anger were the most frequent types of emotions; erotic feelings occurred in only 8 of the 635 reports (1.3%).[47] The dreams were rated as having a low degree of bizarreness. Focusing here on the longest reports because they were more frequently rated as bizarre, 50% were rated as having no bizarreness, 30% as having a low degree of bizarreness, 8% as having a medium degree, and 2% as having a high degree.[47]

A low degree of bizarreness was also reported in a highly detailed laboratory study of that issue.[39] The authors "emphasize the rarity of the bizarre in dreams" because major distortions of actual waking experiences reach a high of only 17% of all the activities and social interactions and of 6% and 8% for all characters and physical surroundings.[39] When they carried out global ratings of each dream for overall novelty, they found that most of them contained very little novelty: only 9% were highly improbable by waking standards; another 26% showed large but plausible differences from previous waking experiences.

The issue of emotions in REM dream reports was first investigated in great depth in the sleep laboratory where participants were quizzed in detail after each awakening as to the presence of emotions and the appropriateness of the emotion to the content. Drawing on ratings by both participants and naïve judges, it was concluded that about 70% of the dream reports had at least some affect.[48] The study further found that there were no differences in the ratings of emotions by the independent judges and the participants.[33] A study in a Swiss sleep laboratory came to very similar conclusions about the frequency and intensity of emotions in dreams.[42]

Several early laboratory studies probed for any changes that might occur in dream content from REM period to REM period, uncovering very few replicable differences. Employing categories for settings, characters, activities, social interactions, and emotions, both quantitative and qualitative analyses find few or no differences from REM to REM when corrections are made for the length of report.[34] In the most comprehensive study of this issue, there were two minor differences among 26 analyses employing HVDC categories for the first four REM periods, whether they were nights with single or multiple awakenings, and there were no differences with spontaneously recalled dreams that came from night or morning REM self-awakenings.[49] However, there may be some degree of thematic continuity from REM to REM on a few nights.[42]

REM and NREM Dream Reports

Although there were indications in early laboratory studies that dreaming occurs almost exclusively in REM sleep and that there were differences in the content of REM and NREM reports, many later studies suggest that the differences in recall are not black and white, especially late in the sleep period, and that some of the content differences disappear when there is a control for word length.[50,51] Still, most studies conclude that dreams are more frequent and longer during REM periods and that many NREM reports seem to be "thoughts," not dreams. In fact, NREM reports are more often a continuation of waking thoughts and memories, whereas there are few episodic memories in REM or home dream reports.[52,53]

The differences in content relate to a greater character density in REM reports, which in turn leads to the possibility of social interactions.[38,50] Then, too, there is evidence that NREM reports late in the sleep period are more similar to REM reports than are NREM reports from the first few hours of sleep.[54] In the most recent studies of this issue, the thought-like nature of NREM decreased by 56% and the hallucinatory nature increased by 62% over the course of the night, leading to the conclusion that "as the night progresses, NREM approaches the neurocognitive characteristics of REM."[55] In two separate studies it was found that the major difference between late-night REM and NREM dreams is on aggressive interactions.[56]

Laboratory and Home Dream Comparisons

Several careful investigations reveal that there are relatively few differences between home and laboratory dream reports even when the dreams are obtained by tape recorders in the sleep laboratory and by written reports at home.[38,42] Furthermore, most of these differences disappear when the proper controls are introduced.[57,58] The one exception to this generalization seems to be hostile and aggressive dream elements, which occur more frequently in the home dream reports of young adults in three different studies.[38,58]

These findings on the relatively small differences between home and laboratory dreams may be explainable in terms of the results from laboratory studies that compare what is reported from REM awakenings with what is still remembered in the morning.[59,60] Such studies reveal that recency and length of report are the primary factors in later recall, which at home would lead to a representative sample of nightly dream content given the lack of content differences from REM to REM and between REM and late-night NREM. However, some of these studies also show that intensity can be a tertiary factor in morning recall, which suggests there is some selection bias toward the everyday recall of more emotionally salient content.

Normative Dream Content in Home Dreams

As might be expected from the results of the laboratory-home comparisons, studies of large samples of dream content collected from young college-educated adults outside the laboratory show many similarities with the laboratory results when the same or comparable content categories are employed. Dreams mostly occur in commonplace settings, contain a large number of familiar characters, and revolve around family concerns, love interests, and activities engaged in during waking life.[61] This point is best seen in a study of several hundred dream reports from German college men and women in which the dream content was coded for at least one instance of several simple ad hoc categories constructed to determine the degree to which the dreams involve people and activities

from everyday life.[62] There were four categories for familiar characters, five categories for commonplace leisure activities, and a single category for involvement in work, school, or politics. The everyday nature of most of these dreams is seen in the fact that 75% of the women's dreams and 62% of the men's have at least one instance of one of the four categories of familiar characters. Similarly, 42% of the women's dreams and 27% of the men's have at least one instance from one of the five leisure-time categories. The routine matters of work, school, or politics appear in 20% of the women's dreams and 29% of the men's dreams. Overall, only 13% of the women's dreams and 20% of the men's have no instance of any of the above categories. In keeping with other findings on gender differences in dream content, the men's dreams are less likely to have familiar characters and familiar leisure time activities and more likely to have instances of school, work, or politics, but the important point for purposes of this chapter is that only a minority of dreams for either gender involves unknown characters and activities that are out of the ordinary.

Given the longstanding clinical and popular interest in dreams with erotic or sexual content, this dream content category has received surprisingly little attention. Questionnaire studies indicate that approximately 80% of adults answer positively to the question, "Have you ever dreamed of sexual experiences?"[63] with men reporting sexual dreams more often than women. The normative data from HVDC indicates that 12% of mens' dreams and 4% of women's dreams contained sexual content, including having or attempting intercourse, petting, kissing, sexual overtures, and fantasies. However, one study[64] of more than 3500 dream reports found no gender differences, with approximately 8% of dream reports from both men and women containing sexually related activity. The differences with the HVDC data may be partially due to sample composition (college students versus student and nonstudent adults). Alternatively, it is also possible that women actually experience more sexual dreams now than they did 40 years ago, or that they now feel more comfortable reporting such dreams because of changing social roles and attitudes, or both.

Age Differences

There appear to be major changes in dream content from the preschool to teen years, but few changes from the late teens to old age. Dream content thus seems to parallel cognitive and emotional development during childhood as well as the stability of adult personality. Much of what is known in a systematic way about children's dreams comes from a classic longitudinal laboratory study of children between the ages of 3 and 15 years, supplemented by a cross-sectional laboratory replication a few years later with children ages 5 to 8 years.[65] More recently, a 5-year longitudinal laboratory study of Swiss children ages 9 to 15 years has provided additional supporting information.[7,66] Detailed summaries of the methods, samples, and findings can be found elsewhere.[22]

The most unexpected finding in the first study was the low amount of recall from REM periods in the 3- to 5-year-olds (only 27% of the REM awakenings yielded any recall that could reasonably be called a dream), and the static, bland, and underdeveloped content of the few reports that were obtained. The reports became more "dreamlike" (in terms of characters, themes, and actions) in the 5- to 7-year-olds, but it was not until the children were 11 to 13 years old that their dreams began to resemble those of adult laboratory participants in frequency, length, emotions, and overall structure, or to show any relationship to personality.[67]

A cross-sectional replication of these results[65] with children ages 5to 8 years supported all of the main original findings. The median rate of reporting was only 20% for all age groups. The imagery in the dreams was more static than dynamic until age 7 years, and the child's "self" character did not tend to take an active role in the dreams until age 8 years.[65] As with young adult dreams, there were more characters in the girls' dreams, and there was the same gender difference in the percentage of male and female characters. There were no failures, few negative emotions, and very few misfortunes. There were few aggressive or friendly interactions, with more friendliness in the girls' reports.[23]

The results from the longitudinal study of Swiss children ages 9 to 15 years were generally similar to those for preadolescents and adolescents in the earlier longitudinal study, and there were only relatively small changes in most categories over the 6-year period. The largest change was a decline in bizarreness for both boys and girls, as defined by degrees of deviation from waking experience and social norms; just over 60% of dream reports had at least some degree of bizarreness at ages 9 to 11 and 11 to 13 years, but the figure fell to 41% at ages 13 to 15 years.[7]

In contrast to the changes in dream content from childhood to adolescence, dream content is extremely stable in terms of characters, social interactions, and most other dream elements after age 18 years according to cross-sectional studies in the United States, Canada, and Switzerland that are summarized in Domhoff.[23] Elderly people recalled fewer dreams in one large longitudinal study,[68] but separate studies suggest their dream content remained generally the same—except perhaps for aggression, where studies suggest a decline.[69,70]

Dream Content and Well-Being

Considerable research efforts have been expended trying to establish dream content correlates of standardized personality variables, measures of psychological well-being in nonclinical samples, and indexes of psychopathology in clinical populations. Taken as a whole, there is mixed evidence that psychometrically defined personality traits (e.g., neuroticism, extraversion) are related to everyday dream content.[71] Robust relations, however, have been demonstrated between waking levels of well-being and specific types of dreams such as nightmares[5] and recurrent dreams,[72] as well as between dream content and various dimensions of waking life, including people's general waking concerns.[24,45,73] Several studies[74,75] have shown that dream content is reactive to the experience of naturalistic and experimental stressors, but whether or not dreams play a role in people's actual adaptation to stress remains an open question. In a series of longitudinal studies of REM dream reports from depressed and nondepressed adults undergoing marital separation or divorce, Cartwright and her collaborators[73,76] provide suggestive evidence that dream content variables centered around affect and the representation of the ex-spouse are associated with how well people adapt to their situation over time. Similarly, one longitudinal study[77] of normal adults found that participants' dream content from home logs was moderately to strongly correlated to their scores on measures of psychological well-being both at fixed points in time and over a 6- to 10-year period, with content variables of dream affect and social

interactions showing the strongest relations. Dream content in severe psychopathologic conditions such as schizophrenia has been reviewed elsewhere,[78,79] and with few exceptions, little by way of consistent findings has emerged from this literature. In addition, many studies in this field suffer from methodologic problems, including unclear diagnoses, inadequate controls, unknown effects of medications, few dream reports per patient, and the use of untested coding systems. However, the HVDC system has been shown to be useful in studies of patients with Parkinson disease and those with REM sleep behavior disorder[80,81] as well as women who have had mastectomies.[82]

In addition, unique features of dream content have also been better documented in relation to conditions typically accompanied by distinct waking thoughts and concerns, such as pregnancy,[83] bereavement,[84] and exposure to trauma.[85] A better understanding is also emerging regarding the frequency and contents of specific kinds of dreams, including typical dreams[63,86] as bad dreams and nightmares.[87]

Individual Case Studies

Within the context of the many well-established group findings, individual case studies can be of great value for both research and possible clinical applications. The dream journals on which such studies are based have value as nonreactive measures that have not been influenced by the purposes of the investigators who later analyze them. The conclusions drawn from nonreactive archival data are considered most reliable and useful when they are based on a diversity of archives likely to have different sources of potential biases.[23]

Studies of more than a dozen different dream journals first proved their usefulness for scientific purposes by revealing an unexpected consistency in dream content when several hundred dream reports were studied.[23] This consistency begins in the late teens and continues to old age. Two studies of discontinuous dream series show that the consistency revealed in continuous dream journals is not the result of practice effects.[23,88]

Individual dream journals also provided the basis for the most rigorous work to date on the lawfulness of dreams and their relationship to waking conceptual processes. This work[89] shows that the social networks in dreams—that is, the pattern of direct and indirect relationships among the characters—have the same properties as waking social networks in that the paths between characters are short and the clustering of characters is high. Moreover, the frequency distribution of the characters is consistent with Zipf's law, a power law for describing frequency distributions in which the top few entities occur very frequently and most other entities appear very rarely. In a recent extension of this work,[90] the dream and waking-life social networks of a middle-aged woman were compared using 4254 dream reports and information from the dreamer concerning her relationships with her dream characters in waking life. Results showed that people important in one network tend to be important in the other, but that people with different relationships to the dreamer (e.g., family, friends, and coworkers) are mixed together much more in the dream than the waking network.

Blind analyses of dream journals[23,91] also have led, through the formulation of inferences that can be accepted or rejected by the dreamer and other respondents, to the conclusion that some dream content is continuous with the dreamers' waking conceptions, concerns, and interests. The most direct continuities involve the main people in a dreamer's life and the nature of the social interactions with them. There also is good continuity for many of the dreamer's main interests and activities. However, these findings on continuity have to be qualified in two ways. First, the continuity is with general concerns, not day-to-day events, as shown by three studies (two based on REM awakenings, one based on morning recall at home) in which judges could not match detailed waking reports of daily concerns with dream reports.[92] This finding is consistent with studies showing low levels of episodic memory in dreams.[52,93] Second, the continuity usually is with both thought and behavior, but sometimes it is only with waking thought. For example, people who have highly aggressive dreams are not always aggressive people in waking life, but they usually admit to many aggressive thoughts and fantasies during the day.[23]

Sensory Experiences and Dreams of Blind Subjects

Although the overwhelming majority of dream reports contain visual and, to a lesser extent, kinesthetic elements, the presence of other sensory modalities has also been noted in both laboratory and home dream reports.[47,94] More than 50% of dream reports contain auditory experiences, whereas explicit references to olfactory, gustatory, and pain sensations occur in less than 1% of all dream reports. One study[94] found that women's dream reports were more likely to contain olfactory or gustatory sensations, whereas references to auditory and pain experiences occurred in a higher percentage of men's dreams. That the more infrequent modalities of smell, taste, and pain occur at all in dreams is an important demonstration of the representational capacities of dreaming.

Perhaps because of the highly visual nature of dreaming, people always have wondered if blind people dream, so some of the earliest systematic interview studies on dreams dealt with this topic, showing that people who are born blind or become blind before age 4 or 5 years do dream even though they do not see images in their dreams,[95] a finding that was then supported by laboratory studies.[96] Nor is there much if any difference in dream content, except that there may be less aggression in their dreams.[8,95] There is also much greater mention of touch, taste, and smell in blind people's dreams.[8] It is noteworthy that people who become blind after age 5 or 6 years often have visual imagery in their dreams, which suggests that there is a window for the development of the capacity to have visual dreams that parallels what was found in longitudinal studies of children aged 3 to 7 years.[97]

Implications for Theories of Dreaming

The array of systematic results presented here suggests that a considerable amount of psychological information can be extracted from dream reports. This conclusion provides support for the core idea of all twentieth century dream theories, but it must be stressed that much dream content is not yet understood. The findings also suggest that most dreams focus on a handful of personal concerns revolving around social interactions with family, friends, and coworkers. The greatest variability in dream content seems to concern the appearance of aggression, especially physical aggression.

Despite the originality and creativity that is displayed in the cognitive production of dreams, and even given the aspects

of dream content that are not understood, most dreams are more realistic and based in everyday life than is suggested by most traditional dream theories. In addition, much dream content seems more transparent than might be expected by older clinical theories that emphasize disguise or symbolism in understanding dreams. Finally, a significant minority of dreams may not be as emotionally based as theories imply, especially before the adolescent years.

As a starting point, perhaps dreams are best understood as embodied simulations that enact the person's main conceptions and concerns, including emotionally salient interpersonal preoccupations. This type of conceptualization is at the heart of the continuity hypothesis, which posits a relationship between everyday dream content and general waking states and concerns. Although most of the research findings reviewed here are consistent with the continuity hypothesis, much work remains to be done to clarify which specific dimensions of waking life (e.g., particular learning tasks, daily mood, major life events, ingrained behaviors, sustained fantasies, cognitive styles) are most robustly associated to what kind of dream content and the nature of these relationships over time.[52] In fact, instead of referring to *the* continuity hypothesis, it may be more appropriate to consider multiple levels of continuity between various waking and dream parameters and to be alert for discontinuities as well. The observations that dreams rarely depict episodic memories and that the nature of temporal references in dreams can take many forms add a layer of complexity to an already difficult problem. In the end, dreaming may or may not have a function, but data convincingly show that dream content is a unique and meaningful psychological product of the human brain, and as such, dreams will continue to interest and challenge clinicians and researchers alike.

CLINICAL PEARLS

- Examining a series of dreams often yields more meaningful information about a patient's psychological state than focusing solely on one particularly salient dream.
- Changes over time in the frequency and content of repetitive dream themes, especially those involving strong affect and social interactions, are most likely to reflect a patient's clinical progress or deterioration.

SUMMARY

This chapter reviews methodologic issues in dream research and systematic findings on the content of people's dreams, and it presents the implications of key findings on normative dream content. Quantitative data on the content of laboratory and home dream reports converge in depicting a reliable picture about the nature of dream content in the general adult population as well as its development in children. At the most general level, these findings indicate that dreams show systematic relationships to various dimensions of the dreamer's waking life and suggest that many dreams are the embodiment of thoughts through dramatizations of life concerns and interests. That a wide range of psychological information can be extracted from dream reports has implications for clinical, theoretical, and empirical approaches to the study of dreams.

Selected Readings

Brown RJ, Donderi DC. Dream content and self-reported well-being among recurrent dreamers, past-recurrent dreamers, and nonrecurrent dreamers. *J Pers Soc Psychol* 1986;**50**:612–23.

Bulkeley K. Digital dream analysis: a revised method. *Conscious Cogn* 2014;**29**:159–70.

Dale A, Lortie-Lussier M, De Koninck J. Ontogenetic patterns in the dreams of women across the lifespan. *Conscious Cogn* 2015;**37**:214–24.

Domhoff GW. *The scientific study of dreams: neural networks, cognitive development, and content analysis*. Washington, DC: American Psychological Association; 2003.

Domhoff GW, Schneider A. Similarities and differences in dream content at the cross-cultural, gender, and individual levels. *Conscious Cogn* 2008;**17**:1257–65.

Gauchat A, Séguin JR, McSween-Cadieux E, Zadra A. The content of recurrent dreams in young adolescents. *Conscious Cogn* 2015;**37**:103–11.

Han HJ, Schweickert R, Xi Z, Viau-Quesnel C. The cognitive social network in dreams: transitivity, assortativity, and giant component proportion are monotonic. *Cognitive Science* 2015:1–26.

Malinowski JE, Horton CL. Memory sources of dreams: the incorporation of autobiographical rather than episodic experiences. *J Sleep Res* 2014;**23**:441–7.

Robert G, Zadra A. Thematic and content analysis of idiopathic nightmares and bad dreams. *Sleep* 2014;**37**:409–17.

Sándor P, Szakadát S, Kertész K, Bódizs R. Content analysis of 4 to 8 year-old children's dream reports. *Front Psychol* 2015;**6**:534.

Schredl M. Questionnaires and diaries as research instruments in dream research: methodological issues. *Dreaming* 2002;**12**:17–26.

Sikka P, Valli K, Virta T, Revonsuo A. I know how you felt last night, or do I? Self- and external ratings of emotions in REM sleep dreams. *Conscious Cogn* 2014;**25**:51–66.

A complete reference list can be found online at ExpertConsult.com.

Brain Correlates of Successful Dream Recall

Luigi De Gennaro; Michele Ferrara

Chapter 50

Chapter Highlights

- The scientific study of dreaming grew up rapidly in the 1950s, after the discovery of rapid eye movement sleep, which was considered to be a neurophysiologic marker of dreaming. Subsequent and more accurate investigations, however, showed that qualitatively different mental experiences are developed during all sleep stages.
- The recent progress with electrophysiologic (high-density electroencephalogram) and imaging techniques seems promising to elucidate the functioning of the brain structures presumably involved in the elaboration of dreaming, both in healthy subjects and in patients with acute or chronic brain damage, neurodegenerative diseases, or sleep disorders.
- This chapter summarizes the established findings and emerging knowledge on the neural correlates of successful dream recall. Lines of recent evidence converge in indicating that dream recall is related to specific electroencephalographic frequencies and to particular topographic features of the scalp electroencephalogram. These results are consistent with neuropsychological data, pointing to an overlap between functional and structural cerebral substrates of waking and rapid eye movement sleep mental imagery.

This chapter surveys the main findings and the rising knowledge of the neural correlates of successful dream recall. Owing to the intrinsic impossibility of direct access to dream *generation*, necessity forces recourse to indirect access through the collection of data on dream *recall*. In this context, the specific question to be addressed in this chapter is twofold. On the one hand, a common *general* question in cognitive neuroscience concerns the definition of the neural substrates of a specific cognitive process (in this case, dreaming). On the other hand, a second focus is on the *specific* problem of how tight the association is between functional/structural brain measures and dream experience itself. The lack of a direct measurement of the physiologic scenario in which dreams are generated implies that correlative findings on successful recall may be interpreted in terms of basic individual characteristics—defined in this chapter as *trait-like* differences—or as a direct consequence of neural functioning in a specific physiologic state—that is, *state-like* differences. For these reasons, different approaches to the question may be useful and produce different insights on the general issue: (1) quantitative scalp electroencephalogram (EEG) studies provide elecctrophysiologic data, with good temporal resolution; (2) intracerebral recordings in patients with pharmacoresistant epilepsy are informative regarding the spatiotemporal dynamics of subcortical electrencephalographic correlates; (3) microstructural analyses by magnetic resonance imaging (MRI) brain scans and by diffusion tensor imaging (DTI) analysis of MR images allow measurement of interindividual differences in brain tissue; and (4) these last measures can be considered as a sort of counterpart, in healthy subjects, of the neuropsychologic evidence, which describes behavioral and cognitive changes in patients with brain injury. Converging evidence and new research techniques from across the neurosciences can be expected to contribute strategies to overcome some intrinsic difficulties in directly accessing dream generation.

ELECTROENCEPHALOGRAPHIC CORRELATES

Scalp Measures

The study of neural correlates of dream recall has coincided for a long time with evaluating the association between the presence or absence of recall on awakening and the cortical electrophysiologic findings in sleep intervals preceding the awakening, as measured by quantitative scalp EEG studies. The assumption is that dream reports and failure to recall after awakening are linked by temporal contiguity to the scenario in which the dream experience happens, supporting the correlative hypothesis. On the other hand, no independent control study has confirmed that sleep mentation, as evaluated by a postawakening protocol, actually refers to a dream experience obtained in close temporal proximity to the awakening. As a further limitation of this approach, the empiric findings have been obtained both during naps and during nocturnal sleep. Independently from the actual evidence of chronobiologic effects on dream production,[1,2] this aspect of the study undoubtedly implies a possible confounding role of circadian/ultradian factors.

Given these caveats, in recent years multielectrode recordings have allowed a gain in spatial resolution, along with the possibility of comparing EEG findings with those provided by other neuroimaging techniques. Accordingly, the

investigation of EEG correlates of successful dream recall has evolved by addressing two general questions:

1. Are differences in dream recall–no recall related to specific EEG frequencies?
2. Does dream recall–no recall show scalp EEG topographic differences?

The alpha EEG activity on the one hand in the temporoparietal area on the other turned out to provide a plausible answer to these questions. Accordingly, increased EEG activity in the alpha range (between 11.72 and 13.67 Hz) is associated with the absence of sleep-onset rapid eye movement (REM) period dreams and the appearance of sleep-onset non-REM (NREM) period dreams.[3] Moreover, a negative correlation between alpha activity at the right central site and a measure of visual activity in dream reports has been reported after nocturnal awakenings, irrespective of sleep stage of awakening.[4] This association was partly confirmed in nocturnal NREM and REM awakenings.[5] In both sleep stages, dream recall was associated with lower alpha power, although of lesser magnitude in REM sleep. Independently from sleep stage, the alpha reduction was mainly localized over temporoparietal areas temporoparietal areas. The specific involvement of alpha activity in stage 2 NREM dream mentation was subsequently confirmed by Marzano and colleagues,[6] who also found that frequency of dream recall after awakenings from stage 2 was linearly correlated with a lower level of alpha oscillatory activity in the right temporal area (Figure 50-1, *left*). On awakening from REM sleep, a higher level of frontal theta activity was linearly correlated with dream recall rate (Figure 50-1, *right*). By this evidence, two stage-specific EEG activities were predictive of a successful dream recall at morning awakening, and the relations were topographically specific.

One of the limitations of the quantitative EEG approach is that the presence of a spectral peak resulting from a fast Fourier transform analysis of the EEG signal does not necessarily imply an underlying oscillatory activity at that frequency.

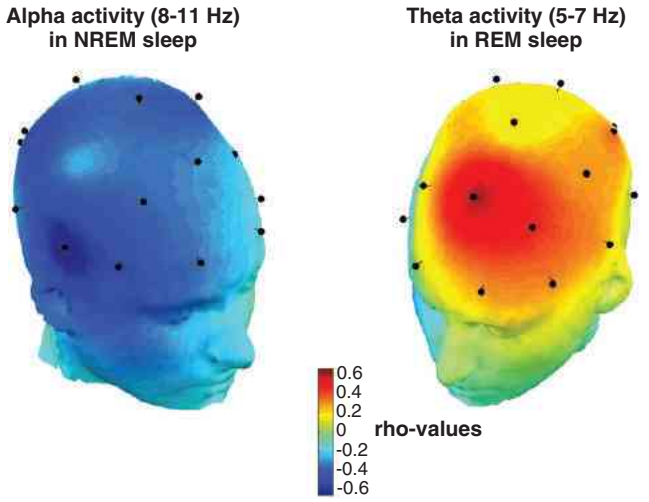


Figure 50-1 A three-dimensional topographic distribution of correlation values (rho) between the number of dreams recalled upon morning awakenings and (1) the amount of EEG theta activity in REM sleep (*right side*) and (2) the amount of EEG alpha activity in NREM sleep (*left side*). EEG, Electroencephalogram. (Data from Marzano C, Ferrara M, Mauro F, et al. Recalling and forgetting dreams: theta and alpha oscillations during sleep predict subsequent dream recall. *J Neurosci* 2011;31:6674–83.)

This limitation may indeed be relevant in light of findings, obtained by scalp recordings during wakefulness, showing that low-frequency brain oscillations increase during encoding of episodic memories.[7-9] The hypothesis that the encoding of dream content during sleep shares some electrophysiologic mechanisms with the successful encoding of episodic information during wakefulness implies a further (different) approach to scalp EEG during sleep. The application to the EEG signals of a method (Better OSCillation [BOSC][10]) to detect oscillatory activity within a signal containing a nonrhythmic portion provided empiric support for that notion. The results of Marzano and associates[6] showed an involvement in dreaming of those regions that control successful memory encoding in waking by detecting higher oscillatory theta activity at frontal areas after awakening from REM sleep and lower temporo-occipital alpha oscillations after awakening from stage 2 sleep. It is intriguing that the cortical regions whose oscillatory activity predicts dream recall are reminiscent of those involved in the global cessation of dreaming after brain injuries (see later under Neuropsychological Correlates).

Although predictive relationships between EEG oscillations and dream recall have been reported as a function of the proximity to the time of awakening, the possibility that stable interindividual EEG patterns characterizing particular subjects explain these findings cannot be ruled out completely. In fact, high recallers are characacterized by larger neurophysiologic responses, as assessed by evoked potentials to auditory stimuli, than low recallers.[11] Of interest, this larger brain reactivity was found during both wakefulness and REM and NREM sleep stages, somehow with the general notion that brain correlates of dream recall share some electrophysiologic mechanisms of the awake brain. However, a definitive response to the question of whether EEG correlates of successful dream recall reflect state- or trait-like differences can be provided only by multinight within-subject studies, whereby preawakening sleep periods followed by successful versus failed dream recall could be compared.

In conclusion, most empiric evidence goes in the direction of a frequency- and topography-specific relation between EEG oscillations in the minutes preceding awakening and dream recall. Some evidence suggests that successful dream recall is associated with the specific electrophysiologic pattern in the minutes preceding awakening.[6] Theta and alpha oscillations are correlated with successful recall, with involvement of temporoparietal and frontal areas. This evidence is consistent with models of episodic memory in which theta oscillations allow a top-down control from the frontal cortex to the hippocampus, modulating the encoding and retrieval of episodic memories,[12] whereas the alpha activity may be functionally related to the activation of stored information.[13]

Therefore EEG studies suggest that the neurophysiologic mechanisms underlying the encoding and recall of episodic memories are the same across different states of consciousness.

Intracerebral Measures

If the EEG correlates of dream recall point to a crucial role for the temporoparietal and frontal cortices, the available knowledge on subcortical areas is limited. The recent application of intracerebral recordings (using stereo EEG) obtained in patients with pharmacoresistant epilepsy has provided some

insights regarding the spatiotemporal dynamics of subcortical EEG correlates of dream recall. In particular, mediotemporal lobe activity associated with dream recall has been investigated under the assumption that dream recall is a peculiar form of episodic memory encoded during sleep.[14] Thus the encoding of dream content could share some electrophysiologic mechanisms with successful episodic memory encoding of the awake brain, because phase synchronization of rhinal-hippocampal EEG activity in the gamma range and increased EEG coherence in the theta range are predictive of successful memory formation.[15,16] Actually, comparison of rhinal-hippocampal and intrahippocampal EEG connectivity between good and poor dream recallers awakened from REM sleep shows a strong relationship between rhinal-hippocampal and intrahippocampal EEG connectivity and the capability to recall dreams.[14] In particular, patients who recollected dreams showed a higher EEG coherence for all of the frequency bands investigated (from 1 to 44 Hz) than that in patients who did not, and this difference holds across wakefulness and NREM and REM sleep. Indeed, this higher connectivity was more evident in the low-frequency theta range, confirming previous data on waking memory formation.[15,16] Thus successful dream memorization and declarative memory formation seem to be associated with increased mediotemporal connectivity, independent of the physiologic state (i.e., wakefulness and REM and NREM sleep). Again, the question of whether a higher EEG connectivity associated with successful dream recall reflects state- or trait-like differences remains open.

Morphologic-Anatomic Measures

Individual volumetric and ultrastructural measures of subcortical nuclei, as determined by DTI, are stable over time and refelct interindividual differences in brain tissue.[17,18] Such measures may permit evaluation of the existence of trait-like individual characteristics of dream experience. Microstructural analyses—that is, of volume and diffusivity (magnitude of neuronal water diffusion)—by MRI brain scans and by DTI analysis of MR images have been used to investigate relationships between deep gray matter structures and dreaming.[19] In particular, this novel approach has been introduced for evaluating the relationships between dreaming and some anatomic measures of the hippocampus and amygdala, according to their possible role in the processing of mnestic and emotional sources of dream content. Although the hippocampus should mediate the partial reproduction of memories of events occurring during wakefulness for inclusion in dream content,[19] concomitant interest in the amygdala is justified by its involvement both in control of the encoding and retrieval of emotional memories and in the physical expression of emotions during wakefulness.[20] Furthermore, hippocampus and amygdala seem likely to affect the emotional quality of dream mentation, because they are involved in the processing and execution of fear memories,[21] and specific forms of learning involve the amygdala and hippocampus at different stages.[22-25] Finally, MRI/DTI analyses in subjects capable of reporting a dream on awakening from REM sleep show bilateral amygdalar activation.[26]

Volume of gray matter and microstructural alterations of gray matter in the hippocampus and in the amygdala, as expressed by reduced cellular barriers that restrict the free diffusion of water molecules in tissues, are not associated with successful dream recall. DTI measures of these nuclei do not show any correlation with dream recall rate.[19] Neuroanatomic measures are indeed related to some *qualitative* features of the recalled dreams (emotional load, bizarreness, and vividness) and, to some extent, with the length of dream reports. A decreased microstructural integrity of the left amygdala is associated with shorter dream reports and decreased emotional load, whereas that of the right amygdala with a lower index of bizzarreness. Left amygdala volume also is related to a lower bizarreness index. Although indeed significant, relationships with ultrastructural measures of hippocampus are weaker.[19]

The main point is that functional neuroanatomy of the amygdala and the hippocampus indicates a dissociation between some quantitative and qualitative aspects of dream reports. Although the mean number of dreams recalled per day does not show any significant relationship with the aforementioned neuroanatomic measures, significant associations with some qualitative features of the recalled dreams are observed.[19] Within this framework, successful dream recall does not reflect trait-like differences, at least with respect to gray matter of the amygdala and hippocampus.

NEUROPSYCHOLOGICAL CORRELATES

The study of behavioral and cognitive changes after accidental brain injury or neurosurgery, or associated with neurologic diseases that affect particular brain regions, represents the classical method for establishing brain-mind relationships in humans. This approach, as applied to the phenomenology of dreaming, has been called the "neuropsychology of dreaming." Although initial observations focused on the brainstem structures, on the basis of the erroneous assumption that the high correlation between REM sleep and dream recall rate necessarily implied that brainstem mechanisms also were responsible for dream generation, in the past several decades investigations have been directed at the consequences of cortical damage.

Subcortical Lesions

Because REM sleep is controlled by pontine brainstem areas,[27] and their lesions suppress or reduce human REM sleep,[28] a concomitant suppression or reduction of dream experience was expected. Solms[29] reviewed the published cases of brainstem lesions, which provided some information on the dream experience. In only one case was a drastic reduction in REM sleep accompanied by a cessation of dreaming.[30] In the other 25 cases, the hypothesized correlation was not found (or it was not taken into account).[28,31-34]

However, the preservation of dreaming in cases with obliteration of REM sleep due to brainstem lesions has not been satisfactorily demonstrated. In fact, preserved dreaming after brain trauma involving large pontine lesions has been reported in only four patients, in the absence of polygraphic monitoring.[35] This approach is partly limited by the fact that pontine lesions extensive enough to significantly affect REM sleep usually do not preserve consciousness in the affected subject.[36]

A reasonable conclusion, therefore, is that pontine brainstem lesions in humans do not abolish dreaming, and that REM sleep is not a necessary prerequisite for the occurrence of dreams. This conclusion also is consistent with findings on pharmacologic suppression of REM sleep. Indeed, phenelzine therapy for depressed patients, which eliminates REM sleep

without altering slow wave sleep, does not abolish or coherently affect the frequency of dream recall.[37]

Cortical Lesions

The first indications of a complete (or nearly complete) loss of dreaming, associated with localized lesions in the forebrain with complete preservation of brainstem, date back to approximately 125 years ago. These clinical reports described two patients who dreamed "almost not at all anymore" after sustaining, respectively, a bilateral occipital-temporal[38] and a bilateral occipital[39] lesion.

A century later, subsequent reevaluation of data for 104 patients with brain lesions who reported information about their dreams[40] indicated that the presence of frontal lobe damage was not systematically associated with loss of dreaming, whereas lesions in the parietal lobes and lesions associated with disconnective syndromes could cause a loss of dreaming without notable hemispheric asymmetry. Moreover, dream cessation after unilateral left or right brain damage was as frequent as after bilateral damage. Because a lesion in either hemisphere could be sufficient to cause dream loss, no simple relation between either hemisphere (i.e., unilateral brain lesions) and dreams could be postulated. Accordingly, it was proposed that the right hemisphere provides the core material for the dreams, whereas the left hemisphere provides the means of decoding it.[40]

The foundation of the neuropsychology of dreaming, however, should be ascribed to Mark Solms, who examined 361 patients with brain injuries.[29,35] He found that cessation of dreaming (called *global anoneria*) follows damage to two different systems: (1) a posterior system, centered in or near the parieto-temporo-occipital (PTO) junction, and (2) an anterior system, mostly bilateral, located in the ventromedial prefrontal cortex and including the white matter surrounding the anterior horns of the lateral ventricles. Lesions of this second system, however, were not invariably associated with dreaming cessation.

As far as the posterior system is concerned, unilateral (or, in a few cases, bilateral) injuries in or near the PTO junction were associated with a complete loss of dreaming, suggesting that this area might be essential for dreaming itself.[29] This finding is crucial, because it has been long recognized that the cortical network for spatial representation is centered in the inferior parietal lobe. The PTO junction supports various cognitive processes that are crucial for visual memory and mental imagery.[41]

Of note, lesions of the posterior system may affect dreaming,[35] waking visual mental imagery,[42] and waking visuospatial abilities.[43] Also of note, visual imagery, in turn, shares approximately two thirds of activated brain areas with visual perception.[42,44] Finally, a complete "anoneria" (see further on) with preservation of REM sleep also has been reported, in a case of deep occipital lobe damage (including the right inferior lingual gyrus) secondary to bilateral occipital artery infarction,[45] as well as after left temporo-occipital injuries.[46] These findings suggest that global cessation of dreaming can be associated with temporo-occipital lesions, even without parietal involvement.

The basic idea that dream experience, mental imagery, and late stages of visual perception may share some neural mechanisms also is strengthened by the notion that lesions of more specific regions, such as V4 or V5, may selectively affect dream representation of color[29] or movement.[43] Specifically, patients who are unable to generate facial and color imagery in waking life (as a consequence of V4 lesions) also cannot generate faces or colors in their dreams[29,47-54]; a decrease in the "vivacity" of dreaming was reported by two patients with damage to areas V3, V3a, and V4.[35]

With specific reference to the visual component of dreams, Doricchi and Violani[40] first proposed a second nosologic syndrome, *nonvisual dreaming,* subsequently termed "visual anoneria" by Solms.[35] In this entity, complete or partial loss of visual dream imagery occurs only with lesions in visual association cortex (with preservation of normal waking vision). Of interest, visual anoneria is accompanied by the inability to produce mental imagery in waking (irreminiscence), further supporting the existence of a common neural substrate. Lesions in primary visual cortex, however, have no effect on dreams. For example, visual dream imagery is intact in cortically blind patients (with V1 or V2 lesions). In addition, partial variants of visual anoneria have been described, characterized by selective loss of specific visual elements (e.g., kinematic anoneria, facial anoneria).

By contrast, dream imagery seems unaffected by lesions of primary unimodal sensomotor cortices.[29] Hemiplegic patients (with unilateral perirolandic lesions) show preserved somatosensory and motor imagery in their dreams.[35,55,56] Similarly, aphasic patients (with left perisylvian lesions) experience normal audioverbal and motor speech imagery in their dreams.[35,57,58] These findings suggest that somatosensory, motor, audioverbal, and speech imagery in dreams is generated outside of the respective unimodal cortices, probably in heteromodal paralimbic or PTO cortex.[29]

Lesions of the second (anterior) system are less frequently associated with global cessation of dreaming. Typically, global anoneria is a consequence of bilateral damage to white matter tracts that surround the anterior horns of the lateral ventricles, underlying ventromedial prefrontal cortex. Many of these nerve fibers originate or terminate in the limbic system. The ventromedial white matter contains dopaminergic projections to the frontal lobe, which are disconnected as a consequence of prefrontal leukotomy.[59] Thus a 70% to 90% incidence of global (or near-complete) cessation of dreaming has been reported in patients who underwent prefrontal leukotomy.[29] On the other hand, lesions in the dorsolateral prefrontal cortex that cause waking deficits of self-monitoring and decision making have no effect on dreaming.[60]

Finally, two types of "dreaming excess" also have been described.[35] Lesions in medial prefrontal cortex, the anterior cingulate cortex, and the basal forebrain are associated with increased frequency and vividness of dreams, with intrusion of content into waking life. In particular, patients with frontal-limbic lesions may exhibit the syndrome referred to as *anoneirognosis,* characterized by difficulty in distinguishing between internally generated experiences, such as dreams, from externally driven perceptions. Second, the syndrome of "recurring nightmares" is characterized by frequent nightmares with a repetitive theme. Increased frequency of nightmares often was associated with temporal-limbic seizure activity, with stereotypical nightmares accompanying complex partial seizures in some cases.

To summarize, the evaluation of brain-damaged patients seems to support the hypothesis that dreaming is not an intrinsic function of REM sleep (or of the brainstem

mechanisms that control it). Moreover, neuropsychological evidence confirms that the neural network supporting dreaming has considerable specificity. In fact, the posterior and anterior systems just described are crucial to dream imagery production, whereas brain-damaged patients do not show significant changes in dreaming when lesions are located outside those systems.

The crucial involvement of the PTO junction and of the ventromedial prefrontal cortex in dreaming generation is further supported by recent neuroimaging findings.[61] Regional cerebral blood flow (rCBF), measured with $[^{15}O]H_2O$ positron emission tomography, was higher in the temporoparietal junction in subjects with high dream recall frequencies (HRs) than in "low recallers" (LRs) during REM sleep and deep NREM sleep (stage N3). Moreover, HRs showed a higher rCBF than LRs in the medial prefrontal cortex during REM sleep. These neurofunctional correlates of dream recall in healthy humans are consistent with the aforementioned neuropsychological findings showing that lesions involving the PTO junction and the ventromedial parts of the prefrontal cortex lead to a global anoneria.[35]

Of interest, a different functional organization of the brain in HRs and LRs also was shown while subjects were resting in the scanner, when the rCBF was higher in HRs than in LRs in both the temporoparietal junction and medial prefrontal cortex. Because both of these regions are part of the default mode network,[62] it has been suggested that the neural substrate of dreaming could be a subsystem of this network.[61,63] In particular, the higher level of temporoparietal junction and medial prefrontal cortex activity in HRs may be associated with increased wandering of the mind and increased involvement in episodic memory recall and evaluative processing.[64] In keeping with this notion, an event-related potential study pointed out that HRs might be more reactive to the environment than LRs.[11] Specifically, the amplitude of the cortical response to first names, an attention-orienting brain response known to recruit temporoparietal junction, was higher in HRs and correlated with the total duration of intrasleep wakefulness. The increased duration of intrasleep wakefulness in HRs could in turn explain their high dream recall frequency, as already suggested by Koulack and Goodenough.[65]

Therefore differences in spontaneous brain activity in the temporoparietal junction and medial prefrontal cortex during both sleep and wakefulness may be the neurophysiologic substrate of a low or a high frequency of dream recall. The increased activity in the temporoparietal junction seen in HRs may facilitate the orienting of attention toward external stimuli and promote intrasleep wakefulness, facilitating by that means dream production and/or dream memory.

In conclusion, neuropsychological data suggest that dream experience and waking mental imagery may share a common neural substrate. Although the continuity between neural mechanisms of waking and sleep mentation is particularly clear if the characteristics of the posterior system, directly linked to an advanced stage of visual (and, more generally, sensory) processing, are taken into account, it may appear less evident in considering the involvement of the mesocortical-mesolimbic dopamine system (i.e., the anterior system), which also seems to play a causal role in the generation of dreams. Nevertheless, the increased activation of medial prefrontal cortex, occipitotemporal visual cortex, and anterior cingulate cortex during REM sleep reported by the (few) functional neuroimaging studies of REM sleep[6] is consistent with the results of the previously discussed lesion studies, supporting the importance of those brain areas in dream generation. Moreover, a large portion of the dorsolateral prefrontal cortex is deactivated during REM,[26,60,66] and lesions in this region do not affect dreaming.[35,40] Taken together, these findings are consistent with a continuity model of the cerebral substrates of waking and REM sleep mental imagery (see Chapter 54).

CONCLUSIONS AND FUTURE DIRECTIONS

Within the specific methodologic limits of dream research, successful recall is associated with a relatively well-defined series of functional and anatomic changes. Cortical EEG oscillations in the minutes preceding the awakening can predict dream recall rate, suggesting that the electrophysiologic changes underlying the encoding and recall of episodic memories are the same across different states of consciousness.[6] Similar conclusions can be derived from intracerebral recordings showing an increased mediotemporal connectivity associated with dream recall, as well as with successful declarative memory formation in waking.[14] In addition, evidence coming from studies in patients with cerebral lesions strengthens the view of an overlap between the cerebral substrates of waking and REM sleep mental imagery.[29,35]

Nevertheless, electrophysiologic and neuropsychological findings do not disentangle the state-/trait-like aspects of these neural correlates. Preliminary evidence from structural MRI scans[19] does not support the structural hypothesis, that is, the hypothesis of basic neural ("trait-like") differences, because no relation is evident between dream recall rate and morphologic-anatomic measures of the hippocampus and the amygdala. Longitudinal studies or, at least, multiawakening observations in the same subject are needed to respond to the functional-structural question.

CLINICAL PEARL

Limited evidence is available to support the view of involvement of the dopaminergic system in dream recall. This lack of substantiation should be taken into consideration by physicians otherwise inclined to prescribe compounds such as varenicline (i.e., for smoking cessation) or dopaminergic agents (i.e., in Parkinson disease). In both cases, dream experiences may be exacerbated, and nightmares can be induced. When possible, alternative treatments are preferred.

SUMMARY

Successful dream recall is associated with the specific electrophysiologic patterns during the minutes preceding the subject's awakening. Quantitative EEG analyses indicate that theta oscillations in REM sleep and alpha oscillations in NREM sleep are related to successful recall, with crucial involvement of temporoparietal and frontal areas. Moreover, analysis of intracerebral recordings in patients with pharmacoresistant epilepsy confirmed that phase synchronization of rhinal-hippocampal EEG activity characterizes both successful dream memorization and declarative memory formation. Therefore the encoding of dream content during sleep evidently shares some electrophysiologic mechanisms

with the successful encoding of episodic information during wakefulness.

The continuity between neural mechanisms of waking and sleep mentation is further supported by neuropsychological evidence. The evaluation of brain-damaged patients indicates that a "posterior system," centered in or near the PTO junction, and an "anterior system," located in the ventromedial prefrontal cortex, are crucial to dream imagery production. The critical involvement of the PTO junction and of the ventromedial prefrontal cortex in dream generation is further supported by recent neuroimaging findings showing that differences in spontaneous brain activity in the temporoparietal junction and medial prefrontal cortex during both sleep and wakefulness may be the neurophysiologic substrate of a low or high frequency of dream recall. Altogether, the available evidence is consistent with a continuity model of the cerebral substrates of waking and REM sleep mental imagery.

Selected Readings

De Gennaro L, Marzano C, Cipolli C, Ferrara M. How we remember the stuff that dreams are made of: neurobiological approaches to the brain mechanisms of dream recall. *Behav Brain Res* 2012;**226**:592–6.

Doricchi F, Violani C. Dream recall in brain-damaged patients: a contribution to the neuropsychology of dreaming through a review of the literature. In: John S, Antrobus MB, editors. *The neuropsychology of sleep and dreaming*. Hillsdale (N.J.): Lawrence Erlbaum Associates; 1992. p. 99–129.

Fagioli I. Mental activity during sleep. *Sleep Med Rev* 2002;**6**:307–20.

Hobson JA, Pace-Schott EF, Stickgold R. Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behav Brain Sci* 2000;**23**: 793–842.

Muntean ML, Trenkwalder C, Walters AS, et al. REM sleep behavioral events and dreaming. *J Clin Sleep Med* 2015;**11**(5):537–41.

Nielsen TA. Chronobiological features of dream production. *Sleep Med Rev* 2004;**8**:403–24.

Nir Y, Tononi G. Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn Sci* 2010;**14**:88–100.

Perogamvros L, Schwartz S. The roles of the reward system in sleep and dreaming. *Neurosci Biobehav Rev* 2012;**36**:1934–51.

Scarpelli S, D'Atri A, Gorgoni M, et al. EEG oscillations during sleep and dream recall: state- or trait-like individual differences? *Front Psychol* 2015;**6**:605.

Schredl M. Dreams in patients with sleep disorders. *Sleep Med Rev* 2009;**13**:215–21.

Solms M. *The neuropsychology of dreaming: a clinico-anatomical study*. Hillsdale (N.J.): Lawrence Erlbaum Associates; 1997: Reprinted: London, Psychology Press 2014.

Stickgold R, Hobson JA, Fosse R, et al. Sleep, learning, and dreams: off-line memory reprocessing. *Science* 2001;**294**:1052–7.

A complete reference list can be found online at ExpertConsult.com.

Chapter 51

Neurobiology of Dreaming

Edward F. Pace-Schott; Dante Picchioni

Chapter Highlights

- This chapter reviews the sleep stage, electrophysiologic, and regional brain activity correlates of dreaming, with a focus on rapid eye movement (REM) sleep observations. REM sleep differs from non–rapid eye movement (NREM) sleep in ways that suggest bases for its enhanced dreaming, including increased fast electroencephalography oscillation and decreased anteroposterior electroencephalography coherence. Neuroimaging studies show widespread decrease in brain activity with the transition from waking to NREM, followed by reactivation, in REM, of midline limbic and paralimbic brain areas to levels that equal and even exceed those of waking. Notably, however, lateral multimodal association cortices retain NREM-like low activity in REM, suggesting possible bases for the unique nature of dream cognition.
- The chapter also reviews functional connectivity methods and the relationship between dreaming and internally cued cognition during wakefulness. Key sleep studies, with an emphasis on REM sleep studies, are presented under the relevant network. This is important because information integration across a large number of brain regions has become a central tenet in theories of consciousness, and alterations in this integration resulting from the absence of core nodes or increased randomness may explain the alterations in consciousness that accompany sleep and dreaming.
- Striking changes in the forebrain neuromodulation take place in the transition from wake to NREM and thence to REM. Levels of cholinergic and aminergic modulators diminish from waking to NREM, whereas the levels of acetylcholine alone return to waking levels during REM. Observation of these changes contributed to the first neurobiologic theory of dreaming, activation synthesis, in which an elevated cholinergic-to-aminergic ratio accounts in part for differences between dreaming and waking cognition. Other dream theories focus on dopamine and activation of reward networks in concert with other midline regions in REM.
- Dreaming shares phenomenology with abnormal waking states such as spontaneous confabulation, in which imagined scenarios are accepted as veridical memories, and complex hallucinosis, in which fully formed fictive visual images are perceived. Examinations of underlying neuroanatomy and neurochemistry of these conditions reveal overlap with those of REM sleep.
- Observations from neuroimaging of sleep, cognitive neuroscience, and clinical neuropsychology allow us to construct a putative model of neurobiologic processes that generate dream phenomena, including restored conscious awareness, altered emotion and memory, fictive movement, complex visual hallucinations, fictive space, and impaired executive function.

Dreaming is a universal human experience occurring during sleep in which fictive events follow one another in an organized, storylike manner and into which are woven hallucinatory, primarily visual, images that are largely congruent with an ongoing confabulated plot. Most often, this wholly imaginary experience is uncritically accepted in the same manner as are veridical waking percepts and events.

THE ASSOCIATION OF DREAMING WITH BEHAVIORAL STATE

Early speculation that rapid eye movement (REM) sleep was the exclusive physiologic substrate of dreaming[1] was soon followed by awakening studies showing substantial recall of mental experiences from non–rapid eye movement (NREM) sleep.[2] Nonetheless, REM sleep reports are more frequent, longer, more bizarre, more visual, more motoric, and more emotional than are NREM sleep reports.[3] In an extensive review, Nielsen estimates an NREM sleep mental experience recall rate of 42.5%, contrasting with 81.8% from REM sleep, and suggests that brain activation processes occurring outside polysomnographically scored REM sleep ("covert REM") may account for NREM-sleep dreaming.[4]

Recent Electrophysiologic Findings

Fast and Slow Oscillations and Dreaming

REM sleep shows much more gamma frequency (30 to 80 Hz) fast brain waves ("oscillations" or "rhythms") than does

NREM sleep as measured by scalp electroencephalography (EEG),[5] intracranial EEG,[6] and magnetoencephalography (MEG).[5] In waking, these fast oscillations are associated with attention to stimuli and other forms of active or effortful cognition.[7] During REM-sleep dreaming, fast oscillations have been hypothetically associated with cognitive and perceptual processing,[8] memory processing,[9] and the temporal binding of dream imagery.[10]

NREM sleep is instead associated with slower oscillations produced by recurrent interactions between the thalamus and cortex (intrinsic "corticothalamocortical" rhythms) such as sleep spindles and delta waves and with the cortical slow (<1 Hz) oscillation that groups the other corticothalamocortical oscillations in time.[11,12] The slow oscillation consists of periods of neuronal quiescence (hyperpolarized or "down" states) that alternate with shorter periods of rapid neuronal firing (depolarized or "up" states).[13] Slow intrinsic oscillations may interfere with ongoing mental activity and lead to a lower frequency of dreams in NREM sleep.[3] Although human slow wave sleep (SWS) shows less sustained gamma activity compared to REM sleep and waking, intracranial EEG has shown that gamma oscillations appear during the transient "up" state of the slow oscillation.[14,15]

Electrophysiologic Connectivity in Sleep

Declines in phase synchrony ("coherence") of EEG rhythms between different brain regions occurring during sleep relative to waking may reflect functional disconnections that contribute to the cognitive features of dreaming. Gamma oscillations in REM sleep become desynchronized between anterior and posterior regions of the brain; a disconnection that may contribute to the hypofrontal and bizarre features of REM-sleep dreaming.[5]

Dreaming and Phasic Activity in Sleep

In cats, a close temporal association exists between REM-sleep rapid eye movements (REM-sleep saccades) and ascending potentials that originate in the brainstem and are termed *ponto-geniculo-occipital* (PGO) *waves*.[3] In the activation-synthesis hypothesis of dreaming,[16] Hobson and McCarley suggest that the brainstem's activation of the forebrain in REM sleep allows the forebrain to synthesize dream scenarios based on currently available information. They suggest that the PGO wave, originating in the pons and arriving at primary visual cortex by the dedicated visual pathway through the thalamic lateral geniculate nucleus, might be interpreted by the brain as visual information, thereby leading to the visual hallucinosis of dreams.

Early evidence of human PGO waves was derived from scalp EEG recordings temporally locked to REM-sleep saccades.[3,17] Subsequently, positron emission tomography (PET)[18] and function magnetic resonance imaging (fMRI)[19] studies also correlated REM-sleep saccades with activation of structures corresponding to the feline PGO wave. Compelling evidence for human PGO waves has recently emerged. First, using MEG, Ioannides and colleagues[20] showed that correlated phasic activity in the pons and frontal eye fields begins before an REM-sleep saccade and intensifies with increasing temporal proximity to saccade onset. Second, using depth electrodes in a patient with Parkinson disease, Lim and colleagues[21] described phasic signals, with waveform and temporal characteristics very similar to the feline PGO, originating in the pedunculopontine nucleus—a structure at the pons-midbrain (mesopontine) junction crucial for generating the feline PGO.[3] Third, using fMRI, Miyauchi and colleagues[22] demonstrated that activity in the pontine tegmentum, ventroposterior thalamus, and visual cortex takes place in the few seconds before REM-sleep saccades. Activation in these regions, relative to the REM-sleep saccade, corresponded both to the neural pathway and time course of the feline PGO wave, that is, pontine tegmentum (–4.7 seconds), ventroposterior thalamus (–3.8 seconds), and primary visual cortex (–2.8 seconds).

The scanning hypothesis, which posits a correlation between REM-sleep saccades and the direction of hallucinated gaze in dreams,[23] continues to generate interesting but conflicting findings. For example, in an event-related potentials study, REM-sleep saccades occurred without the readiness potential that preceded waking saccades, whereas a wakelike potential reflecting visual engagement persisted, suggesting that REM-sleep saccades may trigger rather than follow visual experiences.[24] However, Miyauchi and colleagues[22] observed primary visual cortex activation *before* REM-sleep saccades that, therefore, could not have been triggered by efferent copies from neural activity in the frontal eye fields associated with the saccade, as an alternate explanation might suggest, but rather that REM saccades may have been in response to PGO-initiated dreamed visual imagery.

Convergent evidence does, however, suggest activation of limbic structures in concert with REM-sleep saccades. For example, using low-resolution brain electromagnetic tomography, Abe and colleagues[25] observed a pre-REM-sleep saccade potential with current sources estimated to lie in anterior limbic regions. Using MEG, Ioannides and colleagues[20] also described REM-sleep saccade onset linked current sources estimated to lie in the amygdala and orbitofrontal and parahippocampal cortices. Both groups suggest that this pre-saccade limbic activity reflects phasically enhanced emotional processing.[20,25]

In summary, EEG and MEG studies suggest that brain activity accompanying phasic REM sleep may be related to dream phenomena. Such phenomena include visual imagery,[21,22,24,26] enhanced cognitive activity and attention,[5,8] decoupling of executive control and perception,[5] and enhanced emotional processing.[5,20,25]

Dreaming and Brain Activity

Deactivation of frontal cortices is one of the first signs of human sleep observed using EEG, MEG, and functional neuroimaging.[17,27] PET studies of NREM sleep show declines in brain activity relative to waking both globally[27] and in many specific regions of the subcortex and cortex,[28,29] findings now replicated using fMRI.[30] Global and regional cerebral activity further declines with the deepening of NREM sleep.[27,29-31] Following sleep onset, EEG studies show greater slow wave spectral power in frontal versus posterior sites.[32] Synchronization of slow waves then spreads progressively to posterior regions,[33] a trajectory also traveled by the slow (<1 Hz) oscillation.[34]

Compared with these deactivated NREM sleep conditions, in REM sleep there is a prominent increase of neural activity in subcortical brain regions, including the pons and midbrain,[28,35] thalamus,[28,35] basal ganglia,[28] and limbic subcortex comprising the amygdala,[35] hypothalamus, and ventral

striatum.[28] Increases are also seen in limbic-related cortices anteriorly in the rostral and subcallosal anterior cingulate,[28,35] anterior insula, more posterior orbitofrontal and paracingulate Brodmann area (BA) 32 cortices, BA 10 in medial prefrontal cortex (PFC),[28] and more posteriorly in the parahippocampal gyrus and temporal pole.[28] Certain visual association cortices (regions that process higher-order aspects of vision) are also active.[28,36] However, multiple neuroimaging modalities show that lateral prefrontal cortices remain deactivated after the transition from NREM to REM sleep.[28,35-37]

When REM sleep is directly compared with waking, there is relative deactivation of the lateral PFC.[28,35,37] Maquet and colleagues[37] showed that regions most consistently hypoactive in REM sleep compared with waking include middle and inferior frontal gyri as well as inferior parietal and temporal-parietal junction association cortices. However, compared with waking, in REM sleep there is *greater* activation of limbic and paralimbic regions.[28,35,38-40] Nofzinger and colleagues[39,40] have termed this region the "anterior paralimbic REM activation area" and describe it as a "bilateral confluent paramedian zone which extends from the septal area into ventral striatum, infralimbic, prelimbic, orbitofrontal, and anterior cingulate cortex" (p. 192)[40] and includes the hypothalamus, ventral pallidum, hippocampus, and uncus as well as supplementary motor, pregenual and subgenual anterior cingulate, and insular cortices.[39]

Neuroimaging Has Renewed Interest in Lucid Dreaming

Lucid dreaming[41] has reemerged as an exciting new tool in fMRI studies by offering a degree of real-time access to the dreaming brain. Because, unlike skeletal muscles, oculomotor muscles maintain tonus in REM sleep, eye movements can still be produced, allowing the lucid dreamer to "signal out" of the dreaming state.[41] This technique was recently exploited to demonstrate that dreamed hand clenching activated the corresponding sensorimotor cortex.[42] Prior speculation that lateral PFC deactivation in REM-sleep dreaming contributes to the characteristic lack of insight in dreaming was supported by a recent study, using quantitative EEG, that showed a return of wakelike, gamma activity in the lateral PFC when lucidity was achieved.[43] A recent fMRI study has shown activation, during lucidity, of many regions that normally remain deactivated in nonlucid REM, including posterior midline, lateral parietal, prefrontal, and occipitotemporal regions.[44] A number of these regions are associated with self-reflection, self-evaluation, and volition, functions normally weak in dreams[45] but restored with lucidity.[44] Most recently, a remarkable study showed that the level of self-reflective insight could be increased during sleep in lucid dreamers by applying transcranial direct current stimulation to scalp areas overlying frontal regions of the brain.[46] Hence, in addition to emerging as a unique experimental tool, there now is strong evidence that lucid dreaming is indeed a state intermediate between normal REM-sleep dreaming and wakefulness.[43,44,46]

DREAMING AND BRAIN FUNCTIONAL CONNECTIVITY

When you observe the simultaneous activation of more than one region, those regions can be labeled as a functional network. However, mean changes in two regions' activities can occur independently from changes in the correlation between them. One can perform a "seed" region connectivity analysis. Activity in the seed region is correlated with all other voxels in the brain, so the resulting color map represents the strength and direction of each correlation with the seed region. It is critical to remember that a negative correlation indicates the same magnitude of connectivity as the equivalent positive correlation; a negative correlation is not disconnectivity. One can repeat the seed-region approach with many regions and present all pairwise correlations in a cross-correlation matrix. When considering the entire brain, one can statistically separate activity into spatially independent components or quantify the connectedness of the network as a whole using graph theory.[47]

Functional connectivity is typically measured during "rest" (i.e., unconstrained quiet wakefulness) and using blood oxygen level–dependent fMRI because it has greater temporal resolution than PET. When analyzing the spectral power of the spontaneous fluctuations that appear during rest, less than 0.1-Hz oscillations contribute most to functional connectivity correlations.[48] Functional connectivity networks have been shown to bear close correspondence with anatomic networks identified in nonhuman primates, postmortem studies, and MRI-based imaging of fiber tracts using diffusion-tensor imaging.[49-51] They also overlap with the same regions that are activated during execution of the corresponding task.[52] For example, during rest and in the absence of a language production task, a seed region in Broca's area will correlate with other nodes in the language network.

Obtaining REM sleep during fMRI is difficult, largely because of the acoustic noise associated with fMRI,[53,54] so there are only four fMRI studies of functional connectivity during REM sleep, and they all have relatively small sample sizes. The first study ($n = 3$) is unique because the investigators analyzed connectivity differences between tonic and phasic REM sleep.[53] Investigators from the second study ($n = 2$) used a seed in the posterior cingulate cortex.[55] The third study had the largest sample size ($n = 12$) and is unique because the investigators employed an adaptation night and simultaneous eyelid video monitoring to aid REM-sleep scoring.[56] Investigators from the fourth study ($n = 4$) used a seed in the posterior cingulate cortex and a thalamic seed defined by its connectivity with the posterior cingulate cortex during REM sleep.[57] The results from these studies indicate that connectivity during REM-sleep dreaming exists but may differ from the connectivity observed during waking cognition.

Default-Mode Network

During rest, the same regions exhibit decreased activity across a wide variety of tasks. These experiments were designed to use rest as the baseline, so if we instead conceptualize the task as our baseline, these regions exhibit increased activity at rest. Hence these regions were labeled the *default-mode network*.[58] This network includes posteromedial parietal regions (posterior cingulate, precuneus, and retrosplenial cortices), lateral-inferior-parietal/superior-temporal regions, hippocampal formation (hippocampus and parahippocampal cortex) regions, and medial PFC regions.[59,60]

Functional connectivity networks are not merely resting-state networks. Internally cued cognition replaces externally cued cognition, and the former is another adaptive task.[61,62] This is best measured with comparisons of internally versus

externally cued cognition in response to the same stimuli. For example, increased activity in the nodes of the default-mode network is observed when investigators contrast unconstrained quiet wakefulness to an externally cued task, but the same result is obtained when investigators contrast the judgment of whether a picture was pleasant or unpleasant to the judgment of whether the picture was an indoor or outdoor scene.[59] Therefore these networks can be labeled simply as *intrinsic connectivity networks*. Dreaming is another type of internally cued cognition, so it naturally shares many features with waking internally cued cognition.[63,64]

Default-Mode Network (Simulation Subsystem)

The default-mode network includes two subsystems: one centered on the hippocampal formation (the medial temporal lobe subsystem, henceforth referred to as the *simulation subsystem*) and one centered on the dorsal medial PFC (the dorsal medial PFC subsystem, henceforth referred to as the *self-referential subsystem*), activity in both of which correlates strongly with a core network (anterior-medial PFC and posterior cingulate cortex) but weakly with each other.[65,66] The simulation subsystem is active in both retrospective simulation (remembering the past) and prospective simulation (imagining the future),[60] whereas the self-referential subsystem is active during self-relevant tasks and social cognition, including imagining what others are thinking and feeling (i.e., theory of mind).[67]

During SWS, core default-mode network regions are disconnected.[68-70] During REM sleep, these regions become reconnected,[55,57] and simulation subsystem regions show higher connectivity compared with SWS.[56] The overall similarity in default-mode network connectivity during wakefulness and REM sleep was confirmed using meta-analytic techniques,[63] so the default-mode network simulation subsystem is a primary candidate to explain the dreaming that predominates in REM sleep.[17,63,64,71-73]

Default-Mode Network (Self-Referential Subsystem)

Although internally cued cognitions in dreaming and waking share many features, they have notable dissimilarities. Koike and colleagues[56] discovered incomplete reconnectivity in the self-referential subsystem during REM sleep as evidenced by a lack of anterior-posterior connectivity with the dorsal medial PFC. This difference in connectivity in the self-referential network may underlie the high delusional acceptance of dreaming compared with waking internally cued cognition. Combining data from this study with newer data, Watanabe and colleagues[74] discovered that overall default-mode network connectivity progressively decreases within a single bout of REM sleep, and this reaffirms the heterogeneous and dynamic nature of connectivity during REM sleep.

During dreaming, alterations in default-mode network connectivity may cause the simulation to be perceived as occurring in the present.[17] Without a perception of the self, the dreamer may also confabulate the simulations into a story-like theme as an attempt to interpret them.[73,75] This idea is consistent with the importance of the temporal-parietal junction (BA 40) in dreaming. This is another node in the self-referential subsystem,[65] and its damage causes cessation of dreaming.[76] Similarly to how Damasio describes alterations in the dream self,[77] perhaps the lack of the integration of the self into cognition is the reason that we cannot properly perceive events experienced during dreaming as a simulation.

Executive Control and Dorsal Attention Networks

Aside from the default-mode network, several other cortical networks have been identified. The dorsal attention network includes the intraparietal sulci and frontal eye fields. This network and the executive control network are negatively correlated with the default-mode network. As mentioned previously, because a correlation can be positive or negative, a negative correlation indicates the same magnitude of connectivity. Therefore these three networks can be considered part of a larger network in which activity both increases and decreases during internally cued cognition. During NREM sleep stage 1, the connectivity (i.e., the negative correlation) between these three seemingly separate networks decreases,[69,70,78] and the same phenomenon is observed in REM sleep.[57] The decrease in the negative correlation between these networks during NREM sleep stage 1 and REM sleep could therefore be related to features of mentation during sleep onset (e.g., hypnagogic imagery) and REM-sleep dreaming. Perhaps during wakefulness, the switch from externally cued cognition to internally cued cognition requires unique networks based on unique information processing requirements, but during sleep, unique networks may be less important to support cognition because it is typically internally cued.

Salience Network

The salience network—the anterior insula, anterior cingulate cortex, and ventral striatum—is involved in the interoception of the feelings associated with reward.[79] In addition to the internally cued versus externally cued distinction, paying attention to dreamed characters and objects is, in theory, executed by similar networks compared with waking cognition. Interoception could then apply in either waking or sleep cognition to the sensing of bodily states. The salience network has been implicated in drug craving and the associated relapse in addiction.[80] Dreams and the duration and number of REM-sleep periods increase during drug withdrawal.[81] Although the default-mode network may be important for normal dreaming, perhaps the salience network exhibits altered connectivity during this period of dream rebound and perhaps such alterations predict relapse.

Visual Network

The visual network is composed of primary visual cortex and visual association cortices in inferior temporal cortex and posterior parietal cortex. Unlike the paucity of incorporations of waking visual material into REM-sleep mentation,[3] presleep visual stimuli are reliably incorporated into sleep-onset mentation,[82] and this incorporation is correlated with sleep-dependent improvements in visuospatial memory.[83] Connectivity in the visual network increases during NREM sleep stage 1.[84] Further, increased connectivity between visual association and PFC is correlated with sleep-dependent improvements on a paired-associate face/scene task.[84] Although sleep-onset mentation was not measured, the results from this study indicate that visual network connectivity, sleep-onset mentation, and memory consolidation may be tightly linked.

Large-Scale Networks

A regular network has many nodes with short-distance connections (similar to a local train). This is inefficient because the average number of jumps needed to travel between all possible pairs of nodes in the network is very high. A random network has more nodes with long-distance connections (similar to an express train), and in such a network, the probability that two nodes are connected is unrelated to distance. This is inefficient because it results in more total wiring, and information may need to travel a long distance to traverse two nodes despite the fact that they are close to each other. The brain resembles an ideal "small-world" network,[85] which is a compromise between a regular and random network. The importance of information integration across a large number of brain regions has become a central tenet in theories of consciousness,[86,87] and alterations in this integration resulting from the absence of core nodes or increased randomness may explain the alterations in consciousness that accompany sleep and dreaming.[88,89]

During NREM sleep stages 1 and 2, there is a shift from an ideal small-world network to a more random network, and during SWS, there is a shift from an ideal small-world network to a more regular network,[90] although a shift to a more regular network has also been observed in NREM sleep stage 2.[69] The shift to a more random network during sleep onset may be correlated with sleep-onset mention.

Although portions of the thalamus are sometimes included in the default-mode network, it may be more appropriate to consider them part of a larger network because many thalamic nuclei project diffusely throughout the brain. Thalamocortical connectivity is weaker in NREM sleep stage 2 and SWS compared with wakefulness[91] and stronger in REM sleep compared with wakefulness and SWS.[57] Heightened thalamocortical connectivity during REM sleep may orchestrate dissociated heteromodal cortical regions and, correspondingly, the forced, confabulatory narrative in dreams.[57] Phasic REM-sleep episodes arising from a tonic REM sleep background showed characteristic changes in forebrain activity that include increased functional connectivity between the thalamus and a broad cortical-limbic-striatal network.[53] Wehrle and colleagues[53] suggest that these changes represent activation, during phasic REM sleep, of networks important to memory and emotion processing. Therefore like the EEG and MEG studies detailed previously, functional connectivity indicates that phasic REM sleep is associated with brain activity that might reflect intensified dream imagery, attention, and emotion but in the context of network connectivity that differs from wakefulness.

THE NEUROCHEMISTRY OF DREAMING

Three major neurochemical hypotheses have been advanced to explain differences between dreaming and waking consciousness. First, the activation-synthesis and activation-input-modulation models of Hobson and colleagues suggest that the massive increase in cholinergic (relative to noradrenergic and serotonergic) activation from the ascending reticular activating system (ARAS) during REM sleep contributes strongly to the unique nature of dream consciousness.[3,16] Second, Solms[76] suggests that stimulation of limbic and prefrontal reward networks by dopaminergic projections from the midbrain ventral tegmental area (VTA) generates motivational impulses that initiate dreaming, a hypothesis recently expanded in the reward activation model.[92,93] Third, Gottesmann suggests that dopaminergic stimulation of the cortex during REM sleep, in the absence of waking's inhibitory serotonergic and noradrenergic modulation, allows emergence of psychotomimetic (psychosis-like) aspects of dream consciousness.[94]

Acetylcholine

The activation-synthesis[16] and activation-input-modulation[3] models suggest that forebrain activation in REM-sleep dreaming originates in ascending activation of the thalamus by mesopontine cholinergic nuclei. Much evidence exists for cholinergic enhancement of both REM-sleep and dreaming. Higher mesopontine- and brainstem-derived acetylcholine concentrations during wake and REM versus NREM sleep are seen in the thalamus, including the lateral geniculate nucleus.[95] Cholinergic stimulation potentiates REM sleep when microinjected into the animal brainstem or when systemically administered to humans.[3] Cholinesterase inhibitors can induce REM sleep with dreaming[96] and increase nightmares[97] and hypnagogic hallucinations.[98] Transdermal nicotine[99] and the nicotinic receptor partial agonist varenicline[100] intensify dreams.

Dopamine

A key role is assigned to dopamine in psychotomimetic[94] and reward-based[76] theories of dreaming. The latter was recently expanded in the reward activation model, which posits that high-saliency memories are selectively processed during dreaming following their replay and prioritization by hippocampal-ventral striatal circuits in NREM sleep.[92,93] Support for this model comes from selective consolidation of salient memories during sleep[101,102] and the fact that administration of a dopamine agonist eliminated this selectivity.[103] In addition, L-dopa and certain other dopaminergic agents can enhance dreaming in patients with Parkinson disease.[76,104,105] However, psychostimulants are not associated with dream enhancement, neuroleptics do not prevent dreaming, some dopamine agonists reduce dreaming, and some dopamine antagonists enhance dreaming.[105] Thus dopamine's dream effects may depend on dosage as well as receptor type and location.

Roles for dopamine in REM sleep have also emerged from animal studies. For example, enhancing REM sleep intensity increases c-Fos expression in the VTA,[106] and dopamine concentrations in the medial PFC and nucleus accumbens are greater during REM sleep than during NREM sleep.[107] In addition, increased burst firing in the VTA has been observed during REM sleep[108] and may result from increased cholinergic excitation of the VTA by the pedunculopontine nucleus.[109]

Serotonin

Selective serotonin (5-HT) reuptake inhibitors and other serotonergic drugs can intensify dreaming.[110] Animal studies of serotonergic hallucinogens by Aghajanian and colleagues suggest that low or fluctuating cortical 5-HT levels may induce cortical output conducive to hallucinosis.[111] Serotonergic hallucinogens act presynaptically as partial 5-HT_{2A} (serotonin-2A) receptor agonists at glutamatergic inputs to apical

dendrites of layer V cortical pyramidal neurons, causing an "asynchronous" release of glutamate to follow the larger action-potential-mediated ("synchronous") release.[111] This late, slow release of glutamate induces prolonged excitatory postsynaptic potentials (EPSPs) that are hypothesized to underlie the cognitive-perceptual effects of the hallucinogens.[111] Such potentials do not normally occur because of inhibitory effects at 5-HT_1 autoreceptors.[111] However, under conditions of decreasing 5-HT concentration, such synchronous transmitter release–evoked EPSPs do emerge.[111] The naturally occurring lowest levels of 5-HT occur during REM sleep, and fluctuations of 5-HT release occur during sleep-stage transitions.[3] Thus these conditions may promote the natural occurrence of hallucinosis during dreaming.

NEUROPSYCHIATRIC SYNDROMES THAT INFORM THE STUDY OF DREAMING

Confabulation Shares Neural Substrate with Dreaming

Like the dreamer, patients with spontaneous behavioral confabulation believe they have experienced false events and act on false beliefs, believing them with unshakable conviction.[112-114] Confabulation results from lesions of the ventral medial PFC (vmPFC), the orbitofrontal cortex, and their connections with the basal forebrain, amygdala, thalamic mediodorsal nucleus, and hypothalamus.[113-115] These regions broadly overlap with the anterior paralimbic REM sleep activation region.[39]

One theory on the cognitive deficit resulting in confabulation suggests that vmPFC lesions disrupt a reality-monitoring function that preconsciously suppresses spontaneously activated memories not pertaining to present circumstances.[115] Memories of past experiences are thus perceived as related to the present.[115] An alternative theory posits temporal deficits to be a subset of a more general deficit in strategic retrieval and verification of memories.[114]

Moskovitch and Winocur[116] suggest vmPFC regions subserve an early stage in memory verification—a "feeling-of-rightness"—after which further evaluation of a retrieved memory takes place in dorsal lateral PFC (dlPFC). Gilboa and colleagues[114] suggest impaired feeling-of-rightness resulting from lesions to the vmPFC and orbitofrontal cortex is crucial and sufficient for producing confabulation.

Dreaming may represent a potent form of confabulation in which imaginary events are not only created and believed but also vividly experienced as organized, multimodal hallucinations.[17,73,105] For example, confabulators create plausible but false explanations for inconsistencies in their stories[113] that closely resemble ad hoc explanations for improbable dream occurrences.[117] Similarly, "pathological false recognition"[113] in confabulation parallels dreamers' assigning a known identity to dream characters perceptually dissimilar to their waking counterpart,[118] a phenomenon Schwartz and Maquet[119] liken to Fregoli syndrome. In both confabulation[113] and dreaming, altered function of the vmPFC and the orbitofrontal cortex may release from normal inhibitory, reality-monitoring, and executive constraints, innate human tendencies to represent both imagined and experienced events within a narrative structure.[73] Notably, medial PFC regions have also been associated with the production of normal narrative.[120]

Visual Hallucinosis in Waking and Dreaming

Anatomic regions associated with dreaming and with hallucinosis also overlap. Manford and Andermann[121] suggest that hallucinations result when inferior occipital and temporal visual association cortices that identify objects and scenes (the "ventral processing stream") are released from normal restraints under three conditions: (1) loss of exogenous visual input, (2) ARAS damage that alters serotonergic and cholinergic modulation of the cortex, or (3) abnormally excitatory input.[121] Conditions corresponding to each mechanism may contribute to REM-sleep hallucinosis.

First, in conditions such as Charles Bonnet syndrome, waking hallucinosis results when perceptual input to the visual association cortex is lost because of lesions of the primary visual pathway.[121] In REM-sleep dreaming, there is no retinal input. Moreover, primary visual cortices are deactivated,[36] a phenomenon that perhaps decreases their regulatory input on downstream association cortices. (Interestingly, visual deprivation is able to produce hallucinations similar to Charles Bonnet syndrome in normal subjects.[122]) Second, in peduncular hallucinosis, lesions of the rostral brainstem disrupt the amount and balance of serotonergic and cholinergic input to the forebrain.[121] Serotonergic raphe neurons normally fire at their minimum during REM sleep, whereas mesopontine cholinergic systems are reactivated to near-waking levels.[3] Therefore an altered serotonergic-cholinergic balance during REM sleep may also favor hallucinosis. Third, abnormal excitation of visual association cortices (the regions active during hallucinations in Charles Bonnet syndrome[123]) can produce epileptic hallucinosis.[121] Ventral-processing-stream visual association cortices are more active in REM sleep than in either NREM sleep or postsleep waking[28,36] and are a hypothesized source of dream imagery.[3,76] During REM sleep, PGO waves originating in the mesopontine brainstem[3,21] excite many cortical regions, including visual association regions.[124]

Collerton and colleagues[125] suggest that hallucinations result from combined sensory impairment, attentional deficit, and relatively intact scene perception that allows poorly formed "proto-objects" to resolve into erroneous percepts. In REM-sleep dreaming, ascending activation of visual association cortices may evoke hallucinatory proto-objects that, under the attentionally unstable conditions of REM sleep, resolve into hallucinatory percepts congruent with the ongoing dream plot. Internal consistency of dream plots may then arise because the evolving dream context itself biases resolution of ambiguous percepts into plot-congruent images. Dreams may thus evolve by a "boot-strapping" process whereby current images provide the context that, in turn, determines succeeding dream imagery.[126] In the absence of working memory capacities that provide continuity to waking experience, the evolving dream plot can be strongly influenced by immediately prior dream experiences. In a similar manner, top-down influences from limbic cortices may bias posterior association cortices toward generating, attending to, or disambiguating imagery in a manner that is congruent with ongoing dream emotion.[17]

A DESCRIPTIVE NEURAL MODEL OF DREAM PHENOMENOLOGY AND FUNCTION

Behavioral states and cognitive capacities are physically instantiated in widely distributed networks with distinct

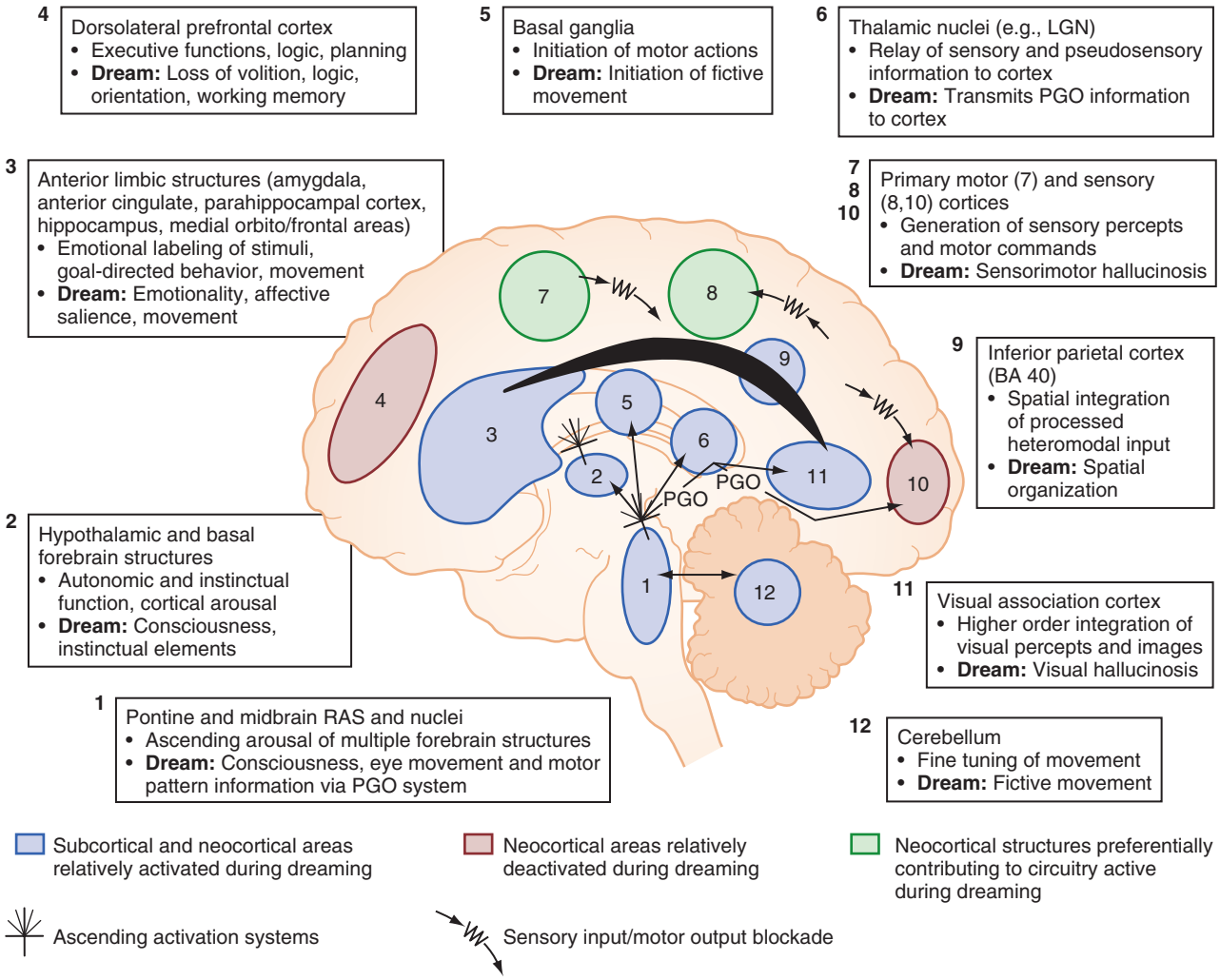


Figure 51-1 Forebrain processes in normal dreaming—an integration of neurophysiologic, neuropsychological, and neuroimaging data. Regions 1 and 2, ascending arousal systems; region 3, subcortical and cortical limbic and paralimbic structures; region 4, dorsal lateral prefrontal executive association cortex; region 5, motor initiation and control centers; region 6, thalamocortical relay centers and thalamic subcortical circuitry; region 7, primary motor cortex; region 8, primary sensory cortex; region 9, inferior parietal lobe; region 10, primary visual cortex; region 11, visual association cortex; region 12, cerebellum. BA 40, Brodmann area 40, the temporal-parietal junction; LGN, lateral geniculate nucleus; PGO, ponto-geniculo-occipital waves; RAS, reticular activating system. (From Hobson JA, Pace-Schott EF, Stickgold R. Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behav Brain Sci* 2000;23:793–842; discussion 904–1121.)

epicenters of critical control in a pattern of "selectively distributed processing."[127] Recent findings on intrinsic connectivity networks described previously have begun to assign distinct neural networks and key nodes to many of these functions. The following working model of neurobiologic structures and networks subserving REM-sleep-dream phenomenology refers to brain regions depicted in Figure 51-1.[3]

Ascending Arousal Systems

Activation of the forebrain in REM sleep, as in waking, occurs through the ascending arousal systems of the brainstem,[12] basal forebrain,[128] and hypothalamus[129] (regions 1 and 2 in Figure 51-1). However, unlike in waking, ascending activation in REM sleep is primarily facilitated by cholinergic systems, whereas aminergic neuromodulation is attenuated.[3] Pedunculopontine nucleus and laterodorsal tegmental cholinergic neurons project to the thalamus and basal forebrain but not the cortex.[130] Therefore cortical activation in REM sleep is preceded by cholinergic activation of the diencephalon[12] and basal forebrain,[128,131] which in turn activate the cortex with glutamate[12] and acetylcholine,[128,131] respectively. In NREM sleep, phasic increases in ARAS activity due to endogenous or exogenous stimulation may transiently stimulate the same forebrain networks activated in REM, the subjective manifestation of which may be NREM-sleep dreaming. The effect of cortical arousal on dream production is always confounded by the effects of such arousal during postsleep wakefulness on dream recall. A recent PET study suggests, however, that forebrain activation during preawakening REM sleep is directly related to dream recall on awakening.[132]

Thalamocortical Relay Centers and Thalamic Subcortical Circuitry

During REM sleep, thalamocortical signaling (region 6 in Figure 51-1) may be interpreted as incoming sensory information by primary and secondary association sensory cortices[16] (region 11) and evoke local activation of stored cognitive representations (hallucinosis of known entities) or novel representations (as in dream bizarreness). The PGO wave may be only one of many pathways for ARAS phasic activation of the cortex by thalamic or basal forebrain intermediaries during REM sleep. For example, in the rat, the pontine p-wave in REM sleep impinges directly on limbic structures such as the amygdala, hippocampus, and entorhinal cortex as well as the visual cortex.[133] Other pathways through the thalamus might include a nonrelay sensory route through the pulvinar nucleus directly to the visual association cortex[36] or to limbic prefrontal regions through magnocellular portions of the mediodorsal nucleus.

During NREM sleep, intrinsic thalamocortical oscillations suppress but do not completely extinguish perception and mentation.[45] NREM sleep oscillatory rhythms reflect endogenous activity of corticocortical and corticothalamocortical circuits grouped by the slow oscillation, the "up" state of which briefly returns cortical neurons to levels of high activity.[12,13,134,135] Steriade[135] has suggested that when this oscillatory pattern is impinged on by phasic thalamocortical bursts, such as the isolated PGO waves in the NREM-REM sleep transitional state, phasic elevation of regional activity may lead to vivid visual imagery. Notably, activation in NREM sleep of primary visual[31,36] and visual association[136] regions may also generate NREM-sleep imagery.

Subcortical and Cortical Limbic and Paralimbic Structures

In REM sleep, selective activation of limbic and paralimbic cortex and subcortex[28,35,36,39,40] (region 3 in Figure 51-1) suggested to PET researchers a role for REM sleep in the processing of emotionally influenced memories,[37,137] integration of neocortical functions with basal forebrain and hypothalamic motivational and reward mechanisms,[40] or internal information processing between visual association and limbic regions.[36] Such processes may underlie the emotionality[3,28,137] and social nature of dreaming.[17,118]

Emotion Regulation and Dreaming

It is often hypothesized that sleep and dreaming play an emotion regulatory function[138-140] that is disrupted in mood and anxiety disorders that, in turn, can alter dreaming, as in nightmares.[139] Indeed, the anterior paralimbic REM sleep activation region includes many of the structures implicated in the experience and expression of emotion.[141] Although dreams may aid in resolution of intrapersonal conflict (e.g., see Cartwright and colleagues[138]), dreaming may also moderate emotional extremes by universal mammalian learning processes such as habituation and extinction.[139,142]

Nielsen and Levin[139] have suggested that, in normal REM sleep, activity in the anterior paralimbic REM-sleep activation region regulates emotion through formation of extinction memories when emotionally salient memories appear in safer contexts during dreaming. Human[143] as well as animal[144] studies link formation, retention, and expression of extinction learning to circuitry linking the amygdala, vmPFC, and hippocampus. Notably, a night's sleep promotes generalization of extinction learning.[145,146] Whereas such circuitry may be recruited by cognitive processes subserved by dlPFC in waking,[147] it may function autonomously during REM-sleep dreaming.[139]

Reward systems are also activated in REM sleep during which positive and negative emotions may be modulated. VTA sources of mesolimbic and mesocortical dopamine are recruited by ascending cholinergic activation[109] and, like their ventral striatal and medial PFC targets, lie well within the anterior paralimbic REM-sleep activation region.[39,40,92,93] In dreaming, hypothalamic-brainstem circuits may initiate instinctively salient behavior[148] that may, in turn, recruit additional forebrain regions to enact appetitive[76] or other adaptive behaviors.[148]

Altered Memory Processing in Dreams

A cholinergically mediated informational barrier between cortex and hippocampus in REM sleep has been proposed to underlie the paucity of episodic memories in dreams.[149] However, despite the inaccessibility of episodic memory, another aspect of declarative memory, "familiarity" or "recognition,"[150] is ubiquitous in dreams. For example, 40% of dream characters may be identified on the basis of "just knowing."[118] Schwartz and Maquet[119] suggest such phenomena result from the sleep-related disconnection of temporal lobe face recognition from prefrontal reality monitoring regions. Alternatively, frequent experiences of familiarity in the absence of accurate replay of episodic memories in dreams may reflect activity of recognition memory mechanisms in anterior perirhinal cortices (BA 35, BA 36) dissociated from hippocampally mediated recall. Indeed, double dissociations between recognition and knowing have been shown in human fMRI studies.[150] Because perirhinal regions of the anterior-medial temporal lobe are proximal to the anterior paralimbic REM sleep activation region,[40] altered interactions of this region with other portions of the hippocampal formation during REM sleep may produce this dissociation.

During dreaming, frontal contributions to memory retrieval may also be altered. Ventral lateral and dlPFC regions that are deactivated in REM sleep[37] subserve cue-specification and search strategies, respectively.[116] In contrast, posterior vmPFC, which is active in REM sleep,[28,35,39,40] subserves "feeling-of-rightness"[116] and "feeling-of-knowing,"[151] for which more anterior PFC regions only later provide cognitive verification.[116] Therefore, during REM sleep, greater activation of posterior-ventral-medial than anterior-lateral PFC regions relative to waking may favor an indiscriminate, emotional confirmation of accuracy for any item in consciousness without the benefit of strategic volitional search or critical verification. In combination, studies of mental simulation,[60] confabulation,[114,115] and memory verification[116,151] predict that restriction of frontal activation to vmPFC would favor an emotionally salient state prone to producing mental simulations that evoke a powerful sense of veracity and familiarity and are uncritically believed.

As discussed previously, a shift to a more random network has been observed during NREM sleep stage 1.[90] This shift could underlie the discontinuous and incongruous memory associations of dreaming in this stage. Sleep-onset mentation is typically short, so it would be difficult to measure the

correlation between these bizarre associations and small-world network changes. No similar studies exist for REM sleep, but the data from NREM sleep stage 1 might lead one to predict a similar change, and these changes could be correlated with the bizarreness of the corresponding dream.

Dreaming as a Simulation

Default-mode network simulation subsystem connectivity during REM sleep as it is related to dreaming could support the memory function of sleep.[72,152] Although entire memories are often not incorporated into REM-sleep mentation, memory elements may be incorporated depending on their emotional salience and/or cued reward value for subsequent memory consolidation,[150] as discussed previously. In addition to consolidating memories, activity in the default-mode network during dreaming would support the idea that dreams rehearse or simulate potential future events so as to prepare for them.[153,154] Imagining future events increases goal-directed actions when the corresponding opportunity later presents itself,[155] so simulation subsystem connectivity and the associated dreaming may increase adaptive behavior, similar to proposals on the functions of waking internally cued cognition.[65]

Social Cognition and Dreams

Brain regions most consistently activated in neuroimaging studies of social cognition include regions that are components of the default-mode network self-referential subsystem: medial PFC (especially the paracingulate cortex, BA 32), superior temporal sulcus, and temporal poles including the amygdala.[156] Incomplete reconnectivity of this subsystem during REM sleep[56] may underlie the high degree of delusional acceptance of dreams compared with waking cognition, and this may depend on an altered integration of the self into cognition. However, theory of mind, a complex aspect of social cognition,[156] is preserved in dreaming despite notable degradation of reasoning about the physical world.[157,158] This may seem like a contradiction, but social cognition has two components: an ability to perceive ourselves as social agents and an ability to perceive the intentions of others.[159] Subtle disconnections in the self-referential subsystem may disrupt the former and lead to delusional acceptance of dreams, whereas residual connectivity in the self-referential subsystem may preserve the latter and produce the ubiquity of interpersonal interactions and emotions in dreams.[17,37,73,157,160]

The function of residual self-referential subsystem activation as it is related to dreaming could be connected to the development and refinement of social skills. The interpretation of existing and new characters during dreaming may serve as practice for the application of social skills during wakefulness. Patients with autism have alterations in default-mode network activity,[161] and it would be interesting to examine whether these alterations persist in REM sleep because these patients exhibit sleep disturbances.[162,163]

Motor Initiation and Control Centers

Strong activation of the basal ganglia[28] (region 5 in Figure 51-1) may mediate the ubiquitous fictive motion of dreams.[164] The basal ganglia are extensively connected not only with motor cortex but also with mesopontine (e.g., pedunculopontine) nuclei[165] that contain gait circuitry and other motor pattern generators as well as REM-sleep regulatory regions.[3] Activation of brainstem vestibular nuclei and the associated cerebellar vermis[28] during REM sleep may additionally contribute vestibular sensations interpreted as flying or falling as well as a sense of motor control.

Visual Association Cortex

Medial occipitotemporal cortices (region 11 in Figure 51-1) are activated in REM sleep.[28,36] These and other visual association regions may generate the visual imagery of dreams.[3,76] As in waking, specific regions of the visual association cortex may process specific visual characteristics of dreaming. For example, the fusiform gyrus both mediates waking face recognition and is activated in REM sleep.[28,36,40] Braun and colleagues[36] suggest that REM sleep constitutes a unique cortical condition of internal information processing (between visual association and limbic cortices), functionally isolated from input from (through primary visual cortex) or output to (through frontal cortex) the external world. Dream image formation may arise as ascending activation impinges on visual and multimodal association regions in occipital, temporal, and inferior parietal cortices.

Inferior Parietal Lobe

The supramarginal and angular gyri of the inferior parietal lobe (BA 39 and 40; region 9 in Figure 51-1), especially in the right hemisphere, are essential for visuospatial awareness.[127] These regions may generate the fictive dream space necessary for the organized hallucinatory experience of dreaming.[76] Destruction of these regions is alone sufficient to produce global cessation of reported dreaming.[76,166] Maquet and colleagues have found right inferior parietal cortex to be relatively activated during REM sleep in some[35] but not all[38] PET studies. In REM sleep, both the previously described visual association cortex and the vmPFC are simultaneously active.[36] Therefore, in REM sleep, self-centric reality simulation, a putative function of the vmPFC,[60] and hallucinatory imagery may arise in concert. Inferior parietal multimodal association cortices may integrate different unimodal inputs and facilitate their incorporation into the emerging plot in the virtual proscenium where the dream is experienced.

Dorsolateral Prefrontal Executive Association Cortex

Lesions of the dlPFC (region 4 in Figure 51-1) do not cause cessation or attenuation of dreaming, an observation that suggests that they are nonessential for the generation of dreaming.[76,166] Unlike vmPFC regions that reactivate, these dorsal lateral prefrontal regions remain deactivated in REM sleep,[28,35-37] and this may explain dream mentation executive deficiencies that include disorientation, illogic, impaired working memory, and amnesia for dreams.[3] Additionally, because the PFC regulates posterior sensory cortices,[167] deactivation of the dlPFC in REM sleep[28,35,37] may promote dreaming by disconnection, release, or disinhibition of sensory association cortices (as also suggested by EEG and MEG[5,168,169]).

This may be consistent with the absence of the dorsal medial PFC from the default-mode network self-referential subsystem during REM sleep.[56] In addition to considering the dorsal medial PFC as part of the default-mode network self-referential subsystem, it and the lateral portions of the PFC can be considered as part of executive association cortex that includes the superior parietal gyrus, and the disruption of this

network may, again, promote dreaming by disinhibition of sensory association cortices.[170] The incomplete reconnectivity of this node in REM sleep might contribute to the bizarre nature of dreaming compared with waking cognition and is closely related to the reconnectivity of the simulation subsystem in REM sleep. Because this reconnectivity co-occurs with a reduction in sensory input, this may elicit the spontaneous and involuntary cognitions of dreaming.[72]

The PFC maintains an online representation of a goal, the means to achieve it, and the ongoing context relevant to this goal to "bias" the functioning of networks elsewhere in the brain toward this particular outcome.[171] For example, top-down influence can sensitize primary and association perceptual cortices to particular stimuli while simultaneously attenuating their sensitivity to competing stimuli.[172] With diminished frontal activation during sleep, goal-directed biases in the regulation of circuits subserving working memory and attention (frontoparietal) and memory encoding and retrieval (frontotemporal) may be impaired during dreaming.[3,17] Nonetheless, top-down influence in REM sleep may bias sensitivity of sensory cortices to hallucinatory percepts related to affective and social processing subserved by the vmPFC, whereas nonemotional imagery may fail to persist in the face of weakened working memory. Additionally, diminished activity in lateral frontal cortices in REM sleep[37] may weaken the ability of frontal, top-down influence to modulate emotional responses to dream imagery.[147] Alteration in dreaming of specific frontal-subcortical networks may underlie dream emotional salience and executive weakness. For example, the "cognitive" fronto-striatal-thalamo-cortical circuit (linking dlPFC with dorsal striatum) may be less active than "affective" (orbito-frontal-ventral caudate) and "motivational" (anterior cingulate–nucleus accumbens) circuits.[173]

CLINICAL PEARL

The physician prescribing selective serotonin reuptake inhibitors, transdermal nicotine, varenicline, beta blockers, dopaminergic agents, or a variety of other medications[97,99,100,104,110] should be alert for possible dream intensification or nightmare induction. Other choices within the same class or a different class of medications may need to be considered if such side effects are intolerable to the patient.

SUMMARY

ARAS, thalamocortical, and basal forebrain-cortical arousal systems activate the forebrain regions involved in dream construction in a manner that is chemically and anatomically different from waking. In REM sleep, such activation may be more frequent and sustained and, perhaps, may proceed through different or more diverse pathways than in NREM sleep. Cortical circuits activated in REM-sleep dreaming are medial circuits linking visual association and paralimbic regions (central crescent in Figure 51-1), but not the primary sensory and lateral frontal executive cortical regions that are active in waking.[36] Therefore dreaming is both positively and negatively emotionally salient (amygdala, ventral striatum, vmPFC), often conflictual (anterior cingulate) and social (vmPFC), while also displaying profoundly deficient working memory, orientation, and logic (lateral prefrontal and parietal deactivation). Subcortical circuits involving the limbic structures, striatum, diencephalon, and brainstem regions are selectively activated in REM sleep. They may contribute to dreaming's emotional (limbic subcortex), motoric (striatum, brainstem, cerebellum), instinctual (hypothalamus), and motivational (midbrain-ventral striatum) properties. Preserved connectivity in default-mode network simulation subsystem during REM sleep may be the neural substrate of dream simulations, whereas altered connectivity in default-mode network self-referential subsystem may underlie the high degree of delusional acceptance of dreams compared to waking cognition.

ACKNOWLEDGMENTS

The authors acknowledge support for this research from the NIH/NIMH R21MH101567, the Intramural Research Program of the National Institute of Mental Health, and the Intramural Research Program of the National Institute of Neurological Disorders and Stroke.

Selected Readings

Braun AR, Balkin TJ, Wesenten NJ, et al. Regional cerebral blood flow throughout the sleep-wake cycle: an H2(15)O PET study. *Brain* 1997;**120**(Pt 7):1173–97.

Dresler M, Wehrle R, Spoormaker VI, et al. Neural correlates of insight in dreaming and psychosis. *Sleep Med Rev* 2015;**20**:92–9.

Han H, Schweickert R, Xi Z, Viau-Quesnel C. The cognitive social network in dreams: transitivity, assortativity and giant component proportion are monotonic. *Cogn Sci* 2015;[Epub ahead of print].

Hobson JA, Pace-Schott EF, Stickgold R. Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behav Brain Sci* 2000;**23**:793–842.

Koike T, Kan S, Misaki M, Miyauchi S. Connectivity pattern changes in default-mode network with deep non-REM and REM sleep. *Neurosci Res* 2011;**69**(4):322–30.

Kusse C, Muto V, Mascetti L, et al. Neuroimaging of dreaming: state of the art and limitations. *Int Rev Neurobiol* 2010;**92**:87–99.

Pace-Schott EF. Dreaming as a story-telling instinct. *Front Psychol* 2013;**4**:159.

Picchioni D, Duyn JH, Horovitz SG. Sleep and the functional connectome. *Neuroimage* 2013;**80**:387–96.

Rechtschaffen A. The single-mindedness and isolation of dreams. *Sleep* 1978;**1**:97–109.

Stickgold R, Hobson JA, Fosse R, Fosse M. Sleep, learning, and dreams: off-line memory reprocessing. *Science* 2001;**294**:1052–7.

Wamsley EJ. Dreaming, waking conscious experience, and the resting brain: report of subjective experience as a tool in the cognitive neurosciences. *Front Psychol* 2013;**4**:637.

A complete reference list can be found online at ExpertConsult.com.

Chapter 52

Lucid Dreaming

Martin Dresler; Daniel Erlacher; Michael Czisch; Victor I. Spoormaker

Chapter Highlights

- During lucid dreaming, the dreamer is fully aware of his or her current dream state. This metacognitive insight often leads to full access to memory and increased volitional control over the dream narrative.
- Lucid dreaming is a rare natural phenomenon; however, several strategies are available to induce dream lucidity or increase the frequency of this sleep state.
- Lucid dreaming is associated with increased neural activity in several neocortical regions, particularly lateral prefrontal, frontopolar, and medial parietal cortices. On the electroencephalogram, 40-Hz gamma power over dorsolateral prefrontal areas is increased during lucid dreaming, and 40-Hz transcranial alternating current stimulation of this region induces dream lucidity.
- Lucid dreaming has a potential clinical application in the therapy of nightmares. In keeping with the view that dreaming can serve as a model for psychosis, lucid dreaming may potentially be of value for the therapy of metacognitive deficits in psychosis.

Conscious experience varies strikingly across the sleep-wake cycle. During wakefulness, humans are alert, aware of external and internal stimuli, able to reflect on their perceptions, emotions, and thoughts, and capable of acting volitionally according to their intentions. Most of these properties of waking consciousness fade during the process of falling asleep; however, they partly reappear during sleep mentation. Conscious experience during sleep is manifold. If not absent as in dreamless sleep, it can include abstract thought fragments, intense emotions, and sensory imagery, up to fully immersive visuomotor hallucinations with a complex interactive dream plot.[1] From the perspective of the dreamer, this virtual reality often feels indistinguishable from waking life. Dream experiences, however, typically show many cognitive peculiarities, with delusional thought, diminished volition, and a complete lack of insight into the true state of mind even in the face of a bizarre dream plot. In this regard, dreaming resembles the psychosis of mental diseases such as schizophrenia, characterized by hallucinations, loosening of associations, incongruity of personal experience, and a loss of self-reflective capacity.[2,3]

In contrast with normal dreaming, the rare phenomenon of lucid dreaming is characterized by the reappearance of many wake-like cognitive capabilities. Minimally defined by the criterion that the sleeper is aware of the current dream state as such,[4] lucid dreaming often leads to full insight into the delusional nature of the dream environment, access to short- and long-term memory, and sometimes even volitional control over the dream narrative.[5] Despite this wake-like cognitive capability, lucid rapid eye movement (REM) sleep comprises all defining markers of REM sleep[4] and all basal dream features such as visuomotor hallucinations[1]—in fact, prototypical aspects of dream phenomenology such as bizarreness might even be more pronounced in lucid dreams.[6] Lucid dreaming is not an all-or-nothing phenomenon but can occur in different degrees.[7,8] Questionnaires such as the Metacognitive, Affective, Cognitive Experience (MACE) questionnaire[9] and the Lucidity and Consciousness in Dreams scale[10] aim to assess this continuum ranging from single-minded sleep mentation over prelucid reflections to full-blown lucid control dreams.

Despite being described in ancient scripture and reported in early modern research literature, lucid dreaming faced considerable skepticism in mainstream sleep research during most of the twentieth century. In the late 1970s, the first systematic validation of lucid dreaming as an objective phenomenon occurring during otherwise normal REM sleep was achieved. In accordance with the scanning hypothesis, according to which rapid eye movements during REM sleep are related to gaze direction during dreaming,[11] lucid dreamers were asked to move their eyes in a left-right-left-right fashion during dreaming as soon as they became aware of their dreaming state.[4] Through this technique, which has since become the gold standard in lucid dream research, lucid dream reports could be objectively verified by eye movement patterns as recorded in the electrooculogram. By providing objective temporal markers of dream content, this method has allowed, for example, investigations into neural correlates of dreamed behaviors[12-14] and comparisons of the passage of time as experienced during dreaming with objective measurements in the real world.[15] Nevertheless, the rarity of lucid dreams and the difficulties of observing them under laboratory conditions hamper research with considerable sample sizes, rendering many results from lucid dreaming studies preliminary. In

addition, lucid dreams documented in the laboratory tend to be much shorter than at home, which further restricts the possibilities of lucid dream research.[16,17]

PREVALENCE AND INDUCTION METHODS

Generally, lucid dreaming is quite rare. Only one half of the general population know the phenomenon from personal experience, approximately 20% have lucid dreams on a monthly basis, and only a minority of approximately 1% have lucid dreams several times a week.[18,19] Some differences across different populations and cultures seem to exist—for example, German students reported a much higher lucid dream frequency than Japanese students,[20,12] and even when representative population samples are surveyed, Germans report more lucid dreams than Austrians.[19,21] Also, age-related differences in lucid dreaming prevalence have been recognized, with young children and adolescents reporting lucid dreams more frequently than adults.[22,17]

Quite often, lucid dreaming spontaneously emerges from nightmares, recurrent dreams, or some peculiarities within a dream. However, lucid dreaming also can be intentionally induced by applying various induction or relaxation techniques, by engaging with topics of dreaming and lucid dreaming, and with use of other deliberate training strategies.[23,24] Sleep-specific circumstances, such as short awakening in the morning or an afternoon nap, as well as stress also can initiate the first lucid dream experience.[17]

Since the onset of lucid dream research, possible induction techniques have always been a pertinent concern. A plethora of strategies to induce lucid dreams have been suggested in the literature, which can be loosely classified into three broad categories: cognitive techniques, external stimulation, and miscellaneous.[25] The first category encompasses all cognitive activities that are carried out to increase the likelihood of achieving lucidity in a dream state. For this category, a large number of different methods have been suggested that can be further divided into methods whereby lucidity is initiated from within a dream, so that the person becomes lucid during a dream (*dream-initiated lucid dreaming*), and methods whereby lucidity is initiated from wakefulness, so that the person retains conscious awareness when falling asleep (*wake-initiated lucid dreaming*). The rationale behind the second category is that an external stimulus presented to a sleeping person can be incorporated into the dream (e.g., spraying water on the sleeping person's face may promote the incorporation of sudden rainfalls in the dream) and that the incorporated stimuli serve as a cue to remind the dreamer about being in the dream state (e.g., someone squirting water signals the dreamer that he or she is dreaming) and thereby triggers dream lucidity.[26] The third category includes miscellaneous aids to gain lucidity, such as drugs (e.g., Donepezil) but also specific practices involving sleep-wake patterns—for example, waking up in early-morning hours and then, after a certain period, going back to bed to take a nap, known as *wake-back-to-bed* (WBTB).[27] WBTB is not a technique per se, because it was empirically tested only in combination with other techniques (*mnemonic induction of lucid dreams* [MILD]) and may boost their efficacy.

Most lucidity induction strategies described in the literature are based on personal and anecdotal accounts. A recent systematic review of evidence found 27 studies that experimentally tested the efficacy of lucidity induction techniques,[25] of which 5 were conducted as sleep laboratory studies and the other 22 were done as field experiments, in some cases with low methodologic quality. None of the induction techniques was verified to induce lucid dreams reliably, consistently, and with a high success rate; some methods, however, proved to be promising—for example, the MILD/WBTB combination. More recently, noninvasive brain stimulation methods yielded encouraging results. Although transcranial direct current stimulation of the dorsolateral prefrontal cortex during REM sleep showed rather modest success in inducing dream lucidity as assessed by the eye signaling technique,[27] transcranial alternating current stimulation in the low gamma range (25 Hz and 40 Hz) led to a robust increase in retrospectively reported dream lucidity as measured by the Lucidity and Consciousness in Dreams questionnaire.[28]

NEUROBIOLOGY

Dream-like mental activity can be observed during all sleep stages; REM sleep dreams, however, are particularly vivid and intense. The specific phenomenologic characteristics of dreaming frequently have been associated with neural activation patterns observed during REM sleep. For example, higher visual areas show strong metabolic activity during REM sleep,[29] which is in line with visuospatial hallucinations as the hallmark of typical dreaming.[1] The amygdala, medial prefrontal cortex, and anterior cingulate cortex also show increased activity during REM sleep.[30,31] All of these brain areas have been implicated in emotional processing, nicely mirroring the intense emotions experienced in many dreams. By contrast, the dorsolateral prefrontal cortex, frontopolar cortex, and parietal areas including the supramarginal cortex and precuneus show low metabolic rates during normal REM sleep.[30,31] In particular, prefrontal deactivations have been postulated to underlie cognitive deficiencies typical of ordinary dreaming such as impaired critical thinking, diminished metacognitive ability, and restricted volitional control.[32]

Although lucid REM sleep dreaming is characterized by the full range of coarse electroencephalographic features of REM sleep according to the classical Rechtschaffen and Kales (1968)[33] or new American Association of Sleep Medicine (AASM)[34] sleep stage scoring, it does incorporate some subtle physiologic changes relative to nonlucid REM sleep, such as higher eye movement density and increases in respiration, heart rate, and skin potential.[16] In addition, brain activity during lucid REM sleep shows distinctive changes from that during nonlucid REM sleep. Early electroelectrogram (EEG) studies observed higher alpha activity,[35,37] and increased beta-1 activity (13 to 19 Hz) over parietal regions during lucid dreaming.[36] A more recent high-density EEG study demonstrated that lucid dreaming is associated with higher activity in the gamma band—the between-states-difference peaking at approximately 40 Hz—and overall EEG coherence compared with nonlucid REM sleep.[37] Both power in the 40-Hz band and coherence levels were observed to be strongest over the dorsolateral prefrontal cortex (Figure 52-1), which has been associated with metacognitive evaluation.[38]

In a combined functional magnetic resonance imaging (fMRI)-EEG approach, activations in a network of purely neocortical regions including the dorsolateral and frontopolar prefrontal cortex were observed during lucid dreaming as

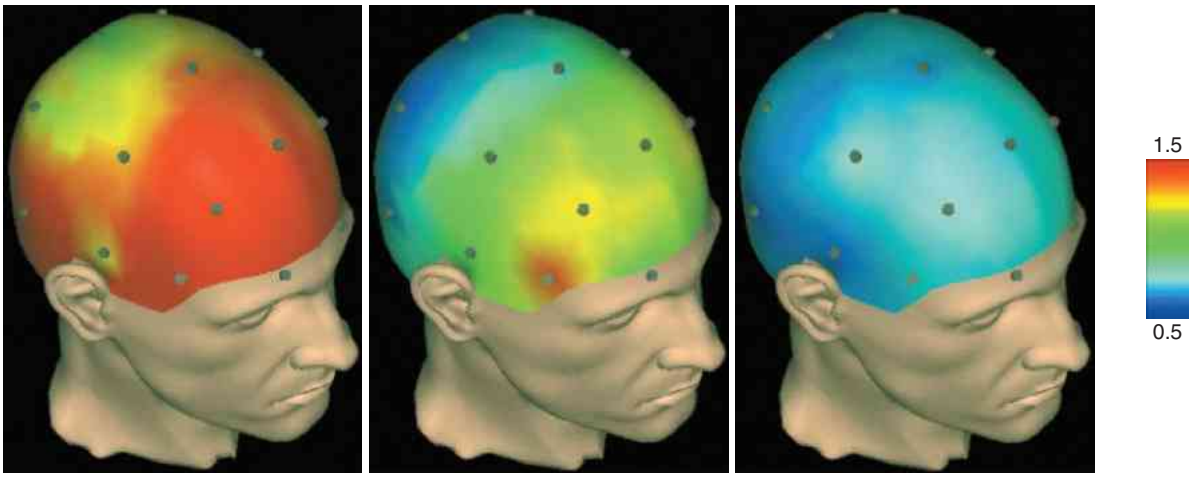


Figure 52-1 **Quantitative EEG Data for Lucid Dreaming.** Shown is gamma 40-Hz standardized current source density power during wakefulness, lucid REM sleep, and nonlucid REM sleep. The dorsolateral prefrontal cortex during lucid dreaming shows similar 40-Hz power to that during wakefulness. EEG, Electroencephalogram. (From Voss U, Holzmann R, Tuin I, Hobson JA. Lucid dreaming: a state of consciousness with features of both waking and non-lucid dreaming. *Sleep* 2009;32:119–200, with permission of the American Academy of Sleep Medicine from Voss et al. 2008; permission conveyed through Copyright Clearance Center, Inc.)

compared with nonlucid REM sleep background (Figure 52-2),[39] thus confirming earlier EEG data. Of note, recent anatomic analyses have demonstrated larger gray matter volume in the frontopolar cortex in dreamers with high self-reported dream lucidity.[40] The frontopolar cortex has been related to the processing of internal states, such as self-evaluation of thoughts and feelings,[41] metacognitive ability,[42] and supervisory modes,[43] which are impaired in normal dreaming but reinstated in lucid dreaming. Strong activation increases during lucid dreaming also were observed in parietal regions including the precuneus, inferior parietal lobules, and supramarginal gyrus.[44] Prefrontal-parietal interactions are involved in many higher-level cognitive processes such as intelligence or working memory,[45] whereas the precuneus has been implicated in self-referential processing such as first-person perspective taking and experience of agency.[46] These findings are in line with the notion that dream lucidity provides increased availability of self-related information, leading to a much higher degree of coherence and stability of the phenomenologic self during lucid dreaming.[47] Taken together, the frontoparietal activation patterns observed during lucid REM sleep nicely mirror the reinstatement of cognitive capabilities experienced during lucid dreaming. Activation increases during lucid dreaming also were found in some occipital and inferior-medial temporal regions.[44] These cortical areas are part of the ventral stream of visual processing, which is involved in several aspects of conscious awareness in visual perception.[48] Although such activations initially may seem puzzling—nonlucid dreams also are characterized by vivid dream imagery—they are consistent with reports of lucid dreamers stating that lucidity is associated with an exceptional brightness and visual clarity of the dream scenery.[49]

In recent years, network analyses of neuroimaging data have been introduced into sleep research.[50,51] Particular interest has been directed at the *default mode network,*[52] which during wakefulness shows increased activity in the absence of processing related to external tasks. Because this activity increase appears to be related to stimulus-independent thought such as internal awareness and daydreaming during wakefulness,[53,54] the default mode network has been proposed to be associated with nonlucid dreaming during sleep.[55,56] A second network shows activation anticorrelated with the default mode system in resting state fMRI analyses: the *dorsal attention system,* which is mainly involved in externally directed perceptual processes.[57] A third network, dubbed the *frontoparietal control system,* has been postulated to integrate information coming from both the default mode and the dorsal attention network by switching between competing internally and externally directed processes.[58,59,60] Phenomenologically, this might be interpreted as a monitoring of and control over mind-wandering and perception by metacognitive processes.[61] Owing to this role as a kind of meta-network, the frontoparietal control system might be seen as an ideal candidate subserving metacognitive aspects of consciousness that are the hallmark of lucid dreaming.[51] Brain regions activated during lucid dreaming indeed comprise substantial parts of the frontoparietal control network.[44]

LUCID DREAMING AS HIGHER-ORDER CONSCIOUSNESS

The contrast between lucid and nonlucid dreaming has been suggested to mirror the conceptual contrast between basal (primary) and higher-order (secondary) aspects of consciousness.[62,63] Although all basal features of consciousness, such as perceptions and emotions, are present in normal dreaming, metacognitive reflections and the insight into the current state of consciousness are, by definition, bound to dream lucidity. Because some reflective thoughts have been reported in nonlucid dreaming, and also because active reflections frequently are absent during daydreaming and other phases of wakefulness, it has been argued that metacognitive activity differs only quantitatively and not qualitatively between dreaming and waking consciousness.[64] This absence, however, is only a

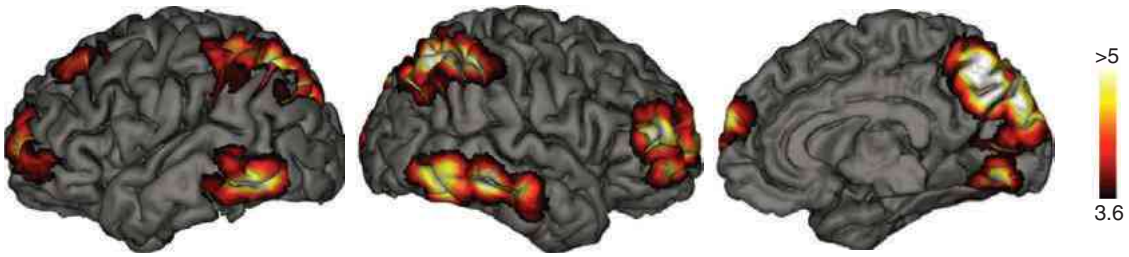


Figure 52-2 Functional MRI Data for Lucid Dreaming. During lucid REM sleep, strong activation is seen for dorsolateral prefrontal and frontopolar regions including the inferior, middle, and superior frontal gyri; parietal regions including the precuneus, inferior parietal lobule, and supramarginal gyrus; and temporal regions including the inferior and middle temporal gyri, as compared with nonlucid REM sleep. The *color bar* represents T-values. MRI, Magnetic resonance imaging. (From Dresler M, Wehrle R, Spoormaker VI, et al. Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: a combined EEG/fMRI case study. *Sleep* 2012;35:1017–20, with permission of the American Academy of Sleep Medicine; permission conveyed through Copyright Clearance Center, Inc.)

"local," not global feature of such phases. It is hardly imaginable, at least in nonpathologic cases, that the daydreaming subject misinterprets the daydream for reality, with full recognition of his or her current state prevailing. In the dreaming state, by contrast, this misperceived "reality" is completely normal—unless the dreamer eventually achieves lucidity by means of such prelucid reflection.[7]

Lucid dreaming may even be critical to achieve full understanding of the neural correlates of higher-order consciousness, because in contrast with coma-wake, anesthesia-wake, or sleep-wake comparisons, no major shift occurs in the vigilance state as defined by formal neurophysiologic criteria. Lucid REM sleep still is REM sleep proper, according to the classical Rechtschaffen and Kales (1968) or newer AASM[34] sleep scoring criteria. When compared with wakefulness, pathologic or pharmaceutically induced loss of consciousness also reduces the brain's basal metabolism, as does deep sleep. Lucid dreaming therefore provides the only known phenomenon that can contrast basal consciousness with full-blown higher-order consciousness within the same arousal level, allowing for comparison of cerebral activity by means of EEG, positron emission tomography, or fMRI without differences in the basal activity state.[51]

Higher-order aspects of consciousness are traditionally thought to be most pronounced in humans.[65,66] If the contrast between ordinary and lucid dreaming mirrors that between basal and higher-order consciousness, data on the neural correlates of dream lucidity might shed new light on this debate. Indeed, it turns out that cerebral regions showing increased activity during lucid dreaming also show extensive volumetric expansion in humans as compared with nonhuman primates.[67,68] In particular, the frontopolar cortex is significantly larger in humans than in other primate species[69] and has even been suggested to be a distinctly human brain structure.[70,71]

CLINICAL APPLICATIONS

Lucid dreaming has been suggested as a therapeutic approach for several clinical conditions, including nightmares, posttraumatic stress disorder, and schizophrenia. Lucid dreaming frequency is moderately correlated with nightmare frequency,[20] and people with frequent lucid dreams have incidentally reported that their nightmares have triggered lucidity. Theoretically, dream lucidity seems a logical solution to the main problem of nightmares, which encompasses a real emotional response to a nonexistent threat.[72] Becoming lucid in a nightmare should therefore take the sting out of it, and once the person realizes the threat is not real, it should disappear—along with the emotional response. Neurocognitive models of disturbed dreaming emphasize a hyperresponsivity of the amygdala in nightmare generation, coupled with a failure of medial prefrontal regions to dampen this activation.[73] Lateral prefrontal regions have been shown to be capable of influencing amygdala function through connections to the medial prefrontal cortex.[74] The neurobiology of lucid dreaming with increased lateral prefrontal activation therefore fits well with potential therapeutic effects of lucid dreaming on nightmares.[14]

Thus far, the theory holds—but does it work as readily in practice? Patients with narcolepsy who frequently suffer from nightmares report that dream lucidity intervention indeed provides relief during nightmares,[75] and a few case studies[76] and one small controlled pilot study[76] have indicated that lucid dreaming therapy was effective in reducing nightmare frequency. In the controlled pilot study, lucid dreaming therapy was superior to a waiting list regarding nightmare frequency but did not have an effect on secondary anxiety and sleep measures; its efficacy was much higher in individual patients than in a group therapy setting, suggesting confounding therapist effects.[77] A larger online self-help study did not find any additional effect of lucid dreaming therapy as an add-on to other effective cognitive-behavioral techniques such as imagery rehearsal therapy,[78] although low power and high dropout rates (>50%) limited the scope of the conclusions.

Lucid dreaming therapy raised some unexpected issues. An important consideration was that people with frequent nightmares may report becoming lucid but then find themselves unable to change the nightmare,[76] presumably because the expectations about the storyline may be too strongly engrained into the brain.[79] Moreover, realizing that one is dreaming does not automatically erase the threat and accompanying (intense) emotions, which could be expected to take some time after complete threat removal. As in all lucid dreams, lucidity in nightmares is not precisely an all-or-none phenomenon but rather a staged process, and a prelucid or half-lucid stage may not suffice to fully tackle a seemingly real threat. Moreover, many subjects with frequent nightmares reported a spontaneous change in their nightmares even without obtaining

lucidity.[77] This finding suggests that control over the nightmare, not lucidity, may be the therapeutic factor in successful nightmare treatment. Last but not least, even if the promising initial findings (which are only partly corroborated in controlled pilot studies) are sustained, the effects of lucid dreaming therapy tend to be stronger on nightmare aspects (frequency, intensity) than on general sleep quality or mental health characteristics. By contrast, the effects of imagery rehearsal have been much broader in scope.[80]

However, one disadvantage of the nightmare treatments that are currently best supported by experimental evidence, such as imagery rehearsal,[81] exposure therapy,[82] and combinations of both,[83] is that they also require a repetitive nightmare or theme to work with. If nightmares are too different from night to night, no story lines are available to rescript, as in imagery rehearsal, nor are repetitive images to systematically desensitize, as in exposure. Here, lucid dreaming therapy has the advantage that although having a repetitive nightmare or theme is beneficial (to allow recognition of the dream state in a future nightmare), it is not a sine qua non, because people can train themselves to become lucid without having nightmares.[72] Moreover, appropriate training can help patients link lucidity with feeling anxiety and fear in therapy and thereby prepare themselves for the next time they feel threatened—as it is likely to occur during a future nightmare. In this manner, lucid dreaming therapy can be useful for people with idiopathic nightmares with very different contents.

One important caveat is that lucid dreaming therapy may not be optimal for treating posttraumatic nightmares, besides being less evidence-based than imagery rehearsal. Because many posttraumatic nightmares may constitute a replication of an original event or parts of an original event,[84] changing the nightmare "on-line" during its occurrence may be much harder to achieve than changing it "off-line" in mentation, and even this issue typically raises questions concerning guilt and "undoing the past." Lucidity may thus have the adverse consequence that patients with posttraumatic stress disorder relive their original traumatic event with full consciousness but without the possibility to change anything.[76] Such an occurrence would be retraumatizing rather than empowering, and although solutions to this inability to change the nightmares have been proposed (e.g., starting by changing small background objects in color and then proceeding from there in small steps), it appears better to avoid experimenting with such inflammable material and to first try the treatments that may work in a majority of cases.

Besides nightmares, lucid dreaming also has been suggested as a therapeutic strategy in the treatment of schizophrenia.[3,28] The idea that normal dreaming can serve as a model for psychosis has a long and honorable tradition; however, it is notoriously speculative. One of the most interesting aspects of the dreaming-psychosis model is the issue of insight. Between 50% and 80% of patients diagnosed with schizophrenia have poor insight into the presence of their disorder,[85] probably owing to ineffective self-reflection processes.[86] Because such deficits are thought to lead to more relapses and rehospitalizations and poorer therapy success in general,[87] the concept of insight is becoming an increasingly important area of investigation in schizophrenia research.[88] On the dreaming side of the model, lack of insight into the current state characterizes almost any dream experience—with the obvious exception of lucid dreaming. This suggests that dream lucidity may be a good model for insight in the dreaming-psychosis model. Of interest, historical approaches to psychosis used the term "lucidity" to denote the patient's awareness of his or her illness.[89] Although the specific composition of the multiple facets of insight in psychosis is still under discussion,[90,91] two crucial dimensions are classically considered to be (1) the affected person's recognition that he or she has a mental illness and (2) the ability to recognize unusual mental events (delusions and hallucinations) as pathologic.[92] Hence, in the dreaming-psychosis model, lucidity during dreaming represents what patients in psychosis lack: full insight into the delusional nature of the current state of consciousness.

In neurobiologic studies, in particular, prefrontal, medial parietal, and inferior temporal cortical regions that are linked to insight problems in psychosis show striking overlap with brain regions associated with dream lucidity.[5] It has been demonstrated that prefrontal cortex function in schizophrenic patients can be improved through cognitive training.[93] Metacognitive training approaches are of particular interest, because skilled lucid dreamers typically gained their frequent insight into the dreaming state by metacognitive training, in particular by developing autosuggestions and the habit of frequently contemplating about their state of consciousness.[24,25] By teaching schizophrenia patients such training regimens, enhancing insight-related prefrontal and medial parietal functions might well lead to enhanced insight capabilities during acute psychosis. In addition, recent advances in dream lucidity induction by electrical brain stimulation methods[94,28] may show generalization effects to insight processes during psychotic wakefulness, or they may potentially serve as direct tools to improve insight during acute psychosis. Of note, a recent case study provides evidence that brain stimulation might indeed transiently attenuate insight problems in psychosis.[95] Lucid dreaming as a model for the successful treatment of psychotic symptoms also may be helpful for developing and testing new antipsychotic medication. If a given pharmacologic agent increases the frequency of lucid dreams in healthy subjects, it can be considered as a promising candidate to enhance insight in psychotic patients as well.[68] Lucid dreaming, therefore, transforms the dreaming-psychosis model from an interesting idea with a long history into a testable scientific hypothesis and a promising new therapeutic approach.

NONCLINICAL APPLICATIONS

Lucid dreaming also is used for several nonclinical purposes. A recognized strategy is to maximize certain behaviors or patterns within the dream state. Among the most popular intended dream behaviors are flying, communication with dream characters, and sexual encounters during dreaming, with lucid dreaming frequency appearing to predict how successful such intentions are recalled and executed in lucid dreams.[17] Besides such purely recreational applications within dreams, use of lucid dreaming also has been reported by many persons to influence aspects of waking life.[96] Two examples for which at least some scientific data are available are creative problem solving and practicing motor skills.

Anecdotal reports on scientific discovery, inventive originality, and artistic productivity suggest that creativity can be triggered or enhanced by sleeping and dreaming. In addition, theoretical considerations and experimental studies suggest that dreams can improve waking-life creativity.[44] Theoretically,

sleep has been suggested to provide an ideal state for creative incubation. The internally generated dream narrative, in the absence of external sensory data, leads to a much more radical renunciation from unsuccessful problem-solving attempts, leading to coactivations of cognitive data that are highly remote in waking life, and both dreaming and creativity have been characterized with primary process thinking, flat associative hierarchies, and defocused attention. In contrast with the more random flow of nonlucid dream narratives, dream lucidity allows for a more goal-oriented use of these creativity-related dream characteristics. Surveys among lucid dreamers and experimental studies demonstrate that lucid dreaming can indeed be used to improve creative thinking and problem solving.[97,96]

Motor practice during lucid dreaming is a novel type of mental rehearsal in which the person uses the dream state to consciously practice specific tasks without waking up.[98] It can be compared to mental practice, which is well established in sports theory and sports practice.[99] For both mental and dream rehearsal, movements are simulated with an imagined body on a purely cognitive level, while the physical body remains still. One advantage that lucid dreaming has over both mental practice and modern virtual reality simulators is that it offers the potential for practice with all kinesthetic sensations of the dream body in an environment that is experienced with as much vividness and realism as would be encountered in waking experience. In addition, the lucid dreamer, being limited only by imagination and attentional stability, has far greater potential for control over his or her own body, actions, and environment than in mental rehearsal, virtual reality environments, or waking life. In contrast with the vast amount of research on mental practice, however, empiric data on practice in lucid dreams are rather sparse.

In several anecdotal reports, amateur and professional athletes have described using lucid dreaming to improve their waking performance, such as in long distance running, tennis, skating, alpine skiing, or martial arts.[100,72] In a more systematic questionnaire study, 840 German athletes from a variety of sports were surveyed about their experiences with lucid dreams.[101] Although lucid dreaming in athletes was similar in prevalence to that in the general population,[19] the percentage of lucid dreams relative to all recalled dreams was found to be nearly doubled in athletes. Approximately 1 in 10 athletes who had lucid dreams (5% of the total sample) used lucid dreaming to practice sports skills, with most of them reporting improved performance.

Few studies have tested possible effects of practice in lucid dreams in controlled experiments. In a qualitative study, subjects were instructed to perform different complex sports skills familiar to them in waking life, such as skiing or gymnastics, in their lucid dreams.[102] Participants reported that they had no difficulties performing these sports skills in their lucid dreams and that their movements improved both in the dream and the waking state. In a quasiexperimental pre-/postdesign study, participants were asked to practice a coin-tossing task in their lucid dreams.[103] Results showed a significant increase in hitting the target from pretest to posttest evaluations for the group that practiced the coin-tossing task in lucid dreams, but no increase was found for the control group. More recently, these results could be replicated with a different motor task (sequential finger tapping). Improvements after lucid dream practice seem to be similar to or slightly less in degree than those obtained with actual physical practice, and similar to or slightly better than those with mental practice in wakefulness.[104]

CONCLUSIONS AND FUTURE DIRECTIONS

Up until the late 1970s, lucid dreaming met much skepticism or was completely ignored by mainstream sleep research, and even in the late 1990s, studying lucid dreaming was not considered to be experimentally advantageous for the neuroscience of consciousness.[105] Since then, an increasing number of studies have elucidated the neurobiologic basis of lucid dreaming and demonstrated its value for clinical and nonclinical applications. Nevertheless, owing to the rarity of the phenomenon, the study of lucid dreaming is still in its infancy, with many preliminary data demanding further confirmation and many details of the neural mechanisms underlying lucid dreaming and its therapeutic effects awaiting thorough investigation. In particular, reliable lucidity induction strategies are needed to boost lucid dream therapy and the relevant research.

CLINICAL PEARL

Lucid dreaming has been proposed as a natural therapy for nightmares, because the insight into the illusional nature of the dreamed threat could prevent the emotional response to it. Lucid dreaming as a therapy for nightmares has indeed some clinical support. It might be particularly suited for addressing nonrecurring nightmares but less so for managing posttraumatic nightmares. Because lucid insight into dreaming and insight into psychosis appear to largely share the same brain basis, lucid dreaming also may potentially be of value in schizophrenia therapy or in the development of novel antipsychotics.

SUMMARY

In contrast with the metacognitive impairments of normal dream mentation, lucid dreaming is characterized by awareness of the current state of mind, often leading to considerable volitional control of the dream narrative. Lucid dreaming is a rare skill; however, it can be learned and reinforced with training using a variety of induction strategies ranging from autosuggestion to transcranial current stimulation. Lucid dreaming as a research topic has faced much skepticism during most of the past century; in recent years, however, interest is acquiring increasing momentum. Lucid dreaming is associated with specific changes in neural activity when compared to nonlucid dreaming, with lateral prefrontal, frontopolar, and medial parietal activation as proposed neural correlates of the increased metacognitive capacity that defines dream lucidity. Lucid dreaming has clinical and nonclinical applications, ranging from nightmare therapy to mental motor skills training and creative problem solving. Reliable induction methods are strongly needed to further explore the potential of lucid dreaming and for its scientific study.

Selected Readings

Dresler M, Wehrle R, Spoormaker VI, et al. Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: a combined EEG/fMRI case study. *Sleep* 2012;**35**:1017–20.

Dresler M, Wehrle R, Spoormaker VI, et al. Neural correlates of insight in dreaming and psychosis. *Sleep Med Rev* 2015;**20**:92–9.

Filevich E, Dresler M, Brick TR, Kühn S. Metacognitive mechanisms underlying lucid dreaming. *J Neurosci* 2015;**35**(3):1082–8.
LaBerge SP, Nagel LE, Dement WC, Zarcone VP. Lucid dreaming verified by volitional communication during REM sleep. *Percept Mot Skills* 1981;**52**:727–32.
LaBerge S, Rheingold H. *Exploring the world of lucid dreaming*. New York: Ballantine; 1990.
Purcell S, Mullington J, Moffitt A, et al. Dream self-reflectiveness as a learned cognitive skill. *Sleep* 1986;**9**:423–37.
Rak M, Beitinger P, Steiger A, et al. Increased lucid dreaming frequency in narcolepsy. *Sleep* 2015;**38**(5):787–92.
Spoormaker VI, van den Bout J. Lucid dreaming treatment for nightmares: a pilot study. *Psychother Psychosom* 2006;**75**:389–94.
Stumbrys T, Erlacher D, Schädlich M, Schredl M. Induction of lucid dreams: a systematic review of evidence. *Conscious Cogn* 2012;**21**:1456–75.
Voss U, Holzmann R, Hobson A, et al. Induction of self awareness in dreams through frontal low current stimulation of gamma activity. *Nat Neurosci* 2014;**17**:810–12.
Voss U, Holzmann R, Tuin I, Hobson JA. Lucid dreaming: a state of consciousness with features of both waking and non-lucid dreaming. *Sleep* 2009;**32**:1191–200.

A complete reference list can be found online at ExpertConsult.com.

Chapter
53

Nightmares and Nightmare Function

Tore Nielsen; Michelle Carr

Chapter Highlights

- The definition of *nightmare disorder* as a clinical entity by the *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5) and the *International Classification of Sleep Disorders,* third edition (ICSD3) suggests that nightmares are symptomatic but not functional.
- The nearly ubiquitous nature of nightmares and dysphoric dreaming in the general population is consistent with the claim that they serve a function related to cognitive-emotional regulation or emotional memory consolidation.
- Theories of nightmare function vary in their emphasis on mechanisms for dealing with emotional regulation and include stress mastery, affect desomatization, fear memory extinction, emotional contextualization, and others.

WHAT IS THE FUNCTION OF NIGHTMARES?

There are diverse opinions—but little consensus—about why dreaming so frequently turns dark and interrupts sleep with fear-filled awakenings. Are these nightmares simply symptomatic of an underlying clinical condition? Or do they serve a homeostatic or cognitive function related to the processing of intense emotion? Or might some nightmares serve a function while others are symptomatic, and if so, how are these nightmare types to be distinguished? The characterization of nightmares remains a major unresolved issue of modern sleep medicine and one with important ramifications for our understanding of mental health and cognitive development. The present chapter addresses the problem of nightmares from the perspective of their possible functionality and reviews new findings that bear on the problem.

We first consider literature demonstrating that nightmares have been, and continue to be, widely accepted as a clinical symptom and thus, possibly, have no adaptive function. We then review studies showing that nightmares and a wide spectrum of other types of negative dreams are highly prevalent in the general population and thus that some nightmares—especially those that are infrequent, or not severe or recurrent in nature—may well play a functional role. After touching briefly on polysomnographic studies, which are inconclusive on nightmares' functionality, and neurocognitive studies, which are too scarce to permit drawing firm conclusions, we review the principal theories that address whether nightmares serve an adaptive function. These theories are multifaceted, propose a variety of hypothetical mechanisms, and for the most part support the notion that nightmares are, indeed, functional.

In the present context we consider the term *nightmare* to refer to dreaming during which intense negative emotion is in play. In this definition we subsume dreams with a variety of dysphoric emotions as well as dreams that do not immediately awaken the dreamer. This broad definition, although not shared by all writers, corresponds closely with current clinical definitions (Table 53-1) as well as with earlier DSM-III and DSM-III-R definitions of *dream anxiety attacks* and *anxiety dreams* and is employed so that a wide swath of nightmare theories may be included in the discussion.

NIGHTMARES ARE A RECOGNIZED CLINICAL ENTITY

Nightmares have for centuries been viewed as pathologic; accordingly, clues to their possible functionality may be found in how they are described clinically. Nightmares are, according to the most authoritative sources, the DSM-5[1] and ICSD3,[2] powerful unpleasant dreams associated with feelings of threat, anxiety, fear, or other negative emotions that occur during late-night REM sleep and that are clearly recalled on awakening (see Table 53-1). The implication of basic fear expression in nightmare genesis is indicated by the fact that fear manifests in 65% to 85% of nightmares, whereas other dysphoric emotions such as anger and sadness prevail in the remainder.[3,4] This clear predominance of fear may mean that nightmares are akin to the symptoms of other fear-dysfunction disorders, such as phobias, generalized anxiety, or social anxiety, but it may also point to a deeper involvement of fear memory, fear extinction, and fear regulation systems that underlie the normal functions of emotional learning and emotional memory consolidation.[5-7] These are not mutually exclusive possibilities of course.

The pathologic context of nightmares is striking and has been reviewed in detail elsewhere.[6,8] Pathologic conditions that are comorbid with nightmares range from the mild to the severe, but causality between nightmares and other pathologies has not yet been clearly established. Nightmares are more frequent among those suffering from impaired sleep quality,[9-11] a variety of sleep disorders,[11,12] depressive and anxiety symptoms and neuroticism,[6,8] and posttraumatic stress disorder (PTSD)[13] than they are among healthy individuals. Nightmares are also reliably associated with suicidal ideation,[14] suicide attempts,[15] and death by suicide[16]—independent of other psychopathologies.[14,17,18] Frequent nightmares are also associated with the eveningness chronotype.[19,20]

Table 53-1 Diagnostic Criteria for Nightmare Disorder from the DSM-5 and ICSD3*

	DSM-5	ICSD3
A. Nature of recalled dream	Repeated occurrence of extremely dysphoric, well-remembered dreams, usually involving threat and occurring in the second half of sleep	Repeated occurrence of extremely dysphoric, well-remembered dreams, usually involving threat and occurring in the second half of sleep
B. Nature of awakening	Becomes alert and oriented on awakening	Becomes alert and oriented on awakening
C. Nature of distress	Causes clinically significant distress or impairment	Causes clinically significant distress or impairment in one of the following areas: mood, sleep, cognition, family, behavior, daytime sleepiness, fatigue, occupation, social
D. Differential diagnosis	Not substance derived	N/A
E. Differential diagnosis	Not due to other mental or medical disorder	N/A
Duration	Acute: <1 mo Subacute: <6 mo Persistent: >6 mo	N/A Note: Nightmare disorder only diagnosed in children in cases of persistent distress
Severity	Mild: <1 per wk Moderate: 1-6 per wk Severe: ≥7 per wk	

**International Classification of Sleep Disorders,* third edition (ICSD3) criteria have changed only slightly from the ICSD2. *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5) criteria have changed from DSM-IV-TR in several respects. DSM-5 introduces categorization of subtypes according to duration and severity (used in ICSD2), whereas ICSD3 drops them. Terrifying hypnagogic hallucinations in DSM-IV-TR are subsumed in DSM-5 as sleep-onset subtype of nightmares. In DSM-5, nightmare disorder may be diagnosed in cases of REM sleep behavior disorder, posttraumatic stress disorder, and acute stress disorder if nightmares preceded the condition and their frequency or severity necessitates independent clinical attention. Both manuals now include any dysphoric emotional tone and recognize the frequent depiction of threat-related content. Awakening from sleep is no longer a defining criterion of a nightmare.

Nightmare-focused treatments can alleviate comorbid symptoms such as anxiety, depression, and PTSD symptoms[21,22] suggesting that nightmares may contribute to the pathology of these conditions.

Altogether, the continued description of nightmare disorder in the DSM-5 and ICSD3 as a primarily fear-based disorder, as well as accumulating evidence linking nightmares to various comorbid affective conditions, supports the position that frequent nightmares reflect a pathologic breakdown in the normal functioning of processes governing fear expression, fear memory, or fear regulation. Nonetheless, despite the distress and suffering caused by nightmares, it remains possible that they also signify—at least up to a certain degree of severity—innate adaptive responses over the long term. Both possibilities, the dysfunctional and functional theories of nightmares, remain viable, albeit unproven, ideas. Nevertheless, in both cases, the presence of multiple comorbid conditions among nightmare sufferers substantially complicates the determination of whether nightmares are a primary contributor to these clinical conditions and whether they support some kind of adaptive function.

NIGHTMARES ARE UBIQUITOUS

Large population studies indicate that nightmares are a prevalent clinical problem but also that dysphoric dreaming is much more ubiquitous than is generally appreciated. Nightmare prevalence at a clinically significant frequency (i.e., about 1/week or more[1]) varies from 0.9% to 6.8% of individuals.[23] The two largest cohort studies (i.e., 69,813 participants from the general Finnish population[23] and 87,408 7th- to 12th-graders from Japan[24]) provide consistent estimates. The former study found that 4.2% reported "frequent" nightmares in the past 30 days; the latter that 6% reported nightmares "always" or "often" in the same time period. However, this is not the complete picture. For one thing, in most of these studies the definition of a nightmare included the requirement that the dream lead to an awakening. In addition, nightmares occur at lower frequencies among many more people (e.g., 40% of the Finnish cohort reported "occasional" nightmares during the past 30 days). A full 85% of adults report at least one nightmare per year.[6] Occasional nightmares are not generally considered pathologic; they may well be evidence of nightmares sustaining a normal, adaptive response. The widespread occurrence of nightmares is also supported by the finding that prospective measures, such as home dream logs, estimate them to be 3 to 10 times more frequent than do retrospective measures, such as questionnaires.[25-27] In addition, there is a much wider spectrum of disturbed and dysphoric dreams, of which nightmares are clearly a part and from which they have not been clearly distinguished.[6,28] To illustrate, many types of disturbed dreams have been described during bereavement,[29] during pregnancy,[30] following trauma or brain surgery,[31] after consuming or withdrawing from various drugs, or in tandem with many mental, physical, and sleep disorders.[28,31] Bad dreams, in particular, have been distinguished from nightmares in that they do not lead to awakenings, are less emotionally intense, but are as likely to contain negative emotions.[32] Bad dreams occur up to four times more frequently than do nightmares.[27,32] Bad dreams and nightmares together occur on average at a rate of about 40 per year among healthy university students[27] and constitute 13.7% of all dreams reported by 572 subjects (9796 dreams).[32] That bad dreams are thematically similar to nightmares (e.g., depict physical aggression) yet more likely to resolve positively (38%) than nightmares (22%),[32] suggests that they may be more functional in

regulating emotions than are nightmares. Beyond bad dreams, negative emotions constitute from 66% to 80% of all dream emotions in home dreams.[33,34]

In sum, although nightmares are clearly recognized as a clinical entity, the ubiquity of nightmares as part of a wider, still insufficiently articulated, spectrum of disturbed and dysphoric dreams supports the possibility that they may play a role in cognitive-emotional regulation or processing of emotional memories; a role that may diminish as the nightmares become more severe and disruptive.

SLEEP POLYSOMNOGRAPHY FINDINGS FOR NIGHTMARE SUFFERERS ARE INCONCLUSIVE

Polysomnography (PSG) has revealed several sleep-dependent memory functions, with memory improvements having been linked to both proportions of sleep stages and microstructural sleep features such as spindles and rapid eye movement (REM) density.[35] REM sleep, in particular, has been linked to the processing of emotional stimuli, such as consolidation of fear and safety memories[36] or of the negative component of complex pictures,[37] and to modulation of emotional reactivity.[35] In light of such advances, the PSG features characterizing nightmare sufferers—and their REM sleep features in particular—may provide clues to nightmare functionality.

Unfortunately, PSG studies pertaining to nightmares and nightmare sufferers are few and inconclusive. Two studies of nightmare episodes[38,39] both revealed signs of REM sleep activation (e.g., heart rate increase), which can be attributed to autonomic arousal that would be expected to accompany fear. On the other hand, in a surprising 60% of cases[39] this expected autonomic arousal was not observed, raising the possibility of a dampening of affect expression during nightmares (see later). Other studies have found more eye movements per minute and shorter respiration rates for dreams that are high compared with low in anxiety.[40] For the habitual sleep of nightmare sufferers, no consistent patterns of PSG abnormalities in either microstructure or macrostructure have been established. Some studies find REM-specific abnormalities, such as increased skipping of early REM periods, increased REM latency and cycle length, more REM periods,[41] and increased spectral power in the high alpha range (10 to 14.5 Hz),[42] whereas others report changes in non–rapid eye movement (NREM) sleep such as low alpha power,[42] reduced cyclic alternating pattern (CAP) A1 but increased CAP A2 and A3 subtypes,[43] and reduced slow wave sleep.[44] Yet other studies report global changes such as more frequent periodic leg movements,[45] more fragmented sleep,[39,46] longer sleep latency, more nocturnal awakenings,[44] and increases in the normalized low-frequency component of the heart rate during recovery sleep after REM sleep deprivation.[47] These findings are discrepant but could be considered to reflect an increase in arousal during the sleep of nightmare sufferers, whether expressed as leg movements, nocturnal awakenings, or alpha oscillations. This notion fits with the subjective experience of increased emotional and physical arousal during nightmares. However, increased arousal in sleep is not a highly specific correlate of known sleep-related functions and also does not account for why sometimes nightmares are triggered and sometimes they are not.

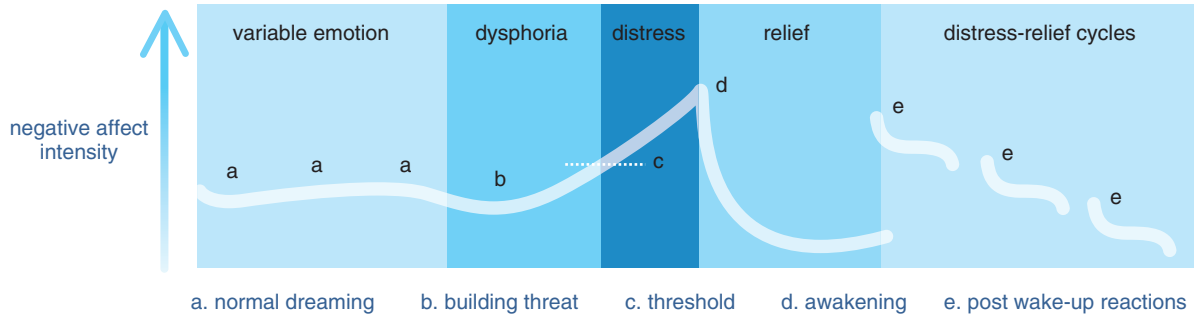
Overall, the inconsistencies and lack of replication studies in the sleep literature do not yet allow reliable links to be made between nightmares and known memory or emotion regulation functions of sleep.

NEUROPSYCHOLOGICAL STUDIES OF NIGHTMARE SUFFERERS ARE SCARCE

The testing of cognitive or neuropsychological attributes might illuminate the functional status of nightmares by revealing specific waking state deficits or advantages that are characteristic of this population. This is analogous to how such testing has revealed relationships between sleep-related memory consolidation and brain conditions like functional memory impairment[48] or mild cognitive impairment[49] or between cognitive competency (IQ) and sleep spindles.[50] However, research on nightmares is scarce. One study[51] of nightmare sufferers and controls revealed some performance deficits and some advantages on prefrontal and frontal-limbic measures of executive functions and emotion regulation. Nightmare sufferers showed general slowing on the Emotional Stroop, longer reaction times on the Emotional Go/NoGo, and slightly lower fluency and substantially higher perseveration on Verbal Fluency. On the Color-Word Stroop, however, nightmare subjects committed *fewer* errors than controls for incongruent word/color pairings, suggesting superior functioning on tasks implicating dorsolateral prefrontal and dorsal anterior cingulate regions.[51] Our attempt to replicate the elevated Verbal Fluency perseveration finding[52] was only partially successful; we failed to find a difference between nightmare and control groups, but we did show a positive correlation between perseveration and nightmare severity. These findings are thus consistent with the possibility of a dose-response relationship between frontal deficits and nightmare severity. Finally, neuropsychological studies of brain injuries and disease[31,53] reveal that nightmares with recurrent stereotypical themes are often associated with seizure disorders and usually involve lesions in the right temporal lobe.[31] Recurrently themed nightmares that replay the trauma are also characteristic of PTSD patients.[54] This particular type of nightmare, then, may reflect a fundamental neural dysfunction.

THERE IS A SPECTRUM OF NIGHTMARE FUNCTIONALITY THEORIES

In the present work we use the term *functional* to refer to a biologic, adaptive advantage attributed to nightmares (see review in Revonsuo[55]). The question of whether nightmares are functional has remained contentious since at least the time of Freud[56] and continues to divide opinion today. Freud, in fact, supported both points of view. He initially considered nightmares ("anxiety-dreams") to conform to his theory that dream function in general is to repress expressions of unacceptable unconscious impulses.[56] However, he also admitted that nightmares challenge this theory by recognizing a repetition compulsion in the more severe "war neurosis" nightmares and, later, other "immoral" dreams for which the dream work mechanisms of censorship had broken down.[57] In more recent times, with growing emphasis on the medicalization of nightmares (DSM),[1] as well as evidence linking nightmares to psychopathologic conditions, beliefs that nightmares have no function or are dysfunctional are widespread. Two such theories[58,59] single out nightmare awakenings as evidence that



Function step(s)	a	b	c	d	e
1. All nonfunctional	✗	✗	✗	✗	✗
2. Normal dreaming	✓	✗	✗	✗	✗
3. Normal dreaming/threat	✓	✓	✗	✗	✗
4. Threat	✗	✓	✗	✗	✗
5. Threat/distress	✗	✓	✓	✗	✗
6. Distress	✗	✗	✓	✗	✗
7. Normal dreaming/threat/distress	✓	✓	✓	✗	✗
8. Wake-up	✗	✗	✗	✓	✗
9. Recall	✗	✗	✗	✗	✓
10. All functional	✓	✓	✓	✓	✓
✗ = nonfunctional; ✓ = functional					

Figure 53-1 *Upper panel:* Five phenomenologic steps of a threat-based nightmare displaying increases in negative affect intensity (*pale solid lines*) and specific affective reactions (*top row of text*). The steps consist of a period of normal dreaming with variable emotions (a); a period of dreaming with building threat and concomitant increases in dysphoria (b) to an arbitrary threshold where the experience evokes increasing distress (c); an awakening, which brings relief and other feelings (d); and a series of postawakening reactions (e) when the nightmare emotions may be recalled, followed by progressively more relief. The latter step may last for minutes, hours, or longer. *Lower panel:* Ten possible theories classified according to where in the five-step sequence functionality is hypothesized to occur (✓) or not occur (✗). *Nonfunctional* theories (1)—at one hypothetical extreme—stipulate that all steps of a nightmare, including prior dreaming, distress threshold, waking up, and postwaking reactions, have no functional value, whereas *all functional* theories (10)—at the other extreme—stipulate that every step is functional. Other, less extreme theories (2 to 9) correspond to most published theories in stipulating that some step or combination of steps of the sequence may be functional. Not all of the 32 possible theories for this five-step sequence are shown. The content of dreaming does not figure in this hypothetical sequence but may nonetheless be decisive; in particular, if the dream replicates a prior trauma, a given step may be hypothesized to be nonfunctional.

dreaming's function has failed. A third[7] considers failures to regulate fear extinction during dreaming as responsible for dysfunction. Apart from these, however, theories formally describing nightmares as a functional failure are few. On the other hand, several functional theories invoke processes of emotion regulation (see later).

Although it may be tempting to view existing theories of nightmare functionality as dichotomous (i.e., that nightmares either are or are not functional), a closer consideration of the literature and of the dynamic phenomenology of nightmares suggests that a number of gradations of functionality falling between these two extremes have been described. As Figure 53-1 illustrates for the case of a simplified typical threat-nightmare, at least five phenomenologic steps with five accompanying emotional reactions unfold during a nightmare experience. These include a phase of normal dreaming, a period of building dysphoric emotion, an arbitrary threshold at which the dream passes over into a realm of distress, an awakening, and postawakening reactions. Theories may attribute functionality or a lack of it to any one or combination of these steps. The lower panel of Figure 53-1 details 10 hypothetical theories, some of which have analogues described in the literature (see later). For example, a theory that considers threatening dreams and nightmares to be functional regardless of affect intensity, such as theory 5,[55] would include steps b and c but not steps a, d, or e. In contrast, a theory that sees nightmares as wholly dysfunctional, such as theory 2, might include steps a or b, but not c, d, or e.[58] Even this detailed breakdown is not exhaustive; functionality or a lack of it can be attributed to specific types of dream content, such as problem-solving (functional) versus recurrent (dysfunctional) content.

Altogether, the empiric literature provides a poor basis on which to determine whether nightmares are functional and, if they are, what parts of the nightmare sequence sustain this function. Problems with how nightmares are conceived and categorized complicate the clear formulation of questions about functionality. The many types of nightmares and other disturbed dreams may or may not all serve the same function—if any at all. Without further taxonomic progress in the field, questions about nightmare function remain very speculative (see Nielsen[59a] for discussion).

A REVIEW OF NIGHTMARE FUNCTION THEORIES

In the following section, we discuss how several nightmare theories pertain to the question of function. Although a number of studies report evidence for genetic[60] and personality risk factors for nightmares, such findings typically have not been formalized into theories of nightmares with implications for function and so will not be considered further here. Also, *The Interpretation of Dreams*[56] heavily influenced many later nightmare theorists (e.g., Jones'[61] theory that anxiety-dreams contain unconscious sexual impulses) and remained influential among psychoanalytic thinkers for half a century. Only three theories that diverge in important respects from classical psychoanalytic thinking—by Kellerman,[62] Fisher,[39] and Palombo,[63]—are reviewed here. Most of the theories we consider share a general assumption that nightmares are implicated in the adaptive modification of emotional responses over time. We present these theories separately to highlight the more specific mechanisms of emotion regulation that they bring forward and to summarize the available empiric findings that support or refute each mechanism.

Emotion-Defense Regulation

One neo-psychoanalytic theory[62] diverged from Freud (1900) in attributing nightmare awakenings to ego defense mechanisms that fail to contain the overactivation of specific types of emotions—and thus fail in the dream's function. Nightmare awakenings are purportedly triggered by eight basic emotion types: joy, acceptance, surprise, expectation, anger, disgust, sorrow, or fear,[64] each of which produces a distinctive nightmare theme and, frequently, its own characteristic behaviors on awakening. For example, fear could lead to a terror nightmare with overt locomotion or speech on awakening; anger could lead to a rage nightmare with fist clenching; sorrow could lead to a grief nightmare with copious crying; and joy—paradoxically perhaps—could lead to a "pleasure" nightmare with nocturnal orgasm. This eightfold structure of nightmare themes was extended to include not only ego defense mechanisms—one type per theme—but also more general attributes of, for example, personality structure, cognitive orientation, and psychosomatic organ systems.[62] An intriguing component of this theory is that every nightmare is thought to contain an element of fear to the extent that there is resistance to fully express a predominant emotion while dreaming, such as a fear of letting anger amplify to overt rage or of letting joy or pleasure amplify to the point of orgasm.

Despite its clarity and accessibility, this theory has not been tested empirically and has fostered surprisingly little research; Google Scholar identified only 10 citations of the work over 27 years, and none was an empiric test of the theory. The theory is, however, generally consistent with current definitions of nightmares as including emotions other than fear as well as studies showing that awakenings from nightmares and other dysphoric dreams are often accompanied by dream-enacting behaviors in both parasomnias like REM sleep behavior disorder[65] and among the general population.[66]

Adaptation to Stress

Although early theories of dream function emphasized roles for REM sleep and dreaming in facilitating adaptation to stress, most of these dealt only marginally with nightmares. Thus, despite evidence that presleep stressors are incorporated into dream content or followed by an increase in dysphoric dream emotions,[67-70] findings were generally interpreted to indicate how normal dreaming was able to "master"[67] or "assimilate"[71] daytime stress rather than to explain how stress might trigger nightmares or how nightmares might modify stress. One exception to this trend was a focus on the possible role of nightmares in "war neurosis,"[72] now known as PTSD, which considered dream function to be mastery of stress and nightmares to reflect continuing attempts to master a trauma.

Another exception was the disruption-avoidance-adaptation theory,[73,74] which hypothesized that dreaming enables stress adaptation by its oscillation between two distinct functions: mastery and avoidance. Dreams about unresolved disturbing events are considered *mastery dreams* and have the potential to disrupt sleep—as in the case of awakenings from nightmares. Mastery is hypothesized to occur by a type of creative emotional problem solving that draws on memories of similar, yet successfully resolved, past situations. Emotional mastery is favored by the lack of interruption, disregard for social acceptability, and free-flow of ideas, thoughts, and emotions unique to dreaming. If dreams are too disruptive, however, awakenings and other sleep disturbances may result. *Avoidance dreams* prevent mastery attempts and thus sleep disruption by various processes (e.g., presenting dream emotions or specific contents that have no apparent relationship to the waking stressor). Oscillation between mastery and avoidance dreams continues, within and across nights, until adaptation is attained. Nightmares, because they disrupt sleep, are evidence of failure of both mastery and avoidance functions.

Evidence supporting stress adaptation theories is mixed. Some studies[75] found that dreaming about a stressful presleep stimulus led to improved mood in the morning, whereas others[76] found essentially the opposite. The disruption-avoidance-adaptation theory was presented specifically to deal with such inconsistencies. However, suggesting that post-stressor dreams may be either related or not to prior stressors amounts almost to an unfalsifiable theory. More detailed predictions are needed about precisely when, and in which sequence, mastery and avoidance dreams are expected following stress.

Desomatization

Fisher and colleagues[39] found that some REM sleep nightmares were not accompanied by the autonomic activation that would be expected for the degree of negative emotion reported in the nightmare. In 60% (12 of 20) of their recorded nightmares, emotion-related autonomic activation, as measured by heart rate, respiratory rate, and eye movement activity, were not seen. In other nightmares, such activity occurred only in the last few minutes of the REM episode. Similar findings were reported in a more recent study.[38] This apparent separation of seemingly fearful dream imagery from its expected autonomic correlates prompted the notion of REM dreaming as a mechanism for "tempering and modulating anxiety, for desomatizing the physiological response to it … [for] abolishing or diminishing the physiological concomitants" (p. 770).[39] This desomatization mechanism was thought to help protect REM sleep, to assuage anxiety during dreaming, and to decrease the degree of disruption that occurs after an awakening. This serves to prevent the self-perpetuation of anxiety and assist in the mastery of traumatic memories—even after

awakening.[39] A breakdown of the mechanism is suggested by intense autonomic activity while dreaming, as occurred in 40% of their recorded nightmares.

Similar desomatization notions have occurred sporadically in the literature, and slightly different desomatizing mechanisms have been proposed. A short note on a desensitization function of dreaming[77] and an empiric study on an anxiety-extinction function of nightmares[78] were both published shortly after Fisher and colleagues' work.[39] These authors suggested that nightmares facilitate extinction through repeated exposure to fear-inducing stimuli (like implosive therapy), but their own findings did not support the hypothesis.[78] Others have pointed to specific physiologic mechanisms of REM sleep that might be responsible for desomatization. These include REM sleep eye movements desensitizing affect by a mechanism similar to that of eye movement desensitization and reprocessing[79,80]; REM sleep atonia desensitizing the somatic component of negative affect by repeatedly blocking kinesthetic feedback during negative dream imagery[81,82]; and, in a related theory, the repeated pairing of dysphoric dream imagery and REM sleep atonia desensitizing anxiety in a manner analogous to systematic desensitization therapy.[83] In all such models, negative affect is considered to be implicated in desensitization up to a certain threshold; this threshold is clear when it is the point of waking up (analogous with flooding therapy) but otherwise remains undefined in most cases.

In addition to evidence replicating Fisher and colleagues' observation of desomatized nightmares,[38] there is evidence that dream emotion may be inhibited by REM sleep processes related to the orienting response[81] or modulated by two REM sleep mechanisms associated with threat-fear and loss-sadness.[29] Desomatization theories imply that negative emotion should be progressively reduced over time, especially within a single dreaming episode. Two studies suggest, in fact, that dream emotion becomes more negative over time.[84,85]

In sum, desomatization theories propose mechanisms of emotion regulation during dreaming and nightmares by which strong emotion is downregulated by its repeated pairing with processes such as autonomic inhibition, muscle atonia, eye movements, or orienting reactions.

Regulation of Negative Moods

The mood regulatory theory posits an emotion regulation function for dreaming[58,86] and is similar in many respects to desomatization approaches. It assumes that a cardinal characteristic of REM sleep is a surge of affective arousal that unfolds over the REM episode. The surge consists of a progressive increase and subsequent plateau in autonomic arousal as indicated by heart and respiratory rates, eye movement amplitude and density, and limbic system activity, among other markers. Dream content "contains" these surges by reducing the intensity and variability of the associated emotions. This is achieved by *progressive-sequential* patterns of dream content that unfold over successive REM periods and that facilitate emotional problem solving. The progressive-sequential dream pattern is distinguished from a *repetitive-traumatic* pattern during which an emotional conflict is simply stated and restated without evidence of adaptive change. Nightmares contribute to problem solving up to a point where the capacity for assimilation of emotional surges is exceeded. Similar mood regulation theories that consider nightmares to function as a coping mechanism for stress have been proposed by others.[87,88]

The physiologic assumption that REM sleep is surgelike in nature remains to be demonstrated empirically. However, dreams are influenced by presleep thoughts and emotions[89] and are related to the next day's mood[90]; the frequency and intensity of nightmares is associated with both increasing daily stress and increasing coping efforts.[88] Stress that is induced either experimentally (e.g., difficult intelligence test[76]) or naturally (e.g., earthquake[91]) leads to an increase in incorporation of the stressor in later dreams and nightmares. Moreover, overnight reductions in negative mood (e.g., unhappiness) correlate with intervening dream content, especially with the number of characters in a dream.[86] Consistent findings were also reported[92,93] by which high presleep depression scores were associated with more dysphoric dreams from the first REM period, but not with sleep physiology variables. Other supporting evidence from this group[87,94] shows that subjects in a marital breakup who report a predominance of early-night negative dreams are more likely to be in remission a year later than are those with more late-night negative dreams. Negative dreams occurring early in sleep may thus reflect a within-sleep mood regulation process similar to the problem-solving pattern that is triggered by a major emotional conflict; a predominance of negative dreams late in sleep may reflect a failure of this regulation function.

In sum, mood regulation models posit mechanisms for downregulating intense emotions by the nature and structure of dream content; this may implicate the regular coupling of emotional surges with a problem-solving dream structure that unfolds over time or the timing of negative dream emotion to occur early in the night.

Sleep to Forget and Sleep to Remember

The sleep to forget and sleep to remember theory[95,96] proposes that REM sleep serves a dual function in prioritizing consolidation of emotional memories and reducing the "affective blanket" that surrounds them. Emotion reduction prevents an unwanted build-up of anxiety in associated memories and thus helps prevent development of affective disorders. Dreaming consciousness is hypothesized to play a role in this function, and recurrent nightmares indicate repeated attempts (and failures) in this process of separating affective tone from emotional memory. The theory is supported by evidence that memory is better for emotional than for neutral experiences,[97] but also by observations that emotional arousal accompanying the remembering of experiences decreases over time. A central claim is that REM sleep reduces amygdala reactivity to memories while retaining the hippocampally held memory, possibly through the theta oscillations characteristic of REM sleep. Downregulation of emotion is attributed to reduced aminergic activity during REM sleep,[98] allowing a decoupling of a memory trace from its associated autonomic reaction.

The sleep to forget and sleep to remember theory is presented as complementing previous dream-based theories of emotion regulation[92,99] with a neurophysiologic explanation. It thus postulates a possible role for dreaming in completing this function, but does not specify the precise dream mechanism. Nonetheless, it is suggested that if emotion depotentiation is unsuccessful on a given night, further attempts on subsequent nights will appear in a pattern of recurrent REM

nightmares, as is characteristic of PTSD.[95] This theory thus associates recurrent nightmares with a failure of dream function.

Corroborating research for the theory is that REM sleep plays a role in emotional memory consolidation[100,101] and the fact that brain regions active during REM sleep overlap with regions implicated in emotional functioning, particularly the amygdala, the anterior cingulate, and the ventromedial prefrontal cortex.[102,103] A key study[104] demonstrates that participants deprived of sleep after memorizing emotional pictures show both reduced recall 72 hours later (poor memory) and lack of reduction in amygdala reactivity (increased emotion) when viewing the pictures a second time. Our recent study of an REM-sleep priming effect[105] supports the theory in that the presence of REM sleep predicts more priming to emotion cue words, although time in REM predicts less priming.

Fear Extinction

A recent theory[6,7] ascribes a fear memory extinction function to dreaming and explains nightmares as a perturbation of this function. Cognitive and neural levels of description are articulated.

Cognitive-Level Explanation

Fear extinction by this theory entails: the activation of fear memory elements such that they are isolated and removed from their episodic contexts, the recombination of these into novel here-and-now simulations of reality, the expression of alternate emotional reactions to this virtual context, and ultimately the production of new *fear extinction memories*. Fear extinction memories provide a sense of safety and thus compete with and, if consolidated, supersede the original fear memories. Fear extinction memories are realized by coupling fear memory elements with realistic nonaversive contexts that are incompatible with fear by a mechanism obeying Pavlov's[106] principles of fear memory learning and extinction.[107] Nightmares occur as a result of disruption of this mechanism. For example, if an entrenched fear memory resists recombining with new contexts, as might be the case for nightmares with recurrent themes, new extinction memories may not be formed. Or, if an extinction memory is not properly consolidated, an original fear memory may be reinstated.[108] Temperament may also interact with these basic extinction processes, especially one's susceptibility to *affect distress,* which could amplify the emotional responses produced.

Neural-Level Explanation

At the neural level, fear extinction is supported by a network of at least four regions that control the representation and expression of emotions in both sleeping and waking states: the amygdala, the medial prefrontal cortex (mPFC), the hippocampal complex, and the anterior cingulate cortex (ACC). Normally, these regions interact with larger integrated networks such as the default mode network[109] or the extended mirror neuron system.[110] Each region controls a set of extinction processes defined at the cognitive level of explanation. During nightmares, the amygdala may be hyperresponsive to fear-related memory elements portrayed in the dream, whereas processes in the mPFC, the hippocampal complex, or the ACC that normally downregulate amygdala activity may be disrupted. The result is the abnormally intense activation of fear. This situation parallels empirically supported models of PTSD pathology.[111] Affect distress, linked to ACC activity, is shaped by the emotional history of the individual (e.g., early adversity) but also by pain-related distress, social rejection, and difficulties with emotional expression.[112]

The fear extinction theory is grounded in a considerable amount of neuroscience research but has not yet generated a great deal of new findings. One test of the theory[51] used neuropsychological tests to assess involvement of frontal brain regions and supported the theory to the extent that frequent nightmare sufferers showed poorer performance on verbal fluency perseverations. These findings were partially replicated with a cohort of French-speaking participants.[52]

Cognitive Avoidance and the Limits of Fear Extinction: Recurrent Nightmares

Spoormaker[113] proposed a variation of the preceding theory to explain fear extinction failures in the subclass of nightmares with recurrent themes. The recurring storylines are stipulated to be "scripts" in memory that are easily activated by ongoing, neutrally toned dreaming and whose specific contents vary as a function of this prior dreaming. Recurring nightmare scripts may be based on real traumatic memories or may develop over time with repeated emotional stress, particularly if stressors are habitually responded to in a way that develops and reinforces the underlying script. Cognitive avoidance is the key mechanism by which occasional distressful nightmares may become recurrent. After awakening from such a nonpathologic nightmare, avoidance strategies (e.g., trying not to think about or remember the nightmare) lead to a failure of fear extinction as well as a reduced likelihood that the root script will be integrated into autobiographical memory or that alternative responses to the nightmare script will be discovered. In detailing the limits of fear extinction processes, this model clarifies the pathologic aspect of nightmare experience.

The theory's emphasis on recurrent themes fits well with the fact that the plot lines of many nightmares are, indeed, recurrent. An unpublished study found that of all nightmares reported by 188 college students, including those that were only occasional, 60% contained a recurrent storyline.[113] Among participants with a clinical nightmare problem (at least 1 per week), 91% claimed that their nightmares possessed recurrent storylines. The inherent dysfunctionality of recurrent nightmares also fits well with the finding that recurrent dreaming more generally is associated with poorer well-being.[114]

Image Contextualization

The image contextualization theory[115,116] highlights the role of emotion in dream formation and considers nightmares—as it does dreaming more generally—to play a role in adaptation to emotional experiences. Dream images are proposed to be driven by the emotions associated with current concerns; the more powerful an emotion, the more powerful and salient the central image of the dream. Recurrent and posttraumatic nightmares are among the clearest examples, but dreams of intense emotions of any type are included. The basic role of dreaming is to *contextualize* the imagery metaphorically within a safe context, that is, during REM sleep when muscular inhibition prevents acting out of the dream. An example of a central contextualizing dream image is the "tidal wave" dream, in which a powerful wave contextualizes feelings of overwhelming fear or helplessness stemming from a waking concern with similarly intense feelings. Contextualizing

depends on a hyperassociativity of neural networks during dreaming, that is, increased cross-connectivity among elements of the emotional concern and past similar experiences. As mnemonic connections increase, emotions become less intense and the concern is progressively integrated. The dream's adaptive function is likened to psychotherapy following trauma in that the therapist provides a safe context within which intense emotions may be expressed and linked constructively to other memories.[117]

Supporting evidence includes findings that powerful and salient central images are more frequent following trauma or abuse,[118] that REM sleep participates in the forming of hippocampus-dependent memories,[119] and that the hippocampus is central to the consolidation of memory for context.[120]

Threat Simulation

The evolutionary theory of threat simulation[55] suggests that the purpose of dreaming—including nightmares—is to provide a realistic (virtual) environment for confronting threatening situations and practicing threat perception and threat-avoidance skills. Repeated threat simulation over time increases the probability of successfully coping with real threats in wakefulness and confers a survival advantage to our species. The threat simulation mechanism is fully activated and thus produces more dreams with threats when the individual is exposed to heightened levels of daytime threat (e.g., living in a war zone). The high prevalence of dysphoric dreams and nightmares supports the theory because these commonly portray threats. Thus nightmares of being attacked or chased are considered functional to the extent that these provide opportunities to identify threatening situations that might be met in real life and to practice adaptive reactions to them. Research supporting the theory includes dream content analyses showing that college students report frequent threatening dreams that are both severe and realistic (e.g., aggression, misfortune themes) and during which appropriate responses are enacted by the dreamer.[121] Further, children subjected to high levels of threat (trauma), in fact, do dream more often and more intensely of threatening events.[122] Arguments against threat simulation theory claim that the threats created in dreams are too often unrealistic[123] and that the dreamer is too often unable to react successfully to the threat.[124] Realistic threats appear in less than 15% of recurrent dreams[125] and in only 8% of undergraduates' home dreams.[123] The experiencing of threat dreams also does not correlate with actual adaption to threatening events, as in the case of PTSD, in which reexperiencing nightmares are often debilitating. Further, the occurrence of nightmares before or after trauma exposure is often a risk factor for developing PTSD.[54]

POSTAWAKENING ADAPTATIONS

A number of approaches consider nightmares to be functional to the extent that reactions after waking up may play a homeostatic or adaptive role. One type of theory suggests that a postawakening function can be automatic. The *information-processing* or *memory cycle* theory[63,126] is a neo-psychoanalytic approach based on the notion that dreaming's function is the incorporation of important new experiences into long-term memory, that is, the affective integration of new with old memories. Anxiety dreams that produce awakenings have a particular functionality in that they signal a failure of normal affective integration but allow waking state processes to modify memory sources of the original anxiety dream. This leads to a modified *correction* dream the following night. Associating an anxiety dream with other thoughts, feelings, and memories essentially integrates the latter new sources of information with the original sources and provides new, more adaptive, memory sources for the correction dream. This postawakening integrative function may be automatic and preconscious, resulting simply from "having the dream in mind" during the day, or it may be deliberate, either by intentional reflection on the dream during waking or by the aid of a therapist. In either case, the corrective feedback leads to a permanent, adaptive reorganization of emotional memory structures. Recurrent anxiety dreams are thought to reflect a failure of affective integration *and* a failure of the correction dream to correct it. The theory has been subjected to very little empiric investigation and is supported only generally by evidence that REM sleep is linked to emotional regulation.

A second type of postawakening theory also considers that using dreams in a self-reflection or therapeutic context leads to biologic adaptations, but the mechanisms for such adaptations are not typically specified. One general goal of some such approaches is to alleviate the suffering associated with the nightmares using pharmacologic or behavioral approaches,[127] but another common goal is to use the nightmares as a source for uncovering focal emotional conflicts that can then be addressed therapeutically. In this case, nightmare-focused therapies have been documented for a diversity of emotional conditions, such as bereavement,[128] drug dependencies,[129] and general psychotherapy.[130]

CONCLUSIONS

Science still remains divided in many ways on whether the darker side of dream experience serves a purpose. Nightmares and other disturbing dreams have been clearly delineated as a pathologic condition that is comorbid with many other illnesses, a view that is broadly consistent with the notion that nightmares are either nonfunctional or dysfunctional symptoms. However, their nearly ubiquitous existence in the general population, including evidence for a much wider spectrum of disturbed and dysphoric dreams (e.g., bad dreams), points to the likelihood that some nightmares may, in fact, be a component of cognitive-emotional regulation or emotional memory processing functions. Nightmares that are neither too frequent, nor too severe, nor too recurrent in nature may be those that play a functional role.

Unfortunately, inconsistent and unreplicated findings from the sleep literature have not yet allowed consistent PSG profiles of either nightmare episodes or the typical sleep of nightmare sufferers to be claimed; reliable links to memory or emotion functions of sleep have thus not emerged. Similarly, a scarcity of neuropsychological findings provides few clues as to whether or how specific brain regions are implicated in nightmare formation or function.

Theories of nightmare function fall along a spectrum varying between functional and nonfunctional, with many differences in which parts of a nightmare progression should be considered as functional. Our analysis of the structure of typical nightmares dovetails well with the multifaceted nature of hypothesized nightmare functionality and demonstrates

how these hypotheses might readily be operationalized. Most functional theories consider nightmares to enable some type of emotion regulation function. Several of these were influenced by Freud's early work. Others attempted to incorporate findings from early investigations of REM sleep physiology. Yet others attempted to found their assumptions on both sleep laboratory findings and a wider understanding of the cognitive neuroscience of emotion. Postawakening theories of function also share the notion of emotion regulation; waking reactions to the nightmare—either spontaneous or therapist assisted—feed back into the nightmare production system with adaptive outcomes. The evolutionary threat simulation theory of nightmares is distinct from these theories in claiming that nightmares enable development of threat perception and threat coping skills.

Many of the theories propose a function for nightmares that is a simple extension of a more general function for dreaming and thus they more easily account for the wider spectrum of dysphoric dreams observed. Most theories differ in the nature of the hypothesized regulatory mechanism. Although they are often lumped together under a single category of emotion regulation, the mechanisms vary, including stress mastery, desomatization of affect, fear memory extinction, emotional contextualization, and others. Despite this variety, however, many nightmare theories remain relatively vague as to the key functional mechanism; for example, it is rare to find cellular- or systems-level explanations, and predictions are as a result not clear. For those theories that propose more detailed explanations, supportive evidence is either still controversial or lacking altogether. Some of the theories are applicable only to certain types of nightmare experience, such as recurrent nightmares,[31,113] fear nightmares,[6,7] or threat nightmares,[55] whereas others[62] deal with such a wide swath of emotions that, paradoxically, even intensely positive dreams are included. All of the theories could benefit greatly from an increase in empiric and comparative investigations.

CLINICAL PEARL

That nightmares are comorbid with many pathologic conditions, such as anxiety, neuroticism, and PTSD, indicates that they are symptomatic, but their ubiquity in the general population suggests that they may serve some functional purpose. If nightmares do facilitate emotional regulation or emotional memory consolidation as suggested by some theorists, they may need to be treated only when they are frequent, severe, and disruptive of daily functioning.

SUMMARY

Opinion is divided on whether nightmares have a function. Although the clinical definition of nightmares as pathologic suggests that they do not, epidemiologic evidence of the widespread occurrence of nightmares and other dysphoric dreams in the general population suggest that they do—at least if they are not frequent, severe, or recurrent in nature. PSG and neuropsychological studies are too inconsistent or scarce to permit drawing conclusions about nightmare function. Theories of nightmare function are multifaceted, attributing functionality to several of the phenomenologic steps of typical nightmares (e.g., preawakening vs. postawakening). Such theories vary in the mechanisms they propose to account for emotion regulation and include stress mastery, desomatization of affect, fear memory extinction, and emotional contextualization. However, supportive evidence is either controversial or absent.

Selected Readings

Brown RJ, Donderi DC. Dream content and self-reported well-being among recurrent dreamers, past-recurrent dreamers, and nonrecurrent dreamers. *J Pers Soc Psychol* 1986;**50**:612–23.

Cartwright R. Dreaming as a mood regulation system. In: Kryger MH, Roth T, Dement WC, editors. *Principles and practice of sleep medicine*. Philadelphia: Elsevier Saunders; 2005. p. 565–72.

Hartmann E. Nightmare after trauma as paradigm for all dreams: a new approach to the nature and functions of dreaming. *Psychiatry* 1998;**61**:223–38.

Hartmann E, et al. Contextualizing images in dreams and daydreams. *Dreaming* 2001;**11**(2):97–104.

Herlin B, Leu-Semenescu S, Chaumereuil C, Arnulf I. Evidence that non-dreamers do dream: a REM sleep behaviour disorder model. *J Sleep Res* 2015;**24**(6):602–9.

Nielsen T, Levin R. Nightmares: a new neurocognitive model. *Sleep Med Rev* 2007;**11**(4):295–310.

Sandman N, et al. Nightmares: risk factors among the Finnish general adult population. *Sleep* 2015;**38**(4):507–14.

Spoormaker VI. A cognitive model of recurrent nightmares. *Int J Dream Res* 2008;**1**(1):15–22.

van Schagen AM, Lancee J, de Groot IW, et al. Imagery rehearsal therapy in addition to treatment as usual for patients with diverse psychiatric diagnoses suffering from nightmares: a randomized controlled trial. *J Clin Psychiatry* 2015;**76**(9):e1105–13.

Walker MP. The role of sleep in cognition and emotion. *Ann N Y Acad Sci* 2009;**1156**(1):168–97.

Zadra A, Desjardins S, Marcotte E. Evolutionary function of dreams: a test of the threat simulation theory in recurrent dreams. *Conscious Cogn* 2006;**15**(2):450–63.

A complete reference list can be found online at ExpertConsult.com.

Incorporation of Waking Experiences into Dreams

Michael Schredl

Chapter 54

Chapter Highlights

- Research on the continuity between waking and dreaming has clearly identified factors that affect the chances that waking-life experiences will be incorporated into subsequent dreams, such as emotional intensity, time interval between waking event and dreaming, and type of daytime activity.
- Because direct empiric tests of possible functions of dreaming are difficult to carry out, research on the continuity hypothesis may potentially provide hints about the functions of dreaming.
- Still unclear is whether dreaming is involved in sleep-dependent memory consolidation.
- The finding that social interactions seem to be integral ingredients of dreams that are closely associated with the waking-life social environment may point to the importance of dreaming in regulating and maintaining social relationships that have been essential for survival in hunter-gatherer societies over the history of mankind.
- Interesting yet unresolved is the question of why people dream about things they have never experienced in waking life.

The question of why people dream has fascinated humankind for centuries. Even though many hypotheses have been formulated over the years, starting from dreams as guardians of sleep[1] to theories that dreaming is a training field for consciousness,[2] the empiric testing of possible functions of dreams faces certain insurmountable obstacles. To elicit the content of a dream, the participant has to report the dream, but if the dream is reported in the waking state, it is not possible to differentiate between the effects of the dream within sleeping and the effects of the narration of the dream, for example, in obtaining creative ideas. In other words, it is not possible to study the effects of nonremembered dreams and, correspondingly, the possible benefits of dreams per se directly. In view of these difficulties, the empiric studies reviewed in this chapter are based on the question of whether the function of dreaming can be elucidated by examining reported dream content and its relation to waking life.

DEFINITION OF CONTINUITY

The occurrence of waking-life experiences in dreams was termed "day-residues" by Freud,[1] whereas the notion of the continuity hypothesis was introduced by Hall and Bell[3] and expanded by Hall and Nordby.[4] The following quotation illustrates the basic idea:

> This [continuity] hypothesis states that dreams are continuous with waking life; the world of dreaming and the world of waking are one . . . . The continuity may be between dreams and covert behavior (thoughts, feelings, and fantasies) or it may be between dreams and overt behavior ("acting out") (p. 104).[4]

Although the definition of dreams as recollections of subjective experiences that occur while sleeping[5] is widely accepted, a controversy has emerged over the years about what aspects of waking life are continuous with dreaming (discussed more fully elsewhere[6]). First, it should kept in mind that dreaming only very rarely includes exact replays of episodic memories[7]—that is, the continuity between waking and dreaming is seldom one-to-one. Second, in the original definition of the continuity hypothesis, Hall and Nordby[4] already specified two areas of waking life that presumably affect dream content, and subsequent researchers elaborated on this notion and studied various aspects of waking in relation to dreaming (Table 54-1). These aspects can be grouped into the two broad categories of "covert" and "overt" behavior, as mentioned earlier. These areas do not exclude each other; according to the model, they contribute together to the relationship between waking and dreaming. Thus, according to Hall[8]: "If the dreamer feels that the world presents a cold, bleak face, he may materialize this conception in the form of a cold climate and a bleak, rocky setting (in the dream)." This example illustrates how conceptions of the world might be reflected in dreams. If you look at the experiential level in waking life, it might be expected that such a person might have very few positive social relationships and feels miserable, so the dream could be seen as a reflection of these waking-life experiences. That is, the conception of the world is mirrored in the daytime experiences of the dreamer.

A similar argument can be made for emotions; feelings of loneliness in the waking state can be directly reflected in dream images of being alone. Stumbrys[9] emphasized a type of continuity rarely looked at by other dream researchers: the

Table 54-1 Aspects of Waking Life that May Affect Dreams the Continuity Hypothesis

Aspect of Waking Life	
Concerns	Actions
Conceptions	Interests
Thoughts	Unfinished business
Emotions	Psychological issues
Experiences	Worries
Preoccupations	Personal significant events
Fantasies	Meta-awareness

aspect of *meta-awareness*. The induction of lucid dreaming by means of the reflection technique ("Am I dreaming or am I awake?") is based on this continuity. Increasing reflective awareness about one's state of consciousness during wakefulness increases the awareness while dreaming and thus can increase lucidity.

Even though empiric research supports the continuity hypothesis of dreaming,[10] a point worthy of mention in this context is that a considerable number of dreams clearly are discontinuous with waking life—for example, flying dreams[11] or walking dreams in congenital paraplegics.[12] The first line of thinking is that such dreams presumably reflect waking-life thoughts, daydreams, and so on. But if the dreamer had never experienced something in waking life, such as with sight in congenital blind persons, it also should be difficult to imagine such experience in the waking state. Even though some continuity might be retained at the emotional level (waking-life emotions and dream emotions)—for example, flying dreams were associated with personality traits associated with positive emotions, whereas falling dreams were associated with traits associated with negative emotions[13]—the concept of discontinuity is still not well understood. Of interest, research in amputees[14] seems to indicate that dreams may reflect innate body images that are not based on experience. In addition, studies looking at pain in dreams,[15] for example, suggest that dreams also may reflect experiences based on mirror neurons, as with seeing someone in pain versus direct experience of the pain. Overall, studies carefully relating waking-life thoughts and fantasies in relation to bizarre dream content can be expected to help clarify whether these topics are still explained within the framework of the continuity hypothesis (reflecting waking-life thoughts) or whether it is necessary to expand this theory by elaborating the concept of discontinuity.

STUDYING THE CONTINUITY BETWEEN WAKING AND DREAMING

Over the years, different approaches have been applied for studying the continuity between waking and dreaming (Box 54-1). Because the methods can have a marked effect on the findings (see further on), a brief outline of the basic ideas of these approaches, with reviews of exemplary studies, is presented next.

The most detailed study looking at temporal references of dream elements was carried out by Strauch and Meier.[16] Fifty dreams stemming from REM sleep awakenings of five

Box 54-1 PARADIGMS FOR STUDYING THE RELATIONSHIP BETWEEN WAKING AND DREAMING

- Assessing temporal references of dream elements
- "Blind" analysis of dream series
- Diary studies
  - Within-subject approach
  - Between-subject approach
- Experimental manipulation of daytime experiences
- Lucid dreaming

subjects included 80 key role characters, 39 "extras" (persons playing a minor role in the dream), 74 settings, and 298 objects. The ability to relate these elements to waking life by the dreamer varied considerably: 25.6% (extras), 30.9% (objects), and 76.3% (key role characters). Thus, for 76.3% of the key role characters, the participants were able to find correspondences to their own waking life—they thought about a particular person, for example, or encountered the person in the workplace. Even though this method seems straightforward (and has also been used in clinical settings), several problems have been recognized. With inclusion of inner processes like waking thoughts, memory restrictions presumably would constitute a major limitation, because it is very difficult to recall all of the thoughts of the previous day, let alone of the previous week. Another issue is that of multiple correspondences of dream elements. Dreaming about one's mother, for instance, might correspond to a telephone conversation a few days ago or to a childhood experience. Owing to these difficulties, the findings obtained by the retrospective approach are necessarily limited in their interpretation regarding the continuity between waking and dreaming.

A very interesting approach called "blind" analysis was developed by Hall[17] and further pursued by Domhoff[18] and others.[19] Dream series are quantitatively analyzed by rating systems[20] or word search algorithms[19] and predictions are then derived as to what should be prominent in the dreamer's waking life. Analyzing the dream series of one subject, for example, Domhoff[18] found strong correspondences between the social interactions within the dream and the real-life social interactions with the person encountered by the subject (mother, sister, brother, friend). That is, if the relationship was full of conflicts in waking life, the aggressiveness/friendliness index for this dream character was high (i.e., many more aggressive interactions than friendly interactions). The major problem with this approach is that the matching between dream content and waking life of the dreamer is rather vague, depending on the data obtained through interviews carried out with the dreamer. If, for example, the participant is a student, false positives would be likely because the waking lives of students are in several aspects very similar.

A simple approach is using a diary in which the participants fill in daytime events and their dreams. In the study of Schredl and Reinhard,[21] the participants were asked after each dream whether the dream had any correspondences with the previous day(s), whereas in the study of Blagrove and colleagues,[22] the participants matched dreams and day reports after the 2-week period during which they kept diaries. The problem with this approach is that the subjectivity of each participant doing the matches can heavily influence the

findings, depending on how correspondences were defined (thematic, literal, metaphorical, and so on).

The diary method also can be used for testing whether interindividual differences in waking life are reflected in corresponding differences in dreaming. Schredl and Hofmann,[23] for example, correlated the amount of time spent engaged in specific activities during the day with the frequency of these activities in dreams. Persons who often drove in their waking lives also had driving dreams more often. The diary approach has the problem, though, that recording daytime experience (and dreams) can affect subsequent dreams and can therefore bias the results of this kind of study. A simplified version was applied by Schredl and Erlacher[24] using a questionnaire eliciting the amount of time spent in particular activities such as sports and reading and correlated these figures with the percentage of dreams including these topics. As expected, the amount of time spent in sports activities during the day correlated directly with the percentage of sports dreams, irrespective of whether the participants were sports or psychology students.[24]

Another option for studying interindividual differences in dreams is to look at specific populations—for example, comparing men and women, or patients with mental disorders and healthy control subjects. The basic idea is that the paired groups differ in some aspects of their waking life and thus should also differ in their dreams. For example, men dream more often about sex than women do,[20] reflecting the greater frequency of sexual fantasies in waking life typical of men.[25]

A further option is to study the effects of experimental manipulations of waking-life experiences on dreams. De Koninck and Brunette,[26] for example, exposed phobic participants to a snake (in a terrarium) and read them different stories before sleep onset. Of interest, the emotional tone of the story affected the participants' dream emotions but not the dream content. In addition, real-life stressors such as awaiting major surgery or an intensive group therapy session[27] also were studied and showed stronger effects than the experimental stressors. Within this context, it should be mentioned that the high percentage of laboratory references in dreams (approximately 20% in more than 2000 dreams collected in 12 studies) obtained on REM or NREM awakenings in the sleep laboratory clearly supports the notion of continuity between waking and dreaming.[28]

The last paradigm included in this brief review is lucid dreaming. Proficient lucid dreamers can carry out prearranged tasks in their dreams.[29] Remembering the instructions received while awake clearly indicates the continuity between waking and dreaming.[30]

EMPIRIC FINDINGS REGARDING THE CONTINUITY BETWEEN WAKING AND DREAMING

Time Course

Freud[1] summarized his own experiences and his literature review regarding the time course of dreams as follows: "Dreams show a clear preference for the impression of the immediately preceding days." This "preference" is embodied by the so-called day-residue effect. Subsequent studies[16,31,32] confirmed that events from the previous day are more often incorporated into dreams compared with events from past days, weeks, or years. Even though a diary study[33] confirmed this decrease in dream incorporation rates with increasing time intervals between waking-life event and dream, it should be considered that the aforementioned studies—as with Freud's observations and the literature he reviewed—applied the approach of assessing in reverse order the temporal references of the reported dream elements. If thoughts, emotions, and the like are included, it then becomes obvious that this approach is limited by the memory capacity of the dreamer; thus the decrease in more remote references might be explained by lack of recalling the corresponding waking experiences.

Several diary studies[22,34-38] found a *day residue effect*—that is, the highest incorporation rates were for events of the previous day. Also observed was a so-called *dream lag effect*, whereby the incorporation rate for events that happened 5 to 7 days before the dreams was higher than for events of the previous 2 to 4 days or those occurring more than 8 days earlier. These data have been interpreted as reflecting adaptive processes, especially in regard to social stressors and processes of memory consolidation.[38]

Of note, however, the study findings should be interpreted with caution, because diary studies are most likely to identify REM sleep dreams before awakening in the morning and thus are not representative for all dreams. Available evidence indicates that the first REM dreams of the night are more often affected by experimental manipulations applied before sleep[39,40] when compared with REM dreams later in the same night, and that dream elements of late-night REM dreams contain more remote references to waking life than do dream elements of early REM dreams.[41,42]

Studying the dream lag in laboratory dreams has shown that the dream lag effect was not present in NREM dreams but was seen only in REM dreams, but the mean time between sleep onset and these REM dreams was 6.16 hours; that is, on average, these were dreams of the latter part of the night. To follow up this line of thinking, it would be necessary to collect larger samples of REM dreams stemming from the first part of the night.

To summarize, according to the available evidence, the time course of incorporating daytime events may not follow a simple exponential function (with lower incorporation rates of more remote waking-life experience). Rather, processes related to memory consolidation or adaption to stressful social events apparently mediate the incorporation of waking-life experiences into dreams.

Effect of Experimental Manipulation of the Presleep Situation on Dream Content

Several studies[39,43-47] showed their subjects different films before the laboratory study night for the recollection of REM dreams. Of interest, in these studies, direct incorporation of the film topics into dreams occurred very rarely. By telling the participants a positively toned version versus a negatively toned version of the same story, De Koninck and Brunette[26] were able to manipulate the emotional tone of the dreams but not the content—that is, neither the phobic object (a snake) nor the control animal (a squirrel) showed up in the dreams, even though the participants were exposed to a snake before bedtime. Formal characteristics of dreams such as colors also have been successfully manipulated by having the subjects wear red goggles during the day, for example.[40]

Using the technique of dream incubation (dating back to ancient Greek traditions), the topics the dreamer wanted to dream about before sleep onset did not show up in the

dreams.[48,49] If, however, the participants were instructed to think about a current personal problem before sleep, the probability of dreaming about this problem increased from 20% to 40%.[16] Similarly, focusing on discrepancies between ideal self and current personality traits before sleep onset increased the probability of dreaming about this trait.[50] In contrast with the findings regarding dream incubation, a carefully designed study[51] indicated that the instruction to not think about the target person yielded more occurrences in subsequent dreams than thinking about the person—supporting the so-called ironic-process theory. Subsequent studies[52-54] using idiosyncratic intrusive thoughts as targets that should be suppressed (not thought about) replicated the findings of the first study. In a sleep laboratory study,[55] the same effect was found for sleep onset dreams, this time using an image of three white bears and the presleep instruction not to think about them.

Research using experimental manipulation of the presleep situation showed that stimuli that are not related to the personal life of the dreamer, even if they are not pleasant, do not exert a strong effect on subsequent dream content. The emotional tone of dreams is more sensitive to stressful stimuli, in keeping with the research findings that nightmare frequency is positively related to current waking-life stress.[56] The closer the topic is related to the personal issues of the dreamer, the higher the chance of an effect of presleep manipulation on dreams.

Effect of Waking Life on Dream Content: Field Studies

The effects of a large variety of waking-life activities and events on dreaming, ranging from media consumption to traumatic experiences, have been studied and are reviewed briefly here.

Approximately 50% to 70% of participants stated that TV viewing or reading books has affected their dreams.[57] In children and adolescents, media figures have been found in nightmares quite frequently.[58,59] Computer games such as Tetris or Alpine Racer and others have shown an effect on subsequent dreams.[60-62]

Spending time in everyday activities such as reading, driving a car, playing sports, or engaging in political discussion is directly related to the percentage of dreams that include these topics.[23,24,63] Similarly, the time spent with people of both sexes during the day is reflected in the ratio of male and female dream characters.[64,65] The romantic partner also plays a major role in dreams.[66,67] Of interest, daytime emotions related to encounters with friends had the strongest correlations with dream emotions, compared with academic studies or personal issues[21]—indicating that social interactions during the day have a considerable impact on nighttime dreaming.

One specific life event that shows a strong influence on dreaming is pregnancy.[68-70] Of interest, the overall emotional tone of dreams in pregnant women is negative, reflecting the fact that pregnancy can be quite stressful to mothers-to-be.[71] Even the dreams reported by the expectant fathers are affected by this life-changing event.[72] Divorce also has a negative effect on dreaming, for both the divorced parents and the children.[73,74]

In patients with insomnia, the occurrence of waking-life problems was correlated with the number of problems within the dream.[75] Overall, the dream studies in patients with mental disorders have shown that psychopathologic symptoms experienced during the day also were present in their dreams.[76-79] For example, the severity of depressive symptoms correlated with negative dream emotions,[80] and dream bizarreness was related to the severity of psychotic symptoms in schizophrenic patients.[81] Patients with eating disorders dreamed more about rejecting food (patients with anorexia nervosa) or more about food in general (patients with bulimia).[82]

Very stressful, traumatic events also affect subsequent dream content very strongly; recurrent nightmares of the trauma are a hallmark symptom of posttraumatic stress disorder (see also Chapter 56).[83] The effects of experiences related to the 9/11 terror attacks,[84] sexual assault,[85] childhood sexual abuse,[86] kidnapping,[87,88] accidents,[89,90] and war[91-93] on dreams can be detected even decades after such experiences in waking life.

Overall, trauma studies indicate that emotional intensity may be one of the factors affecting continuity between waking and dreaming—that is, emotionally more intense waking-life experiences are more likely to be incorporated into subsequent dreams than less intense experiences. Of interest, this plausible hypothesis has rarely been studied directly. From a methodologic standpoint, it seems to be important to use a design in which the emotional intensity of the waking-life event will be rated before the dream occurs; otherwise, the approach of retrospectively rating the emotionality of events that resurfaced in the dream presumably could be biased by hindsight: "I dreamed about this; therefore, it must be important."

In a diary study,[33] this problem was addressed by instructing the participants to keep a diary over a 2-week period to record the five most important events of the day and then to rate the emotional intensity and valence for each. The participants also recorded their dreams and checked whether any of the recorded events of the previous days showed up in the dream. The emotional intensity and valence of the incorporated event were compared with those for the other recorded events of the same day. The findings indicated that emotional intensity was higher for incorporated events, but emotional valence was similar for the events not incorporated.[33] That is, no bias toward incorporating more negatively toned waking life events was detected. The study of Malinowski and Horton[94] replicated this finding: Emotional intensity but not stressfulness was crucial for incorporation of content into subsequent dreams.

TYPE OF WAKING-LIFE EXPERIENCE

Already in 1909, Meumann[95] observed that reading and writing occurred very rarely in his dreams, although he was engaged in these activities up to 6 hours/day. Hartmann[96] compared the frequencies of reading, writing, and typing with other waking activities like walking, talking with friends, and sexuality-related behaviors and also found that the so-called three R's are less likely than other activities to be found in dreams. On the basis of his findings that emotional intensity of the waking-life activity does not necessarily explain these differences (e.g., comparing walking with reading), Hartmann[96] formulated the hypothesis that during cholinergic-driven REM sleep, focused thinking processes are not that easy for the brain to handle compared with the waking state. As a point of interest, "thinking" in general, such as about what other people think or what to do next, is quite common in dreams.[97]

The findings of Hartmann[96] were confirmed by several studies.[23,24,98] In a diary study[23] the so-called cognitive activities (reading, writing, calculating, working with a computer) accounted for 41.6% of the elicited waking-life activities (including talking with friends, driving a car, watching TV, using the phone, and being in nature), whereas only 18.6% of the dream activities fell into this cognitive area. The hypothesis of Schredl[98] that dreams are more likely to reflect "archaic" themes such as social interactions or being in nature was not supported by subsequent work from Schredl and Hofmann,[23] who found that driving a car, for example, was significantly overrepresented in dreams. The additional finding that talking with friends was also very common in dreams is in line with evidence showing briefer lag times for persons or characters and emotions until their incorporation into dreams[16,99] and might point to a preference of dreamers for social topics.

IMPLICATIONS FOR POSSIBLE FUNCTIONS OF DREAMING

Based on the well-documented findings that sleep contributes to memory consolidation,[100] the first question regarding the potential functions of dreaming that comes to mind is whether dreaming is reflecting the processes involved in sleep-dependent memory consolidation. In a nap study using a maze learning paradigm, performance gains were higher if task-related dreams were reported.[101] However, an overnight study[102] using mirror tracing as a procedural learning task did not show any relation between task-related dreams and next-day performance gain in the participants. In view of the fact that memory consolidation involves processes on both cellular (long-term potentiation) and network levels, the second question is whether dreaming (as experience that can be recalled on awaking) is connected to or can reflect these processes—similar to waking life, in that the brain is doing much more than thinking and experiencing. Of interest, training in a specific task during a lucid dream can enhance performance in the morning,[103] although this finding should be viewed with caution because it is uncertain whether simply recalling a dream of training for the task might enhance performance as opposed to the training within the dream (as described early in this chapter). In any case, whether dreaming is associated with or reflecting processes crucial to sleep-dependent memory consolidation has not yet been ascertained.

The research regarding the temporal references of dream elements (see earlier) indicates that new information is mixed with experiences from the distant past. In addition, dreams are creative—that is, they form new associations instead of just replaying daytime experiences.[104] This scenario parallels processes typically involved in problem solving. First, compare the new incident with previous successfully handled experiences, and second, if no strategy is already available, try something new (brainstorming). In view of this analogy, proposing problem solving and adaptation as possible dream functions[105] seems plausible.

The finding that emotional intensity but not emotional tone or stressfulness affects the incorporation of daytime events into dreams[33,94] renders unlikely those dream functions that are focused solely on negative dream emotions such as the threat simulation theory,[106] even though it does not exclude learning about places that might be dangerous as one of the functions of dreaming.

Finally, several findings indicate a strong continuity between the social environment of the dreamer in the waking state and social interactions in dreams. Emotionally "close" persons often play an important part in dreams,[66] emotions related to close persons have strong effects on dreams,[21] and social interactions are overrepresented compared with academic activities.[23] This continuity may point to another evolutionary aspect of dreaming (in addition to avoidance of being eaten before producing offspring): the importance of the social network for survival in hunter-gatherer societies. That is, it is essential to (1) get along with other group members and (2) not be ostracized by the group.[107]

To summarize, all of these ideas about dream function are, of course, highly speculative, but studying the continuity between waking and dreaming may potentially shed light on this still-unanswered question about the function of dreaming.

FUTURE DIRECTIONS

The review of the evidence presented in this chapter clearly shows that research into the factors that have an effect on the continuity between waking and dreaming is still in its infancy. Schredl[10] proposed a mathematical model to promote more systematic research on the effects of time, emotional intensity, and so on on the incorporation rates of waking-life experiences into subsequent dreams. Methodologic problems concerning how to measure waking-life experiences (e.g., type, frequency, intensity) and how to relate these measures to dream content are not yet solved and require more empiric testing. Another line of research was suggested by Schredl and Reinhard[21] reporting on the so-called second-order effect of continuity—that is, dreams that were affected by the previous day are more likely to affect the mood of the following day. A plausible interpretation of this finding would be that some themes, topics, and concerns are processes that day and night occupy the affected person's consciousness. Finally, studying dreams that are clearly not continuous with waking life (such as, for example, flying dreams) would be most beneficial in elucidating why people can dream about things, such as pain,[15] never experienced in waking life, or in explaining the occurrence of walking dreams in congenital paraplegics.[108]

CLINICAL PEARL

In clinical encounters in which the patient relates the specifics of a dream, it is important to identify the basic pattern of the dream, because this pattern is most likely to be continuous with the person's waking life. For example, if dream content includes being chased by a monster, the basic pattern is one of anxiety and avoidance. The question is then whether avoidance of something or someone is an issue for the patient in his or her waking life.

SUMMARY

Research on the continuity between waking and dreaming has clearly identified factors that affect the chances of incorporating waking-life experiences into subsequent dreams. Because direct empiric tests of possible functions of dreaming are difficult to carry out, the question arises as to whether the research on continuity may provide hints about the functions

of dreaming. Still unclear at this stage is whether dreaming is involved in sleep-dependent memory consolidation. The finding that social interactions seem to be integral ingredients of dreams that are closely associated with the waking-life social environment may point to the importance of dreaming in regulating and maintaining social relationships that have been essential for survival in hunter-gatherer societies over the history of humankind. An interesting yet unresolved issue is the question of why people dream about things they have never experienced in waking life (discontinuity).

Selected Readings

Bekrater-Bodmann R, Schredl M, Diers M, et al. Post-amputation pain is associated with the recall of an impaired body representation in dreams—results from a nation-wide survey on limb amputees. *PLoS ONE* 2015;**10**:e0119552.

Fosse MJ, Fosse R, Hobson JA, Stickgold RJ. Dreaming and episodic memory: a functional dissociation? *J Cogn Neurosci* 2003;**15**:1–9.

Hall CS, Nordby VJ. *The individual and his dreams*. New York: New American Library; 1972.

Hartmann E. *The nature and functions of dreaming*. New York: Oxford University Press; 2011.

Kröner-Borowik T, Gosch S, Hansen K, et al. The effects of suppressing intrusive thoughts on dream content, dream distress and psychological parameters. *J Sleep Res* 2013;**22**(5):600–4.

Malinowski J, Horton CL. Evidence for the preferential incorporation of emotional waking-life experiences into dreams. *Dreaming* 2014;**24**(1):18–31.

Schredl M. Continuity between waking and dreaming: a proposal for a mathematical model. *Sleep Hypn* 2003;**5**:38–52.

Schredl M, Erlacher D. Is sleep-dependent memory consolidation of a visuomotor task related to dream content? *Int J Dream Res* 2010;**3**:74–9.

Schredl M, Reinhard I. The continuity between waking mood and dream emotions: Direct and second-order effects. *Imagin Cogn Pers* 2009-2010;**29**:271–82.

Skancke J, Holsen I, Schredl M. Continuity between waking life and dreams of psychiatric patients: a review and discussion of the implications for dream research. *Int J Dream Res* 2014;**7**:39–53.

Strauch I, Meier B. *In search of dreams: results of experimental dream research*. Albany (N.Y.): State University of New York Press; 1996.

Valli K, Frauscher B, Peltomaa T, et al. Dreaming furiously? A sleep laboratory study on the dream content of people with Parkinson's disease and with or without rapid eye movement sleep behavior disorder. *Sleep Medicine* 2015;**16**:419–27.

Van den Bulck J. Media use and dreaming: the relationship among television viewing, computer game play, and nightmares or pleasant dreams. *Dreaming* 2004;**14**:43–9.

Wamsley EJ, Tucker M, Payne JD, et al. Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Curr Biol* 2010;**20**:850–5.

A complete reference list can be found online at ExpertConsult.com.

Dreams and Nightmares in Posttraumatic Stress Disorder

Chapter 55

Wilfred R. Pigeon; Thomas A. Mellman

Chapter Highlights

- Posttraumatic stress disorder (PTSD) develops in a significant minority of persons who are exposed to severely threatening trauma. Trauma-related nightmares appear to be specific to PTSD and can be a persisting and distressing symptom, although not all dream content reported by persons with PTSD is a direct representation of trauma memories.
- Stress and trauma appear to acutely affect the thematic content of dreams irrespective of whether the dreamer goes on to experience continuing emotional distress. Dreams that more specifically replicate the memory of the trauma during the early aftermath of exposure have been associated with the development of PTSD. Findings from experimental and naturalistic studies suggest that dreams can positively influence adaptation to stress and trauma, also suggesting that persisting trauma-related nightmares may represent a failure of adaptive mechanisms.
- Findings from polysomnographic studies of PTSD have not been entirely consistent. Generally, a majority of nightmare episodes of study subjects with PTSD arise in rapid eye movement sleep, although they also can be accompanied by non–rapid eye movement sleep. Additionally, the rapid eye movement sleep of PTSD subjects tends to be more fragmented than that of healthy sleepers, and this pattern may be predictive of PTSD in the acute aftermath of trauma.
- In contrast with normal dreams, dreams in persons with PTSD often incorporate actual memories of frightening experiences. These features have implications for understanding neurocognitive substrates of dreaming and the development of integrative PTSD pathophysiology models.
- Evidence supports that pharmacologic antagonism of noradrenergic receptors ameliorates nightmares in PTSD, as do psychological treatments that use exposure to, and/or cognitive restructuring of, nightmares.

REPLICATIVE-TRAUMA NIGHTMARES: HALLMARK OF A DISORDER?

Posttraumatic stress disorder (PTSD) is a psychiatric condition that develops in approximately 10% of persons who experience severely threatening traumatic experiences.[1] The diagnosis is based on a set of four persisting symptom clusters: intrusive symptoms such as reexperiencing the trauma with intrusive images, flashbacks, or nightmares; avoidance of trauma-related stimuli; negative alterations in mood or cognition; and heightened arousal and reactivity. The course of PTSD can be self-limited, but in approximately a third of cases the disorder persists for many years, and it often manifests with comorbid psychiatric conditions (see also Chapter 136). The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* specifically delineates "recurrent distressing dreams in which the content and/or affect are related to the traumatic event" among the intrusive symptoms.[2] In an influential review and theoretical paper, Ross and colleagues emphasized the occurrence of "repetitive replicas" of trauma scenes as a feature of dreams that is virtually specific to PTSD, further referring to rapid eye movement (REM) sleep disturbance and trauma nightmares as a "hallmark of the disorder."[3] A recent review of this hypothesis maintains that nightmares constitute a signature feature of PTSD.[4]

Whereas the overall rate of frequent nightmares (one or more per week) is 2% to 5% in the general adult population,[5-7] approximately half of all patients with PTSD have frequent trauma-replicating nightmares, and another 20% to 25% experience nightmares that are thematically or symbolically related to their traumatic event.[8-10] Among children, rates of nightmares range widely, from 7% to 20% in the general population and 20% to 81% in trauma-exposed samples, with the highest rates observed among those with diagnosed PTSD[11] and extreme living situations such as in war zones.[12] In addition, even compared with patients with idiopathic nightmares, patients with PTSD have an elevated frequency of nightmares.[8,9,13] Treating patients who have PTSD often leaves the clinician impressed with the impact of distressing dreams featuring traumatic experiences. Recurrent dreams, in general, are associated with more negative dream content, increased distress, and a lower degree of psychological well-being compared with nonrecurrent dreams[14]; the content of PTSD nightmares, however, also tends to

be more replicative of traumatic experiences and even more distressing.[15]

Several studies support the specificity of the relationship between trauma-related nightmares and PTSD. Van der Kolk and colleagues[16] reported that patients with combat-related PTSD were more likely to indicate that their nightmares exactly or almost exactly replicated an actual event, as compared with nightmare sufferers who did not have PTSD. Mellman and associates[17] surveyed combat veterans from clinical and nonclinical settings and found that nightmares about combat experiences were more specifically associated with PTSD than were nightmares on other topics. In an analysis of a large epidemiologic database on Vietnam veterans, a measure that combined instances of nightmares related to military experiences and of distress from dreaming was strongly associated with PTSD.[18] The relationship between replicative nightmares and higher PTSD severity has been observed in patients as long as 40 years after a traumatic event.[19]

The available evidence thus provides consistent support for the theory that continuing representation of trauma memories in dreams is a feature of PTSD but not necessarily of trauma exposure absent the diagnosis. The aforementioned studies relied on retrospective and global, categorical assessments of dreams. Esposito and coworkers[20] performed content analysis of dreams elicited from morning diaries in a group of combat veterans receiving treatment for PTSD. Approximately half of the subjects' dreams contained direct references to combat experiences, and almost all of the dreams featured threat. A majority of these reported dreams also were similar to normal dreams in that they contained implausible elements that were not representations of actual memories. Similarly, among a group of combat veterans experiencing PTSD symptomology, approximately half of the dreams elicited after awakenings in the laboratory setting referred to military experiences.[21] These studies all focused on data for combat veterans many years after their combat experiences.

Several studies have examined the influence of trauma on dream content during an acute phase after the traumatic event. In the 2 months subsequent to a hurricane, dream questionnaires from 22 primary care patients were compared with those for a larger sample surveyed before the hurricane (n = 265). A significantly higher percentage of the posthurricane subjects reported that their dreams were related to general stressors (74% versus 48%) and to "especially stressful life experiences" (67% versus 37%), although only 13% reported dreams specific to the hurricane.[22] Similarly, evacuees of an urban fire were more likely than control subjects to have recorded dreams in their home diaries with content related to death, disasters, and fires.[23] Among children and adolescents, a similar pattern of trauma event incorporation into early posttrauma dreams has been observed. This pattern was seen among all 23 children kidnapped and buried underground in a trailer,[24] in 8 of 10 children who witnessed rapes,[25] and in 63% of those on a playground during a sniper attack,[26] with higher rates of trauma dreams correlated with proximity to the threat or reflecting direct versus indirect association with the trauma.[26,27] In review of the pediatric literature, these types of findings are more mixed among children than among adult studies.[11,28] In both the children and adult studies, most reports do not relate dream content to PTSD status. In a sample of 13- to 18-year-old Afghan unaccompanied youths seeking asylum in the United Kingdom, PTSD symptoms were measured and found to be correlated with nightmare frequency, although dream content was not assessed.[29]

Mellman's group elicited dream content from study participants during the acute aftermath of traumatic injury.[30] A subgroup of the sample described dreams that were distressing and "highly similar" to the traumatic experience (17% of the sample; 56% of those who reported dreams). Subjects in this group had more severe concurrent PTSD symptom ratings than those in the groups with other categories of dreams, and they had higher subsequent PTSD severity than the subjects who did not recall dream content.[30] Wittman and colleagues observed a similar pattern in children immediately after motor vehicle accidents, with acute replicative nightmares predicting PTSD symptoms 2 and 6 months later.[31]

Thus, in trauma-exposed populations, dreams with specific trauma-related or thematically related content are typically reported during the acute aftermath of disasters and other traumas, and dreams that are similar to the memory of the actual traumatic event are associated with the development of PTSD. Chronic PTSD is associated with recurring dreams that represent specific memories of a traumatic experience. However, the more comprehensive evaluations of dreams during chronic phases of PTSD document that salient dream content is not limited to representations of traumatic memories.

DREAM CONTENT WITH STRESS AND TRAUMA: BEYOND REPLICATION

Although the issue of trauma replication has received considerable attention, other aspects of content have been described in the literature on dreaming, nightmares, and PTSD. The boundaries of traumatic stress are not always clear, and the topic seems embedded in broader questions of stress and dreaming, including observations of the impact of stressful experiences on dreams, the influence of dreams on adaptation to stressful experiences, and content themes related to trauma exposure and PTSD beyond replication of trauma.

Stress and Dream Content

It has been consistently observed that dream narratives incorporate elements of current concerns including the incorporation of daily stressors as distressing dream content.[32] Studies of naturalistic stressors have suggested an impact of stressful waking experiences on dream content. For example, health-related stressors have been evidenced in the home dream content of patients awaiting surgery,[33] in hospitalized burn-injured patients,[34] and in pregnant women.[35,36] In addition, the dreams of women with stressful menstruation are more likely to contain emotional content and relationship themes during the peak hormonal phase than those during other days in the menstrual cycle.[37] Academic and occupational stresses also have been associated with dream content, including increased dream recall in college students a week before final examinations compared with that during a control week.[38] Correlations also have been found between recurrent nightmares, dreams of being chased, and dreams of falling with both graduate students' financial stress[39] and stockbrokers' stress levels during market downturns.[40] Living in threatening

environments is correlated with increased threatening dreams in traumatized Kurdish children.[41] However, dream reports of South African people living in areas of high crime did not differ in the frequency of threatening content from that for a control Welsh sample.[42] Of interest, several dream-reporting studies were undertaken around the time of the September 11, 2011 terrorist attacks in the United States. Although their methodologies varied, each study concluded that the attacks were associated with increased intensity of dreams compared to pre-9/11 dreams,[43-45] but only two of the studies found that post-9/11 dream narratives contained explicit references to the terror attacks.[45,46]

It has been noted that recurring dreams, which have more negative content than nonrecurring dreams, tend to be activated by stressful life experiences.[14,47,48] In a community sample, subjects with active recurring dreams reported greater life stress in the previous 6 months and had more negative dream content than both former recurrent dreamers and dreamers who had never had recurrent dreams.[49] These findings were replicated in two college student samples.[38,50] In addition to these naturalistic observations, several older studies used experimental induction of stress and examined dream content. Disturbing movies,[51,52] sham intelligence examinations,[53,54] and experimentally induced pain[55] all have been incorporated into dreams collected in the sleep laboratory. Overall, experimental stressors are associated with subsequent dreams that are characterized by a negative tone. Although difficult to quantify across studies, both naturalistic and experimental stressors are incorporated into dream content with several caveats. If they are not severe enough and/or salient to the dreamer, they may not appear as precisely replicative of the stressor, and their incorporation is likely to be influenced by individual subjects' characteristics.

Although available evidence suggests that relatively innocuous daytime events and concerns can be incorporated into dreams, it seems reasonable to infer from the studies just reviewed that stressful waking experiences are more likely to be incorporated into dreams and have an impact on dream emotional content. This relationship of dream incorporation to emotional saliency has led several dream theorists to invoke an emotional information processing function for dreaming.[33,56-58] The threat simulation theory of dreaming further posits an evolutionary function of dreaming, one that simulates realistic threats within the safe space of the dream world to permit adequate rehearsal and preparation for future threats in waking life.[59] Although it has been difficult to confirm that threat rehearsal occurs in dreams, several lines of evidence support a relationship between dreams and adaptations to emotional stress.

Dreams and Adaptation to Stress

In one of the experiments that used viewing a disturbing film as a probe, volunteer subjects experimentally deprived of REM sleep were subsequently more distressed by the film than either an non–rapid eye movement (NREM) sleep interruption group or an uninterrupted sleep group.[52] Naturalistic studies suggesting that dream incorporation can aid adaptive processing include several that found that references to drugs and alcohol in dreams are a positive predictor of abstinence.[60-64] Cartwright and colleagues[56] elicited dream reports after laboratory awakenings from REM sleep in men and women going through divorce, the first near the time of the initial breakup and a second 1 year later. Those who incorporated the ex-spouse into their dreams at the time of the breakup were less depressed and better adjusted at the 12-month follow-up evaluation than those who did not. These two sets of observations initially might seem to contradict the observations reviewed earlier of an association between trauma-replicating dreams and the outcome of PTSD. The observations just reviewed, however, refer to incorporation of a reference or representation of a stressful situation and not necessarily to the replication of events.

For dreaming to influence emotional adaptation to stress, the neurocognitive activity of sleep must have an enduring influence on memory representations. Support does exist for sleep's role in the consolidation and reprocessing of memory (i.e., learning), which is reviewed in Chapter 22.

Trauma, Posttraumatic Stress Disorder, and Content Themes

Dow and colleagues[65] studied depressed combat veterans with and without PTSD. They examined dreams collected from both groups in the sleep laboratory after awakenings from REM sleep. The PTSD group's dreams were rated higher for anxiety and were more likely to be set in the past. Another laboratory study of veterans found more frequent aggression in the dreams of subjects with combat-related PTSD compared with the dreams of healthy control subjects.[66] Ratings for anxiety, aggression, and interpersonal conflicts were greater in the dreams of a symptomatic subgroup of Holocaust survivors.[67] In a study of combat-related PTSD in which dream recall was stimulated using diary records filled out in the morning, threatening content was observed in a majority of dreams (83%) with and without the presence of combat references.[20]

Studies relating dream content to trauma exposure in children also have described content themes beyond representation of the trauma. The dreams of Palestinian children exposed to ongoing civil and military violence, elicited by home diaries, contained more themes of aggression, persecution, and negative emotions than dreams recorded by children living in a more peaceful region.[68] Kurdish children with trauma exposure evidenced more dreams with threat and aggression than either those in a nonexposed group or children in a control group from a peaceful country. Unfortunately, neither of these studies determined the PTSD status of their subjects.[41]

In a study of patients with recent traumatic injuries, dreams of those in whom PTSD subsequently developed had more content related to general and physical misfortune, as well as more negative emotions, than those in whom PTSD did not develop.[69] Thus, compared with other dreams, dreams after trauma have more general negative emotions, anxiety, threat, and aggression, and these appear to be most pronounced with PTSD.

POLYSOMNOGRAPHIC CORRELATES OF NIGHTMARES IN POSTTRAUMATIC STRESS DISORDER

An association, albeit not an absolute one, between dreams—particularly those with content that is more elaborate, visual,

emotional, and bizarre—and REM sleep stage has been recognized.[70] Other sleep phenomena, such as night terrors and related parasomnias, arise in NREM sleep stages, particularly slow wave sleep.[71] Owing to the atypical aspects of dreams from patients with PTSD, interest has emerged in determining their relationships to stages of sleep. It has been asserted that nightmares in PTSD are associated with NREM sleep. In one study, body movement occurred during stage 2 sleep when nightmares were reported in the morning.[16] In another study in a group of 24 subjects evaluated in the sleep laboratory, Kramer and Kinney[72] identified a subgroup of 8 combat veterans with nightmare symptoms who reported a majority of their nightmares after awakening from stage 2 sleep.

By contrast, the one spontaneous nightmare awakening that occurred during sleep recordings in a study by Ross and coworkers[3] and three recorded by Mellman's group were preceded by REM sleep.[73] Hefez and colleagues[74] described a pattern of REM interruption insomnia in a group of patients with PTSD reporting nightmares. The largest sample of polysomnographically recorded spontaneous nightmares ($n = 17$) for patients with PTSD was reported in an abstract by Woodward and colleagues.[75] These workers indicated that the probability of REM sleep occurring in the 10 minutes preceding spontaneous awakenings with nightmares was 57%, whereas only 17% of total sleep was REM. The probability of sleep preceding nightmares featuring stage 2 (27%) was less than its percentage of total sleep. None of the nightmares emerged during slow wave sleep. Thus, on balance it seems that nightmares in PTSD tend to be preceded by REM sleep, but they sometimes emerge in other sleep stages (1 and 2). The characteristics of nightmares in PTSD, however, can contrast with normative features of dreams associated with REM sleep, such as a lack of correspondence with actual events, bizarreness, and mixing of time frames and contexts.[70,76]

The observations reviewed in previous sections of this chapter suggest that associative memory functions linked to REM sleep may facilitate emotional adaptation. As previously mentioned, a relationship has been noted between the development of PTSD and dreams that were rated as being very similar to the trauma memory.[30] In a subsequent study of patients hospitalized for traumatic injuries, polysomnographic recordings were obtained close to the time of medical and surgical stabilization. Patients in whom PTSD subsequently developed did not differ from those in whom PTSD did not develop with respect to general measures of sleep maintenance, amount of REM sleep, and increased eye movement density (compared with uninjured control subjects). The group in which PTSD developed, however, spent significantly less uninterrupted time in REM sleep before shifting to waking or other EEG sleep stages,[73,77] which also was observed among a community sample of subjects with PTSD.[78] Other disrupted sleep characteristics reported during REM sleep of persons with PTSD are hypothesized to be indices of the heightened arousal that is symptomatic of PTSD and include higher sympathetic heart rate variability indices,[79] more periodic leg movements,[80] higher beta frequency power,[77] and lower theta frequency power.[81,82] The most consistent polysomnographic finding in patients with PTSD, however, appears to be one of increased REM fragmentation.[83] We hypothesized that this observed fragmentation of REM sleep may compromise REM sleep's potentially adaptive memory-processing functions.[73]

CONCLUSIONS AND THEORETICAL IMPLICATIONS

The representation of the memory of a traumatic experience in dreams is a distinguishing feature of PTSD. Although having characteristics different from those generally ascribed to dreams arising in the REM sleep stage, PTSD dreams appear most often to be associated with REM sleep, although they can emerge in stages 1 and 2 as well. Traumatic and stressful experiences appear to influence dream content whether or not the affected person is successfully adapting emotionally.

Findings suggest that nightmares that are replicative of the actual memory of the trauma experience are associated with the development of PTSD. When PTSD enters a chronic phase, trauma memories continue to be represented in recurring nightmares. Dream content of patients with established PTSD is not limited to trauma memories, however, and it might reflect the negative and restricted emotional state of the dreamer in other ways as well. The literature on stress and dreaming suggests that dream content reflects the emotional processing of the dreamer and that dreams potentially have a positive influence on emotional adaptation. By contrast, recurring nightmares in patients with PTSD appear to reinforce the memory of the trauma and contribute to the dreamer's distress. The distinguishing feature of replicating or representing the memory of the trauma over time might provide important clues regarding what is not working in the dream life and the more general emotional memory processing in people with PTSD.

Fosse and colleagues[84] have found that dreams recalled by healthy college students often contain references to recent experiences, but the events' representation in the dreams rarely corresponds to actual events. The researchers concluded that normal dreams are not episodic memories. Noncorrespondence to coherent whole-memory representations and other characteristics of dream mentation are thought to be consequent to the selective activation of neural structures during REM sleep. In the generally accepted model of sleep state regulation, neural activation during REM sleep is mediated by firing of cholinergic brainstem nuclei and is further facilitated by inhibition of noradrenergic firing.[76]

As has been discussed, the dreams that characterize PTSD are not necessarily unaltered replays of the traumatic event. Yet a tendency has been noted for dreams occurring with PTSD to contain more representations of events that include or are closer to unaltered memories (of traumatic events) than in normal dreams. Experimental evidence indicates that declarative memory for emotionally arousing stimuli is mediated by noradrenergic mechanisms.[85] Impaired inhibition of noradrenergic tone during sleep, therefore, could be a mechanism underlying the presence of episodic-specific, fear-enhanced memories in dreams. Direct support for this hypothesis comes from evidence that pharmacologic blockade of noradrenergic stimulation with prazosin, an $alpha_1$-noradrenergic antagonist, ameliorates nightmares in PTSD.[86]

A further consideration regards an adaptive memory-processing function for REM sleep and REM dreaming that may be impaired with PTSD. Normal dream mentation has a hyperassociative quality. Characters, places, and sequences that typically are not linked in waking conscious thought tend to be juxtaposed in dreams.[76] Foa and colleagues[87] have

suggested that one of the mechanisms of successful emotional processing during exposure therapy is the development of a new network of associations with the traumatic memories. Such a process may be facilitated by the normal neurocognitive characteristics of REM sleep and impaired by a more selective activation of trauma memories.

More recent neurocircuitry models of PTSD posit that the ventromedial prefrontal cortex fails to inhibit the amygdala, leading to attentional bias toward threat, increased fear responses, impaired extinction of traumatic memories, and deficits in emotion regulation.[88,89] These models inform, and can be informed by, the growing specificity of theories and models of how sleep neurobiology contributes to adaptive and maladaptive emotional processing of traumatic memories. This area of investigation includes an interesting proposition that eye movement desensitization reprocessing, a successful form of PTSD treatment, may have as a mechanism of action induction of an optimal REM sleep–like state in which hippocampal mediation of episodic trauma memories and their amygdala-dependent affective valence can be integrated into broader cortical networks.[90] Nielsen and Levin have updated a neurocognitive model of nightmare production incorporating findings from sleep and PTSD neurobiology to explain how extinction (and consolidation) of fear memories can occur.[91,92] Finally, Germain and colleagues have proposed that amygdala hyperactivity and attenuated medial prefrontal cortex activity are tied to whole-brain neuronal activity in NREM sleep, which would be permissive of insomnia,[93] and that REM sleep amplifies activity of the amygdala and the medial prefrontal cortex, which would be permissive of nightmares.[93] Conceptual models such as these, which integrate new neurobiologic findings from both the sleep and PTSD fields as well as their respective clinical research findings, will be needed to guide empirical inquiry into the role of recurrent traumatic dreams in PTSD pathophysiology. At the same time, trauma-related nightmares, as well as their response to various treatment strategies, can inform models of central PTSD mechanisms.

TREATMENT OF POSTTRAUMATIC NIGHTMARES AND RELATED SLEEP DISTURBANCES

A number of treatments, particularly selective serotonin reuptake inhibitor antidepressant therapy and cognitive-behavioral psychotherapeutic techniques, have been documented to be effective for treating PTSD and are included in treatment guidelines.[94-96] Reports of the treatment studies typically indicate reduction in the severity of reexperiencing symptoms but do not specifically parcel out the treatment effects on nightmares. It is likely that nightmares tend to decrease in frequency and intensity as PTSD generally abates. Clinical experience and literature on the treatment of residual sleep symptoms, however, suggest that especially insomnia and/or nightmares can be inadequately responsive to first-line interventions for PTSD.[97-101]

Formulations discussed in preceding sections suggest pharmacologic and psychotherapeutic approaches for targeting PTSD nightmares. Clinical research evidence now supports efficacy of treatment strategies that are consistent with the previously stated theoretical formulations. For instance, a therapeutic effect of postsynaptic blockade of noradrenergic neurotransmission is consistent with the role for noradrenergic activity in mediating traumatic nightmares that was postulated in the previous section. In this regard, prazosin has now been demonstrated to have a placebo-controlled effect on nightmares in several populations.[102-105] Evaluations of other pharmacologic approaches to target nightmare symptoms of PTSD have mostly not been controlled or not been promising in double-blind trials. The 2010 American Academy of Sleep Medicine's best practice guide for the treatment of nightmare disorders recommends only prazosin based on its designation of "level A" evidence; a number of agents (atypical antipsychotic medications, cyproheptadine, fluvoxamine, gabapentin, low-dose cortisol, nitrazepam, phenelzine, topiramate, trazodone, triazolam, and tricyclic antidepressants) were deemed to have "level C" evidence for their use.[106] A meta-analysis confirmed that prazosin is an effective agent for nightmare treatment, with a moderate effect size similar to that of psychological therapies.[107]

In addition, the idea that dreams are related to both successful and impaired patterns of emotional processing suggests that recurrent nightmares related to trauma might respond to psychotherapeutic interventions. In fact, several clinical trials have demonstrated good efficacy for nightmare interventions that are based on principles of exposure therapy or cognitive restructuring. Imagery rehearsal therapy (IRT) has the largest evidence base to date.[108] In this technique, patients first learn cognitive-behavioral strategies to manage unpleasant images during wakefulness before writing out the content of a distressing dream, rescripting the content any way they wish, and then rehearsing the images of the altered dream scenario. Successful variations on this approach include combining relaxation training and direct, repeated exposure to disturbing dream content with rescripting of the dreams (exposure, relaxation, and rescripting therapy [ERRT]),[109] relaxation and exposure without rescripting,[110] and an approach that uses lucid dreaming techniques wherein patients learn to alter dream content as the dream actually occurs.[111] Overall, this set of interventions results in significant reductions in nightmare severity and frequency, with more modest improvements in insomnia and other symptoms of PTSD. A meta-analysis found that less complex approaches such as relaxation alone were not effective, but that psychological interventions for nightmares were effective overall, with a moderate effect size comparable to that for pharmacologic interventions.[107] At present, none of these related approaches can be recommended as a first-line comprehensive treatment for PTSD. However, IRT and ERRT appear to be particularly effective in targeting nightmare symptoms and can be applied as stand-alone interventions for nightmares with moderate to large effects.[107,112-114] IRT also has been combined with cognitive-behavioral therapy for insomnia, with good success.[115-117] It also may be a useful adjunct to first-line PTSD treatments.

These psychotherapeutic nightmare interventions are consistent with principles from the established cognitive-behavioral treatments of PTSD that apply exposure or cognitive restructuring, or both, to trauma memories.[64] That such techniques can be effectively applied directly to nightmare content (which in PTSD often features trauma memories) has important implications. One is that clinicians can be realistically hopeful about the potential for alleviating distressing nightmare symptoms. As previously mentioned, an apparent paradox is that stress-related references in dreams

are associated with positive outcome, whereas trauma-replicating nightmares are associated with PTSD. This paradox may be reconciled if the former dreams represent an adaptive emotional processing, possibly related to modification of associative networks, and the latter represent a failure of such a process. Studies of imagery rehearsal suggest that conscious exposure to the content of recurring distressing dreams, along with instruction to modify the dream scenario, facilitates movement toward the more adaptive response.

That benefits of nightmare-focused interventions are generalizable to other symptom domains lends support to the idea of continuity between waking emotional life and dreaming. Dreams also may offer clinicians a unique window into patients' emotional adaptation to trauma and stress, along with the status of their processing of traumatic memories. The apparently robust association between PTSD and replays or representations of trauma memories in dreams further provides an important clue to understanding abnormalities of emotional memory processing that differentiate those who suffer with PTSD from those who are more resilient to adverse psychobiological consequences of trauma.

It has been shown that poor sleep quality has a negative impact on the severity and trajectory of PTSD symptoms[118,119] and that sleep disturbance, including nightmares specifically, may even contribute to suicidal thoughts and behaviors.[120,121] The growing evidence from the literature, as reviewed in this chapter, indicates that treatment of nightmares not only can ameliorate sleep disturbances but also can reduce overall symptoms of PTSD, although not to the point of remission.

CLINICAL PEARL

The phenomenon of trauma replication in dreams is a distinguishing feature of PTSD. Recurrent trauma nightmares contribute to considerable distress in patients with PTSD. Evidence supports the use of prazosin as a pharmacologic treatment for PTSD-related nightmares. Psychological treatments recommended for such nightmares include IRT and EERT.

SUMMARY

Trauma-related nightmares continue to warrant their designation as a hallmark of PTSD. Although not entirely consistent or linear, the preponderance of evidence indicates that stressors of increasing strength are incorporated into dream narratives with severe traumas and PTSD nightmares occupying the extreme end of a possible continuum. Evidence also suggests that dreaming of life stressors and traumatic events may be emotionally adaptive in acute post-trauma periods, but that persistent and trauma-replicating dreams are maladaptive. Findings from sleep neurobiology, PTSD neurobiology, and polysomnographic studies of trauma and PTSD patients have informed increasing refinements in neurocognitive models of nightmares. At the same time, during the past decade efficacy of some pharmacologic and nonpharmacologic nightmare treatments have been established, which are consistent with the contention that incorporation of actual trauma memories and their affective components into dreams is indicative of neurocognitive substrates.

Selected Readings

Augedal AW, Hansen KS, Kronhaug CR, et al. Randomized controlled trials of psychological and pharmacological treatments for nightmares: a meta-analysis. *Sleep Med Rev* 2013;**17**:143–52.

Berardis DD, Marini S, Serroni N, et al. Targeting the noradrenergic system in posttraumatic stress disorder: a systematic review and meta-analysis of prazosin trials. *Curr Drug Targets* 2015;**16**(10):1094–106.

Bernert R, Nardoff MR. Sleep disturbances and suicide risk. *Sleep Med Clinics* 2015;**10**:35–9.

Germain A, Nielsen TA. Sleep pathophysiology in posttraumatic stress disorder and idiopathic nightmare sufferers. *Biol Psychiatry* 2003;**54**:1092–8.

Kobayashi I, Boarts JM, Delahanty DL. Polysomnographically measured sleep abnormalities in PTSD: a meta-analytic review. *Psychophysiology* 2007;**44**:660–9.

Malinowski J, Horton C. Evidence for the preferential incorporation of emotional waking-life experiences into dreams. *Dreaming* 2014;**24**:18–31.

Mellman TA, Pigeon WR, Nowell PD, Nolan B. Relationships between REM sleep findings and PTSD symptoms during the early aftermath of trauma. *J Trauma Stress* 2007;**20**:893–901.

Nielsen T, Levin R. Nightmares: a new neurocognitive model. *Sleep Med Rev* 2007;**11**:295–310.

Ross RJ, Ball WA, Sullivan KA, Caroff SN. Sleep disturbance as the hallmark of posttraumatic stress disorder. *Am J Psychiatry* 1989;**146**:697–707.

Schredl M. Nightmare frequency and nightmare topics in a representative German sample. *Eur Arch Psychiatry Clin Neurosci* 2010;**260**:565–70.

Stickgold R. EMDR: a putative neurobiological mechanism of action. *J Clin Psychol* 2002;**58**:61–75.

A complete reference list can be found online at ExpertConsult.com.

Emotion, Motivation, and Reward in Relation to Dreaming

Chapter 56

Sophie Schwartz; Lampros Perogamvros

Chapter Highlights

- The activation of emotional and reward brain networks during non–rapid eye movement and rapid eye movement sleep is supported by robust evidence.
- Specific patterns of brain activity during sleep probably determine key features of dream content. Conversely, dream characteristics such as intense emotional experiences offer valuable insights into information processing during sleep.
- The activation of limbic regions, in particular, the amygdala, and of the reward system during sleep seems to promote the reactivation of affectively relevant memories, as well as the experience of emotions and the expression of approach and avoidance behaviors in dreams.
- Emotional and reward processing during dreaming may explain its contribution to waking cognitive and emotional processes, including emotion regulation and performance improvement. This hypothesis is supported by recent experimental evidence.

Recent neuroimaging, neurophysiological, and clinical studies converge to suggest that emotional and reward networks are activated during sleep. Because these networks are active primarily while the organism interacts with its external environment, of critical relevance is how and why such systems are also activated during sleep. A main hypothesis is that their recruitment during sleep relates to the affective and motivational components of the dreaming experience and has beneficial consequences on waking behavior and emotional well-being. This chapter examines how the activation of emotional and reward circuits during sleep may determine the affective and motivational attributes and functions of dreaming.

DREAMING AND EMOTIONAL PROCESSING

Sleep is an active yet variable and heterogeneous state of the brain. Over the past two decades or so, neuroimaging studies in humans have demonstrated that distinct brain regions display specific patterns of activity during rapid eye movement (REM) sleep and non–REM (NREM) sleep. As briefly reviewed subsequently, these new data provide unprecedented insights into the neural circuits whose recruitment during human sleep plausibly determines typical features in dreams. In particular, the activation of limbic and mesolimbic systems presumably promotes affective and memory processing during sleep and dreams.

REM sleep is characterized by the activation of brain regions involved in sensory-motor, memory, and emotion processing, including associative visual areas, motor cortex, medial temporal regions, amygdala, anterior cingulate cortex, and medial prefrontal cortex.[1-4] During wakefulness, the amygdala is involved in the detection of emotional stimuli and also contributes to emotional learning and extinction, in close interaction with the medial prefrontal cortex.[5] In addition, amygdala activity was found to correlate with heart rate variability during REM sleep, suggesting a role for the amygdala in autonomic regulation during this sleep state.[6] The particular recruitment of the amygdala and medial prefrontal cortex during REM sleep would thus provide a favorable condition for the reprocessing and regulation of emotionally relevant information during sleep and may elicit intense emotions in dreams.[7-11] Furthermore, a recent brain structural study in humans demonstrated that preserved microstructural integrity in the left amygdala was associated with longer and more emotional dream reports.[12]

By contrast, several regions implicated in executive and attentional functions during wakefulness are significantly deactivated during REM sleep, including the dorsolateral prefrontal cortex, orbitofrontal cortex, and inferior parietal cortex.[1,2] These deactivations are compatible with several cognitive characteristics of dream reports such as illogical thinking, disorientation, impaired working memory, and limited cognitive control.[13,14] Reduced activity in medial prefrontal regions is also consistent with relatively poor and labile dream recall, whereas an increase in frontal theta activity (5 to 7 Hz) during REM sleep underlies successful dream recall.[15] Moreover, limited monitoring from frontal and parietal regions may explain the intriguing tolerance to bizarre aspects in dreams. Objects, places, or characters are recognized despite their modified physical appearance in dreams,[14] and discontinuities in the dream plot or events that are highly unlikely or

impossible during wakefulness may not be identified as such[16]; furthermore, full-blown basic and social emotions (e.g., fear, sadness, sexual emotions, embarrassment) may be experienced without disrupting the dreamer's sleep. Dreaming may thus involve this particular conjunction of reduced control and criticism over the illusory nature and peculiarities of the dream world. Recent experimental evidence has shown that transcranial alternating current stimulation in the low gamma range (25 to 40 Hz) applied to frontotemporal regions during REM sleep could enhance "executive ego functions" such as self-reflective awareness in dreams by promoting insight into the fact that one is currently dreaming, third-person perspective-taking, and control over the dream plot.[17]

Early neuroimaging studies of human NREM sleep consistently reported decreases in regional brain activity, in accordance with a homeostatic need for brain energy recovery. In particular, activity is decreased in the brainstem, basal ganglia, and several cortical areas including the prefrontal cortex (as reviewed by Dang-Vu and associates[18]). Recent functional magnetic resonance imaging studies, however, have started to unveil transient increases in neural activity across emotion and memory circuits during this sleep stage. Specifically, NREM sleep spindles have been associated with increased activity in the insula, anterior cingulate cortex, and superior temporal gyrus,[19] while slow waves have been associated with increased activity in the precuneus, parahippocampal gyrus, and posterior cingulate cortex.[20] Bilateral increases in regional metabolism from waking to NREM have also been found in the ventral striatum, anterior cingulate cortex, amygdala, and hippocampus.[21] Increased activity in memory-related regions, in particular, those time-locked to slow oscillations, supports the hypothesis that memories are reactivated during NREM sleep.[22]

Overall, existing neuroimaging data suggest that emotional networks are activated during all stages of sleep, especially during REM sleep. These activations may explain the emotional characteristics of REM and NREM dreams. Indeed, dream content analyses most often report a general intensification of emotions and predominance of negative emotions in REM compared with NREM dreams.[23] More than 50 years ago, the French neurophysiologist Michel Jouvet observed that when muscle atonia was abolished during REM sleep, cats displayed negatively toned behaviors (e.g., those corresponding with anger or fear).[24] Similarly, aggressive behaviors are frequent in patients with REM sleep behavior disorder, who act out their dreams.[25] On the other hand, positive emotional states like friendliness are frequent in NREM dreams.[26] Dream data thus confirm that emotional processes may be active during both REM and NREM sleep stages. They also suggest, however, that distinct emotions may prevail during specific sleep stages, with more intense negative emotions and aggressiveness predominating in REM dreams. The possible implications related to the processing of emotions during sleep are considered further on, under Implications for Emotion Regulation and Learning.

DREAMING AND REWARD PROCESSING

It has been suggested that an elevated dopamine level in the mesolimbic dopaminergic (ML-DA) system during sleep plays an important role in the generation of dreams.[27] Lesions in the white matter surrounding the frontal horns of the lateral ventricles can cause a total cessation of dreaming, without affecting REM sleep.[27,28] These lesions disrupt the mesolimbic circuit, which connects dopaminergic neurons of the ventral tegmental area (VTA)—the source of dopamine activity—with the nucleus accumbens (NAcc), amygdala, hippocampus, anterior cingulate cortex, and frontal cortex (insular and medial orbitofrontal cortex, medial frontal cortex, ventromedial prefrontal cortex), and which forms the so-called SEEKING system. Introduced in the 1980s by the neuroscientist Jaak Panksepp, the SEEKING system is a "curiosity-interest-expectancy" command system related to instinctual appetitive states.[29,30] It was proposed that this psychobehavioral emotional and motivational system may be implicated in the generation of dreams.[28,31] Activation of the ML-DA system would induce a SEEKING disposition, which could promote approach-like behaviors and emotional anticipation during both wakefulness and in dreams. Of note, Panksepp's conceptualization of the SEEKING system is compatible with other models of the reward system subserving reward prediction error,[32] reinforcement-learning,[33] or "incentive salience."[34]

Pharmacological studies have provided further evidence that dreaming relates to the activation of the ML-DA system (and associated SEEKING disposition) by showing that the administration of dopaminergic (D_2) antagonists is associated with a reduction in dreaming[35] and in nightmares,[36-38] whereas the administration of dopaminergic agents (e.g., pramipexole) causes vivid dreams.[39-41] Furthermore, total or partial sleep deprivation leads to disturbed reward brain functions.[42-44]

Activation of reward networks in sleep (and not necessarily in dreaming) also is suggested by some recent human studies showing that the consolidation of declarative memory benefits from NREM sleep and is influenced by motivational biases.[45-47] For instance, recently learned declarative memories (word-pair associates) are better retrieved after a period of sleep, but only if the participants had been informed that a retrieval test would occur.[46] Thus increasing the future relevance of memories enhances their consolidation, and this enhancement is correlated with NREM oscillatory activity (slow waves and spindles). Such a preferential consolidation of highly rewarded memories during sleep seems to involve the activation of the dopaminergic reward system in human sleep.[47] More specifically, in a placebo group, memory consolidation was enhanced after sleep only for those memories associated with high reward; after administration of the dopaminergic agonist pramipexole, memory was improved for both high and low reward conditions. These observations converge with findings in animal studies showing that during sleep, a spontaneous reactivation (replay) of neuronal firing patterns occurs in neurons of the ventral striatum after reward-searching behavior.[48,49] This "off-line" replay is orchestrated by hippocampal ripples and may strengthen the association between a memory trace and a specific motivational value.[50,51] The spontaneous reactivation of reward-related neuronal firing patterns in the ventral striatum during NREM sleep may involve a transfer of novelty or relevance signal from the hippocampus to the VTA,[49] and contribute to the activation of the VTA in REM sleep.[31,52]

Indeed, during REM sleep as well, key structures of the ML-DA reward system are activated, including the VTA and the NAcc. In rodents, VTA bursting activity is elevated during REM sleep,[52,53] up to levels observed during reward or

punisher anticipation at waking.[54] This activity is significantly higher in REM sleep than in the awake state or in NREM sleep; it is comparable in intensity and duration to activations during waking behaviors such as feeding, punishment, or sex.[52] Moreover, extracellular levels of dopamine are increased in the NAcc during REM sleep in rats.[55] In both animal and human studies, other reward-related regions, including the ventromedial prefrontal cortex and the anterior cingulate cortex, are activated during REM sleep.[2,3,55] Finally, the orexin neurons in the lateral hypothalamus, which are involved in the regulation of sleep-wake states and in reward-seeking behaviors,[56-58] display bursts of activity during REM sleep.[59,60] Further support for an activation of reward-related circuits during REM sleep comes from the observation, in both animals and humans, that the hippocampus exhibits a theta rhythm,[61,62] which is associated with novelty-seeking, exploratory, and instinctual behaviors during wakefulness.[30]

How does the activation of reward-seeking mechanisms during sleep influence, at least partly, the content of dreams? Recent research has shown that approach behaviors (e.g., behaviors of exploration, curiosity, or engagement) may be as frequent as "avoidance" behaviors (e.g., avoiding a threat through fleeing, freezing, or hiding) both in dreams[63] and as overtly expressed in parasomnias.[64,65] Moreover, dream reports are biased toward content with strong motivational value (e.g., socializing, fighting, sexual content) and less oriented toward content with no such particular value (e.g., typing, washing dishes, buying food at the supermarket).[66] Of importance, frequent approach behaviors are not incompatible with the high prevalence of negative affect in dreams,[67] because both the ML-DA system and the amygdala-limbic system are activated during sleep.[68]

IMPLICATIONS FOR EMOTION REGULATION AND LEARNING

The links between dreaming and emotional or motivational processes, as previously expounded, offer new insights into the possible functions of dreaming. NREM and REM dreams appear to encompass specific emotional/motivational states; NREM dreaming would predominantly simulate friendly interactions, self-related information, and actual waking life events, whereas REM dreams contain comparatively more aggressive social interactions[26] and less integration between self-referential and social cognitive reasoning with autobiographical memory.[69] It has been proposed that the variety of emotions that the dreamer experiences may contribute to emotion regulation processes. Evidence in favor of this claim comes from the demonstration that different REM dream characteristics relate to specific daytime affective functions, such as protection against depression after divorce[70,71] and response to antidepressant treatment.[72] Moreover, being exposed to diverse emotional stimuli (objects, situations, thoughts, memories, and physical sensations) during dreaming, including feared ones, in a safe context that the dream state represents, resembles and may act like desensitization therapy.[73] This assumption is indirectly supported by studies demonstrating that sleep, in general, promotes the retention and generalization of extinction learning.[74]

Along the same lines, the Finnish psychologist and philosopher Antti Revonsuo proposed the threat simulation theory, according to which dreaming promotes the development and maintenance of threat-avoidance skills during wakefulness by simulating threatening events during sleep.[10,67] This mechanism would rely on the activation of limbic regions, in particular the amygdala, during REM sleep. Consistent with this hypothesis, persons suffering from posttraumatic stress disorder simulate threatening events in their dreams more often than controls.[75] Moreover, REM sleep deprivation impairs threat avoidance skills in waking life,[76] whereas sleep selectively prioritizes the consolidation of perceptual features of aversively conditioned stimuli, ultimately enhancing the detection of potential dangers in the environment.[77] Also, recent experimental evidence demonstrated that sleep does not only protect memory for emotional pictures but also preserves emotional reactivity, as measured by subjective ratings of valence and arousal elicited by the pictures.[9] Of note, the latter results relate to the effects of sleep on emotional processing; whether they illuminate a possible function of dreaming still remains speculative. On the other hand, nightmares may correspond to the failure of fear memory extinction, when temporary (e.g., daily concerns) or more persistent (e.g., trauma) emotionally disturbing memories are being reactivated in sleep.[78]

Extending the threat simulation theory, we recently put forward the reward activation model for sleep and dreaming,[31] which supports that dreaming favors the activation of stimulus representations or behaviors of high emotional and motivational relevance, including instinctual and approach/avoidance behaviors (such as feeding, mating, fighting, or fleeing). In this model, emotionally relevant experiences have a higher probability of being activated during dreaming and exhibit a preferential access to sleep-related memory consolidation processes. Plausible functions of dreaming may thus implicate off-line emotion regulation processes by exposing the sleeper to aversive or rewarding stimuli, as well as enhanced memory consolidation during sleep, and improved performance in real life situations.

Indeed, a second major contribution of dreaming, apart from emotion regulation, seems to pertain to memory processes. Dreaming is a particular state of consciousness during which some memory representations are reexperienced and then potentially reorganized. Evidence for the role of dreaming in the consolidation of memory was recently demonstrated by Wamsley and associates.[79] In this study, participants first were trained on a virtual navigation task and then either took a nap or rested while awake. Task-related imagery during the sleep/rest period improved subsequent task performance significantly more than when task-related thoughts occurred during wakefulness. Such off-line processes of memory reactivation and reorganization would also explain why sleep may favor creative insights.[80] Using a semantic priming task, Stickgold and colleagues[81] demonstrated that participants awakened from REM sleep showed greater priming by weak primes (than by both unrelated primes and strong primes), consistent with a hyperassociative state of the sleeping mind, reminiscent of the unusual associations of features or events in REM dreams.[14] In a similar experiment, Walker and coworkers[82] demonstrated that subjects awakened from REM sleep exhibited a 32% advantage on an anagram-solving task, compared with the number of correct responses after NREM awakenings. In line with these observations, compared with quiet rest and NREM sleep, REM sleep (but not necessarily dreaming) enhances the integration of initially unassociated information

resulting in more creative problem solving.[83] More experimental studies are needed in order to better delineate how dreaming per se influences memory consolidation, creativity, and insight.

Taken together, these findings demonstrate that distinct sleep stages contribute to the remodeling of memory stores, and that memory reprocessing may be enhanced in the dream state, with important implications for waking performance. On the basis of robust neurophysiological evidence of reward activation during sleep, it has been suggested that dreams relate not only to known past events but also to an unexpected, novel, or probabilistic future.[31,69] Dream content analysis provides support for this idea by showing that although past and current waking concerns are common in dreams,[71,84] dreams rarely provide exact replicates of past events; instead they most often represent novel combinations of memory elements.[85,86] Dreaming would thus offer an off-line cognitive and emotional preparation of the dreamer for future waking life events. By suggesting that subjective experiences in dreams fulfill important biological and psychological functions, this theoretical proposal also has clinical implications because dream characteristics may represent biomarkers of important brain functions, such as emotion regulation processes.[7]

METHODOLOGICAL ISSUES

The difficulty of understanding the functions of dreams is related in part to the fact that the neural correlates of the dreaming state still remain largely undetermined. This shortcoming may be due to methodological difficulties pertaining to the commonly used techniques in dream research (e.g., limited number of EEG electrodes, neuroimaging studies specific to sleep but not to dreams) and the lack of studies correlating the dreaming experience, which is by definition subjective, with its neurobiological substrates. Moreover, studying dreams implies that subjective parameters are collected *after* the experience (i.e., the dreamer remembers of the dream after being awakened). Therefore recent methodological improvements in the integration of self-report data with neuroimaging data, with the potential for decoding mental content from ongoing brain activity, represent a promising avenue for future developments in the science of dreaming.[87-89]

Although it is tempting to relate some aspects of dreaming to motivational and emotional functions, as in work summarized in this chapter and elsewhere (threat simulation theory,[67] protoconsciousness theory,[90] default-mode activation theory[91]), more studies specifically addressing such functional relationship are needed.[71,72,79,92] Moreover, distinct sleep stages and dream states (associated with NREM or REM sleep) may differ in their contribution to off-line reprocessing of emotional and reward information. NREM sleep and dreams may be more specialized in linking memory traces with motivational values,[46,49] whereas REM sleep and dreams may be responsible for emotional memory consolidation and synaptic consolidation.[93] Nevertheless, studies combining dream content analysis, sleep recordings, and an assessment of emotion and/or memory functions are still very scarce.[79]

Progress in dream research will thus benefit both from new neuroimaging techniques that allow the identification of brain states associated with specific mental contents ("brain decoding" methods) and from a better integration of these physiological data with subjective dream data.

CLINICAL PEARL

The demonstration that dreaming affects emotion regulation processes may be useful to promote measures preventing sleep (and dream) restriction. This is particularly important for the most vulnerable populations, such as psychiatric patients or children, whose brains may be in elevated need of the reprocessing and simulation of emotions occurring during dreaming. Conversely, dream characteristics constitute biomarkers of important brain functions, such as emotion-regulating processes.[92]

SUMMARY

This chapter summarizes the evidence for activation of emotional-limbic and reward-related circuits during sleep. Dreaming seems to be closely related to these specific patterns of neural and behavioral activations, a link that may explain its contribution to important functions such as emotion regulation, associative learning, and social cognition. By simulating defense and approach behaviors, subserved by limbic regions and the mesolimbic dopaminergic reward system, dreaming provides a virtual and safe environment, in which the dreamer can be exposed to an important load of aversive or rewarding stimuli. This mechanism provides the person with enhanced learning capacities, which can be potentially used in the waking life. However, more empirical studies are needed to better characterize these proposed roles of dreaming.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported by grants from the Swiss National Science Foundation, the Swiss Center for Affective Sciences, the Boninchi Foundation, and the BIAL Foundation.

Selected Readings

Cartwright R, Agargun MY, Kirkby J, Friedman JK. Relation of dreams to waking concerns. *Psychiatry Res* 2006;**141**(3):261–70.

Dang-Vu TT, Schabus M, Desseilles M, et al. Functional neuroimaging insights into the physiology of human sleep. *Sleep* 2010;**33**:1589–603.

Diekelmann S, Born J. The memory function of sleep. *Nat Rev Neurosci* 2010;**11**:114–26.

Dumel G, et al. Infrequent dream recall associated with low performance but high overnight improvement on mirror-tracing. *J Sleep Res* 2015;**24**: 372–82.

Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y. Neural decoding of visual imagery during sleep. *Science* 2013;**340**:639–42.

Lansink CS, Goltstein PM, Lankelma JV, et al. Preferential reactivation of motivationally relevant information in the ventral striatum. *J Neurosci* 2008;**28**:6372–82.

Marzano C, Ferrara M, Mauro F, et al. Recalling and forgetting dreams: theta and alpha oscillations during sleep predict subsequent dream recall. *J Neurosci* 2011;**31**:6674–83.

Perogamvros L, Schwartz S. The roles of the reward system in sleep and dreaming. *Neurosci Biobehav Rev* 2012;**36**:1934–51.

Schwartz S, Maquet P. Sleep imaging and the neuro-psychological assessment of dreams. *Trends Cogn Sci* 2002;**6**(1):23–30.

Solms M. *The neuropsychology of dreams: a clinico-anatomical study*. Hillsdale (N.J.): Lawrence Erlbaum Associates; 1997.

Valli K, Revonsuo A. The threat simulation theory in light of recent empirical evidence: a review. *Am J Psychol* 2009;**122**:17–38.

Wamsley EJ, Tucker M, Payne JD, et al. Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Curr Biol* 2010;**20**:850–5.

Wilhelm I, Diekelmann S, Molzow I, et al. Sleep selectively enhances memory expected to be of future relevance. *J Neurosci* 2011;**31**:1563–9.

A complete reference list can be found online at ExpertConsult.com.

PART II

Practice of Sleep Medicine

Section 8

Impact, Presentation, and Diagnosis

Chapter 57

Approach to the Patient with Disordered Sleep

Beth A. Malow

Chapter Highlights

- This chapter emphasizes the clinical approach to the patient with disordered sleep, focusing on specific aspects of the history and physical examination.
- Patients who complain of disturbed sleep usually describe one or more of three types of problems: insomnia; abnormal movements, behaviors, or sensations during sleep or during nocturnal awakenings; or excessive daytime sleepiness.
- Taking a systematic history that includes medication use, family history, social history, and review of systems can provide important clues regarding the diagnosis.

Patients who complain of disturbed sleep usually describe one or more of three types of problems: insomnia; abnormal movements, behaviors, or sensations during sleep or during nocturnal awakenings; or excessive daytime sleepiness. These sleep complaints are not mutually exclusive, and a given sleep disorder may be associated with more than one type. For example, patients with sleep apnea may complain of insomnia, excessive daytime sleepiness, choking or gasping during the night, or all three. Those with narcolepsy may complain of sleep paralysis and hallucinations at sleep onset or on awakening, disrupted sleep, and daytime sleepiness.

CHIEF COMPLAINT AND HISTORY

Evaluation begins with the chief complaint, which provides a focus for delineating the patient's concerns and eliciting the history. It is often useful to ask why the patient is seeking help at the present time, particularly if the problem has been of long standing. If the chief complaint is from the spouse or bed partner, it is important to determine whether the patient recognizes the problem, is unaware of it, or denies its existence. Many clinicians also obtain a brief patient profile during the interview that includes the patient's age, sex, occupational or academic status, marital status, and living arrangements. The profile often includes valuable information about how the sleep concern is affecting the patient's daily functioning (e.g., difficulty performing job responsibilities or participating in leisure activities with family). After the chief complaint is delineated, details concerning the sleep problem are sought, including its duration, the circumstances at its onset, the factors that lead to exacerbation or improvement, and any associated symptoms.

The patient's daily schedule is reviewed, including the usual bedtime and estimated time to sleep onset, the number and timing of awakenings, and the time of final awakening. Morning symptoms should be elicited, such as increased nasal congestion, dry mouth, or morning headaches. These symptoms may support a diagnosis of obstructive sleep apnea.

Daytime symptoms, including during passive or repetitious activities (e.g., watching television or riding in a car), should be investigated to characterize the severity of sleepiness. A comprehensive sleep history also includes questions about the frequency and duration of daytime naps and the presence or absence of cataplexy, hypnic hallucinations, sleep paralysis, and automatic behavior.

Insomnia

Patients with insomnia usually complain that their nocturnal sleep is inadequate in some way. They may describe difficulty falling asleep, frequent awakening, or early morning awakening with inability to return to sleep. It is important to distinguish among these patterns of insomnia because they may have different causes. For example, awakenings from sleep because of obstructive sleep apnea may result in sleep maintenance insomnia but would not result in a patient's complaining of lying awake for hours not being able to fall asleep. The description of insomnia and its course may help determine cause, as outlined in Chapter 84.

Excessive Sleepiness

Patients with daytime sleepiness typically complain of drowsiness that interferes with daytime activities, unavoidable napping, or both. Falling asleep while driving or at other particularly inappropriate or dangerous times is often the impetus that brings the patient to the clinician. Some of these patients complain that they need more sleep at night or that daytime drowsiness occurs regardless of how much sleep is obtained at night. Patients may also complain of difficulty with concentration or memory or increased irritability. Children may exhibit hyperactivity rather than sleepiness.

The differential diagnosis of excessive daytime sleepiness ranges from insufficient sleep to sufficient sleep that is disrupted by pathologic events, such as apneas, or neurologic disorders, such as narcolepsy. Inquiring about sleep routines and bedtimes and wake times is essential in excluding insufficient sleep as a cause of sleepiness. Asking patients who complain of sleepiness about other associated symptoms provides essential information. Loud snoring, gasping, snorting, and episodes of apnea suggest the diagnosis of obstructive sleep apnea syndrome (see Chapter 116). A history of episodic muscle weakness with buckling of the knees, laxity of the neck or jaw muscles, or complete loss of muscle tone associated with laughter, anger, or hearing or telling a joke suggests cataplexy and a diagnosis of narcolepsy (see Chapter 92). Questions assessing mood are needed to identify patients with sleep disorders associated with depression (see Chapter 140). Circadian rhythm sleep disorders should be considered in patients with complaints of nocturnal insomnia and daytime sleepiness (see Chapter 41).

Nocturnal Movements, Behaviors, and Sensations

Information from collateral sources is needed for evaluation of episodic movements and behaviors during sleep. The bed partner should be asked to describe behaviors and vocalizations during the episodes, to relate episodes to sleep onset and time of night, and to note the degree of the patient's responsiveness during the episode. The patient's ability to recall the events is also significant. Episodes of inconsolable screaming and amnesia during the first third of the night suggest sleep terrors (see Chapter 104); episodes of dream-enactment behavior associated with dream recall that occur toward the end of the sleep cycle REM sleep behavior disorder (see Chapter 105). Epileptic seizures may occur at any time of the night and should be strongly considered if a history of stereotyped behavior or dystonic posturing is elicited (see Chapter 99).

MEDICATION USE AND MEDICAL HISTORY

Assessment of medication use, including nonprescription medications, herbal supplements, and illicit drugs, is critical because of the wide variety of medications that alter sleep, wakefulness, and sleep disorders (see Chapter 46).

The history of current or past medical, surgical, and psychiatric illnesses is a source of important information. Seizure disorders, parkinsonism and dementia, arthritic conditions, asthma, ischemic heart disease, migraine or cluster headache, compressive neuropathies, and almost any painful illness can cause significant sleep disturbance. Anemia, renal disease, and pregnancy may cause or exacerbate restless legs syndrome or periodic limb movement disorder. Anxiety disorders, including panic disorder, and mood disorders are psychiatric disturbances that are often accompanied by insomnia, and some patients with depression complain of excessive daytime sleepiness.

FAMILY HISTORY

A history of disordered sleep in family members is important information. Specific inquiry should be made about the existence in family members of previously diagnosed sleep disorders or symptoms suggestive of narcolepsy, obstructive sleep apnea, periodic limb movements, enuresis, sleep terrors or sleepwalking, or insomnia. There is a strong genetic contribution to the development of narcolepsy (see Chapter 91), and genetic and familial influences sometimes have a role in the development and expression of obstructive sleep apnea (see Chapter 115) and some of the parasomnias.

SOCIAL HISTORY

Assessment of psychosocial, occupational, and academic functioning as well as of satisfaction with personal relationships can yield valuable information about the impact of disordered sleep on the patient's life. Alcohol, caffeine, nicotine, and illicit drug use should be determined. Alcohol use or abuse may intensify snoring and obstructive sleep apnea, may be a contributor to insomnia, or may produce long-lasting changes in sleep patterns. Caffeine use produces significant sleep disturbance in susceptible persons, and nicotine dependency may lead to nocturnal awakenings.

REVIEW OF SYSTEMS

The review of systems may uncover symptoms of medical illnesses that can cause or contribute to sleep disorders (Box 57-1). Recent weight gain or increase in collar size increases the likelihood of obstructive sleep apnea. Particular attention should be paid to the cardiovascular and pulmonary systems because of their relation to breathing and oxygenation during sleep. Angina, orthopnea, paroxysmal nocturnal dyspnea, and wheezing may indicate that sleep disturbance is

Box 57-1 SAMPLE SLEEP REVIEW OF SYSTEMS

Sleep Habits
Bedtime on weekdays
Bedtime on weekends
Wake time on weekdays
Wake time on weekends
Time it takes to fall asleep
Awakenings during the night
Time it takes to fall back asleep

Morning Symptoms
Dry mouth
Refreshed
Morning headaches
Nasal congestion

Daytime Functioning
Sleepiness (specific situations when patient has fallen asleep)
Falling asleep while driving
Any accidents caused by sleepiness
Memory problems
Difficulty concentrating
Fatigue
Irritability
Naps (how many, how long, dreams?)

Bed Partner's Observations or What Patient Has Been Told
Loudness of snoring (mild, moderate, severe)
Witnessed apneas
Choking or gasping
Arousals

Sleep-Related Movements
Periodic leg movements
Leg cramps
Restless legs symptoms

Narcoleptic Symptoms
Cataplexy
Hallucinations
Sleep paralysis
Automatic behaviors
Disrupted sleep

Genitourinary System
Nocturia
Impaired sexual functioning

Other
Weight gain in recent years
Sleepwalking
Dream-enactment behavior

due to cardiac or pulmonary disease. Heartburn and reflux of gastric contents into the throat when the patient is recumbent may cause nocturnal choking episodes. Leg cramps and neuropathic pain may be accompanied by sleep disruption. Nocturia is a common cause of disturbed sleep, particularly in older men. Depression or anxiety can contribute to insomnia.

PHYSICAL EXAMINATION

Examination of the head and neck is particularly important in patients with suspected obstructive sleep apnea. Auscultation of the chest may reveal expiratory wheezes in patients with nocturnal asthma attacks. Thoracic abnormalities such as kyphoscoliosis may compromise ventilatory capacity, leading to hypoventilation and nocturnal breathing difficulties. Auscultation may reveal a prominent fourth heart sound originating from the enlarged right ventricle and murmurs related to pulmonary or tricuspid valve insufficiency. On abdominal examination, hepatomegaly may suggest that alcohol abuse is contributing to sleep disturbance or, in conjunction with other findings, that congestive heart failure is a factor. Examination of the extremities may reveal joint swelling or deformity, decreased range of motion across affected joints, and thickening of synovial tissue in patients with disordered sleep due to arthritis.

Findings on mental status testing and neurologic examination may indicate the presence of a psychiatric or neurologic disease that causes or contributes to disturbed sleep. Impairment of short-term memory, judgment, language functions, and abstract reasoning suggests the presence of a dementing illness that may cause insomnia or nocturnal confusion. Assessment of mood may suggest the presence of mania or depression, either of which may be associated with insomnia. Delusional thoughts and agitation may indicate that acute psychosis is the cause of insomnia. Reduced alertness with slurred speech and nystagmus may be signs of hypnotic or sedative abuse. Impaired sensation and reduced or absent tendon reflexes may indicate peripheral neuropathy, sometimes accompanied by nocturnal paresthesias or burning pain. Elements of the physical examination relevant to the sleep patient are covered in greater detail in Chapter 59.

CLINICAL PEARL

A complete sleep and medical history often yields a specific cause of the patient's sleep complaint. For example, in patients complaining of excessive daytime sleepiness, the cause can often be pinpointed by close attention to the patient's (and bed partner's) account of nighttime symptoms, bedtime and wake time schedules, medications, and coexisting medical disorders.

SUMMARY

The evaluation of a patient with disordered sleep begins with the chief complaint, which can be classified as insomnia, daytime sleepiness, episodic nocturnal movements or behaviors, or a combination of these concerns. A thorough characterization of these concerns, coupled with a comprehensive sleep history that includes the daily schedule, bedtime routine, and morning and daytime symptoms, forms the foundation for diagnosis. As in other fields of medicine, it is essential to consider other medical and psychiatric conditions, medication use, family history, social history including the psychosocial situation, review of systems, and physical examination before formulating a differential diagnosis and performing diagnostic studies. This systematic approach allows accurate diagnosis and specific interventions for many treatable sleep disorders.

Chapter
58

Cardinal Manifestations of Sleep Disorders

Bradley V. Vaughn; O'Neill F. D'Cruz

Chapter Highlights

- Sleep disorders include a wide range of conditions that impair health and quality of life. To identify patients afflicted with these disorders, to institute effective treatment, and to optimize patients' health and quality of life, early clinical recognition of the fundamental symptoms (insomnia, hypersomnia, and unusual sleep-related behaviors), as well as the more subtle signs of these conditions, is essential.
- Insomnia is a common symptom, and as a diagnostic entity it can be related to many contributing factors. Features that predispose to, precipitate, and perpetuate the insomnia can be identified in persons who suffer from chronic insomnias. Insomnia can have an intricate relationship with other medical and psychiatric disorders.
- Hypersomnia often is a presenting feature of other sleep disorders or may reflect lifestyle choices. The difficulty in diagnosis is in distinguishing sleepiness from fatigue, and in determining the underlying contributing features. The pattern of sleepiness, characteristics of the sleep period including length, and response to these sleep periods give clear clues. Other manifestations, such as snoring, witnessed apnea, unrefreshing sleep, and morning headache, may indicate potential sleep-related breathing disorders or cataplexy, prompting an evaluation for narcolepsy.
- Unusual nocturnal sensations or events can provide a window into sleep issues. Features of restless legs syndrome, periodic limb movements in sleep, and parasomnias, such as sleepwalking, sleep terrors, and dream enactment, require detailed description of the sleep-related behaviors to help differentiate among potential causes. These events may indicate other underlying sleep disorders and brain disease.
- Sleep issues also may manifest as systemic problems. Findings of hypertension, unexplained weight gain, or mood or cognitive issues may be important flags signaling a need to inquire about underlying sleep issues. In addition, sleep issues may be early signs of brain disease.

Sleep is essential to health, restoring properties that promote wakefulness and a sense of well-being. Sleep disturbance, however, frequently disrupts this sense of well-being and can result in a wide range of systemic and neuropsychological symptoms. Sleep disruption due to intrusion of components of the sleep state into periods of wakefulness may manifest as hypersomnia. Similarly, intrusion of components of wakefulness into the sleep period may manifest as insomnia. Beyond the medical manifestations, sleep disruption also has consequences for societal and public health issues by impairing work performance and psychosocial interactions. As the connection of sleep to good health is increasingly appreciated, sleep disruption is now recognized to exacerbate symptoms of other diseases. These pathologic effects may manifest as worsening of a preexisting disorder or as impairment of the patient's ability to cope with its symptoms. The challenge for physicians is to recognize these manifestations and appropriately delineate them as being related to dysfunction of sleep.

Most patients referred to sleep centers present with one or a combination of three classic complaints: excessive sleepiness, difficulty attaining or sustaining sleep, or unusual events associated with sleep. These symptoms can be easily recognized as related to sleep and are not mutually exclusive in nature. Patients may note more than one problem, such as difficulty sleeping at night and excessive sleepiness during the day. Others may complain of unusual events at night with daytime sleepiness or inability to sleep. Each of these symptoms conveys clues to the underlying pathologic process (Figures 58-1 to 58-3). This chapter reviews the cardinal manifestations of sleep disorders and describes some of the key features that guide the clinician to pursue further diagnostic evaluations.

INSOMNIA

The diagnosis of insomnia is dependent on the complaint of difficulty initiating or maintaining sleep combined with daytime sequelae. This combination of poor nighttime sleep with an adverse effect on daytime activities is important to establishment of the complaint as insomnia. The daytime manifestations of insomnia may take the form of excessive fatigue, impairment of performance, or emotional change. Individual sleep need may vary significantly. Some people may feel fine and note no impairment of performance with 5 hours

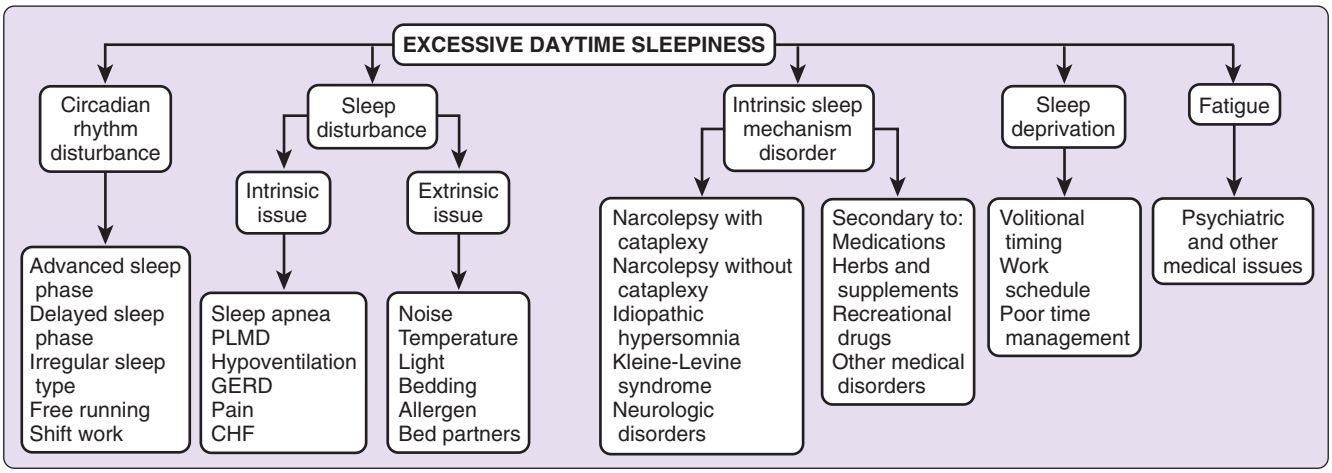


Figure 58-1 Diagnostic flow chart to approach excessive daytime sleepiness. CHF, Congestive heart failure; GERD, gastroesophageal reflux disease; PLMD, periodic limb movement disorder.

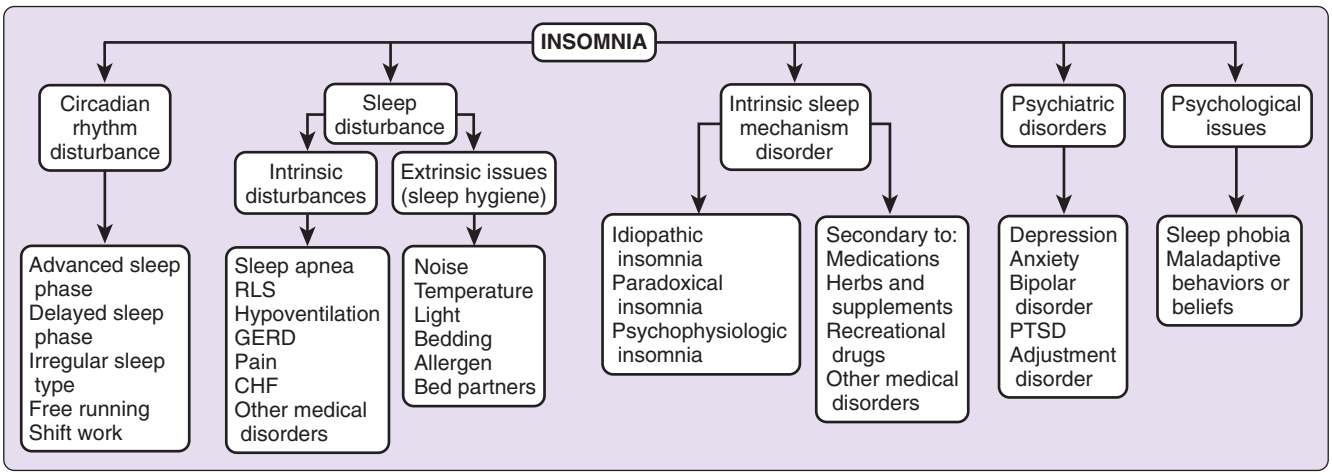


Figure 58-2 Diagnostic flow chart to approach insomnia. CHF, Congestive heart failure; GERD, gastroesophageal reflux disease; PTSD, posttraumatic stress disorder; RLS, restless legs syndrome.

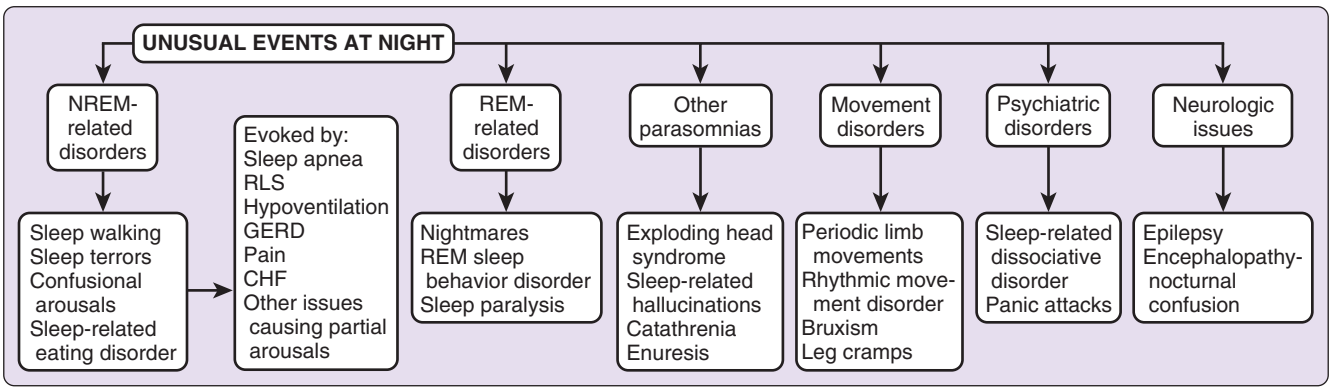


Figure 58-3 Diagnostic flow chart to approach unusual nocturnal events. CHF, Congestive heart failure; GERD, gastroesophageal reflux disease; RLS, restless legs syndrome.

of sleep per night, whereas others may need more than 9 hours to preserve daytime functioning. Thus the requirement of daytime sequelae differentiates individual sleep need from the complaint of insomnia.

Most people have an occasional night fraught with difficulty falling asleep or trouble maintaining sleep. These occasional nights may be closely linked to the surrounding events of the day, psychological challenges, or sudden changes in environment or onset or exacerbation of a medical condition. Surveys have shown that approximately one third of people complain that their sleep is disrupted on occasion and that a smaller group, of approximately 1 in 10, have a more persistent insomnia.[1] For these patients, lack of "good-quality" sleep produces a greater disruption of life and may lead to more significant medical or psychological symptoms.

As a symptom, insomnia is directly related to the patient's perception of poor sleep. Patients with insomnia believe that the sleep disruption produces their excessive sleepiness, fatigue, lack of concentration, muscle aches, and depression, and that a good night's sleep would reverse these symptoms. Patients with insomnia frequently describe themselves as tense, anxious, nervous, tired, irritable, unable to relax, obsessively worried, and depressed. Many of these traits may predate the onset of the insomnia, but others may arise after the onset of the poor sleep. Patients with insomnia frequently give historical clues directed toward the mechanisms behind their insomnia. The symptom complex may indicate an underlying disorder related to primary failure of sleep mechanics or one in which sleep disruption is the byproduct of another disorder. Because sleep is an active process, neuronal networks involved with sleep induction must be engaged and networks involved in wakefulness must be diminished for sleep to occur. Rarely is just one factor responsible for chronic insomnia (defined as lasting longer than 3 months). In most patients, multiple factors contribute to the risk of developing and maintaining insomnia. The presence of predisposing, precipitating, and perpetuating factors emphasizes the nature of insomnia as an ongoing process, and clinicians need to search for these contributing factors to outline an effective treatment course (see Chapters 80 to 88).

As indicated by epidemiologic studies, insomnia appears to be more common in women, older persons, and those with psychiatric or chronic medical illness. Insomnia also is more common in persons of lower socioeconomic status and with less education. Behavioral traits such as obsessive-compulsive nature, frequent rumination, and poor coping strategies and a "hyperalert" baseline state are correlated with greater risk for insomnia. So-called hyperarousal documented on recent neuroimaging studies may explain the neurophysiologic basis for these associated factors predisposing to chronic insomnia.[1]

Insomnia may be initiated by sudden changes in environment or challenges to the body or mind. These challenges may come in the form of acute medical illness, psychological or psychiatric events, shift in schedule, or changes in medications or diet, including supplements. Although these events provide good clues to preventing further recurrence of insomnia, initiating events may play little role in the patient's current, ongoing process..

Many patients, in attempting to improve their sleep, may adopt behaviors that actually perpetuate the insomnia. Patients may employ rituals and "remedies" that convert the short-term insomnia into a chronic form. During this evolution, the patient may institute changes in sleep schedule, resort to certain somnogenic substances, or develop secondary medical or psychological issues. Many of these behaviors conflict with typical sleep hygiene practices, producing an environment detrimental to sleep. Such maladaptive habits may occur during the day or night and include issues such as heavy caffeine or alcohol use, watching television or playing video games while in bed, and even eating or exercising during the usual sleep period. Some patients may claim that television or radio distracts them from intrusive thoughts. In others, the development of sleep associations may be a mechanism to counterbalance negative experiences. A subgroup of patients actually fear going to bed or experience performance anxiety over the oncoming sleep period. This expectation of poor sleep promotes apprehension toward sleep and may perpetuate counterproductive sleep rituals. These maladaptive behaviors become the predominant feature of psychophysiologic insomnia. Many patients with these types of negative associations temporarily improve once placed in a new environment.

The timing of the insomnia during the sleep period also may be helpful in the evaluation. Circadian rhythm sleep-wake disorders can masquerade as complaints of insomnia or excessive sleepiness, and patients with insomnia may develop dysfunction of their circadian rhythm. Difficulty with the onset of sleep suggests an underlying delayed sleep phase or occasionally depression in younger adults. Insomnia with early-morning arousal raises the possibility of underlying depression or advanced sleep phase. Schedule changes, such as from jet lag or shift work, are important clues, and sleep diaries of bedtime and wake time can be useful in determining potential links to schedule or circadian rhythm issues. Timing also may correlate with other issues such as restless legs syndrome or medication or caffeine intake. Specific questions should explore the patient's daily routine including the timing of activities that may be stimulating, such as exercise or work or gaming on a computer.

Perception of good sleep is an important factor in evaluating the complaint of insomnia. Some patients exaggerate their symptoms; still other patients may not even recognize that sleep is occurring. Paradoxical insomnia is one subtype of chronic insomnia in which people do not recognize that they have slept despite the recording of normal physiologic parameters of sleep. Other patients may endorse unrealistic expectations or unobtainable goals. Patients may assume that sleep should not be interrupted by any arousals or that it is essential to sleep for a set number of hours. These beliefs can be easily addressed with appropriate education about these issues.

For some people, insomnia may start in childhood and be lifelong. The subtype of chronic insomnia known as idiopathic insomnia is not associated with clear inciting factors. In affected persons, insomnia persists despite a change in environment or other measures, and significant family history may be identified. These factors are discussed further in Chapters 82 and 83.

Medical or neurologic disorders may precipitate and perpetuate insomnia. Derangement of almost any system in the body can disrupt sleep. Patients with heart, liver, or renal failure or disturbances of the gastrointestinal system or pulmonary disease commonly complain of insomnia. Patients with fulminant rash or significant burns frequently note disturbed sleep, and urologic issues such as nocturia may provoke frequent arousals. Neurologic disorders also promote sleep

disruption. Neuromuscular disorders, for example, may be associated with discomfort or inadequate ventilation at night that provokes insomnia. Some patients who suffered a stroke have noted insomnia or sleep disruption after the vascular event. Paralysis from central or peripheral nervous system disorders can result in nighttime discomfort from inability to move. Patients with Parkinson disease may exhibit akinesis, tremor, or medication effects, and patients with dementia may have circadian rhythm abnormalities that promote awakenings at night.

Pain can disturb sleep and promotes insomnia. Musculoskeletal discomfort may become worse with periods of rest. Arthritis and other rheumatologic disorders frequently can disrupt sleep with increasing nighttime pain and stiffness. Pain from headaches, such as cluster headache, and even pain related to increased intracranial pressure or brain mass lesions can become more intense during sleep, and symptoms from entrapment neuropathies, such as carpal tunnel syndrome, typically are worse at night. Restless legs syndrome produces a classic urge to move that is worse in the evening.

Nearly all the psychiatric illnesses have some link to poor sleep. Patients with depression or anxiety disorders may endure insomnia for years before presentation of the affective component. Although the cause and effect are still in debate, the association is clear. Insomnia may herald the onset of psychosis or mania.

The clinician may uncover few physical findings in patients with insomnia. Anxious or hyperalert persons may demonstrate mild tachycardia or rapid respiratory rate; cold hands may be an associated feature. These patients may easily startle or be distracted during the interview. The clinician should look carefully for signs of obstructive sleep apnea, narrow airway, and obesity because these too can be manifested as insomnia. Signs of Cushing syndrome (round face and buffalo hump) or hyperthyroidism (tachycardia and excessive sweating) are important clues to an endocrine disorder. Each patient with a complaint of insomnia should undergo a complete neurologic examination to look for potential neurologic lesions impairing sleep. This examination should include an assessment of cognition, mood, and affect. Insomnia or interrupted sleep can occur in many neurologic disease states, including several forms of dementia. The Mini Mental State Examination is a tool that helps assess cognitive abilities, with findings monitored for follow-up over time.[2] Clinicians also can use the Minnesota Multiphasic Personality Inventory to identify personality and affect issues, and the Hamilton Anxiety and Depression Scales may be helpful tools in clinical follow-up for persons so affected.

EXCESSIVE DAYTIME SLEEPINESS

Sleepiness is a common symptom noted by 5% to 20% of people.[3,4] Most people can relate some instances of falling asleep when they intended to be awake. Sleepiness is a normal feeling on approaching a typical sleep period or after prolonged wakefulness. Excessive sleepiness may manifest as sleep in an inappropriate setting or as episodes of unintentional sleep. Excessive sleepiness can occur in various degrees of severity. In mild sleepiness, the affected person may fall asleep while reading a book or while sitting quietly. This degree of sleepiness may produce only limited impairment in the person's perceived quality of life. Greater degrees of sleepiness may be associated with bouts of irresistible sleep or sleep attacks intruding on such activities as driving, participating in a conversation, or eating meals. This degree of sleepiness may place the patient at significant risk for accidents and have a major impact on the person's health and sense of well-being.

As with other subjective symptoms, the person's perception of sleepiness influences the nature of the complaint. Some patients may overreport the degree of sleepiness and note this feeling even during periods of normal wakefulness. Others may underreport and not recognize periods of daytime sleepiness. For some people in the latter group, sleepiness may be described as periods of lapse of attention or diminished cognitive abilities, such as missing an exit on the highway or brief delay in performing a task. Perception of sleepiness also is reduced with continued sleep deprivation. People who are chronically sleep-deprived become accustomed to their impairment and are less likely to recognize their degree of sleepiness.

Clinicians should always question their hypersomnic patients for clues of potential sleep debt, dyssomnia, or medical or psychiatric causes. Sleep deprivation is common in today's society, and patients should be queried about their schedule during the week and weekends. Information elicited about sleep habits and environment may disclose important factors contributing to the sleepiness.

Excessive sleepiness may result from a wide range of medical disorders and medication. Patients with heart, kidney, or liver failure and rheumatologic or endocrinologic disorders such as hypothyroidism and diabetes may note sleepiness and fatigue. Similarly, a wide range of medications may cause daytime sleepiness even when taken at night. Neurologic disorders, such as strokes, tumors, demyelinating diseases, and head trauma, can be associated with excessive sleepiness. Sleepiness frequently is the cardinal symptom of many sleep disorders. Patients with sleep apnea, narcolepsy, idiopathic hypersomnia, or even parasomnias may identify excessive daytime sleepiness as their main complaint. Historical features of snoring, observed apneas, morning headaches, cataplexy, sleep paralysis, hypnagogic hallucinations, and confusion on arousals suggest contributions of a specific sleep disorder. Persons with idiopathic hypersomnolence experience unrelenting daytime sleepiness despite prolonged periods of sleep, which differentiates this disorder from sleep deprivation. Many adults with idiopathic hypersomnia find naps are not refreshing, whereas patients with narcolepsy (type 1 and type 2) note that brief naps actually improve their daytime sleepiness.

Physical findings are few in patients with sleepiness. Frequent pauses, slowed responses, drooping eyelids, and repetitive yawning support the complaint of sleepiness. Patients may be asleep when the clinician enters the examination room, and some patients may show signs of chronic sleepiness, such as dark circles under the eyes. The neurologic examination may yield findings of inattentiveness or even brief "microsleeps."

Sleepiness can be quantified subjectively by questionnaires or by objective assessments such as a multiple sleep latency test. The Epworth Sleepiness Scale is one example of a quantifiable subjective measure of sleepiness and has been translated into several languages (Table 58-1).[5] For scoring on this scale, the patient is asked to rate on a scale of 0 to 3 (0, no chance; 3, high likelihood) the chance of dozing in a series of eight situations. This score has a modest correlation with

physiologic measures of sleep but has a better correlation with the respiratory disturbance index in patients with obstructive sleep apnea (Table 58-2).

Two quantitative tests are available to measure ability to fall asleep and to stay awake: the Multiple Sleep Latency Test (MSLT) and the Maintenance of Wakefulness Test. The MSLT quantifies objective sleepiness on the basis of the time to onset of physiologic changes associated with sleep across five trials separated by 2 hours each across the typical wake period. Sleep is determined by the loss of the posterior dominant rhythm on the electroencephalogram or, in the absence of posterior dominant rhythm, slow eye movements, vertex sharp waves, and slowing of the background electroencephalographic activity. The MSLT uses these physiologic markers to quantify the time to sleep. Unfortunately, the MSLT is not well correlated with daytime function, and significant overlap has been documented between persons deemed "normal" and those deemed to have sleep disruption. Although the MSLT can "quantify" the degree of sleepiness on a particular day, the test is validated only for the diagnosis of narcolepsy (type 1 and type 2). The Maintenance of Wakefulness Test quantifies the propensity to stay awake during four attempts in a dimly lit room. This test has not been extensively tested in relationship to daytime function. These tests are covered in greater detail in Chapter 169.

Table 58-1 The Epworth Sleepiness Scale

Name: _____
Today's date: _____ **Your age (years):** _____
Your sex (male = M; female = F): _____
How likely are you to doze off or fall asleep in the following situations, in contrast with feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you. Use the following scale to choose the most appropriate number for each situation:
0 = would never doze
1 = slight chance of dozing
2 = moderate chance of dozing
3 = high chance of dozing

Situation*	Chance of Dozing
Sitting and reading	_____
Watching TV	_____
Sitting, inactive in a public place (e.g., a theater or a meeting)	_____
As a passenger in a car for an hour without a break	_____
Lying down to rest in the afternoon when circumstances permit	_____
Sitting and talking to someone	_____
Sitting quietly after a lunch without alcohol	_____
In a car, while stopped for a few minutes in traffic	_____
Thank you for your cooperation.	

*The numbers for the eight situations are added together to give a global score between 0 and 24.
From Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.

FATIGUE

The complaint of fatigue is a complex symptom typically related to the perception of lack of energy. Many patients with excessive daytime sleepiness note fatigue or decreased energy. Patients may be aware of the lack of energy but not perceive the degree of sleepiness, or may confuse the symptom of fatigue with excessive sleepiness. Although frequently noted in combination, fatigue is distinct from excessive sleepiness. Patients with fatigue alone may not have an increased ability to fall asleep but believe that a good night's sleep would correct their lack of energy. Distinguishing sleepiness from fatigue can be difficult even for the most astute clinician. Detailed questioning regarding ability to fall asleep can be helpful. Patients with insomnia also frequently complain of fatigue related to disrupted sleep, as may patients with immunologic, endocrinologic, or organ failure. Similarly, patients with depression frequently have associated fatigue without sleepiness.

SNORING

Snoring is the sound created by turbulent airflow vibrating upper airway soft tissue. Usually more prominent during

Table 58-2 Sleep-Related Conditions and Sleepiness: Epworth Sleepiness Scale Score in Experimental Subjects

Condition/Diagnosis	Total No. of Subjects (Males/Females)	Age (Years): Mean ± SD	Epworth Sleepiness Scale Score	
			Mean ± SD	Range
Healthy control subjects	30 (14/16)	36.4 ± 9.9	5.9 ± 2.2	2–10
Primary snoring	32 (29/3)	45.7 ± 10.7	6.5 ± 3.0	0–11
Obstructive sleep apnea syndrome	55 (53/2)	48.4 ± 10.7	11.7 ± 4.6	4–23
Narcolepsy	13 (8/5)	46.6 ± 12.0	17.5 ± 3.5	13–23
Idiopathic hypersomnia	14 (8/6)	41.4 ± 14.0	17.9 ± 3.1	12–24
Insomnia	16 (6/12)	40.3 ± 14.6	2.2 ± 2.0	0–6
Periodic limb movement disorder	18 (16/2)	52.5 ± 10.3	9.2 ± 4.0	2–16

SD, Standard deviation.
From Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.

inspiration, snoring occurs in approximately 32% of adults and in more than 7% of children.[6,7] Many adults have little knowledge or recognition of their snoring habits, and accounts from bed partners may be more helpful for the clinician.

Snoring usually is worse with the patient in the supine position, after sleep deprivation or alcohol ingestion. Loud snoring may not disturb sound sleepers, but some patients may report complaints from family members and even neighbors. Snoring may continue for decades. Persistent loud snoring is a classic symptom of obstructive sleep apnea syndrome, but the absence of snoring does not exclude the diagnosis of apnea. In some cases, the airway dynamics are not conducive to snoring. This is especially true in patients who have had upper airway surgical procedures that eliminated flaccid tissue. Other patients, such as those with neuromuscular disorders, may not generate enough force to produce turbulent airflow.

Snoring, for many people, produces little disruption in their lives, but snoring may have implications for overall health. People who snore are at greater risk for vascular disease. Witnesses may describe the snoring occurring in bursts or associated with snorts, gasps, choking, body jerks, and movements. Patients may recall being awoken by their own gasps and relate symptoms of gastroesophageal reflux. These associated symptoms raise the clinical suspicion of obstructive sleep apnea.

SLEEP APNEA

Apnea is the absence of ventilation. In the sleep laboratory, an apneic event is defined as the cessation of breathing for more than 10 seconds and usually is associated with oxygen desaturation and arousal.[8] Although snoring is very common, witnessed apneic events and nocturnal gasping or choking are the most reliable subjective indicators of sleep apnea.[9] Some patients may experience hundreds of events in a single night and are unable to obtain good-quality sleep because of the frequent arousals. These patients typically are unaware of the arousals, but some may report being aware of occasional awakening with a gasp or snort. Sleep apnea is classified in two major forms: obstructive and central.

Obstructive apnea is the most common form of sleep apnea. These apneas are due to obstruction in the upper airway and more commonly are noted during stage N1 or N2 or rapid eye movement (REM) sleep. Snoring is a frequent associated complaint, but a wide variety of symptoms may be present. Some questionnaires, such as the Sleep Apnea section of the Sleep Disorder Questionnaire, Berlin Questionnaire, or the STOPBANG, a combination of items regarding snoring, witnessed apneas, body habitus, and associated disorders such as hypertension[10,11] (Table 58-3), provide a summary score that correlates with presence of obstructive sleep apnea, but the questionnaires have been tested only in selected populations. Thus the questionnaires themselves serve as rough guides and do not confirm the diagnosis of sleep apnea. Astute clinicians should not rule out the presence of apnea on the basis of a low score on a questionnaire. The physical examination may show structural evidence for airway obstruction. Many patients are obese, with a thick neck or crowded upper airway, yet some have a normal body habitus. Common structural abnormalities, such as a narrow nasal passage, long soft palate, large tonsils, or retroflexed mandible leading to a small airway, contribute to airway obstruction. Sleep apnea is associated with significant health risks and lowers the quality of life of both patient and bed partner. Mounting evidence from multiple studies such as the Sleep Heart Health Study indicates that hypopneas with oxygen desaturation correlate with greater risk of vascular disease.[12] These are discussed in greater detail in Chapters 126 to 128.

Central apnea is the absence of ventilation due to an absence of contraction of the lower respiratory musculature.

Table 58-3 Key Features of the Berlin Questionnaire

Height _____	**Age** _____
Weight _____	**Gender** _____
Has your weight changed in the last 5 years?	Increased Decreased No change
Do you snore?	Yes No Do not know
Your snoring is:	Slightly louder than breathing As loud as talking Louder than talking Very loud
How often do you snore?	Nearly every day 3 to 4 times per week 1 to 2 times per week 1 to 2 times per month Never or almost never
Has your snoring bothered other people?	Yes No
Has anyone noticed that you quit breathing during your sleep?	Almost every day 3 to 4 times per week 1 to 2 times per week 1 to 2 times per month Never or almost never
Are you tired or fatigued after your sleep?	Nearly every day 3 to 4 times per week 1 to 2 times per week 1 to 2 times per month Never or almost never
During your wake time, do you feel tired, fatigued, or not up to par?	Nearly every day 3 to 4 times per week 1 to 2 times per week 1 to 2 times per month Never or almost never
Have you ever fallen asleep while driving?	Yes No If so, how often does it occur? Nearly every day 3 to 4 times per week 1 to 2 times per week 1 to 2 times per month Never or almost never
Do you have high blood pressure?	Yes No Do not know

Modified from Reprinting of the Berlin questionnaire. *Sleep Breath* 2000;4(4): 187–92, with the permission of Kingman P. Strohl.

Affected patients experience pauses in respiration, also associated with oxygen desaturation and arousals. Central apneas can be caused by narcotics or a neurologic abnormality in the brainstem or other areas involved in regulation of respiration. Cheyne-Stokes breathing can manifest features of both central and obstructive apnea. The classic pattern of crescendo-decrescendo breathing with a central apnea can be seen in persons with heart failure, neurologic lesions, and metabolic or toxic encephalopathies. This pattern may be present only in sleep and may signal underlying disease. Apneic episodes may follow other neurologic events, such as nocturnal seizures, and are more prevalent after acute strokes and seizures. Central apnea is reviewed in Chapters 109 and 110.

CATAPLEXY

Cataplexy is the abrupt loss of muscle tone triggered by strong emotional stimuli or physical exercise.[13] Patients are aware of their surroundings and have clear memory for the complete events. Events can be triggered by a joke, surprise, anger, fear, or enthusiastic athletic endeavors. Individual experiences range in intensity from a mild feeling of weakness to severe weakness precipitating a fall. Cataplectic attacks arise over several seconds and may start with brief waves of loss of tone that may appear like jerks initially. In most cases, symptoms initially affect the face and neck and then progress to involve the rest of the body. Patients may describe more subtle events as a feeling of slowness to respond or slurring or speech. The attacks generally are brief, lasting less than a few minutes. Patients then regain muscle control and exhibit no postictal confusion or memory deficits. More prolonged attacks may be ended with the patient's entering sleep and then awakening. Examination of the patient during the cataplectic attack will demonstrate paralysis with diffuse hypotonia, absence of deep tendon reflexes, diminished corneal reflexes, preservation of pupillary responses, and in many instances, phasic muscle twitching. Phasic muscle twitching can occur as single jerks or repetitive muscle twitching and most frequently is seen in the face, sometimes being confused with seizure activity.

The combination of excessive daytime sleepiness and cataplexy is nearly always related to narcolepsy type 1. Cataplexy can rarely be seen as an isolated symptom suggestive of an underlying neurologic lesion in the brainstem. Maintenance of consciousness and memory helps differentiate these events from most seizures and syncope. The historical feature of clear emotional triggers differentiates cataplexy from vertebral basilar insufficiency and the group of neuromuscular disorders known to produce periodic paralysis. Cataplexy also is differentiated from myasthenia gravis by the abrupt onset and absence of muscle fatigue with repetitive stimulation that is typical of myasthenia gravis.

SLEEP PARALYSIS

Sleep paralysis is an inability to move during the transition into or out of sleep. The association with intentional sleep distinguishes these events from cataplexy. Patients may describe complete awareness of their surroundings or feeling partially asleep with awareness but being unable to move even the fingers or to vocalize. Patients may try to scream but produce only a whisper. Some affected persons may describe a feeling of suffocation, with resumption of breathing only when the event has passed. Patients frequently describe a strong feeling of impending doom, being chased, or having to escape imminent danger. On occasion, patients may note the feeling that someone else is in the bedroom. Auditory and tactile hallucinations may accompany the events, and patients may recount dramatic stories. These events can be emotionally profound and leave a lasting memory that patients vividly recall years later. Most sleep paralysis episodes last a few minutes and usually end after the patient is touched or is alerted. If the event is allowed to persist, the patient usually reenters sleep and awakens later. These events are experienced by many people in association with severe sleep deprivation, schedule disruption, or ingestion of alcohol and may be more frequent in patients with narcolepsy.

HYPNAGOGIC AND HYPNOPOMPIC HALLUCINATIONS

Hallucinations can occur with sleep onset (hypnagogic) or at the end of sleep (hypnopompic). Such hallucinations may include visual, auditory, or tactile components and may last seconds to minutes. The events occur at the transition between wake and sleep and incorporate some dream-like features. They can be relatively pleasant or very terrifying and difficult to distinguish from reality. Patients may note a feeling of weightlessness, falling, or flying or describe an out-of-body experience; the episode may sometimes terminate with a sudden jerk (hypnic jerk). Visual hallucinations may be described as poorly formed colors and shapes or well-formed images of people and animals. The events are terminated once the patient awakens.

Exploding head syndrome is characterized by the subjective experience of a loud, painless explosion that occurs near sleep onset. This parasomnia is benign and may be associated with a hallucinatory flash of light.

In the face of excessive daytime sleepiness, patients with hypnagogic hallucinations should be evaluated for narcolepsy. These events may be repetitive but usually are not stereotypical. This lack of stereotypical features distinguishes these events from seizures. Affected patients may experience these events after sleep deprivation or a change in the sleep schedule. Alcohol ingestion or withdrawal of REM suppressants may also provoke these events. The relationship of sleep to these hallucinations distinguishes them from hallucinations of psychosis and dementia. Hypnagogic hallucinations are shorter in duration than peduncular hallucinations. Some people with dementia experience hallucinations at night. These types of hallucinations are associated with cognitive impairment during the day and most commonly are seen in patients with Lewy body dementia but can occur with other forms of dementia. Small people or animals predominate in these hallucinations, and many patients experience these events while awake.

AUTOMATIC BEHAVIOR

Automatic behavior consists of purposeful but inappropriate activities that occur with the patient partially asleep. Patients relay stories of putting milk containers in the microwave oven, cereal bowls in the dryer, or even missing an exit on the highway. Sleep-deprived soldiers have been reported to

continue marching in the wrong direction. Patients appear drowsy or groggy during the event and usually are partially or totally amnestic for the actual happenings. Events may last minutes to up to an hour. Automatic behavior and "sleep inertia," the persistence of profound sleepiness into the awake state, are more common in persons with idiopathic hypersomnolence but also are common in patients with a delayed sleep phase.

These events are distinguished from seizures by the lack of stereotypical behavior. Automatisms associated with seizures usually are stereotypical and repetitive, such as picking, rubbing, or lip smacking. Patients with sleep-related automatic behavior appear sleepy but can be alerted and answer questions appropriately, in contrast with those with postictal confusion and metabolic or toxic encephalopathy. The quick return of orientation with lack of bewilderment and anxiety also differentiates this entity from transient global amnesia.

EXCESSIVE MOVEMENT IN SLEEP OR PARASOMNIA

Patients and their bed partners may complain of frequent movement during the sleep period. This complaint may be more concerning to the bed partner than to the patient. Excessive sleep movement also is a common complaint among patients who complain of insomnia and in persons with sleep apnea. Some patients may complain of being active sleepers and describe a corresponding high level of mental activity or an inability to turn off their mind. These patients need an evaluation focusing on features of insomnia. Those who are physically active need evaluation for the movements or a parasomnia.

Parasomnia refers to undesirable physical or behavioral phenomena that occur predominantly during sleep. They include disorders of arousals, such as sleepwalking and night terrors; sleep-wake transition disorders, such as bruxism or rhythmic movement disorder (e.g., head banging); and REM parasomnias, such as REM sleep behavior disorder. These behavioral events may mimic epileptic seizures or other psychiatric events, and a clear description from a keen observer is very helpful in leading to the correct diagnosis. The parasomnias are covered in greater detail in Chapters 101 to 106.

Key features of age at onset, time of night of the events, memory for the events, and family history are important in determining the etiology of parasomnias. Stereotypical behavior—recurrence of the same behavior with each event—also can help in categorizing the events. Events such as periodic limb movements, rhythmic movement disorder, and epileptic seizures are associated with stereotypical behavior, whereas sleeptalking, sleepwalking, night terrors, and dream enactment incur different behavior with each event. Although historical features can be useful in distinguishing among these disorders, most patients require polysomnographic recording to delineate the cause.

Sleeptalking

Sleeptalking is a relatively frequent event, ranging from occasional utterances to coherent conversation during sleep. This usually occurs in the lighter stages of NREM sleep but can occur in REM ("stage R") sleep. Patients have no memory for such events, during which they may convey information that may have little resemblance to the truth. Many people talk in their sleep, and this is considered a normal variant. In the absence of other sleep disturbances, sleeptalking is of little medical concern.

Sleepwalking

Sleepwalking events usually are part of the disorders of arousal, indicating incomplete arousal typically from slow wave sleep (stage N3) occurring during the first half of the sleep period. The events can consist of minor behaviors and movement or elaborate behaviors including dressing, unlocking locks, minor housekeeping or work tasks, and even driving. Patients usually have little or no memory of the event. They typically can recall various associated feelings or impressions, however, and some imagery is more common in adults. Affected patients do not exhibit significant tachycardia, sweating, or manifestations of fear. The lack of screaming and autonomic features differentiates sleepwalking from sleep terrors. Both children and adults with a history of recent sleepwalking should be questioned regarding signs of other sleep disorders. Any disorder evoking arousals may increase the likelihood of these events, so a careful assessment to uncover symptoms of other sleep disorders is essential. Patients typically exhibit normal neurologic examination findings during wakefulness.

Sleep Terrors

Sleep terrors are a more intense form of disorder of arousal with a predominance of autonomic expression. Witnesses rarely forget the patient's sudden arousal accompanied by a piercing scream or cry, autonomic output, and behavioral manifestation of intense fear, but the patient has little to no memory of the event. The onset of the episode is abrupt, accompanied by tachycardia, tachypnea, flushing, diaphoresis, and mydriasis. The affected patient is confused and disoriented, and attempts to intercede may result in prolongation of the event and potential harm to the person trying to wake the patient. Patients can become violent, resulting in injury to the patient and bed partners. Less than 1% of adults may experience these events.[14] They usually occur in the first third of the night, and the events are nonstereotypical. Diurnal neurologic examination typically yields normal findings, and as with sleepwalking, a careful assessment for the presence of other sleep disorders is indicated.

Confusional Arousals

Confusional arousals can occur during any arousal from NREM sleep. These events are characterized by disorientation, slowed speech and mentation, or inappropriate behavior. Affected patients have memory impairment for the event, and the events can be induced with forced arousal. These events usually become less frequent with age, but rate of occurrence may remain stable in adulthood.

Patients may exhibit other complex sleep-related behaviors. In a well-characterized variant of such behavior, eating is the sleep-related event, as seen in sleep-related eating disorder. Affected persons consume quanitities of high-calorie, sometimes bizarre foods and have no or little memory of the episode. Morning anorexia and unexplained weight gain are typical. Another reported complex sleep behavior consists of engaging in multiple episodes of sexual intercourse during sleep. Again, the affected person relates no memory of the event.

Sleep-Related Groaning (Catathrenia)

Rarely, patients or families may present because of repetitive nocturnal groaning. Bed partners usually express concern because the patients voice long expiratory groans that sound mournful. Patients usually have no recollection of the sound or feeling distressed but may have morning hoarseness. No other detectable abnormalities are found on physical examination. These groaning events have been compared with central apneas, and the disorder is now classified under sleep-related breathing disorders.[15]

Dream Enactment

REM sleep ("stage R") is characterized by diffuse muscle atonia. Normally, only brief phasic muscle activity is noted during REM sleep, but pathologic dream enactment behavior can include punching, kicking, leaping, running, talking, yelling, and any behavior that could occur during a dream. Bed partners frequently are injured, and patients may go to great lengths to protect themselves and bed partners. This dream enactment commonly is seen as part of REM sleep behavior disorder (RBD). Patients usually have a vivid recall of the actual dreams that correlates with the witnessed behavior, and many dreams involve fleeing or defending themes. Dream recall is not uniformly noted, and patients may not be willing to talk about the dream that led them to seek medical attention. These events occur more commonly in the latter half of the night but can occur any time on entry into REM sleep. Patients may experience multiple events during a single night. In most cases, the dream enactment behavior begins in late adulthood, but children with symptoms of RBD have been described. RBD can be induced by medication, and tricyclic antidepressants, monoamine oxidase inhibitors, and serotonin reuptake inhibitors all have been implicated in causing RBD-like behavior. Acute forms of RBD also can occur during alcohol withdrawal and potentially benzodiazepine withdrawal. In the more chronic form of RBD, behaviors may occurr for years before the patient presents for medical evaluation. RBD has been linked to alpha-synucleinopathies. This group of disorders includes Parkinson disease, multiple system atrophy, and Lewy body dementia. Patients need a complete neurologic evaluation to look for signs of degenerative disorders. Other identifiable neurologic disorders, such as strokes, posterior fossa tumors, and demyelination, have been reported in association with RBD.[16]

Nightmares

Nightmares or recurrent disturbing dream mentation can be a presenting symptom of sleep disturbance. The hallmark of nightmares is emotionally intense dreaming associated with fear, anxiety, anger, sadness, or other negative emotions. Affected persons awaken from stage R or light NREM sleep to full alertness and usually recall the event immediately. Nightmares most commonly are associated with a psychologically disturbing event but may also be a result of medications, such as antihypertensives, antidepressants, or dopamine agonists. Nightmares can occur in patients with narcolepsy as well as in patients with sleep apnea.

Sleep-Related Rhythmic Movement Disorder

Rhythmic movement disorder can be manifested as a variety of distracting behaviors that occur before sleep onset. The movements are stereotypical, usually involving large muscles, and are sustained into light sleep. Movements and related behaviors may include head banging, body rocking, leg rolling, humming, and chanting and frequently are more concerning to the bed partner and family than to the patient. Many patients are relatively unaware of the movement, and others may describe such movement as inducing a calming effect or as a compulsion before sleep. This behavior frequently is seen in infants and young children, and the prevalence diminishes with age. It is more commonly seen in persons with mental challenges or autism and is more prevalent in males. Emotional stress may provoke it. Patients can be easily alerted during the events, which helps differentiate these events from seizures.

Sleep-Related Bruxism

Sleep bruxism also can occur as a rhythmic or repetitive movement during sleep.[17] Grinding or clenching of the teeth during sleep may produce bizarre sounds, and patients rarely may even vocalize with the episodes. Patients may exhibit abnormal wear of the teeth and complain of jaw pain, headache, facial pain, or tooth pain. They may experience hundreds of events per night, and the events increase with emotional stress. Some studies suggest that as many as 85% of the population grinds the teeth to some degree during the day or night. These events often occur in children, and persistence of symptoms occasionally is associated with a familial tendency.

RESTLESS LEGS SYNDROME AND PERIODIC MOVEMENTS OF SLEEP

Patients may complain of an unpleasant crawling, deep aching sensation in the legs or arms that is relieved in some measure by motion of the extremities. Diagnostic criteria focus on four main symptoms: the sensation or urge to move the limbs, worsening of symptoms with rest, improvement with movement, and higher frequency in the evening.[18] In addition, the patient must note a symptom of concern, distress, sleep disturbance, or some impairment related to the sensations. Patients with restless legs syndrome may relay that the discomfort can be debilitating at times, and some people are driven to pursue extreme measures to decrease the symptoms. Most patients experience the symptoms while sitting or lying down and may complain of the need to walk or to have continuous movement of their legs. These symptoms can lead to the affected person's walking until the early morning hours or trying to sleep despite the continuous leg motion. Other afflicted persons may use a combination of medication and alcohol to reduce the symptoms. Some patients note that their legs move or dance on their own, indicating periodic limb movements in wakefulness. Restless legs syndrome usually occurs along with periodic movements of sleep.

Periodic movements of sleep are repetitive stereotypical movements, typically of the lower extremities, that occur during sleep; in the legs, they consist of the extension of the great toe with dorsiflexion of the ankle and flexion of the knee and hip. The patient or a bed partner may complain of kicking or arm movements at night. These movements may occur as periodic events or appear random. Movements also can occur in the arms and axial muscles. The individual movements are relatively brief, lasting 0.5 to 5.0 seconds, and occur at 5- to

90-second intervals. Although a majority of people with restless legs also experience periodic limb movements of sleep, only a minority of patients with periodic movements of sleep will experience excessive daytime sleepiness or insomnia. Patients may be unaware of the movements, but bed partners usually are unable to ignore them. Similar factors that provoke periodic limb movements increase the likelihood of restless legs syndrome. Periodic movements of sleep have been associated with uremia, peripheral vascular disease, anemia, arthritis, peripheral neuropathy, spinal cord lesions, antidepressants, antiemetics, and caffeine use.

MORNING HEADACHE

Morning headache is a common symptom.[19] Almost three fourths of the population has occasional headache; this symptom is relatively nonspecific. Morning headache is more specifically linked to sleep dysfunction but also may indicate elevated blood pressure at night. Characteristics of the headache, such as location, quality, and nature of pain, and potential associations can aid in determining the etiology. Approximately one half of the patients with obstructive sleep apnea and hypoventilation note morning headache, typically dull and generalized in nature, which usually clears within an hour of waking. Patients with chronic obstructive pulmonary disease and obstructive sleep apnea may develop morning headache from the increase in carbon dioxide, low oxygen saturation, or vascular changes. Patients with sinus disorders, muscle contraction headache, post–alcohol intake, and withdrawal from medication (rebound headaches) may have distinct patterns. Cluster headaches are noted to occur in REM sleep. Hypnic headaches, or "alarm clock headaches," so called because of their regularity, are throbbing or sharp pains lasting 15 to 60 minutes that awaken the patient typically between 1 A.M. and 3 A.M. Headaches that routinely awaken patients from sleep should be further evaluated, potentially including head imaging. Patients with brain tumors frequently note worsening of headache at night. Headaches in sleep are discussed in greater detail in Chapter 98.

SYSTEMIC FEATURES

Recognition of the connection of good sleep to good health continues to expand the current understanding of the importance of sleep. Sleep plays a role in endocrine regulation, weight maintenance, and metabolism and has been hypothesized to improve neuronal network proficiency and function. Therefore sleep disorders may influence both regulatory processes and compensatory mechanisms.

Sleep disorders may influence systemic disorders by three general mechanisms: (1) they may directly cause the primary physiologic changes that result in systemic disease; (2) they may exacerbate a preexisting disorder by altering a normal compensatory mechanism; or (3) they may be the hallmark symptom of the systemic disease, sharing a common pathophysiologic mechanism. Sleep disorders may result in systemic manifestations, as noted in sleep-related breathing disorders. Sleep-related disordered breathing can result in a variety of vascular and autonomic changes that increase the likelihood of hypertension and other vascular disorders.

In patients with high-risk disorders, including those with hypertension, vascular disease, heart disease, diabetes mellitus, and obesity, the clinical assessment should include specific evaluation for symptoms of sleep dysfunction that may indicate sleep disorders acting as a cofactor. Sleep disorders (e.g., obstructive sleep apnea or hypoventilation) may exacerbate underlying conditions such as hypertension, diabetes mellitus, congestive heart failure, epilepsy, and depression. Thus the clinician must pay attention to aggravation of symptoms of medical, neurologic, or psychiatric dysfunction as a clue to disordered sleep. The interplay of sleep and brain function makes symptoms of neurologic and psychiatric disease natural manifestations of sleep dysfunction. This association has been documented in persons with epilepsy as well as in those with symptoms of depression and anxiety. Patients may not be aware of the connection of sleep to their other medical problems or may not place emphasis on their sleep symptoms. Therefore the clinician needs to consider the potential of sleep disturbance as an aggravating factor and to recognize the relationship of systemic findings to sleep disruption. Patients also may present with a systemic illness that may share a common pathophysiologic mechanism with a sleep disorder. In persons with anemia and restless legs, iron deficiency may be the common link, so appropriate treatment of the underlying cause may address the symptoms of both. Finally, the sleep symptom may be the hallmark of another process, as observed with REM sleep behavior disorder preceding the development of other neurologic disorders such as Parkinson disease. Although these symptoms may not be considered the typical cardinal manifestations of sleep disorders, their clinical identification does represent an important aspect of diagnosis and may constitute an appropriate entry point for patients into the medical system. Recognition of the systemic manifestations of sleep disorders is an important step in understanding the full relationship of sleep to the body.

PEDIATRIC CARDINAL MANIFESTATIONS

Symptoms of sleep disorders in children can be strikingly different from those in adults and are likely to be overlooked or misinterpreted.[20] Moreover, disruption of family dynamics, psychosocial factors that influence the family unit, and temperamental differences between parent-child pairs may be reported as childhood sleep disorders. Sleep-onset associations, cultural norms, and parental expectations can influence the perception of sleep problems in infants and toddlers. Children can present with a range of physical and behavioral manifestations. In young children, sleep disturbances may manifest as poor growth, learning difficulties, persistent fussiness, inconsolability, or increased oppositional behavior. School-aged children may exhibit suboptimal academic performance, inattention or hyperactivity, or daydreaming behavior in sedentary settings. Adolescents typically fall asleep in class and may present with affective symptoms that need to be differentiated from primary psychiatric disorders. In children with symptoms of RLS (often reported as "growing pains"), a strong family history often is present, suggesting a hereditary predisposition with age-dependent expression. In all pediatric age groups, unrefreshing nocturnal sleep often is a clue to sleep disturbance. Although many of these symptoms are nonspecific, the clinician must be aware of the potential role of sleep dysfunction in the genesis of the symptoms.

CLINICAL PEARL

Patients may present with initial complaints that may seem unrelated to sleep, such as morning headache, "fogginess" during the day, or elevated blood pressure. The patient initially may not have any concerns about his or her sleep or may not even be aware of sleep symptoms. Clarifying functional questions (e.g., asking patients to describe how they feel during certain activities such as on getting up in the morning or after lunch) may give additional clues. Also, family members may be aware of sleep issues before the patient notices a problem, thus lending insight into the impact of sleep on the clinical picture.

SUMMARY

Sleep provides the benchmark for many aspects of daily life. The restorative powers of sleep improve baseline status in wakefulness and maximize the ability to attain higher levels of functioning, whereas poor sleep has a negative impact on health, sense of well-being, and performance. Obvious manifestations of poor sleep, such as daytime sleepiness, insomnia, and sleep-related events, are the hallmark signs indicating the need for further investigation. A proper understanding of the intricate relationship of sleep and health must go beyond the most apparent manifestations of sleepiness. The astute clinician should be alert to less obvious and discrete signs. Predominance of negative over positive memories, trouble with creative solutions, obesity, and poor healing all may suggest sleep disruption. In both research and clinical practice, it is essential to continue to search for more clues that delineate the connection of sleep to health. The insights thus obtained can be expected to expand the clinical recognition of these cardinal manifestations of sleep disorders. Identification of persons who need further evaluation and application of appropriate therapies, as described in the following chapters of Section 8, will lead to optimal management of individual patients and also address the societal and public health issues associated with these disorders.

Selected Readings

Alibhai FJ, Tsimakouridze EV, Reitz CJ, et al. Consequences of circadian and sleep disturbances for the cardiovascular system. *Can J Cardiol* 2015;**31**:860–72.

American Academy of Sleep Medicine. *International classification of sleep disorders*. 3rd ed. Darien (Ill.): American Academy of Sleep Medicine; 2014.

Baumann CR, Mignot E, Lammers GJ, et al. Challenges in diagnosing narcolepsy without cataplexy: a consensus statement. *Sleep* 2014;**37**:1035–42.

Boeve BF, Silber MH, Ferman TJ, et al. Clinicopathologic correlations in 172 cases of rapid eye movement sleep behavior disorder with or without a coexisting neurologic disorder. *Sleep Med* 2013;**14**:754–62.

Buysse DJ. Insomnia. *JAMA* 2013;**309**:706–16.

Hoban TF. Sleep disorders in children. *Ann N Y Acad Sci* 2010;**1184**:1–14.

Howell MJ. Parasomnias: an updated review. *Neurother* 2012;**9**:753–75.

Ioachimescu OC, Collop NA. Sleep-disordered breathing. *Neurol Clin* 2012;**30**:1095–136.

Leschziner G, Gringras P. Restless legs syndrome. *BMJ* 2012;**344**:e3056.

Mannarino MR, Di Filippo F, Pirro M. Obstructive sleep apnea syndrome. *Eur J Intern Med* 2012;**23**:586–93.

Picchietti DL, Bruni O, de Weerd A, et al. Pediatric restless legs syndrome diagnostic criteria: an update by the International Restless Legs Syndrome Study Group. *Sleep Med* 2013;**14**:1253–9.

Zeitzer JM. Control of sleep and wakefulness in health and disease. *Prog Mol Biol Transl Sci* 2013;**119**:137–54.

A complete reference list can be found online at ExpertConsult.com.

Chapter 59

Physical Examination in Sleep Medicine

Alon Y. Avidan; Meir Kryger

Chapter Highlights

- Because many patients who present with sleep disorders have sleep-disordered breathing, assessment of vital signs (heart rate, blood pressure, pulse rate), along with height and weight to calculate the body mass index (BMI), is key.
- Head and neck physical examination is critical for evaluation of any directly visualized anatomic factors that could hinder airflow at the level of the upper airway. This chapter reviews the Mallampati and Friedman classification systems used to assess upper airway patency.
- In addition to increased BMI, other factors that predispose to sleep apnea include increased neck circumference, macroglossia, retrognathia, tonsillar adenoid hypertrophy, overjet, and decreased cricomental space. This chapter provides some specific examples of how these factors are measured and contribute to sleep apnea.
- Specific phenotypes that contribute to sleep-disordered breathing include systemic diseases such as systemic amyloidosis and mucopolysaccharidosis, which lead to abnormal upper airway tissue infiltration and airway restriction.
- Craniofacial disorders related to Down syndrome, acromegaly, and primary mandibular deficiency result in abnormal reduction of upper airway size and contribute directly to obstructive sleep apnea.
- Facial and body habitus phenotype in specific endocrinopathies such Graves disease, Cushing disease, and polycystic ovarian syndrome can predispose patients to both sleep apnea and insomnia and should prompt the clinician to review for these complaints in the appropriate patients.
- Patients with neurologic disorders have significant sleep comorbidities. For example, those with neuromuscular disorders and motor neuron disease are especially vulnerable to nocturnal hypoventilation. Specific neurologic signs such as bulbar weakness, Gower maneuver from sitting to standing, and hypophonia are critical and should prompt the clinician to ask about breathing patterns, sleepiness, and nighttime sleep disruption.
- Patients with parasomnias and nocturnal seizures may present with unexplained bruises, ecchymosis, and nonspecific injuries the next day.
- Symptoms of rapid eye movement sleep behavior disorder often precede the onset of Parkinson disease. Patients with Parkinson disease present with neurologic examination findings of bradykinesia, cogwheel rigidity, masked facies, and resting tremor.
- Patients with bruxism, when severe, often demonstrate dental consequences, including teeth fractures and masseter muscle hypertrophy.
- Patients with narcolepsy often lack characteristic features on physical examination. However, cataplectic facies is a physical finding that was recently used to describe facial weakness and grimaces among patients with cataplexy in the setting of narcolepsy type 1.

The physical examination of the patient presenting with sleep complaints often provides important supporting information for the diagnosis of a sleep disorder. In this chapter, the examination findings characteristic of the major categories of sleep disorders are described and illustrated. These include findings observed in obstructive sleep apnea (OSA), central sleep apnea, hypoventilation syndromes, narcolepsy, Willis-Ekbom disease (WED), parasomnias, and bruxism.

After obtaining the narrative medical history, the sleep care provider performs the physical examination, a key and necessary element in evaluating patients with sleep disorders. The examination may provide important clues that lead to elucidation of the etiology and pathophysiology of the sleep disorder. These will help guide the clinician in determining what diagnostic tests will be ordered, what comorbidities require management, and ultimately what therapy will be

employed. The sleep physical examination is a critical component of monitoring outcome of treatment of many sleep disorders. See also Chapter 143, Oropharyngeal Growth and Skeletal Malformations.

SLEEP APNEA

OSA is associated with multiple anatomic and physical risk factors. Some of these require elaborate measurements of nasopharyngeal anatomy using fiberoptic visualization or cephalometric radiographic techniques, whereas others measure changes in response to maneuvers. The most commonly used signs are static, anthropometric measurements from simple examination of oropharyngeal and craniofacial structure.[1] However, at the time of the initial evaluation of the sleep apnea patient, the main ones are obesity, as reflected by elevated body mass index (BMI), and increased neck circumference.[2]

Figure 59-1 summarizes the key anatomic changes that result from increasing age and BMI. The airways become restricted, and the soft palate, which becomes longer and thicker, is now closer to the posterior and lateral pharyngeal walls, restricting the retropalatal space even further.[2] Increased adipose volume in the floor of the mouth displaces the tongue superiorly and posteriorly, thus decreasing the retroglossal airway space. Chronic allergic rhinitis expands the turbinate tissue, leading to diminished intranasal airway space (see Figure 59-1). The spread of adipose volume in the submental triangle and in the supraplatysmal regions produces the so-called double-chin appearance with a full neck phenotype.[2]

Anthropometric Measurements in Patients with Suspected Sleep Apnea

Obesity, as measurement of BMI, is strongly associated with OSA. However, recent data indicate that general body adiposity and regional adiposity are also important risk factors in the evolution of OSA. Besides BMI, other anthropometric obesity indexes, such as increased waist circumference and neck circumference, are significant risk factors for the evolution of OSA.[3]

Body Mass Index Calculations

The following BMI criteria are used to quantitate weight phenotype[4]: Patients with a BMI below 18.5 are considered underweight, those between 18.5 and 24.9 are classified as normal weight, those between 25.0 and 29.9 are classified a overweight, and those with a BMI higher than 30.0 are obese (Table 59-1).

Figure 59-2 illustrates the obesity phenotype in two brothers. Visceral fat accumulation, as depicted in Figures 59-3 and 59-4, is an important risk indicator for sleep apnea, particularly in male patients.[1,2,4] Expansion of regional fat deposition is believed to compromise airway space, whereas visceral or

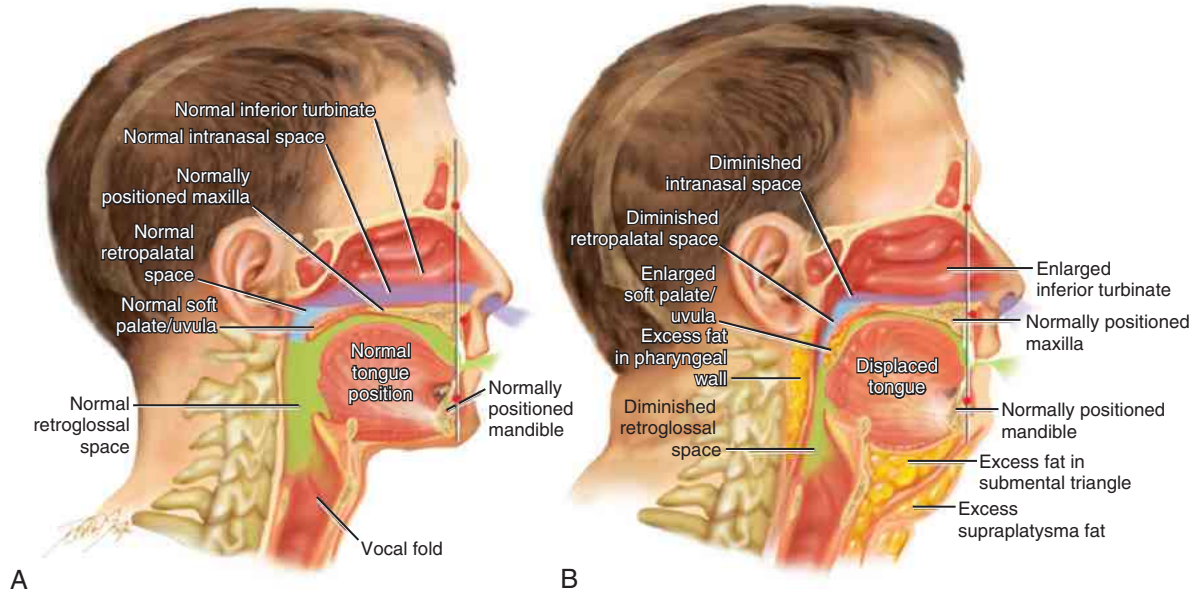


Figure 59-1 Normal and abnormal airway anatomy. **A,** This illustrated midline sagittal cross-section of the head and neck depicts the normal upper airway and maxillofacial spaces and anatomy of a healthy, normal-weight 20-year-old man. The patient has a normal upper and lower facial skeleton and normal soft tissue indicators (soft palate, tongue, tonsils, and adenoids), without any compromise of the intranasal cavity. **B,** With increasing age and weight gain, the same individual, three decades later has an elevated body mass index. Although the anatomy of the upper and lower facial skeleton remains fixed without any changes, fatty tissue consisting of adipose cells has expanded and infiltrated the crevices and space in the upper airway. Particularly compromised are the retropharyngeal and the lateral pharyngeal tissues, the soft palate, and the floor of the mouth, culminating in restricted airflow. At age 20 years, the patient had normal upper airway space (the intranasal, retropalatal, and the retroglossal sites were all well visualized and with appropriate space for air flow to proceed smoothly and unimpeded. At age 50 years, he has developed obstructive sleep apnea: The normal airspace (*green*) is severely compromised because of restriction of the upper airway, intranasal space, and retropalatal and the retroglossal spaces. A *perfect storm* indeed. (From Posnick JC. Obstructive sleep apnea: evaluation and treatment. In: Posnick JC, editor. *Orthognathic surgery: principles and practice.* Philadelphia: Elsevier; 2014. p. 992–1058.)

abdominal obesity reduces lung volume and therefore caudal traction on the pharynx.[5]

Neck circumference at the superior border of the cricothyroid membrane can be measured to evaluate excessive adiposity in the upper body. This measurement is performed with the patient in the upright position (Figure 59-5). Recent data suggest that in adults with metabolic syndrome, measurement of neck circumference is associated with OSA and should be considered in the definition of metabolic syndrome (Figure 59-6).[6] In pediatric and adolescent patients, neck circumference percentile, particularly that greater than the 95th percentile for age and sex, may be an additional screening tool for OSA.[7] In adults, having a large neck circumference in the context of OSA can predict difficult intubations in the anesthesia setting[8] and has been documented in several metabolic derangements, as in the patients presented in Figure 59-7.[9]

Table 59-1 Body Mass Index Calculations

Measurement Units	Formula and Calculation
Kilograms and meters (or centimeters)	Formula: Weight (kg)/height (m^2) Using the metric system, the formula for body mass index (BMI) is weight (expressed in kilograms) divided by height in meters squared. Because height is commonly measured in centimeters, divide height in centimeters by 100 to obtain height in meters For example: Weight = 68 kg, height = 165 cm (1.65 m) Calculation: $68 \div 1.65^2 = 24.98$
Pounds and inches	Formula: weight (lb)/height (in^2) × 703 Calculate BMI by dividing weight in pounds (lb) by height in inches squared (in^2) and multiplying by a conversion factor of 703 Example: Weight = 150 lb, height = 5′5″ (65″) Calculation: $[150 \div 65^2] \times 703 = 24.96$

However, many patients with OSA are not obese but may exhibit reduced oropharyngeal airspace, retrognathia, or micrognathia, which put patients are risk for OSA (Figure 59-8). In contrast, central sleep apnea usually presents with abnormalities reflective of impaired respiratory effort, including the manifestations of heart failure, central nervous system (CNS) disease, or neuromuscular disease. Hypoventilation may be secondary to obesity but may also reflect pulmonary disease or neuromuscular and chest wall disorders. We review the manifestations of sleep apnea on the basis of anatomic site.

Overall Inspection

As noted in the previous section, sleep apnea often presents in association with obesity, which increases the prevalence 10-fold (20% to 40%).[10] Obesity and, in particular, the central type of obesity (see Figure 59-4) are significant risk factors for OSA.[11] They impose increased pharyngeal collapsibility through mechanical compression of the pharyngeal soft

Figure 59-2 Obesity is strongly associated by body mass index (BMI). The example depicts two brothers with sleep apnea with elevated BMI, in the morbidly obese range. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-3, A.)

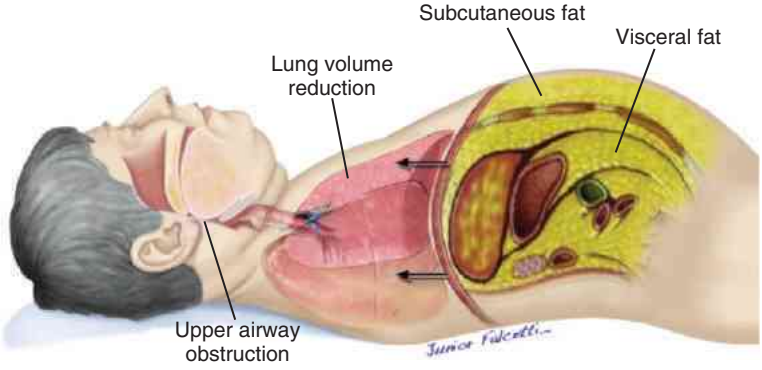


Figure 59-3 The contribution of obesity to obstructive sleep apnea. The illustration depicts the principle anatomic factors that place obese patients at significant risk for obstructive sleep apnea. (From Drager L, Togeiro SM, Polotsky VY, Lorenzi-Filho G. Obstructive sleep apnea: a cardiometabolic risk in obesity and the metabolic syndrome. *J Am Coll Cardiol* 2013;62[7]:569–576.)

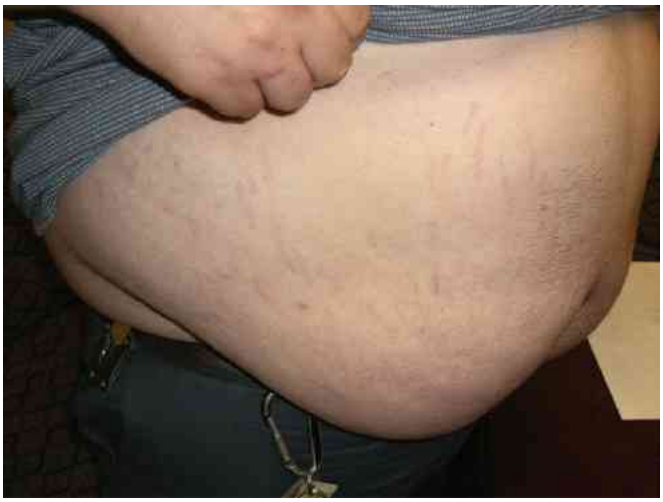


Figure 59-4 Central obesity in obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-42.)

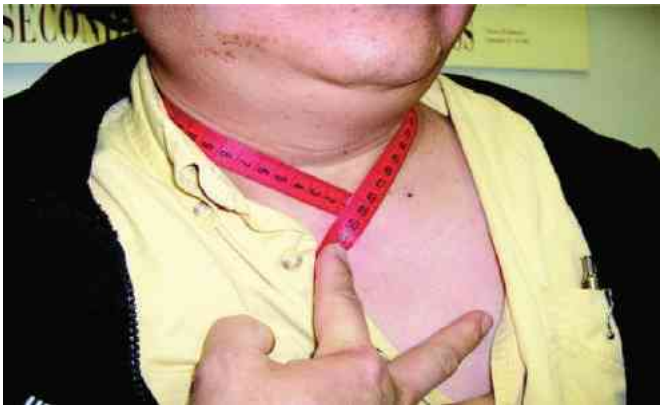


Figure 59-5 Measuring neck collar size. Neck circumference (NC) values of 17 inches or greater in men and 16 inches or greater in women are strongly correlated with the risk for obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-40.)

tissues and decreased lung volume through CNS-acting signaling proteins (adipokines) that may alter airway neuromuscular control.[11,12] OSA may independently predispose individuals to worsening obesity as a result of sleep deprivation, hypersomnia, and disrupted metabolism.[13]

Sleep apnea is also associated with endocrinopathies such as hypothyroidism[14,15] and acromegaly.[16] Hypothyroidism is a known cause of secondary OSA; oropharyngeal airway myopathy, edema, and obesity predispose patients to upper airway collapse and obstruction. Acromegaly as depicted in Figure 59-9 results from excessive growth hormone, resulting in enlarged growth of the craniofacial bones, enlargement of the tongue (macroglossia as shown in Figure 59-10), and thickening and enlargement of the laryngeal region; all of these factors can contribute to upper airway obstruction.[17] Goiter, which is associated with acromegaly and hypothyroidism as well as a euthyroid state,[18] can contribute to OSA (Figure 59-11). Patients with Down syndrome (Figure 59-12) regularly experience snoring and obstructive apneas, two common manifestations of upper airway obstruction in this condition, which independently predict neurocognitive impairment.[19] Recent data show significant prevalence of OSA in Down

Figure 59-6 The typical facial features of a patient with obstructive sleep apnea. Notice the width of the neck. (From Venn PJH. Obstructive sleep apnoea and anaesthesia. *Anaesth Intens Care Med* 2014;12:313–8.)

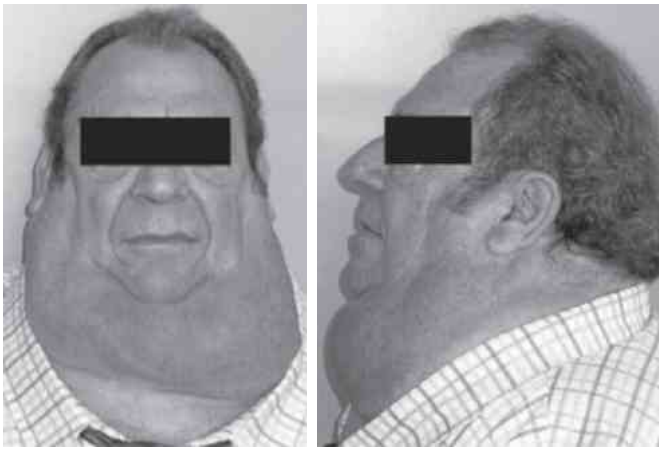


Figure 59-7 Significantly large neck size due to symmetrical accumulation of adipose tissue in a patient with multiple symmetrical lipomatosis, a conditions characterized by a diffuse, symmetrical accumulation of adipose tissue, primarily around the neck. (From Esteban Julvez L, Perello Aragones S, Aguilar Bargallo X. Sleep apnea-hypopnea syndrome and multiple symmetrical lipomatosis. *Arch Bronconeumol* 2013;49:86–7.)

syndrome patients, conferred through several factors, including craniofacial anatomy, high BMI, adenotonsillar hypertrophy, and muscle hypotonia.[20]

Metabolic derangement such as deposition disorders, including mucopolysaccharidosis (Figure 59-13) and amyloidosis (Figure 59-14),[21] are strongly correlated to OSA. In fact, in patients with mucopolysaccharidosis the prevalence of OSA syndrome can be as high as 70%.[22]

Specific endocrinopathies, in particular polycystic ovarian syndrome, are extremely common among women of reproductive age but often go undiagnosed.[23] Polycystic ovarian

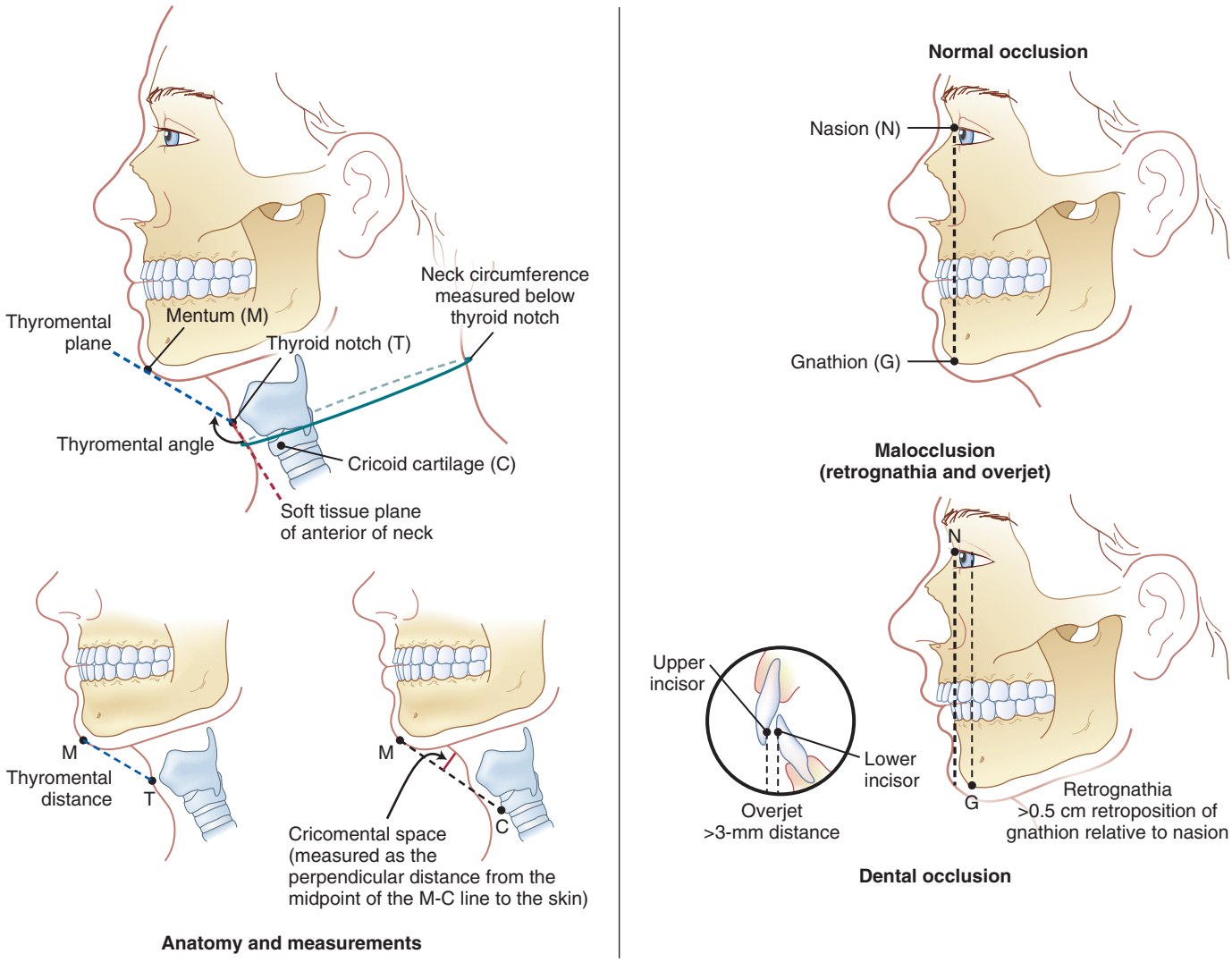


Figure 59-8 Anatomy and surface measurements in the assessment of a patient with suspected obstructive sleep apnea (OSA). Retrognathia, overjet, and reduced cricomental space are key craniofacial properties that are predictive of OSA. (From Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea? The rational clinical examination systematic review. *JAMA* 2013;310:731–41.)

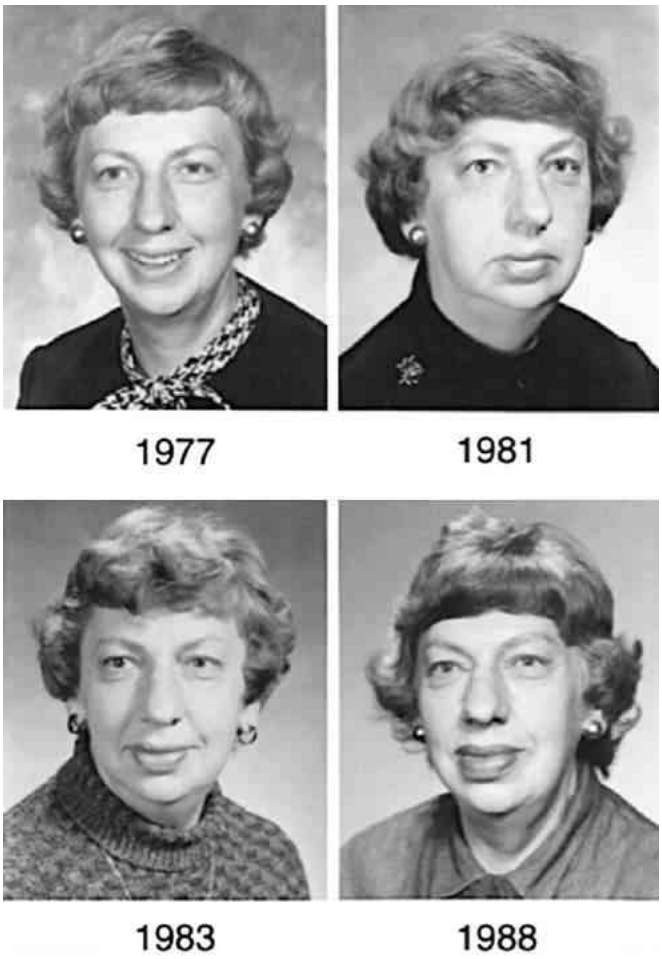


Figure 59-9 Progressive change in facial features in a patient with acromegaly. The onset of physical changes is sometime insidious, and patients may not present with specific complaints relating directly to these distinguishing signs of acromegaly. However, patients may be more likely to present with symptoms referred to other conditions such as diabetes, hypertension, and obstructive sleep apnea. At the advanced stages of the condition, patients exhibit more dramatic physical characteristics, such as enlarged hands, feet, lips, and tongue; prominent supraorbital ridges; and lower jaw protrusion. (From Molitch ME. Clinical manifestations of acromegaly. *Endocrinol Metab Clin North Am* 1992;21[3]:597–614.)

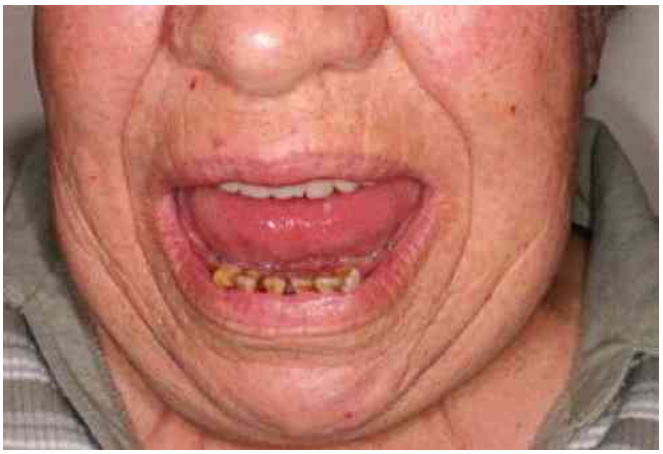


Figure 59-10 Patient with acromegaly showing the coarse facial features, macroglossia, and interdental separation typically seen in this condition, which lead to airway restriction and contribute to the development of obstructive sleep apnea. (From Burke G. Endocrine disease. In: Sprout C, Burke G, McGurk M, editors. *Essential human disease for dentists*. Edinburgh: Churchill Livingstone; 2006. p. 99–119.)

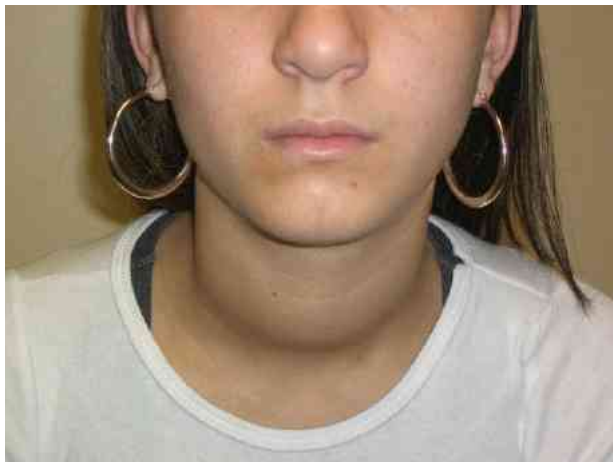


Figure 59-11 Goiter. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 15.1-8, B.)

Figure 59-12 Two patients with the characteristic phenotype of Down syndrome. Contributing factors to obstructive sleep apnea in Down syndrome include alteration in craniofacial anatomy, macroglossia, adenotonsillar hypertrophy, and muscle hypotonia. (Courtesy Dr. Meir H. Kryger.)

Figure 59-13 Hunter syndrome (mucopolysaccharidosis type II) depicting significant macroglossia, a risk factor for OSA. In this 8-year-old boy with Hunter syndrome, infiltration of the macroglossia can be seen. Other features include macrocephaly, coarse hair, abnormally short neck, hairy face, puffy eyelids, depressed nasal bridge, upturned nose, full lips, and thick skin texture. (From Chou W-C, Weng C-Y, Lin S-P, Chu S-Y. Postenzyme replacement therapy era for type 2 mucopolysaccharidosis. *Tzu Chi Med J* 2013;25:128–9.)

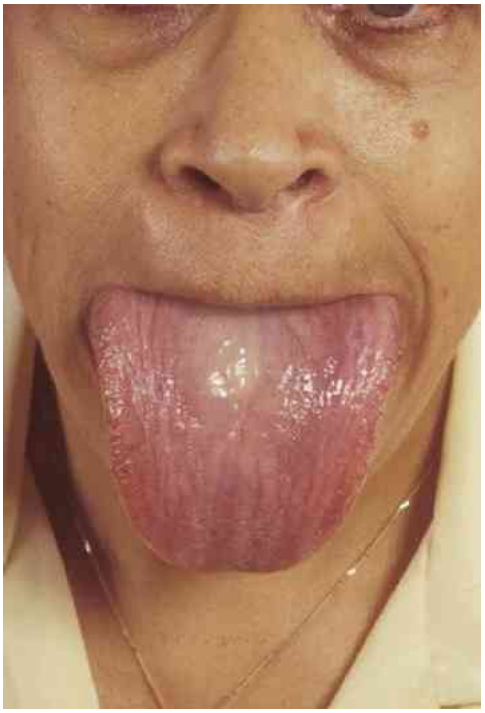


Figure 59-14 Profound enlargement of the tongue (macroglossia) as a result of amyloid infiltration. The patient had severe obstructive sleep apnea. The tongue fills the oral cavity completely, contributing to profound hypopharyngeal and oropharyngeal airway blockade. (From Hoffman R, Benz EJ, Silberstein LE, et al. *Hematology: diagnosis and treatment*. Philadelphia: Elsevier Science; 2013. p. 1352, Figure 87-3.)

syndrome is associated with metabolic syndrome and carries a greatly increased risk for obesity with metabolic syndrome, OSA, impaired glucose tolerance, type 2 diabetes mellitus, and cardiovascular disease (Figure 59-15).[23]

Craniofacial Factors

As summarized in Figure 59-8, cephalometric measurements reveal that subjects with OSA have significant changes in the size and position of the soft palate and uvula, volume and position of the tongue, hyoid position, and mandibulomaxillary protrusion compared with controls. Mandibular retrognathia (Figure 59-16) and micrognathia (Figure 59-17), which cause the tongue to rest in a more superior and posterior position, impinging on the upper airway, can be detected on examination, especially by observing the patient from the side. As seen in Figure 59-17, *B*, the cricomentalis space defined by the distance between the neck and the bisection of a line from the chin to the cricoid membrane, when the head is in a neutral position, is extremely limited.[24] A scalloped tongue (Figure 59-18) may accompany micrognathia. Men with retrognathia or micrognathia may grow a beard to compensate for this anatomic variant. Crowded teeth (Figure 59-19) and overjet (Figure 59-20), with the mandibular teeth excessively posterior to the maxillary teeth (Figure 59-20, *B*), often accompany retrognathia or micrognathia. Figure 59-21 depicts the global consequences of primary mandibular insufficiency on the patency of the upper airways, leading to compromised retronasal, retropalatal, and retroglossal spaces.[2] Commonly encountered craniofacial features predisposing to sleep apnea consist of mandibular deficiency syndrome, an inferiorly placed hyoid bone relative to the mandibular plane, narrowing of the posterior airspace, and elongation of the soft

Figure 59-15 A 31-year-old woman with polycystic ovary syndrome. Shown here is the particular phenotype of increased central fat distribution among these patients. (From Magnotti M, Futterweit W. Obesity and the polycystic ovary syndrome. *Med Clin North Am* 2007;91:1151–68, ix-x.)

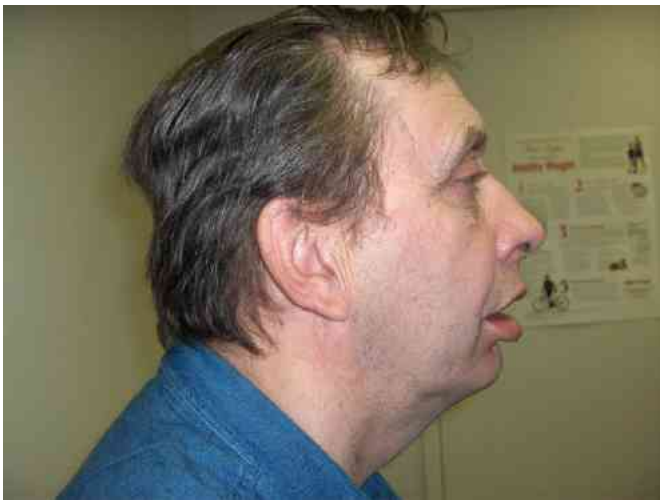


Figure 59-16 Mandibular retrognathia contributing to obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-16.)

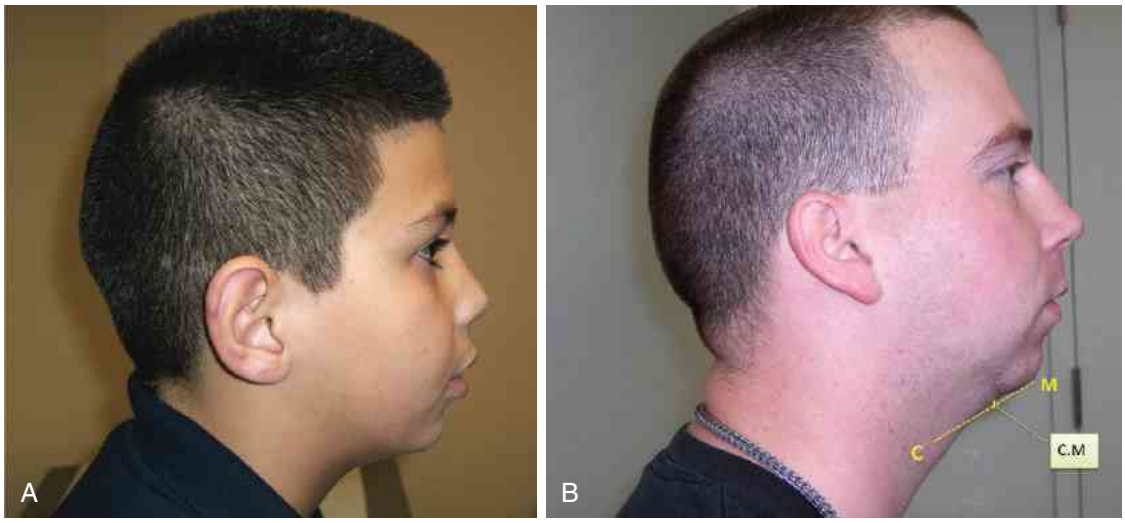


Figure 59-17 A child **(A)** and an adult **(B)** with significant mandibular micrognathia contributing to obstructive sleep apnea. The cricomentalis (C–M) space (as delineated by the *dotted line*) is severely reduced in the adult with mandibular micrognathia. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-12, A.)

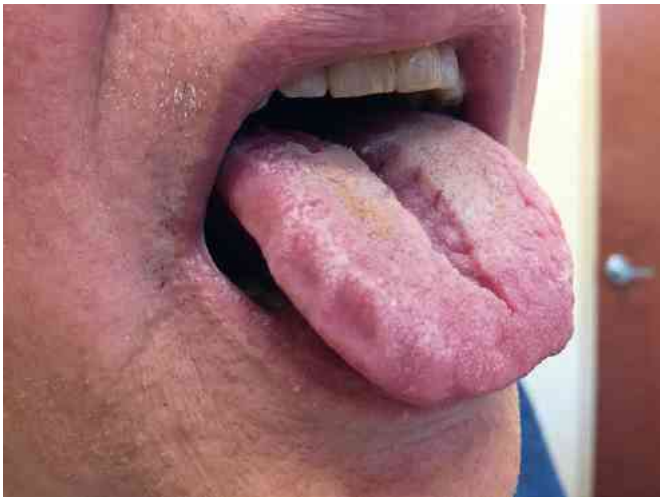


Figure 59-18 Scalloped tongue in a patient with obstructive sleep apnea and micrognathia. The scalloping and furrow result from the tongue's pressing against teeth (especially on the right side). In addition, the tongue is atrophied, raising the possibility of iron or vitamin B_{12} deficiency. (Courtesy Dr. Meir H. Kryger.)

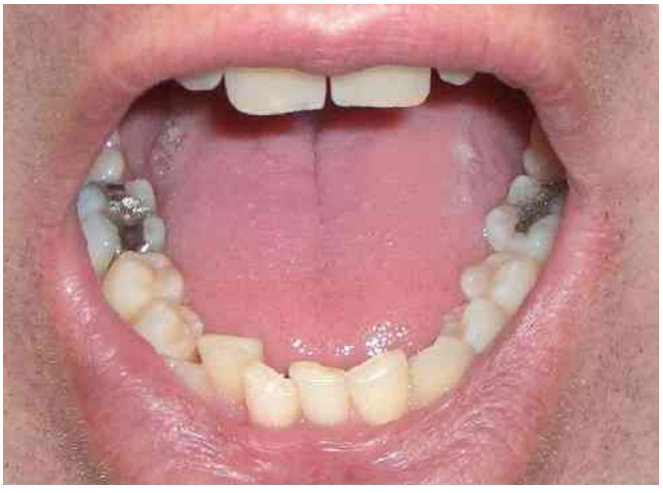


Figure 59-19 Crowded teeth indicate a small mandible, contributing to obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-12.)

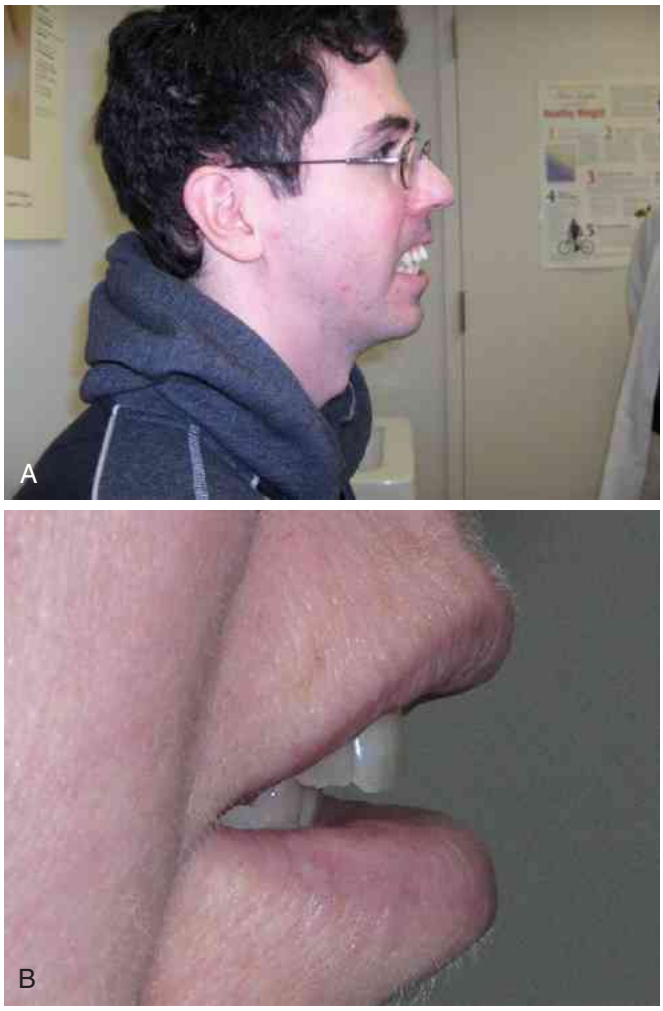


Figure 59-20 Overjet contributing to obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-14.)

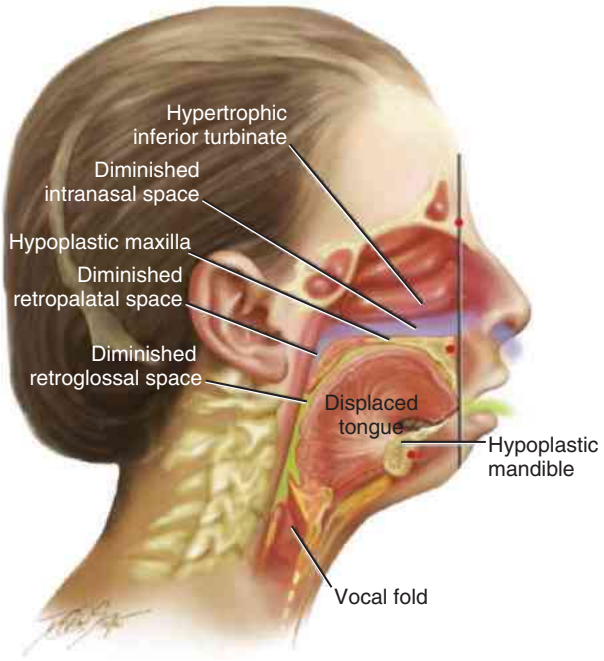


Figure 59-21 An illustration of a sagittal cross-sectional head and neck view from a 16-year-old patient with primary mandibular deficiency and overjet malocclusion predisposing to obstructive sleep apnea. (Courtesy Dr. Meir H. Kryger.)

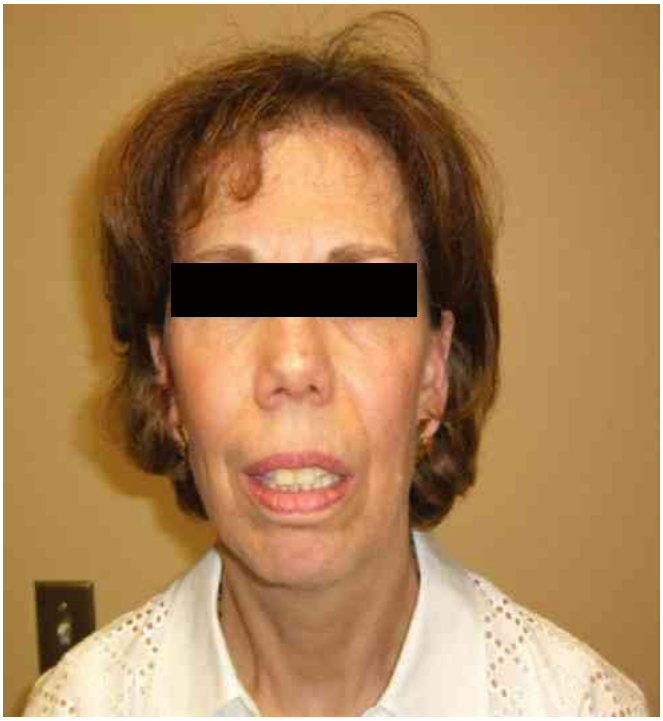


Figure 59-22 A patient with the long face syndrome, a well-established risk for obstructive sleep apnea, that is conferred through an increase of anterior facial height generally associated with retrognathia. (Courtesy Dr. Meir H. Kryger.)

palate.[25] In addition, marfanoid habitus, including the long face phenotype (Figure 59-22), leads to upper airway restriction, thereby predisposing to OSA.[26] Indeed when the well-established role of obesity in the development of OSA is taken into account, a model of OSA emerges in which the degree of craniofacial abnormalities determines the extent of obesity required to produce OSA in a given individual.

Racial differences in cephalometric properties probably play a major role in conferring risk for OSA in the absence of obesity. For example, in Chinese patients with OSA, a more retropositioned mandible was associated with more severe OSA, after controlling for obesity.[27] In Japanese patients with OSA, micrognathia was a major risk factor.[28] Children and adults with Down syndrome (see Figure 59-12) frequently have sleep apnea most likely related to a combination of craniofacial abnormality and macroglossia.

Patients with OSA have an increased pharyngeal narrowing ratio, which is defined as a ratio between the airway cross section at the hard palate level and the narrowest cross section from the hard palate to the epiglottis.[29]

Nasal Factors

Examination of the nasal airway should focus on anatomic abnormalities that may contribute to nasal obstruction. These may be congenital, traumatic, infectious, or neoplastic in etiology (Figure 59-23).

Neck Circumference

Increased neck circumference (see Figure 59-6) is an important risk factor for OSA. Patients with a neck circumference

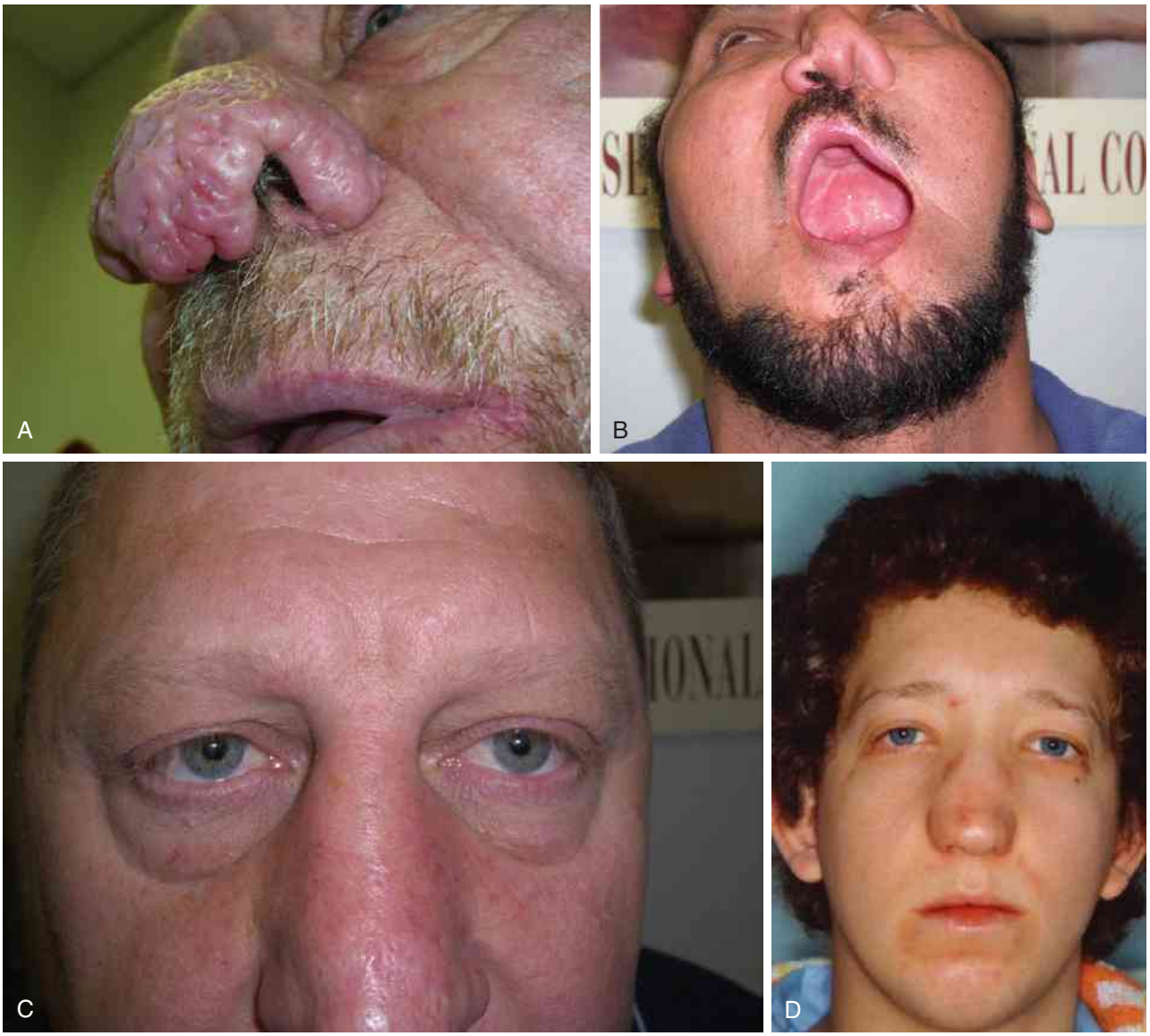


Figure 59-23 Deformity of the nose can be a very important contributor to sleep disordered breathing. **A,** Rhinophyma ("bulbous nose" or "phymatous rosacea") is a nodular hypertrophy characterized by progressive thickening of the nose and leading to compromise of the nasal orifice airflow. **B,** Gun shot wound leading to significant facial injury and nasal collapse, requiring maxillomandibular reconstruction and plastic surgery. **C,** Nasal deviations due to remote fool injury. **D,** Nasal deformity due to the presence of nasal polyps. (**A, B,** and **D,** Courtesy Dr. Meir H. Kryger. **C,** From McGurk M. ENT disorders. In: Sprout C, Burke G, McGurk M, editors. *Essential human disease for dentists.* Edinburgh: Churchill Livingstone; 2006. p. 195–204.)

greater than 48 cm (19.2 inches) have a 20-fold increased risk for OSA.[30]

Examination of the Pharynx

There are two well-established classifications to determine the relation of the tongue to the pharynx. The Mallampati classification was first described as a method for anesthesiologists to predict difficult tracheal intubation (Figure 59-24).[31] The Friedman classification identifies prognostic indicators for successful surgery for sleep-disordered breathing, combining palate position with tonsillar size.[32] The Mallampati classification can be seen in Table 59-2 and Figures 59-25 through 59-28. The Friedman classification is illustrated in Figure 59-29.

Table 59-2 Mallampati Classification

Class I	Soft palate, fauces, uvula, and posterior and anterior pillars are visible (Figure 59-25)
Class II	Soft palate, fauces, and uvula are visible (Figure 59-26)
Class III	Soft palate, fauces, and only base of uvula are visible (Figure 59-27)
Class IV	Soft palate is not visible (Figure 59-28)

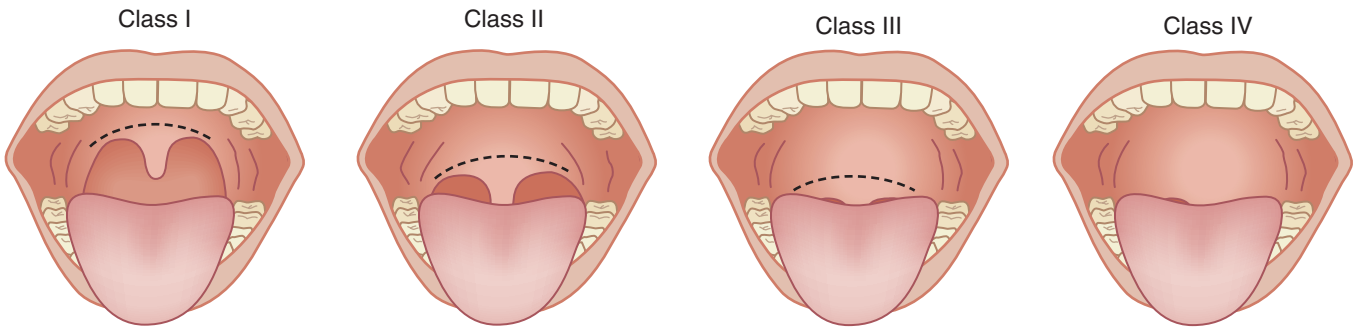


Figure 59-24 The Mallampati classification system is visualized with the tongue protruded, but without the patient phonating. A *modified* form of the Mallampati system is measured with the tongue remaining on the floor of the mouth. The system was initially developed to predict ease of intubation but was later adopted by sleep medicine to help forecast the severity of obstructive sleep apnea in the ambulatory setting. It can also be used to help predict the appropriateness of upper airway surgery in certain patients by delineating the relationship of the various upper airway structures and noting the tongue size in relation to the uvula, tonsils, soft palate, and oropharyngeal wall. The standard for tongue size measurement involved the patient holding his or her head in a neutral position, opening the mouth as wide as possible, and sticking out the tongue. Class I is characterized by direct visualization of the soft palate, uvula, palatine tonsils, and pillars. However, as these structures become obscured, so does the Mallampati class, until only the hard palate is visible (class IV). (From Townsend CM Jr, Beauchamp RD, Evers BM, et al. *Sabiston textbook of surgery*. 19th ed. Philadelphia: Elsevier; 2012.)

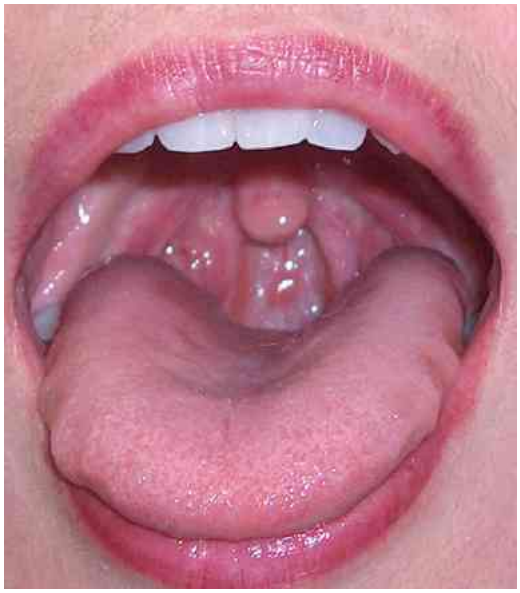


Figure 59-25 Mallampati class I. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-28.)

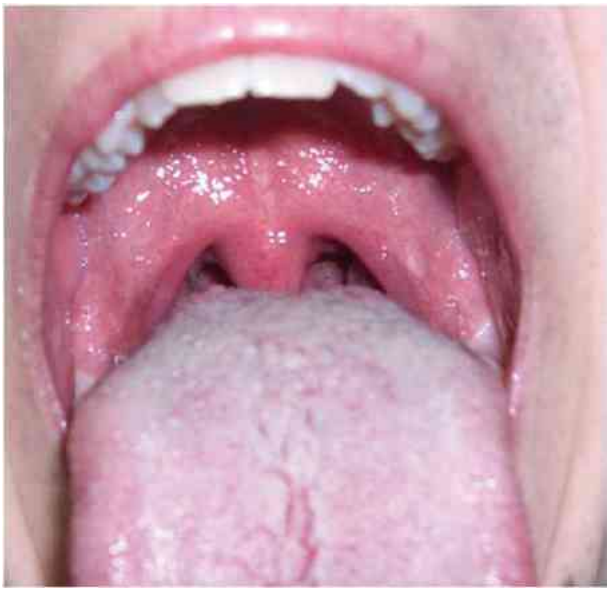


Figure 59-26 Mallampati class II. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-29, A.)

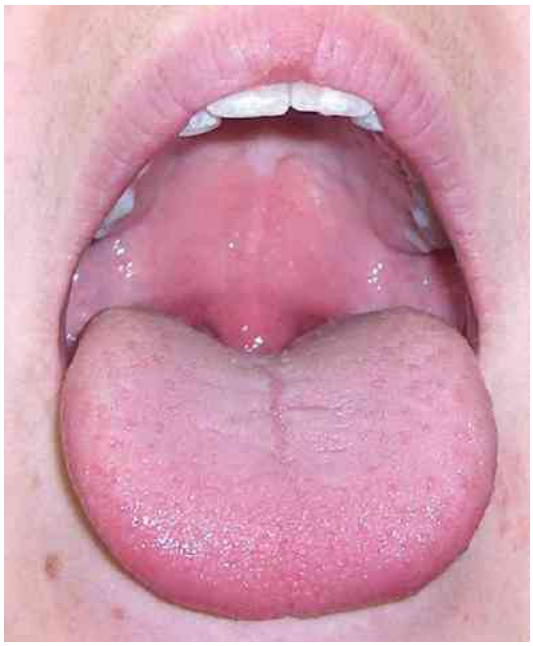


Figure 59-27 Mallampati class III. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-30, A.)

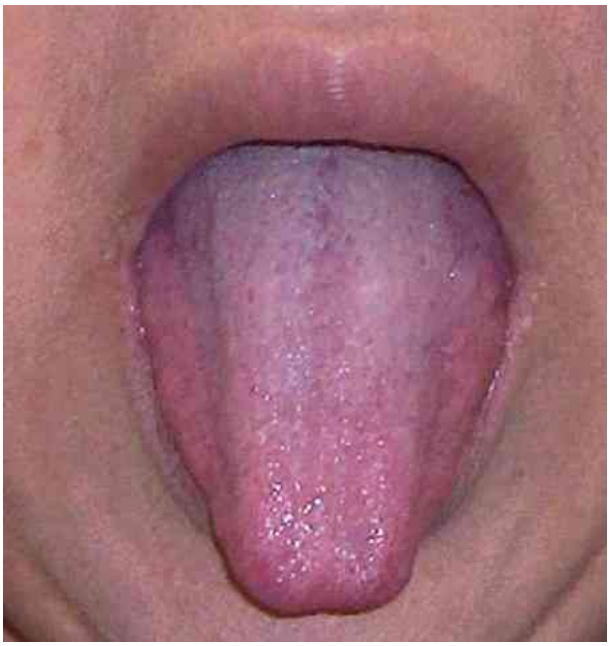


Figure 59-28 Mallampati class IV. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014: Fig. 13.1-31.)

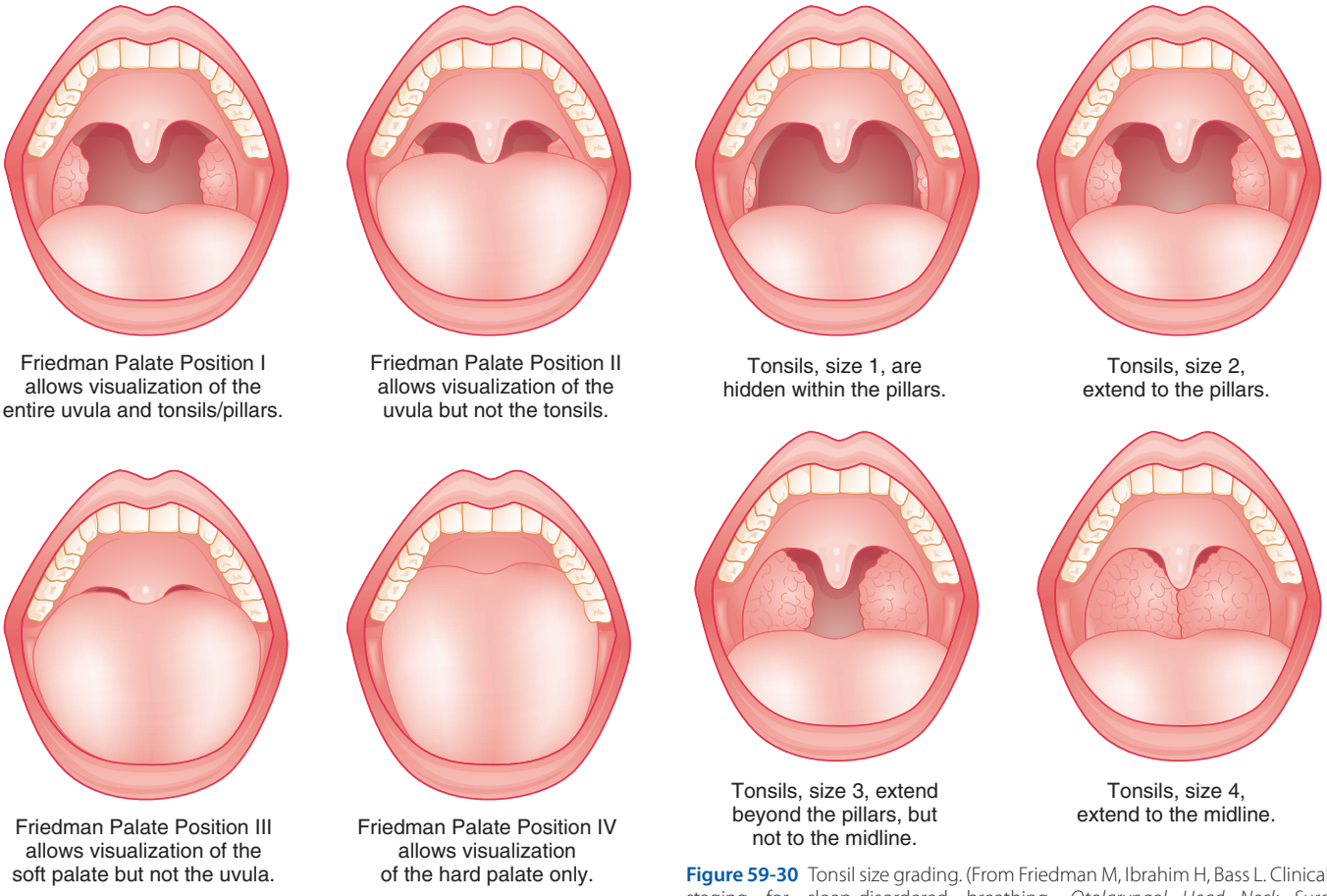


Figure 59-29 Friedman classification. This grading is based on the tongue in a natural position inside the mouth (I). Palate grade I allows the observer to visualize the entire uvula and tonsils or pillars. Palate grade II allows visualization of the uvula but not the tonsils. Palate grade III allows visualization of the soft palate but not the uvula. Palate grade IV allows visualization of the hard palate only. (From Friedman M, Ibrahim H, Bass L. Clinical staging for sleep-disordered breathing. *Otolaryngol Head Neck Surg* 2002;127:13–21.)

Figure 59-30 Tonsil size grading. (From Friedman M, Ibrahim H, Bass L. Clinical staging for sleep-disordered breathing. *Otolaryngol Head Neck Surg* 2002;127:13–21.)

Examination of the Tonsils

Enlarged tonsils and adenoids are a major cause of airway obstruction and sleep apnea in children, but a minority of adults may also have enlargement of these structures contributing to airway obstruction.[33] Adenoids cannot be visualized in a routine physical examination, and the examination of tonsils may require use of a tongue blade. Tonsillar size is graded on a scale of 1 to 4 (Figure 59-30). Children with marked adenotonsillar hypertrophy and nasal obstruction have been noted to have a peculiar "dull" expression (i.e., "adenoid facies"), as shown in Figure 59-31, *A* and *B*. Children with chronic sinus allergies may present with an "allergic salute" sign (see Figure 59-31, *C*).

Neurologic Examination

The neurologic examination may hold important clues to the presence of obstructive or central sleep apnea and hypoventilation syndromes. Features of neuromuscular disease evident on physical examination may indicate these syndromes. For example, progressive muscle atrophy and fasciculations of the hand (Figure 59-32) or tongue may indicate amyotrophic lateral sclerosis. In amyotrophic lateral sclerosis, phrenic nerve dysfunction is common and results in diaphragmatic paralysis, with prominent hypoventilation during rapid eye movement (REM) sleep. In addition, coexisting OSA may occur in amyotrophic lateral sclerosis with bulbar involvement. Weakness of thoracoabdominal or respiratory accessory muscles, often with accompanying kyphoscoliosis, may be observed in poliomyelitis. Postpolio syndrome, muscular dystrophies, myasthenia gravis, and metabolic myopathies may also manifest with weakness of the chest wall musculature[34] and diaphragm weakness. Myasthenia gravis (Figure 59-33) may also involve facial structures, resulting in OSA. Craniofacial abnormalities may occur in myotonic dystrophy (Figure 59-34) or muscular dystrophy; macroglossia may also occur (e.g., Duchenne muscular dystrophy).[35]

Figure 59-35 depicts a patient with facial weakness in the setting of progressive muscular dystrophy. Figure 59-36 shows a patient with myotonic dystrophy with tightness of the muscles (called myotonia), leading to difficulty relaxing certain muscles after using them, such as being able to release grip in a handshake or on a doorknob, or as in the example provided. Figure 59-37 depicts the classic Gower maneuver in Becker muscular dystrophy. Defects in upper airway neuromuscular control in many of the patients with dystrophinopathies play a critical role in sleep apnea pathogenesis, and the sleep care provider must maintain a vigil eye on sleep disturbances in this group of patients.[36]

Finally, obesity (e.g., from steroid use, as in Figure 59-38, or inactivity) may also contribute to sleep apnea in neuromuscular disease.

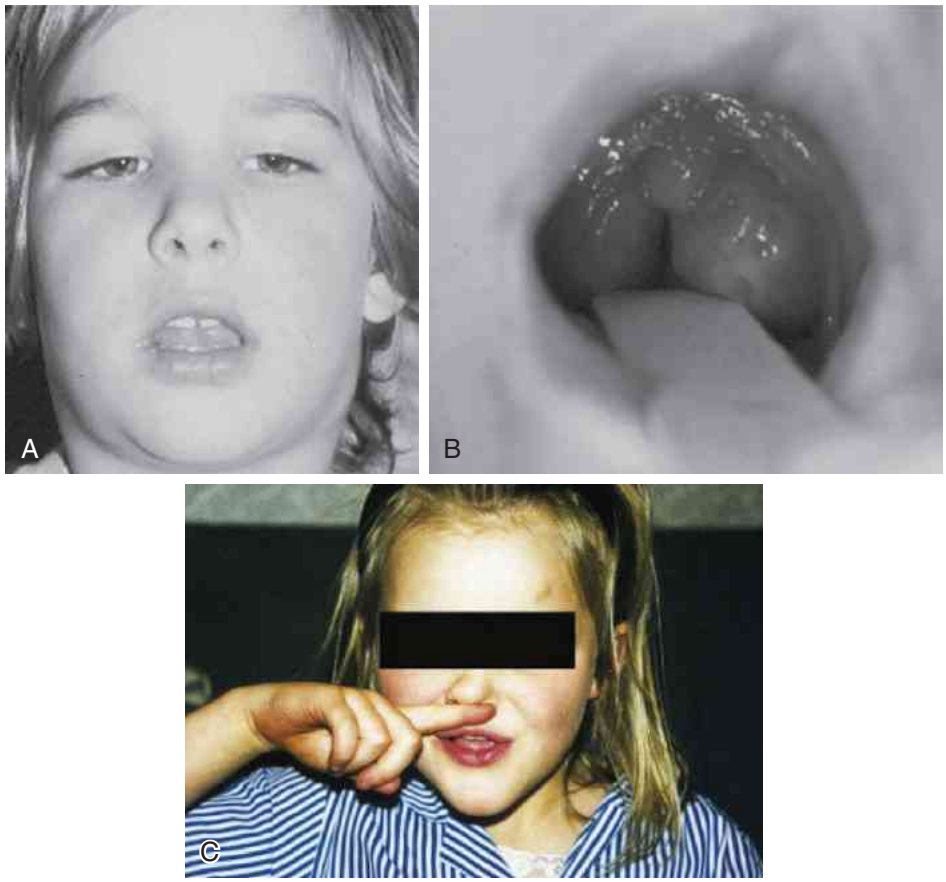


Figure 59-31 A pediatric patient with adenoidal and tonsillar hypertrophy and sinus allergies. **A,** The patient has a "dull expression" of a child with marked adenotonsillar hypertrophy and nasal obstruction (i.e., "adenoid facies"). He must keep his mouth open to breathe and shows signs of fatigue as a result of the result of sleep disruption as a consequence of obstructive sleep apnea. **B,** A severely crowded oropharynx due to tonsillar hypertrophy. **C,** Allergic salute in a patient with chronic allergic rhinitis. (**A** and **B,** From Landsman IS, Werkhaven JA, Motoyama EK. Anesthesia for pediatric otorhinolaryngologic surgery. In: Davis PJ, Cladis FP, Motoyama EK, editors. *Smith's anesthesia for infants and children*. Philadelphia: Mosby; 2011. p. 786–820. **C,** From Scadding GK, Church MK, Borish L. Allergic rhinitis and rhinosinusitis. *Allergy* 2012;203–226.)

Figure 59-32 Hand atrophy (*arrow*) in amyotrophic lateral sclerosis. (From Goldman L, Ausiello DA, editors. *Cecil medicine*. 23rd ed. Philadelphia: Elsevier; 2008.)

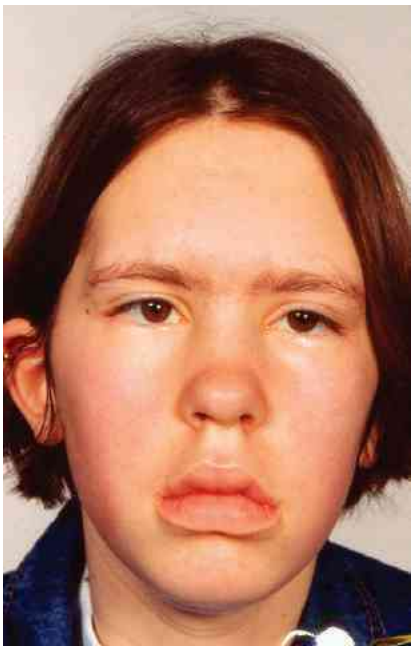


Figure 59-33 Facial muscle weakness in myasthenia gravis. (From Goldman L, Ausiello DA, editors. *Cecil medicine*. 23rd ed. Philadelphia: Elsevier; 2008.)

Cardiopulmonary Examination

The presence of congestive heart failure (see Chapter 129) indicates a high likelihood of central sleep apnea. Peripheral edema (Figure 59-39) is a common finding in patients with obesity-hypoventilation syndrome (as a manifestation of cor pulmonale) and in some patients with OSA who also have left ventricular cardiac failure. Resolution of peripheral edema with treatment correlates with clinical improvement. Chronic obstructive pulmonary disease (see Chapter 111) and asthma (see Chapter 111) are also seen in association with OSA. In patients with cardiopulmonary insufficiency, clubbing of digits and nails may be a cardinal sign (Figure 59-40).

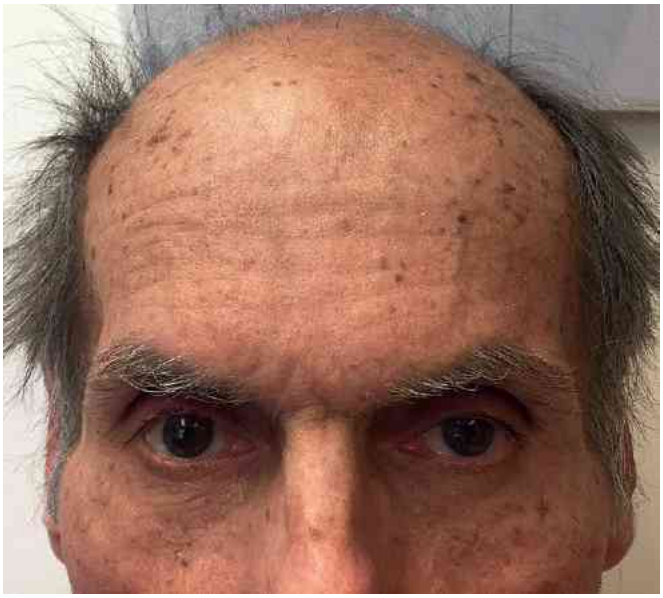


Figure 59-34 Myotonic muscular dystrophy (DM1). Findings in a patient with myotonic muscular dystrophy include wasting of the temporal muscles (shown) and male-pattern baldness that began at an early age. These patients may also have weakness of other facial muscles and mycrognathia. (Courtesy Dr. Meir H. Kryger.)

CENTRAL NERVOUS SYSTEM HYPERSOMNIA

Narcolepsy

Physical findings in patients with narcolepsy are nonspecific and may be subtle, infrequent, and absent during the clinic visit. During cataplectic spells, patients present with muscle atonia, absence of deep tendon reflexes, and decrease in the H-reflex.[37] Cataplexy attacks may range from partial episodes characterized by sagging of the jaw and mild dropping of the head and shoulders, to generalized spells leading to loss of muscle tone with unbuckling of the knee. However, it is rare to encounter cataplexy during the actual clinic visit or physical examination, which makes it difficult to describe during routine clinic visits. In general patients with narcolepsy tend to be obese, with increased predilection to type 2 diabetes mellitus, and have a lower basal metabolism compared with controls.[38,39] Children with obesity and precocious puberty should be screened for narcolepsy and cataplexy.[40]

Narcolepsy related to medical conditions (symptomatic narcolepsy) is seen in disorders such as CNS tumors, head trauma, multiple sclerosis, neurosarcoidosis, acute disseminated encephalomyelitis, CNS vascular disorders, encephalitis, and neurodegeneration.[41] An abnormal neurologic examination can be an important sign that hypersomnia may be due

Figure 59-35 A and **B,** Bilateral facial weakness due to progressive muscular dystrophy. The patient exhibits the classic signs, including bilateral ptosis. Facial weakness involves the orbicularis oculi, orbicularis oris, and zygomaticus muscle, producing the characteristic myopathic facial. Weakness of muscles of the thoracic region often leads to respiratory insufficiency, and many patients also present with bulbar symptoms (dysarthria, dysphagia). (From Laina V, Orlando A. Bilateral facial palsy and oral incompetence due to muscular dystrophy treated with a palmaris longus tendon graft. *J Plast Reconstr Aesthet Surg* 2009;62[11]:e479–81.)

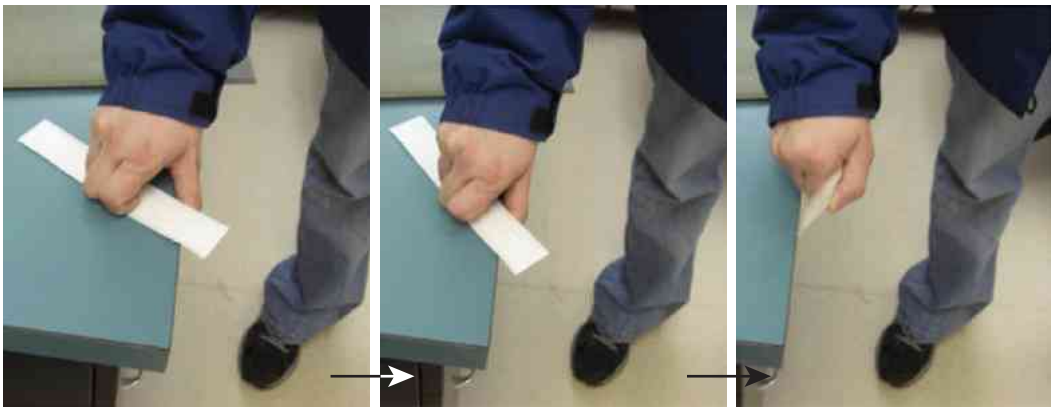


Figure 59-36 Muscular dystrophy. An attempt at grasp by the patient with difficulties relaxing his muscles following the grasp. (Courtesy Dr. Meir H. Kryger.)

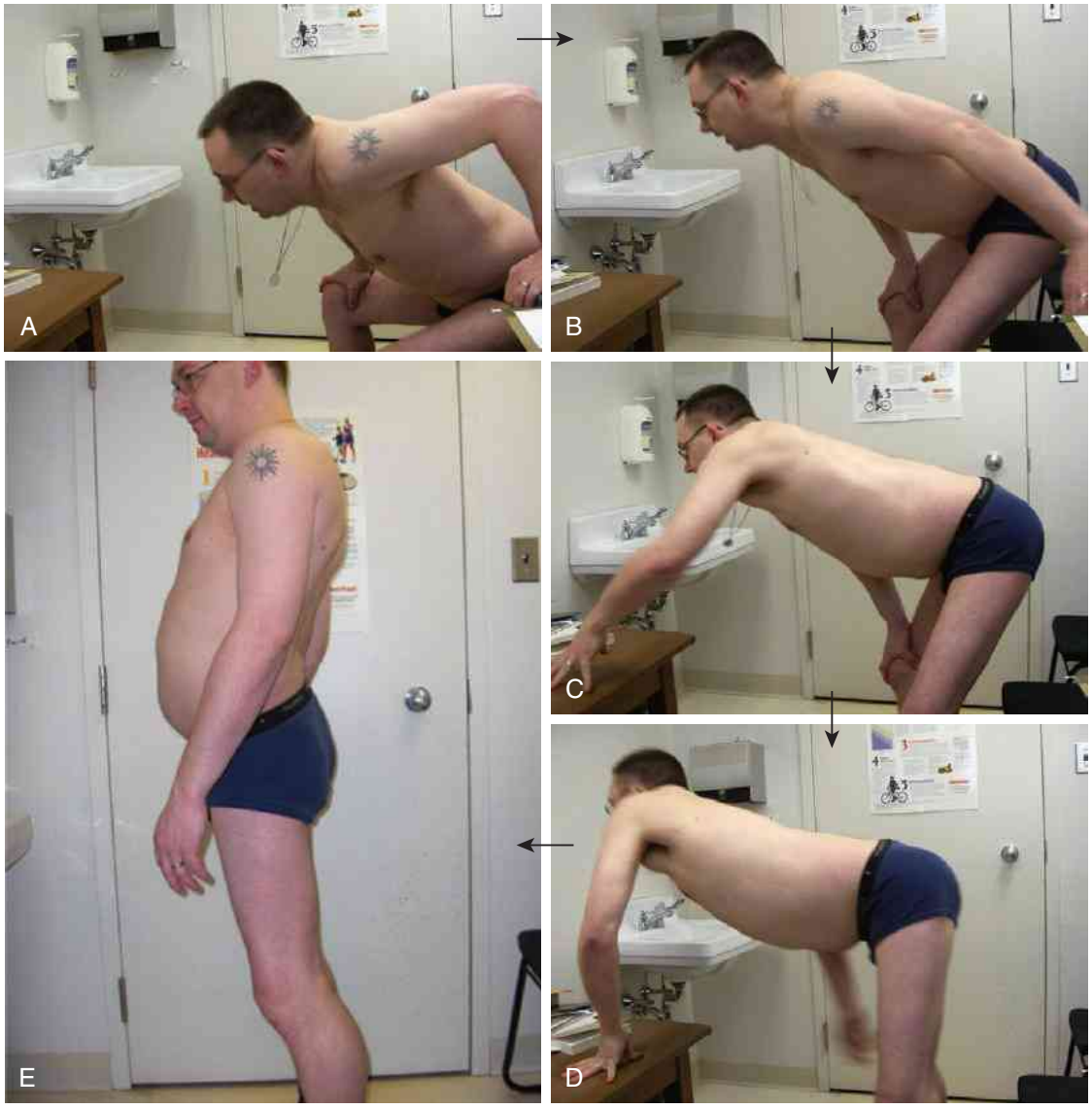


Figure 59-37 Patient with Becker muscular dystrophy caused by an in-frame deletion of exons 45 to 47 in the dystrophin gene. The patient is using the Gower maneuver to rise from sitting to standing position: While sitting **(A),** he uses the force of his hands to stand **(B, C, D)**. In **E,** using his thighs, he pushed himself upright, leading to the characteristic hyperlordosis posture. (Courtesy Dr. Meir H. Kryger.)

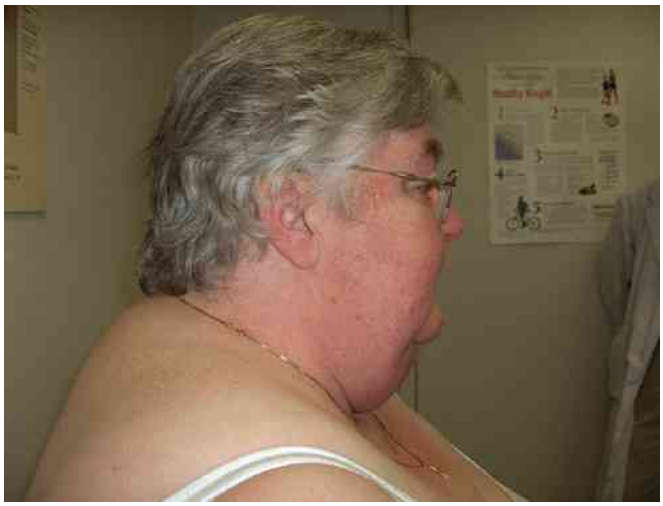


Figure 59-38 "Buffalo hump" in a patient with chronic steroid use. (Courtesy Dr. Meir H. Kryger.)

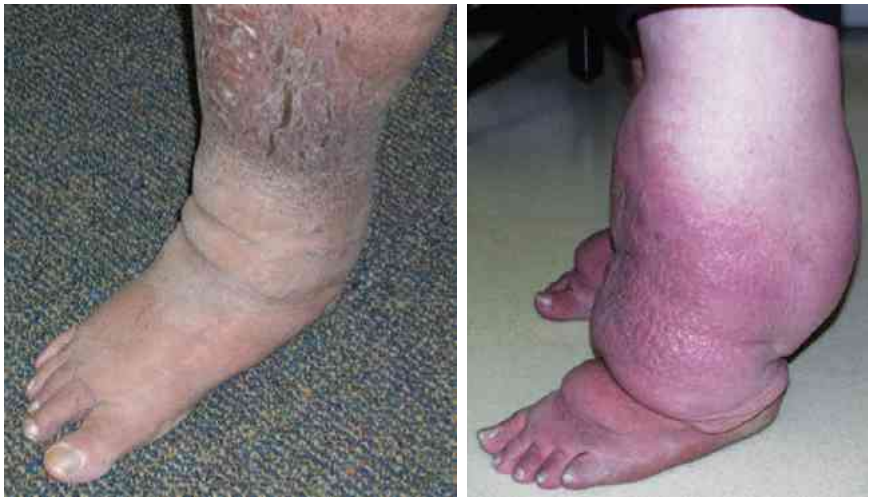


Figure 59-39 Chronic peripheral edema is a common finding in patients with obesity-hypoventilation syndrome. (Courtesy Dr. Meir H. Kryger.)

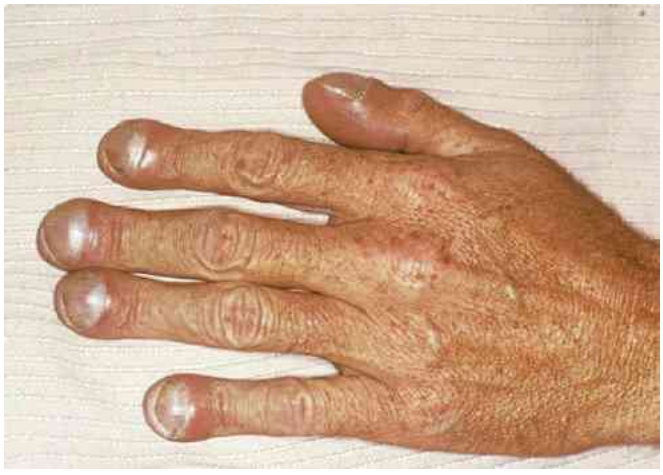


Figure 59-40 Clubbing of digits and nails may be associated with cardiopulmonary insufficiency. (Courtesy Dr. Meir H. Kryger.)

to a CNS etiology. In lesions of the diencephalon due to inflammatory changes such as neurosarcoidosis, one finds additional physical findings associated with panhypopituitarism such as orthostatic hypotension, temperature fluctuations, and other finding of autonomic dysregulation.

Patients with narcolepsy type 1 who experience cataplexy are sometimes observed to have a state of peculiar semipermanent ptosis and jaw weakness, on which partial and complete cataplectic attacks were superimposed cataplectic facies,[42,43] as depicted in Figure 59-41.

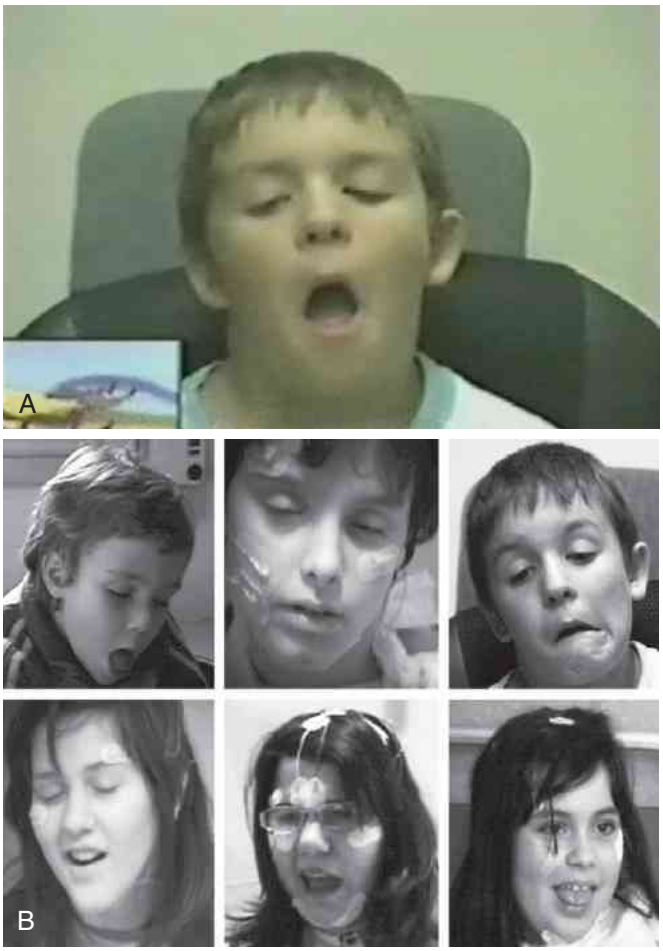


Figure 59-41 **A,** Patients with cataplexy are shown responding to the trigger stimulus (a cartoon). The facial weakness is also present during normal activity without stimulus. **B,** The patient experiences facial muscle weakness, as noted by bilateral, facial grimaces while attempting to keep the eyes open. Facial slackening and tongue protrusion with the mouth opened and a quasi "drunken or droopy look" phenotype, characterize the "cataplectic facies." (From Leonardo S, Pasquale M, Emmanuel M, et al. Cataplexy features in childhood narcolepsy. *Mov Disord* 2008;23:858–65.)

PARASOMNIAS

Nocturnal Eating Disorder and Sleep-Related Eating Disorder

In nocturnal eating disorder, patients often manifest compulsive food-searching behaviors and a return to sleep after food ingestion. Body mass index was abnormally high in 6 of 10 patients after careful exclusion of both anorexia nervosa and bulimia.[44] Sleep-related eating disorder, which also occurs in the setting of WED, is characterized by recurrent episodes of eating after an arousal from nighttime sleep with or without amnesia[45] and may also result in obesity.

REM Sleep Behavior Disorder

Patients with idiopathic REM sleep behavior disorder (RBD) can develop dramatic and aggressive dream enactment events sometimes leading to serious injury. Figure 59-42 depicts a patient who presented at the author's sleep clinic together with his wife who complained that he was dreaming about golfing, was in an argument, and fell to the floor. In the process he hit his neck on the corner of the bedside table and bruised his ear and cheek on bedside table. Although the condition is unlikely to produce severe injury to the patient, the bed partner paradoxically ends up suffering severe sleep interruptions, is more likely to experience sleepiness, and is at risk for injury. Patients with RBD are frequently at risk for development of α-synucleinopathies such as Parkinson disease, and most present with hyposmia (impaired smell), which is a potential preclinical nonmotor sign of the disease.[46] Odor identification was also found impaired in Japanese patients with idiopathic RBD and Parkinson disease.[47]

Cardinal features of Parkinson disease are shown in Figure 59-43. Patients with multiple system atrophy may present with inspiratory stridor, which along with RBD may serve as a clue to the disease in a patient with autonomic failure.[48]

SLEEP-RELATED MOVEMENT DISORDERS

Willis-Ekbom Disease

The prevalence of WED, also known as restless legs syndrome, in patients with type 2 diabetes is 17.7%,[49] and the prevalence may be higher in patients with hereditary neuropathy.[50] WED occurs in about one third of patients with polyneuropathy,[51] with preferential involvement of small sensory fibers. Electrophysiologic studies demonstrate that axonal neuropathy is common in WED, which further necessitates comprehensive peripheral nerve evaluations in these patients.[52]

Reduced iron stores can also cause WED. With iron deficiency, examination of the pharynx may reveal inflammation (redness) or loss or atrophy of the lingual mucosa, indicating glossitis (Figure 59-44). The patient may complain of a sore or tender tongue.

On neurologic examination, symptoms include sensory loss, often described by patients as a sense of numbness or tingling. In the generalized polyneuropathies, symptoms frequently begin in the most distal aspect of the longest sensory fibers, which produce disturbances in sensation in the toes and feet. In addition to sensory loss, patients frequently complain of paresthesias and dysesthesias, often characterized by numbness, tingling, prickling, and pins-and-needles sensations. The sensory examination will often disclose a distal to proximal loss of the various sensory modalities. In certain polyneuropathies, pain predominates in the clinical picture, and the sensory examination tends to disclose deficits predominantly of pain and thermal sensation. When significant proprioceptive deafferentation occurs, patients may present with altered joint position sense that can manifest as an ataxia or tremor of the affected limbs and an imbalance of gait and station.

Pain may be a significant symptom for many patients with WED in which the etiology is related to a polyneuropathy. It

may be described as a dull aching sensation, an intense burning sensation, or, occasionally, intermittent lancinating pulses of pain. On occasion, patients notice that their skin is hypersensitive to tactile stimulation, such as from the touch of bed sheets or clothing or standing on the feet. Some patients note an exaggerated painful sensation resulting from any stimulus to the affected area, a form of pain termed allodynia. Various limb deformities and trophic changes may be observed in chronic polyneuropathies. Pes cavus, characterized by high arches and hammertoes, and the clawfoot deformity are typical foot deformities in hereditary polyneuropathies with childhood onset. These deformities are due to progressive weakness and atrophy of intrinsic foot muscles. A similar clawlike deformity may be observed in the hand. Autonomic involvement of a limb may cause the affected area to appear warm, red, and swollen at times and pale and cold at other times owing to dysregulation of small vessels due to autonomic denervation. Various trophic changes, including tight, shiny skin, may occur. In patients who have had severe sensory

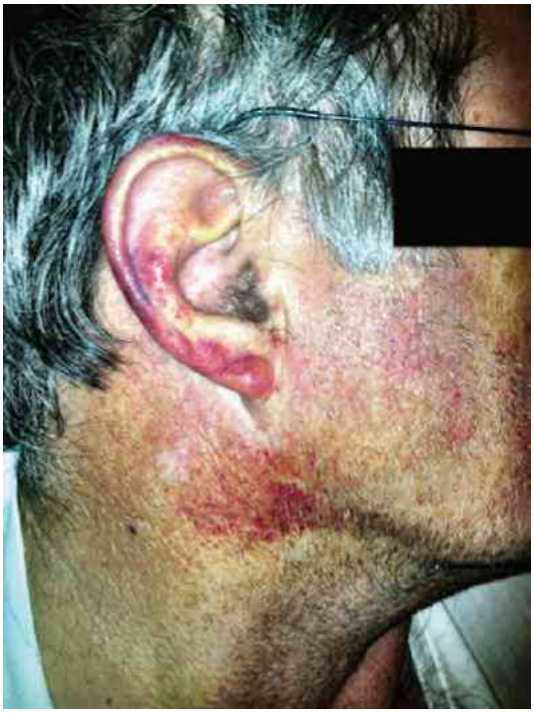


Figure 59-42 A patient with aggressive dream enactment behavior who experienced severe injury during one of his nocturnal episodes in the setting of REM sleep behavior disorder. (Copyright Alon Y. Avidan, MD, MPH.)

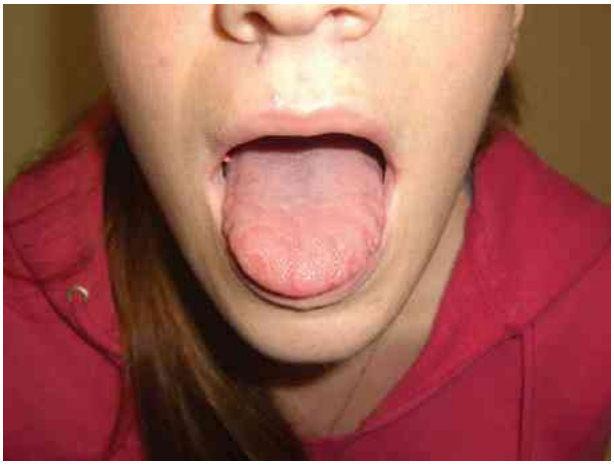


Figure 59-44 Tongue glossitis in iron deficiency. (Courtesy Dr. Meir H. Kryger.)

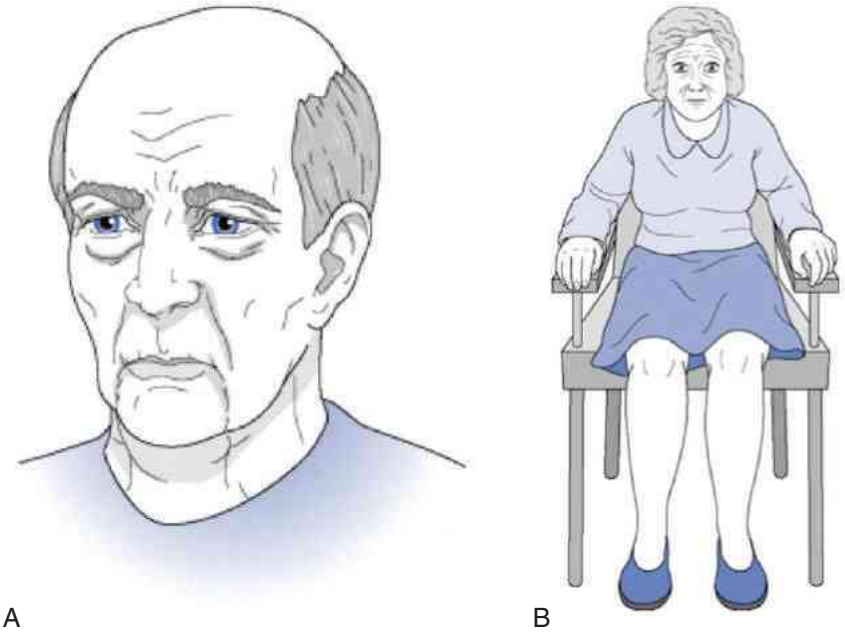


Figure 59-43 A, Patients with Parkinson disease (PD) blink less frequently and make fewer facial expressions, with less frequent head movements ("masked facies"). **B,** Patients with PD rarely participate in normal gestures or repositioning movements. They are observed to sit motionless with their legs uncrossed and their feet flat. The upper extremities remain motionless on the chair or in their lap. (From Kaufman DM: Involuntary movement disorders. In: Kaufman DM, editor. *Clinical neurology for psychiatrists.* 6th ed. Philadelphia: Elsevier; 2007. p. 401–64.)

loss in the limbs, the affected areas may be subject to incidental traumas, including burns, pressure sores, and other injuries that are not perceived by the patient, in whom repeated injuries and traumas may result in chronic infections and, when severe, lead to osteomyelitis. A clinical evaluation of peripheral neuropathy is provided by Kelly.[53,54]

Bruxism

Bruxism (see Chapters 144 and 145) represents a stereotyped movement disorder clinically characterized by grinding or clenching of the teeth during sleep. The sounds made by friction of the teeth are usually perceived by a bed partner as very unpleasant.[55] The condition is typically brought to the attention of the medical or dental practitioner in efforts to eliminate the disturbing sounds. Bruxism can lead to abnormal wear of the teeth (Figure 59-45), periodontal tissue damage, or jaw pain. Other symptoms include facial muscle and tooth pain and headache. Bruxism induces dental damage with abnormal wear to the teeth and damage to the structures surrounding the teeth. Chronically, over time and when untreated, this leads to recession and inflammation of the gums, alveolar bone resorption, muscles of mastication hypertrophy (Figure 59-46), and temporomandibular joint disorders, often associated with facial pain. Additional physical findings include tenderness of the muscles of mastication (masseter, temporalis, pterygoid, sternocleidal), temporomandibular disorders, tongue indentation, subjective appreciation of a tense personality, and hypervigilant patient.[56]

Case reports in patients with bruxism demonstrate bilateral enlargement in the region of the mandibular angle, corresponding with the masseter hypertrophy (see Figure 59-46).[57] Children with bruxism have a significantly longer and higher palate in the sagittal plane and bigger dental arches compared with normal children.[58] Psychiatric patients have a higher prevalence of bruxism and signs of temporomandibular disorders, possibly related to neuroleptic-induced phenomenon.[59]

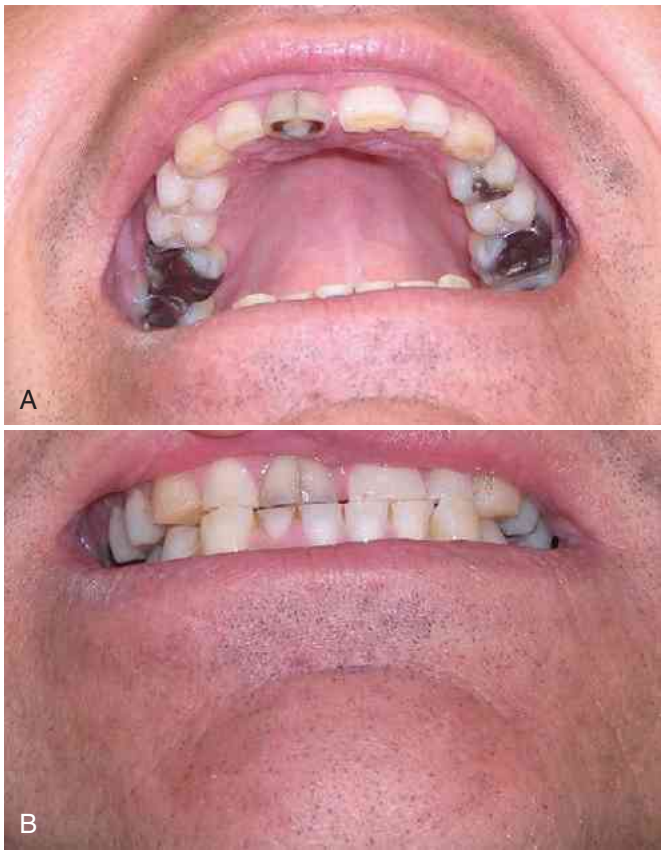


Figure 59-45 Bruxism with abnormal wear of the teeth.

Insomnia

Insomnia, especially comorbid in type, is often seen in the context of endocrinopathies, mood disorders, anxiety disorders (Figure 59-47), rheumatologic conditions, pain, and a

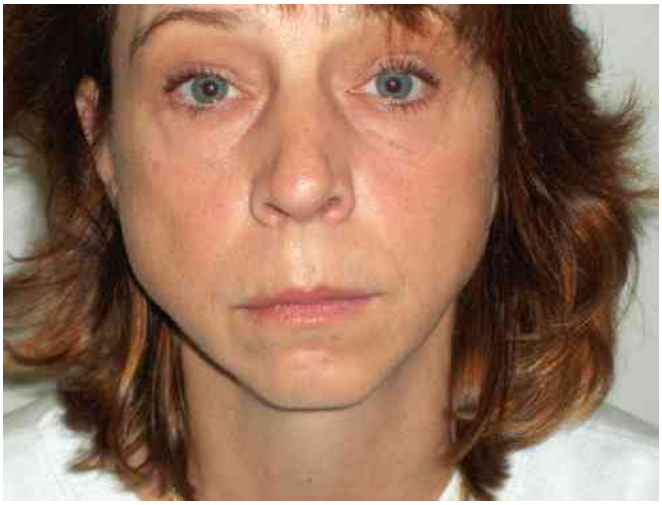


Figure 59-46 Hypertrophic masseter in bruxism. The masseter muscle bulk is markedly increased over the mandibular angle region. (Courtesy Dr. Meir H. Kryger.)

Figure 59-47 The painting of anxiety, a common comorbidity in chronic insomnia. (From Gross M. Shining new light on the brain. *Curr Biol* 2011;21[20]:R831–3.)

long list of other medical, psychiatric, and primary sleep disorders. Patients with Graves disease (Figure 59-48), an autoimmune disorder and a common cause of hyperthyroidism, is characterized by the presence of autoantibodies that bind and stimulate the thyroid-stimulating hormone receptor, resulting in hyperfunction of the thyroid. Graves disease is characterized by a phenotype of heat intolerance, involuntary weight loss, thyromegaly, tremor, and hyperactivity manifesting as restless sleep and insomnia.[60] Finally, floppy eyelid syndrome (Figure 59-49) is sometimes confused with tiredness, a thyroid disorder, or a neuromuscular disorder and is characterized by flaccid and easily everted upper lids, occurring spontaneously or with minimal manipulation.[61] It is can seen in middle-aged men who are overweight and has been associated with OSA.

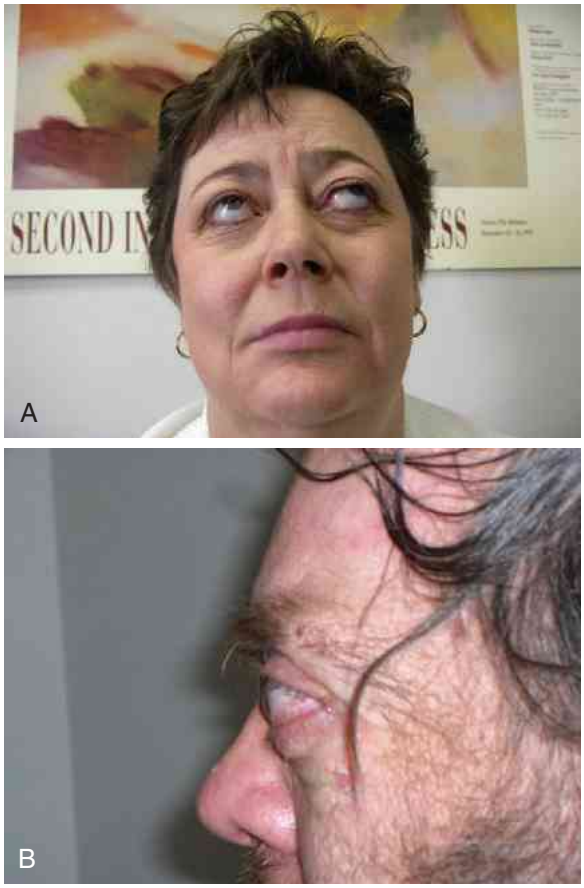


Figure 59-48 Proptosis seen in the setting of Graves disease. **A,** Eye signs in Graves disease. **B,** Severe proptosis in Graves disease. (**A** and **B,** Courtesy Dr. Meir H. Kryger.)

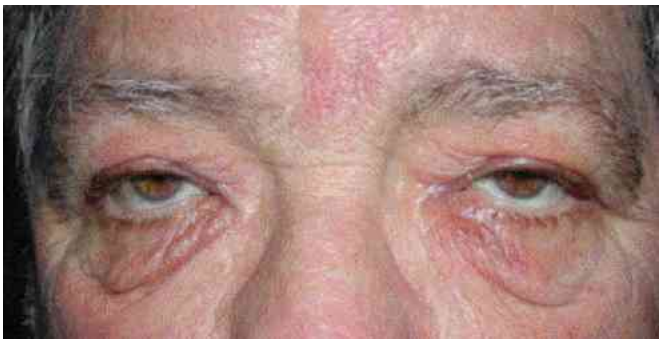


Figure 59-49 Bilateral upper lid ptosis in a patient with floppy eyelid syndrome. (From Leibovitch I, Selva D. Floppy eyelid syndrome: clinical features and the association with obstructive sleep apnea. *Sleep Med* 2006;7[2]: 117–122.)

CLINICAL PEARLS

- Overall inspection of the patient, coupled with observation of craniofacial, nasal, and pharyngeal factors, allows detection of key risk factors for sleep apnea.
- Examination of patients with insomnia should focus on the potential associated comorbidities, including hypothyroidism and rheumatologic disorders.
- Patients with motor disorders of sleep and parasomnias also have clinical findings conferred by the underlying medical, neurologic, and associated psychiatric comorbidities. For example, anosmia, orthostatic fluctuations in the setting of dream enactment behavior, and loss of electromyographic tone on polysomnogram may be predictive of an evolving α-synucleinopathy.
- Clues to the presence of abnormal nocturnal events such as parasomnias or nocturnal seizures may include unexplained bruising, lacerations in the former, and tongue laceration in the latter. However, even tongue biting, which is believed to be a clinical sign of epilepsy, can occur in syncope and nonepileptic seizures. These difficulties highlight the importance of a well-tailored approach in which the clinical history, physical examination, and supportive laboratory and polysomnographic data are used to arrive the most plausible clinical diagnosis.

SUMMARY

The physical examination of any patient with sleep disorders is the cornerstone for making critical decisions about the possible clinical diagnosis, determining the need for formal polysomnography, and ensuring that treatment is successful. Given that medial trainees often do not receive formalized sleep medicine education in medical school, appreciating the fundamental phenotypical patterns responsible for sleep-disordered breathing is critical. A basic appreciation of the abnormal neurologic examination is important for nonneurologists who may encounter patients with parasomnias and motor and movement disorders of sleep. Finally, no clinical examination of a sleepy patient should conclude without a comprehensive review of the patient's medical, endocrine, metabolic, genetic, and disease background, given that may phenotypes contribute directly to disruptive sleep.

Selected Readings

Friedman M. *Sleep apnea and snoring: surgical and non-surgical therapy*. Philadelphia: Elsevier Science; 2008.

Kryger MH, Avidan AY, Berry R. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014.

Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea? The rational clinical examination systematic review. *JAMA* 2013;**310**:731–41.

Posnick JC. *Orthognathic surgery: principles and practice*. Philadelphia: Elsevier Science; 2014.

Wilhelm CP, deShazo RD, Tamanna S, et al. The nose, upper airway, and obstructive sleep apnea. *Ann Allergy Asthma Immunol* 2015;**115**(2): 96–102.

A complete reference list can be found online at ExpertConsult.com.

Use of Clinical Tools and Tests in Sleep Medicine

Chapter 60

Cathy A. Goldstein; Ronald D. Chervin

Chapter Highlights

- A clinician confronted with a sleep-related complaint combines symptoms, signs, and test results to make a diagnostic assessment.
- Information about test performance characteristics, such as sensitivity, specificity, and predictive value, can be used more formally to decide on optimal approaches.
- In addition to the history and physical examination, tools and tests used in the evaluation of a patient with sleep-related symptoms may include questionnaires, sleep diaries or logs, actigraphy, nocturnal polysomnography, an out-of-center sleep test, or a multiple sleep latency test.

This chapter focuses on the comparative value of different approaches to clinical assessment of sleep-related problems. The symptoms, signs, and test results relevant to particular sleep disorders are described in detail in chapters that focus on specific clinical entities. This chapter, instead, highlights the clinical reasoning process by which a clinician challenged with a sleep complaint can combine information from different sources, appropriately weigh available evidence, and arrive at sound diagnoses and treatment plans. Here we review the value of tests in evaluations of suspected obstructive sleep-disordered breathing, hypersomnolence, insomnia, suspected circadian rhythm sleep-wake disorders, restless legs syndrome (RLS), and suspected parasomnias. A selection of evidence-based practice parameters and reviews produced by the American Academy of Sleep Medicine (AASM) can be accessed at http://aasmnet.org/practiceguidelines.aspx (Table 60-1) to supplement overviews presented in this chapter.

EVALUATION FOR SLEEP-RELATED BREATHING DISORDERS

History and Questionnaires

Sleep-related breathing disorders are by far the most common disorders diagnosed at sleep centers, and obstructive sleep apnea (OSA) alone accounts for nearly 70% of all patients evaluated.[1]

Subjective clinical impressions of OSA tend to have inadequate sensitivity (probability of a positive test result or assessment given that the disorder is present) and specificity (probability of a negative test result given that the disorder is absent).[2] Combinations of some signs and symptoms can have sensitivity above 0.90, but specificity is usually poor. Performance of these models in clinical practice can be worse than originally reported.[3] However, sensitivity and specificity data suggest that the negative predictive value (NPV) of some symptom combinations may be good, whereas the positive predictive value (PPV) is probably poor, especially when the prevalence of OSA in the tested population is not high.[4] Accordingly, patients without a history suggestive of OSA usually do not receive further testing for it. Among patients referred for suspected OSA, models based on historical information may accurately classify a minority as apnea free without further tests.[5] In practice, patients who do have symptoms of OSA generally are tested.

Although, in general, the PPV of symptoms alone is not high, a minority of patients have a clinical presentation so convincing as to be essentially diagnostic. However, the *International Classification of Sleep Disorders*, third edition (ICSD3), requires minimal objective criteria, in addition to symptoms, to establish a diagnosis of OSA.[6] Tests in such cases may also serve to define the severity of OSA. In a patient with a history strongly suggestive of OSA, the diagnosis must still be suspected when it is not confirmed by a single polysomnogram (PSG) and especially when it is not confirmed by a more abbreviated home sleep test.[7]

The diagnostic values of specific symptoms are difficult to judge on the basis of studies with significant methodologic differences. For example, among patients referred specifically for possible OSA, the symptom of excessive daytime sleepiness may[8] or may not[9] be useful in making the diagnosis, and a history of hypertension may be better than a report of snoring as an indication that OSA is present.[10] Among patients referred to a sleep center, snoring has high sensitivity (80% to 90%) and low specificity (20% to 50%) for the diagnosis of OSA, whereas nocturnal choking or gasping is less sensitive (52%) and more specific (84%).[2] In patients referred for suspected sleep-disordered breathing, the presence of nocturnal choking or gasping yields a PPV for OSA of 35%, which is greater than the PPV for morning headache, reported apnea, excessive daytime sleepiness, or snoring.[2] In contrast, in the community, the symptom with the highest predictive value for OSA is habitual snoring, although excessive daytime sleepiness and observed apneas are also useful.[11]

Physical Examination

In the community, among variables related to body weight, neck circumference and body mass index (BMI) correlate well with the presence and severity of OSA.[12] Among patients referred for possible OSA, these variables still may be useful, but their predictive value is not large except in extreme

Table 60-1 Practice Guidelines for the Use of Tools and Tests in Sleep Medicine from the American Academy of Sleep Medicine (AASM)

Type	Published (mo/yr)	Subject Matter	Paper Title
P	1/2008	Sleep-related breathing disorders	Practice Parameters for the Use of Autotitrating Continuous Positive Airway Pressure Devices for Titrating Pressures and Treating Adult Patients with Obstructive Sleep Apnea Syndrome: An Update for 2007
CG	6/2009	Sleep-related breathing disorders	Clinical Guideline for the Evaluation, Management, and Long-term Care of Obstructive Sleep Apnea in Adults
CG	2/2008	Sleep-related breathing disorders	Clinical Guidelines for the Manual Titration of Positive Airway Pressure in Patients with Obstructive Sleep Apnea
P	11/2007	Circadian rhythm sleep disorders	Practice Parameters for the Clinical Evaluation and Treatment of Circadian Rhythm Sleep Disorders
R	11/2007	Circadian rhythm sleep disorders	Circadian Rhythm Sleep Disorders: Part I, Basic Principles, Shift Work, and Jet Lag Disorders
R	11/2007	Circadian rhythm sleep disorders	Circadian Rhythm Sleep Disorders: Part II, Advanced Sleep Phase Disorder, Delayed Sleep Phase Disorder, Free-Running Disorder, and Irregular Sleep-Wake Rhythm
CG	10/2008	Insomnia	Clinical Guideline for the Evaluation and Management of Chronic Insomnia in Adults
P	11/2012	Pediatrics	Practice Parameters for the Non-Respiratory Indications for Polysomnography and Multiple Sleep Latency Testing for Children
P	3/2011	Pediatrics	Practice Parameters for the Respiratory Indications for Polysomnography in Children
R	11/2012	Pediatrics	Non-Respiratory Indications for Polysomnography and Related Procedures in Children: An Evidence-Based Review
R	3/2011	Pediatrics	Executive Summary of Respiratory Indications for Polysomnography in Children: An Evidence-Based Review
P	4/2007	Diagnostics	Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007
P	4/2005	Diagnostics	Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005
P	1/2005	Diagnostics	Practice Parameters for Clinical Use of the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test
P	9/2003	Diagnostics	Practice Parameters for Using Polysomnography to Evaluate Insomnia: An Update
R	10/2011	Diagnostics	Obstructive Sleep Apnea Devices for Out-Of-Center (OOC) Testing: Technology Evaluation
R	1/2005	Diagnostics	A Review by the MSLT and MWT Task Force of the Standards of Practice Committee of the American Academy of Sleep Medicine: The Clinical Use of the MSLT and MWT
CG	12/2007	Diagnostics	Clinical Guidelines for the Use of Unattended Portable Monitors in the Diagnosis of Obstructive Sleep Apnea in Adult Patients

MSLT, Multiple Sleep Latency Test; MWT, Maintenance of Wakefulness Test.
Practice guidelines are written by the Standards of Practice Committee of the AASM. Practice Parameters (P), Systematic Reviews (R), Clinical Guidelines (CG), and Best Practice Guides (BPG) are generally available at http://aasmnet.org/practiceguidelines.aspx.

ranges.[9,13,14] Patients with OSA who are not obese often have pharyngeal crowding, obstructed nasal passages, or other craniofacial abnormalities associated with narrowing of the upper airway.[15] The predictive value of such findings may differ somewhat between men and women.[16] The Mallampati score, which reflects oropharyngeal crowding on a 4-point scale, was found to predict OSA.[17] Each 1-point increase in the score was associated with an odds ratio of 2.5 (95% confidence interval [CI] [1.2, 5.0]) for OSA and predicted a 5-point higher apnea-hypopnea index (coefficient = 5.3 [0.2 to 10]), independent of many other physical findings and symptoms.

Physical findings can also be combined into predictive quantitative models to aid in the diagnosis of OSA. Models based on measures that can be obtained during the physical examination, such as BMI, neck circumference, craniofacial measurements, pharyngeal scores, and tonsil size, demonstrate excellent PPV (90% to 100%) but less strong NPV (49% to 89%).[18-20] Other physical findings may also have value in the diagnosis of OSA. High blood pressure increases the chance that OSA will be present, especially among persons who are less obese.[21] Signs of neuropathy or neuromuscular disease also may increase the likelihood of OSA.

Table 60-2 Value of Specific Questionnaire Instruments that Combine Symptoms and Physical Findings to Diagnose Obstructive Sleep Apnea

Instrument	Study	Subjects	Gold Standard	Sensitivity	Specificity	Predictive Value
Berlin Questionnaire	Subramanian et al., 2011[154]	Referral based	PSG RDI ≥15	0.93	0.14	
			PSG RDI ≥5	0.92	0.18	
	Hrubos-Strøm et al., 2011[155]	Population based	PSG AHI ≥15	0.43	0.80	PPV = 0.34, NPV = 0.86
			PSG AHI ≥5	0.37	0.84	PPV = 0.61, NPV = 0.66
	Sun et al., 2011[156]	Referral based	PSG AHI ≥15	0.97	0.48	
	Kang et al., 2013[157]	Population based	PSG AHI ≥5	0.69	0.83	
	Cowan et al., 2014[158]	Referral based	Home PG AHI ≥15	0.94	0.08	PPV = 0.44, NPV = 0.67
			Home PG AHI ≥5	0.93	0.06	PPV = 0.75, NPV = 0.22
STOP-BANG	Chung et al., 2008[159]	Preoperative population	PSG AHI ≥15	0.93	0.43	PPV = 0.52, NPV = 0.90
			PSG AHI ≥5	0.84	0.56	PPV = 0.81, NPV = 0.61
	Ong et al., 2010[160]	Referral based	PSG AHI ≥15	0.91	0.40	PPV = 0.61, NPV = 0.82
			PSG AHI ≥5	0.85	0.53	PPV = 0.84, NPV = 0.53
	Cowan et al., 2014[158]	Referral based	Home PG AHI ≥15	1.0	0.21	PPV = 0.5, NPV = 1.0
			Home PG AHI ≥5	0.95	0.3	PPV = 0.81, NPV = 0.64
NAMES	Subramanian et al., 2011[154]	Referral based	PSG RDI ≥15	0.91	0.23	PPV = 0.62, NPV = 0.63
			PSG RDI ≥5	0.88	0.29	
NAMES2	Subramanian et al., 2011[154]	Referral based	PSG RDI ≥15	0.92	0.34	
			PSG RDI ≥5	0.85	0.42	
Snoring Severity Scale with BMI	Morris et al., 2008[161]	Referral based	PSG RDI ≥15	0.97	0.40	PPV = 0.82, NPV = 0.84

AHI, Apnea-hypopnea index (number of apneas or hypopneas per hour of sleep); BMI, body mass index; RDI, respiratory disturbance index (number of apneas, hypopneas, or respiratory effort related arousals per hour of sleep); NPV, negative predictive value; PPV, positive predictive value; PSG, polysomnography; SDB, sleep-disordered breathing.
The NAMES instrument assesses neck circumference, airway classification, comorbidities, Epworth scale, and snoring. The NAMES2 instrument contains the same variables in NAMES with the addition of BMI and gender. The STOP-BANG instrument assesses snoring, tiredness, observed apneas, blood pressure, BMI, age, neck circumference, and gender. The Snoring Severity Scale (SSS) assesses snoring loudness, frequency, and duration.

Instruments that use a combination of symptoms, comorbidities, and physical findings to assess risk for OSA include the Berlin Questionnaire, STOP-BANG, NAMES, NAMES2, and the Snoring Severity Scale combined with BMI. Table 60-2 shows sensitivity, specificity, and predictive values for these tools. Of these tools, the Berlin Questionnaire and the STOP-BANG instrument (Figure 60-1) are most frequently encountered in clinical practice.

Nocturnal Polysomnography

A nocturnal, laboratory-based PSG is commonly used to objectively test for OSA. The PSG often is considered a gold standard for OSA diagnosis, assessment of severity, and identification of some other sleep disorders that can accompany OSA. The PSG allows direct monitoring and quantification of respiratory events and physiologic consequences—such as hypoxemia, arousals, and awakenings—that are suspected to cause daytime symptoms. A single-night PSG is usually sufficient to diagnose or to exclude OSA. However, the test is not infallible. Accuracy may be reduced by variability in biologic severity, laboratory equipment, human scoring, or scoring protocols. Night-to-night variability may be particularly high in subjects with low but clinically significant rates of apneas and hypopneas during sleep. A repeat PSG may confirm OSA in 20% to 50% of individuals who have symptoms suggestive of OSA but initial PSG negative for OSA.[22,23]

SNORING?
Do you **snore loudly** (louder than talking or loud enough to be heard through closed doors)? Yes No

TIRED?
Do you often feel **tired, fatigued, or sleepy** during the daytime? Yes No

OBSERVED?
Has anyone **observed** you **stop breathing** during your sleep? Yes No

PRESSURE?
Do you have or are you being treated for **high blood pressure?** Yes No

BODY MASS INDEX more than 35 kg/m²? Yes No

AGE older than 50 years? Yes No

NECK circumference?
Neck circumference greater than 40 cm? Yes No

Gender=male? Yes No

Figure 60-1 The STOP-BANG questionnaire. Low risk for OSA: yes to 0 to 2 questions. High risk for OSA: yes to 3 or more questions. (From Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology* 2008;108[5]:812–21, with permission.)

Publication by the AASM in 2007 (since updated twice) of new sleep scoring guidelines also included recommendations for polysomnographic equipment; scoring of abnormal respiratory events, electrocardiographic findings, movements, and arousals during sleep; and modifications necessary for children.[24] These guidelines serve to improve uniformity of procedures between laboratories.

Arguably, the change in the updated manual with the highest potential to influence clinical practice is the modification to the scoring rules for hypopneas. The scoring manual recommends that the technician scores a hypopnea when the nasal pressure transducer signal (or positive airway pressure flow during a titration study) drops by at least 30% from the preevent baseline, the duration of this amplitude reduction is at least 10 seconds, and the event results in either an arousal or oxygen desaturation of at least 3%.[25] The recommended hypopnea definition aims to increase the sensitivity of PSG to detect OSA in patients with sleep fragmentation and daytime impairment but without significant oxygen desaturations.[25] The scoring manual also defines an acceptable rule as an alternative method to score hypopneas that requires an oxygen desaturation of 4%.[25]

The distinction between the recommended rule and acceptable rule for scoring hypopneas is important from a diagnostic standpoint. For example, PSGs were rescored in a group of lean patients with known OSA based on use of a scoring rule that did not require oxygen desaturation to score hypopneas.[26] The PSGs were rescored with a hypopnea definition that required a 4% oxygen desaturation but not necessarily an arousal. This change resulted in a reduction of the apnea-hypopnea index (AHI) and classified 40% of these symptomatic patients as negative for OSA.[26] These findings highlight the potential benefits of the recommended hypopnea definition set forth in the 2012 scoring manual.

Most laboratories report an AHI that represents the sum total of apneas and hypopneas per hour of sleep. Additionally, the respiratory disturbance index (RDI) calculates the sum total of apneas, hypopneas, and respiratory effort related arousals (RERAs) per hour of sleep. Nasal or esophageal pressure monitoring allows for the scoring of RERAs. Although additional data are still needed, the importance of scoring RERAs is likely to depend in part on which hypopnea definition is used. With the less sensitive definition that focuses on 4% desaturations and ignores arousals, scoring RERAs makes a much larger difference in the computed total rate of respiratory events and may frequently make the difference between diagnosis and failure to diagnose OSA. If the more sensitive definition for hypopneas is used that includes arousal but does not require oxygen desaturation, fewer additional events are detected when RERAs are scored; RERAs then have less impact on the total rates of apneic events and therefore less clinical impact.

As a result of these challenges and the imperfect reliability of PSGs, interpretation of PSG reports remains more complicated and may not be definitive, particularly in borderline cases. The patient's clinical presentation should be considered when interpreting the PSG to help mitigate underdiagnosis and overdiagnosis. Although many clinicians believe that an AHI above 5 indicates OSA, the PSG finding of an AHI greater than 5 may not be associated with symptoms. For example, a large population-based epidemiologic study found that only 22.6% of women and 15.5% of men who met this criterion clearly complained of daytime hypersomnolence.[12] Conversely, some patients with an AHI less than 5 may still have OSA that merits treatment to improve symptoms and morbidity.[27,28]

Further research is needed to define and to improve the ability of PSGs to measure those aspects of sleep-disordered breathing that most affect health and daytime sleepiness. The AHI and minimum oxygen saturation do not correlate strongly with daytime sleepiness,[29] although the AHI may correlate better with cardiovascular morbidity.[22] Esophageal pressure monitoring is the gold standard to assess respiratory effort and may identify increased respiratory effort and RERAs in patients without significant apneas and hypopneas.[30,31] However, criteria for abnormal esophageal pressure recordings, as defined by association with poor outcomes, remain to be studied more definitively. Nasal pressure monitoring may provide a well-tolerated alternative; however, despite increased sensitivity, studies have yet to demonstrate improved prediction of outcomes, and initial comparisons to thermistor results show correlations high enough (e.g., 0.90 or higher)[32,33] to suggest redundancy of information. Other polysomnographic measures that may (or may not) prove to enhance the ability of PSGs to predict outcomes of sleep-disordered breathing include end-tidal or transcutaneous carbon dioxide monitoring,[34] pulse transit time,[35] peripheral arterial tonometry,[36] scoring of arousals,[37,38] and analysis of respiratory cycle–related electroencephalographic changes.[38,39]

In short, the PSG is the single most useful and definitive test in the diagnosis of sleep-related breathing disorders, but the information it provides cannot be reliably interpreted by persons without experience in sleep medicine, summarized by any single number, or applied to patient care without careful use of additional clinical data. Failure to recognize these limitations, by health care policy makers or clinicians, could trigger unnecessary intervention or deprive a patient of effective treatment.

Modified Forms of the Polysomnogram

In comparison to the standard PSG, daytime and split-night studies may reduce costs and expedite evaluation. Studies of daytime PSGs have sometimes found a high NPV, with lower PPV, but inconsistent results and the lack of sufficient data explain why daytime PSGs have not generally been recommended.[4] A successful split-night study may save a patient from a second night in the sleep laboratory. Studies of diagnostic accuracy and treatment outcomes appear promising.[40] Concordance is high between AHI measured in the first 2 hours of PSG recording and AHI measured in a full-night PSG (concordance correlation coefficient = 0.93).[41] Although the traditional gold standard has been separate, full-night studies for diagnostic assessment and then positive airway pressure titration, split-night studies may be adequate alternatives in most cases.[42,43]

Home Sleep Tests

Many different devices exist to assess OSA at home, although most do not record sleep. These "portable" recordings usually are less costly than laboratory-based PSGs, and patients often prefer home studies to laboratory studies. However, the diagnostic value of unattended portable monitoring is often reduced by the inability to make behavioral observations, standardize recording conditions, address technical problems,

make interventions during the night, or monitor variables equivalent to those recorded in the laboratory setting. Home sleep tests that do not monitor signals necessary to identify sleep stages or leg movements only evaluate for sleep apnea. Additionally, scored respiratory events may not have occurred during sleep and thus may result in inaccurate sleep apnea severity.

A Portable Monitoring Task Force of the AASM recommended home studies only after a comprehensive sleep evaluation by a clinician board-eligible or certified in sleep medicine, and then interpreted by someone with the same level of specialty training.[7] This recommendation is based on the fact that studies demonstrating effectiveness of home sleep tests were conducted in the context of thorough clinical evaluations by sleep specialists. Further, home studies have significant limitations and therefore must be interpreted by specialists aware of these constraints. Under these conditions, home studies can be used as an alternative to laboratory-based PSGs when clinical judgment suggests that pretest probability of moderate to severe OSA is high. Home studies should generally not be used when the patient is a child or older person, has significant health comorbidities (e.g., severe pulmonary disease, neuromuscular disease, or congestive heart failure), or in whom other additional sleep disorders are suspected.[7] Home studies may be indicated for patients who do not have access to laboratory-based PSG, cannot tolerate the procedure, or need follow-up assessment of response to non–positive airway pressure treatments of OSA.

Published guidelines for home studies recommend that at minimum they monitor airflow, respiratory effort, and blood oxygenation and that the equipment should be applied by a sleep technologist or health care practitioner with appropriate training.[7] Home studies can underestimate sleep-disordered breathing. Many home studies do not monitor an electroencephalogram (EEG) and as such do not allow for hypopneas to be scored when they terminate in cortical arousal but do not result in oxygen desaturation. Additionally, the AHI on a home study that does not monitor sleep is calculated with total recording time (as opposed to total sleep time) as the denominator.[42] Therefore, if a home study in an appropriately selected individual does not demonstrate OSA, a more definitive laboratory-based sleep study should be considered. A suggested algorithm for use of home studies is shown in Figure 60-2.

In carefully selected patients, portable recording devices are effective tools to diagnose OSA. A meta-analysis of level 3 portable monitoring devices demonstrated sensitivity from 0.79 to 0.97 and specificity from 0.60 to 0.93 for OSA, depending on AHI cutoffs.[44] However, these validation studies, generally performed in a controlled laboratory setting, do not take into account the potential for technical failure of portable devices in an ambulatory environment. Despite these limitations, initial investigations demonstrate similar outcomes in patients randomized to home sleep tests versus attended PSG.[45,46] Portable devices that have low costs, high sensitivities, and high specificities have the potential to be cost-effective in comparison to PSGs. However, cost-effectiveness analyses thus far suggest that full-night PSG is superior to portable tests for OSA[47,48] (discussed under Beyond Sensitivity, Specificity, and Predictive Value in this chapter). In some situations, home studies could increase costs, delay confirmatory laboratory testing, encourage treatment of patients with false-positive results, or allow development of medical morbidity from undiagnosed and therefore untreated OSA.

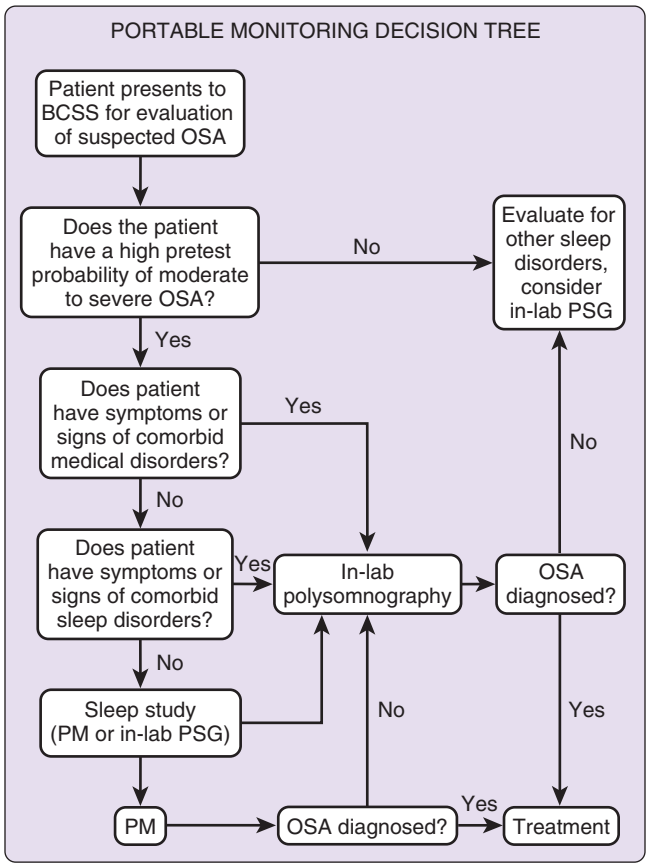


Figure 60-2 Recommended use of home studies. BCSS, Board-certified sleep specialist; OSA, obstructive sleep apnea; PM, portable monitoring; PSG, polysomnography. (From Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2007;3:737–47, with permission.)

Studies of Airway Morphology

Although imaging of the upper airway for research purposes has led to a better understanding of OSA pathophysiology, such studies are not routinely performed in diagnostic evaluations of patients, in part because findings that predict OSA or its severity with sufficient accuracy to allow use in management of individual patients have not been identified. However, cephalometric radiography and pharyngoscopy may be useful in preoperative identification of sites of obstruction and in selection of appropriate surgical procedures. The diagnostic value of cephalometrics may be limited in part because only sagittal plane dimensions are provided while coronal plane dimensions or volume may be more pertinent to OSA.[49] Pharyngoscopy allows three-dimensional anatomic characterization, but whether airway collapse with Müller's maneuver predicts response to uvulopalatopharyngoplasty is debated. Pharyngoscopy may be particularly valuable if it is performed during supine sleep.[50] Techniques for quantitative computer-assisted video endoscopic airway analysis are also being developed and have shown, for example, correlation between the

extent of anatomic change after uvulopalatopharyngoplasty and improvement in the AHI.[51]

Computed tomography and magnetic resonance imaging (MRI) studies can show upper airway morphology,[49] and some authors suggest the potential for clinical usefulness.[52] Specifically, MRI evaluates the upper airway in multiple planes and can be used during sleep.[53-55] Dynamic sleep MRI protocols may characterize upper airway obstruction better than single-plane images during wakefulness.[53-55] However, the value of these techniques in the clinical setting is not well defined.

EVALUATION OF HYPERSOMNOLENCE

History and Questionnaires

The history provides important clues to the severity of hypersomnolence. Direct inquiry about sleepiness can be supplemented by questions about sleepiness in sedentary situations, such as driving, desk work, reading, or watching television. However, patients may report little of the excessive daytime sleepiness suggested by family members, clinical signs, or objective tests. Words other than *sleepiness* are often used by patients with sleep disorders to describe the chief complaint. Among 190 apneic subjects in one study, preferred terms included *lack of energy* (40%), *tiredness* (20%), *fatigue* (18%), and *sleepiness* (22%).[56] Furthermore, each of these symptoms tends to resolve after use of continuous positive airway pressure. Patients' opinions about their own sleepiness sometimes show no significant association with results of the Multiple Sleep Latency Test (MSLT).[57]

Questionnaires such as the Epworth Sleepiness Scale[58] and the Stanford Sleepiness Scale[59] (see Chapter 169) provide a more formal and perhaps reliable measure of excessive daytime sleepiness. The impact of sleepiness on activities of daily living can be assessed with the Functional Outcomes of Sleep Questionnaire.[60] Epworth results correlate reasonably well with patients' self-ratings for overall sleepiness but not well with MSLT results.[61] Although the Epworth Sleepiness Scale and the Stanford Sleepiness Scale can have clinical utility, for example, in monitoring response to treatment over time, they do not substitute for well-validated objective measures of sleepiness. Unfortunately, the ability of subjective tests of sleepiness to predict future health outcomes remains largely unknown.

In addition to the severity of EDS, it is critical to assess napping patterns and other symptoms associated with hypersomnolence to determine its etiology. For example, cataplexy is the essential feature that distinguishes narcolepsy type 1 from narcolepsy type 2 and other disorders of sleepiness.[6] Cataplexy must be derived from patient report because it is rarely observed during the clinical evaluation for hypersomnolence. Sleep paralysis and hypnagogic and hypnopompic hallucinations are reported in about 50% of patients with narcolepsy; however, these symptoms are often present in individuals without the disease and thus are not specific.[62] Inquiry about nap duration and quality is also useful because patients with narcolepsy, in contrast to other sleep disorders, may experience a greater (although transient) alerting affect from short naps.[63] In contrast, more than two thirds of patients with idiopathic hypersomnia report that naps are nonrestorative.[64-66]

Physical Examination

Although the alerting effect of an examination obscures physical signs of sleepiness in most patients, overt signs of sleepiness—such as the inability to stay awake or to keep eyes open in the examination room—have high PPV and may obviate the need for additional tests. The examination may also help distinguish severe sleepiness from stupor due to neurologic impairment or drugs.

Sleep Logs and Actigraphy

The evaluation of hypersomnolence includes assessment of sleep duration and timing to rule out insufficient sleep syndrome or circadian rhythm sleep-wake disorders. The clinician should ask about sleep schedules at the time of clinical evaluation. Unfortunately, singular point estimates of sleep duration and timing demonstrate poor agreement with longitudinal measures.[67-69] Therefore tools such as sleep logs and actigraphy are valuable to track sleep patterns over days to weeks. Historically, sleep logs have been used to ensure adequate sleep duration before objective assessment of hypersomnolence with MSLT despite the absence of data supporting this use.[70]

An actigram is a device worn on the nondominant wrist that uses accelerometry to detect movement to estimate sleep and wake. Agreement between actigraphy and PSG in the detection of sleep is approximately 90%, and this device is accepted as a valid method to evaluate sleep patterns.[71-73] Actigraphy and sleep logs recorded for 2 weeks before MSLT were compared in a group of patients who underwent evaluation for hypersomnolence. Sleep logs were found to overestimate average nightly sleep duration by 1.43 hours (±1.31 hours) compared with actigraphy.[67] In the subgroup of patients whose MSLT results objectively confirmed excessive daytime sleepiness, average nightly sleep duration was 4.53 ± 1.37 hours by actigraphy, which was 2.55 ± 1.41 hours shorter than that recorded on sleep logs.[67] This study may be difficult to generalize because it was conducted in a military population. However, in a subsequent study at a large academic institution, poor agreement was seen between actigraphy and sleep logs typically owing to increased total sleep time reported on sleep logs compared with actigraphy.[74] Actigraphy is highly sensitive but not specific for EEG-defined sleep and may overestimate true sleep. Therefore the finding that sleep duration reported on sleep logs exceeds sleep duration derived by actigraphy is troublesome when determining whether sleep duration is adequate before MSLT. The ICSD-3 recommends documentation of sleep duration for 7 days on a sleep log and, whenever possible, actigraphy in conjunction with the sleep log, before the MSLT.[6]

Nocturnal Polysomnography

Many patients referred to sleep centers for excessive daytime sleepiness have nocturnal sleep disorders, and PSG is often more notable for the manifestations of such disorders than for signs of excessive daytime sleepiness. The single polysomnographic variable that best reflects sleepiness, as measured by the mean sleep latency on the MSLT, is nocturnal sleep latency.[75] Polysomnographic measures of sleep pathology, such as the AHI and minimum oxygen saturation, show only low magnitudes of correlation with MSLT results.[76] However, a short latency to REM sleep on overnight PSG can provide a valuable clue to the presence of narcolepsy. Among patients referred to a sleep laboratory, REM onset latency less than 15 minutes on nocturnal PSG had poor sensitivity (approximately 40%) but excellent specificity (99.6%) for type 1 narcolepsy.[77] The ICSD-3 now allows a sleep-onset REM period,

on overnight PSG, to account for one of the two REM periods necessary to diagnose narcolepsy with an MSLT.[6]

Multiple Sleep Latency Test

The mean sleep latency on the MSLT is the most commonly used objective measure in the assessment of daytime sleepiness.[78] The MSLT may contribute to diagnosis but is usually not sufficient, alone, to establish a diagnosis. The mean sleep latency is most useful when it is clearly abnormally low. A patient with a mean sleep latency of 2 minutes on a properly performed MSLT is unlikely to be exaggerating a complaint of excessive daytime sleepiness, to suffer from fatigue rather than sleepiness, or to be free of any sleep disorder. The MSLT can help determine the clinical significance of a sleep disorder or assess response to treatment.

As a general guideline, mean sleep latencies shorter than 8 minutes on a properly conducted MSLT are considered abnormal,[78] and latencies shorter than 5 minutes often indicate severe excessive daytime sleepiness. However, proper interpretation of MSLT results requires integration of other factors and especially knowledge of the limitations of this test. Results may be misleading if they are affected by youth (different criteria apply for children), noise, anxiety, or atypical sleep on the previous night. Use of medications such as stimulants or antidepressants, their recent discontinuance, or inability to be weaned off them at least 10 days before testing can complicate interpretation of an MSLT. Sleep apnea and other sleep disorders may make sleep onset more difficult and thereby interfere with the test. In general, the NPV of a long mean sleep latency is less than the PPV of a particularly short mean sleep latency. When an MSLT is normal, clinicians must carefully consider other possible explanations before telling a subjectively sleepy patient that there is no objective evidence of excessive daytime sleepiness. Formal prospective studies of "real-life" outcomes associated with different mean sleep latencies are still needed, but until such data are available, clinicians should realize that MSLT results form a continuum without strictly interpretable cutoffs. Community-based samples of adults show mean sleep latencies of 8 minutes or less in well more than 20% of subjects.[79] High test-retest reliability among normal subjects[80] does not necessarily generalize to patients.[81] In fact, 40% of central hypersomnia patients had mean sleep latencies that crossed to the other side of the 8-minute threshold when MSLTs conducted about 4 years apart were compared.[82] Interrater reliability can be excellent but adds another source of potential variation in test results.[83]

Nocturnal Polysomnography and MSLT in the Diagnosis of Narcolepsy

The diagnostic criteria for narcolepsy—two or more sleep-onset REM periods (SOREMPs) and short mean sleep latency—were once thought to have high sensitivity and specificity. Original case series suggested that all narcoleptic subjects and virtually no normal controls had two or more SOREMPs[84]; the PPV of two or more SOREMPs for the diagnosis of narcolepsy was 98%, and the NPV was 89%.[85] Subsequent studies did not find the SOREMP criteria to provide such diagnostic accuracy, partly because the most common reasons for sleep laboratory referral evolved. Two or more SOREMPs were found in 25% of 187 sleep apneic subjects,[86] 17% of 139 normal subjects,[87] and 83% of 200 narcoleptic subjects who had cataplexy.[88] Among 2083 patients evaluated with MSLTs at one sleep center, the PPV of two or more SOREMPs was 57% and the NPV was 98%.[89] Thus the presence of SOREMPs must be interpreted in conjunction with other clinical and polysomnographic findings. The criterion of two or more SOREMPs cannot be used to diagnose narcolepsy when the patient has untreated OSA. Furthermore, the number of SOREMPs can change enough to alter the diagnosis (idiopathic hypersomnia versus narcolepsy without cataplexy) in up to 30% of patients on repeated MSLT.[82] As an alternative to the MSLT, cerebrospinal fluid hypocretin-1 levels can be used to confirm type 1 narcolepsy. These levels are low (≤110 pg/mL or <one third of the mean for controls) in more than 90% of affected patients but almost never among patients without this diagnosis.[6]

Variations of the Multiple Sleep Latency Test and Other Physiologic Tests

Results of the Maintenance of Wakefulness Test (MWT) can differ markedly from those of the MSLT,[90] but whether the MWT results are more predictive of adverse effects of sleepiness in daily life remains unknown. Results of both the MWT and MSLT can be influenced by the patient's motivation.[91] The MWT results correlate with measures of sleep apnea severity to about the same extent as MSLT results do[92] but may better reflect improvement with treatment.[90] Shorter sleep latencies on MWTs correlate with increased errors on driving simulation tests.[93,94] However, until MWT and MSLT results are shown to differ in a clinically meaningful way, the MSLT continues to offer advantages of more published experience, familiarity among clinicians, and relevance to the diagnosis of narcolepsy. The Federal Aviation Administration and other agencies may at times request or require an MWT, but given the dearth of proven real-life predictive value, the role of this test or the MSLT in predicting workplace safety remains controversial.[95,96]

MSLT modifications, for which limited validity data exist, include addition of performance tasks[97,98]; analysis of sleep stages during naps[99]; focus on the percentage of time spent awake[100]; definition of sleep onset by failure to respond to a repeated signal[101]; and use of survival analysis to better account for failure to sleep on some naps.[102] In the clinical setting, none of these modifications have been adopted; neither have a range of other physiologic tests, including pupillometry and brainstem auditory evoked potentials. A variety of performance-based tests are used, usually in research settings, to assess variables related to sleepiness. Examples are the Psychomotor Vigilance Task[103] and the Steer Clear driving simulation test.[104]

Another available method to assess hypersomnia is the 24-hour PSG. About 40% of patients with idiopathic hypersomnia demonstrate mean sleep latencies greater than 8 minutes on the MSLT despite severe subjective sleepiness.[65,105,106] This finding is more common in patients with long nocturnal sleep durations. In these individuals, prolonged PSG reveals a total sleep duration near 700 minutes over 24 hours.[105] Therefore a 24-hour PSG that documents total sleep time of at least 660 minutes can be used to diagnose idiopathic hypersomnia in patients with symptoms consistent with the disorder but mean sleep latency greater than 8 minutes.[6] The ICSD-3 also allows actigraphy for this purpose; however, it has not been validated in this particular setting.[6]

EVALUATION OF INSOMNIA

History and Questionnaires

Like excessive daytime sleepiness, the complaint of inadequate, insufficient, or nonrestorative sleep can have many different causes. However, causes of insomnia are often diagnosed by history alone.[107,108] In part because the gold standard is not a physiologic test, few data are available with which to assess the relative value of individual symptoms. Predictive values for some symptoms are likely to be high because symptoms define the disorders. When a history does not reveal a cause of the insomnia, PSG may be useful (discussed under Nocturnal Polysomnography in this section). Psychometric tests can reveal cognitive differences between insomniac subjects and normal controls,[109] but these tests are not commonly used for diagnostic purposes in the clinical sleep medicine setting.

The Insomnia Severity Index is a seven-item self-report instrument that is typically used in insomnia research.[110] However, this tool may also be beneficial in the clinical setting. A score of 10 or higher on the Insomnia Severity Index identified insomnia with a sensitivity of 86% and a specificity of 88% in a community sample.[111]

Sleep Logs and Actigraphy

Sleep logs are an important tool in the evaluation of insomnia.[112] Patients record sleep-onset latency (SOL) and wake after sleep onset (WASO) on sleep logs, and investigators have tested the ability of different cutoffs of these quantitative parameters to predict insomnia. In one study, SOL or WASO of 31 minutes or longer identified insomnia with a sensitivity of 64% and specificity of 77% in subjects with insomnia at least 3 times per week for 6 months.[113] A subsequent investigation that also used sleep logs found that SOL or WASO of 20 minutes or more alone identified insomnia with a sensitivity of 94% and specificity of 80%.[114] Logs are not necessary to establish the presence of insomnia but can help define severity and facilitate identification of causes such as inadequate sleep hygiene or circadian rhythm sleep-wake disorders.

Actigraphy should be used with caution in patients with insomnia. The ability of actigraphy to measure sleep parameters (e.g., total sleep time or WASO) deteriorates as sleep efficiency decreases, which limits utility in patients with insomnia. Among patients with insomnia, in an epoch-by-epoch analysis that compared actigraphy with PSG, the accuracy, sensitivity, and specificity of actigraphy were 0.83, 0.95, and 0.35, respectively.[73] SOL is consistently underestimated by actigraphy in patients with insomnia.[115] Actigraphy is an effective tool to assess sleep-wake patterns when insomnia symptoms are thought to be secondary to a circadian rhythm sleep-wake disorder. However, because of the limitations described previously, actigraphy is not used routinely to confirm a diagnosis of insomnia.[108]

Nocturnal Polysomnography

PSG is not indicated for routine evaluation of insomnia; although when a patient's history and physical examination suggest that insomnia may be due to sleep-disordered breathing, periodic limb movement disorder, paradoxical insomnia, or uncertain causes, it can be an important aid to diagnosis.[116] Additionally, PSG may be indicated if insomnia fails to respond to treatment or in patients who have precipitous arousals with violent or injurious behavior.[108] Of note, injudicious use of PSG can sometimes enhance patient's conviction that insomnia is due to physical rather than behavioral causes or lead to diagnoses that eventually prove irrelevant to the main complaint.

EVALUATION OF SUSPECTED CIRCADIAN RHYTHM SLEEP-WAKE DISORDERS

History and Questionnaires

Discrepancies between the desired time for sleep and wake and the circadian propensity for sleep and wake may present as insomnia or hypersomnolence. Approximately 7% to 16% of patients who present to sleep disorders clinics with symptoms of insomnia are ultimately diagnosed with delayed sleep-wake phase disorder.[117,118] To distinguish circadian rhythm sleep-wake disorders from other causes of insomnia and hypersomnolence, examples of useful questions may include, "What time of day do you feel most alert?" and "When do you perform the best?" Comparison of regular sleep schedules to schedules on days free from work or school can reveal discrepancies that help to identify a circadian rhythm disorder.

Questionnaires such as the Horne-Ostberg Morningness-Eveningness Questionnaire (MEQ) and the Munich Chronotype Questionnaire (MCTQ) evaluate circadian preference, also known as chronotype.[119,120] The MEQ is the most widely used instrument to assess chronotype. It is a 19-item self-assessment tool that evaluates personal preference for the timing of sleep and other behaviors.[119] The MCTQ is also self-completed but assesses the actual (opposed to preferred) timing of sleep on work or school days versus free days.[120] The midpoint of sleep on free days and self-rated chronotype derived from the MCTQ both correlate highly with chronotype based on MEQ score (r = –0.7 and –0.8, respectively).[120,121] In addition to correlating with each other, the MEQ and MCTQ correlate with objective markers of circadian phase. The MEQ scores correlate with salivary dim-light melatonin onset (DLMO) (r = –0.4 to –0.5) and peak serum melatonin (r = –0.4).[122-124] The midpoint of sleep on free days derived from the MCTQ correlates with DLMO (r = 0.5).[124] The MEQ has also been validated against core body temperature and cortisol secretion.[119,125,126]

Sleep Logs and Actigraphy

The AASM recommends the use of sleep logs and, whenever possible, actigraphy for 7 to 14 days to evaluate suspected circadian rhythm sleep-wake disorders. Actigraphy has been validated in patients with circadian rhythm sleep-wake disorders.[72] Inclusion of ad libitum sleep-wake times (e.g., days off from school or work) is essential when sleep is tracked with actigraphy or sleep logs and will provide a more accurate estimate of true, endogenous circadian phase.[122,127,128] Vacation times may be particularly revealing; in fact sleep-wake times of individuals on an unrestricted sleep-wake schedule demonstrate higher correlation (r = 0.77) with DLMO than sleep-wake times of those on a fixed schedule (r = 0.40).[127]

Nocturnal Polysomnography

PSG is not necessary to diagnose circadian rhythm sleep-wake disorders, although it may be performed to rule out other comorbid sleep conditions. A PSG conducted at conventional times may demonstrate a delay of sleep onset or

early morning awakening in patients with delayed or advanced sleep-wake phase disorders, respectively.[129-131]

Multiple Sleep Latency Test

The MSLT is not used to diagnose circadian rhythm sleep-wake disorders; however, if obtained, mean sleep latency may be reduced in circadian rhythm sleep-wake disorders in the setting of sleep loss and excessive daytime sleepiness. Notably, in a large epidemiologic study, shift workers (night or rotating) were almost eight times more likely to have a mean sleep latency of less than 8 minutes combined with at least two SOREMPs on MSLT.[132] Additionally, adolescents with a delay in circadian phase can demonstrate SOREMPs during the MSLT (particularly during the first nap) when they wake according to their school schedule.[133]

Objective Markers of Circadian Phase

The *International Classification of Sleep Disorders*, third edition notes that endogenous markers of circadian phase can confirm the diagnosis of certain circadian rhythm sleep-wake disorders.[6] The salivary (DLMO) or urinary (6-sulfatoxymelatonin, aMT6s) melatonin assays are the most commonly used objective markers of circadian phase. These measures objectively document a stable advance, stable delay, or progressive delay of circadian phase in advanced sleep-wake phase disorder, delayed sleep-wake phase disorder, and non–24-hour sleep-wake rhythm disorder, respectively.[6] Although melatonin assays are infrequently used in clinical practice, at-home DLMO assays correlate well with in-lab DLMO assessments (r = 0.85 when a fixed threshold of 3 pg/mL is used).[134] Salivary melatonin assays can capture the onset of melatonin secretion but are not practical to determine the secretion profile overnight in the ambulatory setting. Alternatively, urinary aMT6s can be collected at 8-hour intervals; therefore the first morning void allows for calculation of overnight aMT6s secretion in the home setting.[135] Further investigation is required to determine whether ambulatory measurement of circadian phase markers is reliable and feasible.

EVALUATION OF RESTLESS LEGS SYNDROME

History and Questionnaires

The diagnosis of RLS is made by a clinical history of an urge to move the limbs that is worse at rest, improved with movement, and worse in the evening or night.[6] These four criteria have a PPV of 76% when expert interview is used as a gold standard.[136] Differentiating RLS carefully from leg cramps or positional discomfort improves the specificity of the four criteria from 84% to 94%.[6,136]

Several instruments exist to assist in the evaluation of RLS or its severity, including the International Restless Legs Scale (IRLS), RLS-6, and Johns Hopkins Severity Scale. The IRLS scale is a 10-item questionnaire that assesses the severity of RLS symptoms.[137] This scale has good internal consistency, interexaminer reliability, and test-retest reliability,[137] and a 6-point decrease is considered to be a clinically relevant improvement.[138] Although these scales are used mostly in research, they may be beneficial in the clinic setting to quantify symptom severity; determine the impact of RLS symptoms on patient quality of life, mood, and sleep; measure the progression of RLS symptoms; and evaluate therapeutic response.[139]

Physical Examination

A full neurologic examination is indicated to evaluate for RLS because this condition may arise in the context of other neurologic diseases, such as neuropathy, multiple sclerosis, or Parkinson disease. Assessment of affect and mood to help identify psychiatric disorders is important because mental health morbidity frequently coexists with RLS.

Laboratory Tests

Evaluation of a patient with RLS should include serum iron and ferritin levels. More than one third of individuals with RLS have low serum iron levels, and greater than two thirds have ferritin values of 50 ng/mL or less. Depression, fatigue, and increased severity of symptoms are associated with low serum iron levels in RLS patients.[140] Ferritin levels are inversely related to RLS severity.[141,142] Iron supplementation may reduce symptoms in RLS, although findings are inconsistent.[143] Therefore evaluation of serum iron and ferritin is an integral part of the evaluation of RLS for both diagnosis and treatment.

Nocturnal Polysomnography

PSG is not routinely indicated in the evaluation of RLS and should be performed only if the clinician suspects a comorbid sleep disorder such as OSA. Periodic limb movements during sleep are found in up to 90% of patients with RLS. However, periodic limb movements during sleep are nonspecific because they also occur in approximately 25% of individuals without RLS.[144,145]

EVALUATION FOR SUSPECTED PARASOMNIAS

History and Questionnaires

With the notable exception of REM sleep behavior disorder (RBD), parasomnias often can be diagnosed by history alone.[6] Information obtained from a bed partner may contribute more than that obtained from the patient.

Physical Examination

The physical examination of patients evaluated for parasomnias can be useful, but its value is not well quantified. Some signs may suggest sleep apnea as an underlying trigger for confusional arousals, sleepwalking, sleep terrors, RBD, or nocturnal enuresis. Worn occlusive surfaces of molars can provide key evidence of sleep bruxism. The urogenital examination is important in patients thought to have sleep enuresis. A neurologic examination may suggest a primary cause of sleep enuresis or RBD. Similarly, appropriate laboratory findings may be helpful in some cases; for example, a urinalysis may reveal the cause of sleep enuresis.

Nocturnal Polysomnography

Few studies have examined the predictive value of PSG for parasomnia diagnoses. When the behavior in question occurs during the PSG, the diagnostic value of the test is likely to be high, especially if appropriate additional recording devices, such as extra EEG leads, extra surface electromyogram (EMG) leads, or video monitoring, are used.[146] Additional EEG leads used during PSG, combined with clinical history, may effectively differentiate sleep-related epilepsy from parasomnias.

However, EEG does not reliably diagnose nocturnal frontal lobe epilepsy (NFLE) because more than 60% of patients with

NFLE fail to demonstrate a definite ictal rhythm.[147] Therefore semiology of the events is the key to diagnosis. Derry and colleagues created a rigorous decision tree algorithm based on 120 nocturnal events recorded on video PSG to distinguish NREM disorders of arousal from NFLE.[147] The algorithm included the following characteristics suggestive of NFLE as opposed to parasomnia: complete arousal after the ictus, discrete offset of the behavior, presence of versive head turning or posturing, and persistence of recumbent posture. This decision tree algorithm classified 94% of events correctly. Unfortunately, PSG often fails to document the behavior—especially in cases of suspected RBD, sleepwalking, night terrors, and epilepsy—either because the behavior does not occur on most nights or perhaps because the sleep laboratory is not an environment familiar to the patient. For evaluation of parasomnias, the NPV of a completely normal study is less clear than the PPV of an abnormal study. In one series of 122 patients with suspected parasomnias, one or two nights of PSG with video monitoring contributed useful diagnostic information in more than 50% of cases.[146]

Even in the absence of abnormal behaviors on the night of the PSG, other findings can be valuable. Examples include excessive limb twitching during REM sleep characteristic of RBD and interictal spike and wave complexes that may represent an interictal expression of epilepsy. REM sleep without atonia (RSWA) is the PSG hallmark of RBD and is required to confirm the diagnosis.[6] RSWA may be quantified manually by the eye or by automated computer programs. When scored manually, RSWA is defined by the AASM manual for the scoring of sleep and associated events as either excessive phasic or tonic elevation in EMG tone.[28] Multiple computer algorithms exist to automatically score RSWA, including the REM atonia index and supra-threshold REM EMG activity metric.[148] A supra-threshold REM EMG activity metric cutoff of 15 or higher is able to detect RBD with 100% sensitivity and 71% specificity.[149]

At this time, manual scoring remains the gold standard to detect RSWA.[148] However, the AASM manual does not specify a minimum number or proportion of epochs that must contain RSWA to meet the PSG criteria to confirm suspected RBD.[28] To address this, different cut points have been investigated. When using the submentalis muscle alone, the presence of RSWA (phasic) in 15% of 2-second REM mini-epochs will correctly classify 84% of patients.[150] The Sleep Innsbrook Barcelona group has extensively tested different methods to quantify EMG tone in stage REM sleep and recommends recording EMG in both the submentalis muscle and the bilateral flexor digitorum superficialis muscles to score RSWA.[151] Use of this montage with specificity set at 100% (no false-positive RBD diagnoses permitted) yields a cut point of 32% of 3-second REM mini-epochs to diagnose RBD (area under the receiving operator characteristic curve = 0.998).[151] Upper limb as opposed to lower limb EMG more reliably distinguishes patients with RBD from those without RBD.[148]

BEYOND SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUE: DECISION AND COST-EFFECTIVENESS ANALYSES

Data on sensitivity, specificity, pretest probability, and utility of outcomes can be used to construct a decision analysis. A clinical decision analysis typically models a choice between diagnostic and therapeutic alternatives. Logical rules are used to weigh information and to make the best decision for an individual patient.[152] Decision analysis may be useful, for example, when one procedure has a high probability of a small benefit but an alternative has a low probability of a large benefit.

Beyond utility, economic data on a procedure can include cost studies, cost-effectiveness analyses that compare different methods to achieve the same end, cost-utility analyses that compare costs per common unit of utility (often quality-adjusted life years), and cost-benefit analyses that compare monetary costs with monetary gains.[153] Such studies require quantitative information on costs and outcomes, data that are not abundant for sleep disorders.[95,96] Despite uncertainty of some important data points, cost-utility models have focused on the decision of whether to diagnose OSA with the aid of a full-night PSG, split-night PSG, portable cardiorespiratory monitoring, or no ancillary test.[47,48] The full-night PSG costs less and results in more quality-adjusted life-years than split-night PSG or unattended portable monitoring over the course of an individual's lifetime.[48] Despite the increased upfront cost of full-night PSG, these results reflect the high utility of an accurate OSA diagnosis and the expense of diagnostic mistakes. These findings highlight the importance of performing decision and cost-utility analyses before conclusions are made about relative values of diagnostic tests.

CLINICAL PEARL

Evaluation of common sleep complaints is based on symptoms, signs, and test results, combined with an understanding of the diagnostic value that each type of data contributes.

SUMMARY

Clinical tools and tests must be used carefully in the evaluation of suspected sleep disorders. All patients with sleep-related complaints should undergo a history and physical examination. Evaluations for obstructive sleep-disordered breathing start with a history and physical examination, which generate valuable information. Symptom-based questionnaires alone usually show inadequate specificity. Laboratory-based nocturnal polysomnography is a gold standard but not infallible. Final diagnostic decisions should be based on integration of multiple clinical and objective data points rather than on any specific cutoff for one specific variable, such as the apnea-hypopnea index. Evaluation for hypersomnolence also relies on historical symptoms collected during an interview or by use of a questionnaire. Objective testing with an MSLT is particularly useful when a shortened mean sleep latency confirms excessive daytime sleepiness or a shortened latency in addition to sleep-onset REM periods confirm narcolepsy. Results must be interpreted carefully, especially when they are normal, because of potential confounds. Evaluation for insomnia often relies solely on historical information. Sleep logs can be helpful, but polysomnography is indicated only when other occult sleep disorders may underlie the insomnia. Questionnaires to determine chronotype and actigraphy are valuable tools when symptoms suggest a circadian rhythm sleep-wake disorder. RLS is a diagnosis based on

clinical history, but serum iron studies provide valuable information with repercussions for treatment. Evaluation for a parasomnia starts with a thorough history, obtained whenever possible from a bed partner in addition to the patient. Polysomnography may confirm a diagnosis or distinguish between several possibilities.

Selected Readings

Andlauer O, Moore H, Jouhier L, et al. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. *JAMA Neurol* 2013;**70**(7):891–902.

Aurora RN, Collop NA, Jacobowitz O, et al. Quality measures for the care of adult patients with obstructive sleep apnea. *J Clin Sleep Med* 2015;**11**(3):357–83.

Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2012;**8**(5):597–619.

Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2007;**3**:737–47.

Edinger JD, Buysse DJ, Deriy L, et al. Quality measures for the care of patients with insomnia. *J Clin Sleep Med* 2015;**11**(3):311–34.

Epstein LJ, Kristo D, Strollo PJ Jr, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;**5**(3):263–76.

Frauscher B, Ehrann L, Högl B. Defining muscle activities for assessment of rapid eye movement sleep behavior disorder: from a qualitative to a quantitative diagnostic level. *Sleep Med* 2013;**14**(8):729–33.

Krahn LE, Hershner S, Loeding LD, et al. Quality measures for the care of patients with narcolepsy. *J Clin Sleep Med* 2015;**11**(3):335–55.

Littner MR, Kushida C, Wise M, et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep* 2005;**28**:113–21.

Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep* 2007;**30**(4):519–29.

Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea? The Rational Clinical Examination systematic review. *JAMA* 2013;**310**(7):731–41.

Pietzsch JB, Garner A, Cipriano LE, Linehan JH. An integrated health-economic analysis of diagnostic and therapeutic strategies in the treatment of moderate-to-severe obstructive sleep apnea. *Sleep* 2011;**34**(6):695–709.

Trotti LM, Goldstein CA, Harrod CG, et al. Quality measures for the care of adult patients with restless legs syndrome. *J Clin Sleep Med* 2015;**11**(3):293–310.

A complete reference list can be found online at ExpertConsult.com.

Chapter
61

Classification of Sleep Disorders

Michael J. Sateia; Michael J. Thorpy

Chapter Highlights

- The classification of sleep disorders is necessary to discriminate among disorders and to facilitate an understanding of symptoms, etiology, pathophysiology, and treatment.
- The *International Classification of Sleep Disorders*, third edition (ICSD3),[1] published in 2014, combines a symptomatic presentation (e.g., insomnia) with one organized in part on pathophysiology (e.g., circadian rhythms) and in part on organ systems (e.g., breathing disorders).
- Major sections of the ICSD3 include insomnia, sleep-related breathing disorders, central disorders of hypersomnolence, circadian rhythm sleep-wake disorders, parasomnias, and sleep-related movement disorders.
- The ICSD3 is not only a listing of sleep disorders but also a compendium of major diagnostic features, associated conditions, course, prognosis, developmental features, epidemiology, and pathophysiology.

The classification of sleep disorders has been of particular interest to clinicians since sleep disorders were first recognized. The first major classification, the *Diagnostic Classification of Sleep and Arousal Disorders*,[2] published in 1979, organized the sleep disorders into categories that formed the basis of the current classification systems. The initial *International Classification of Sleep Disorders* (ICSD) was produced in 1990 and revised in 1997. In 2005, the ICSD second edition (ICSD2) was published and included a significant reorganization of the nosology, an approach that has been maintained in the most recent (ICSD3) manual, released in 2014. The *International Classification of Sleep Disorders* is published by the American Academy of Sleep Medicine, with consultation and review by panels of international experts and sleep societies worldwide. The ICSD system, developed primarily for diagnostic, epidemiologic, and research purposes, has been widely used by clinicians and has allowed better international communication in sleep disorder research.

The ICSD3 classification (Table 61-1) lists 59 sleep disorders, each presented in detail and with a descriptive diagnostic text that includes specific diagnostic criteria and coding recommendations. The ICSD3 has seven major sections: (1) insomnia, (2) sleep-related breathing disorders, (3) central disorders of hypersomnolence, (4) circadian rhythm sleep-wake disorders, (5) parasomnias, (6) sleep-related movement disorders, and (7) other sleep disorders.

Two additional sleep disorders classification systems are also in current use. The *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5)[3] of the American Psychiatric Association includes a sleep disorders section that was designed for mental health and general medical clinicians who are not experts in sleep medicine. Efforts to achieve consistency between DSM-5 and ICSD3 were largely successful during the parallel development of these two approaches, although some discrepancies in diagnostic criteria do exist. Notably, the ICSD system is significantly more detailed, as expected in light of the differing target audience. The U.S. clinical modification (CM) of the *International Classification of Diseases,* 9th and 10th revisions (ICD-9-CM and ICD-10-CM),[4] includes a classification and coding for sleep disorders primarily for statistical and epidemiologic purposes, which was substantially revised with the publication of ICSD2. Nevertheless, the crosswalk between ICSD and the ICD system is complex in light of the significant differences that exist between these systems. ICSD3 includes several features not found in ICSD2. All definitions of polysomnography (PSG) findings (e.g., apnea/hypopnea, periodic limb movement, rapid eye movement [REM] without atonia) refer to the most recent version of the *American Academy of Sleep Medicine (AASM) Manual for the Scoring of Sleep and Associated Events.*[5] A separate Developmental Issues section has been added to the text to more effectively address variations in presentation and findings across age ranges, especially in children and adolescents. ICD-9-CM and ICD-10-CM code recommendations are included for each diagnosis. Conditions that were previously included in a separate chapter of ICSD2 (Isolated Symptoms and Normal Variants) have now been assigned to the major chapters to which they are most closely related (e.g., short sleep to insomnia and snoring to sleep-related breathing disorders).

INSOMNIA

Historically, insomnia has been characterized as either primary or secondary (comorbid).[6] The latter was intended to describe insomnia resulting from a medical or psychiatric illness, another sleep disorder, or substance abuse. However, current conceptualization of chronic insomnia has resulted in elimination of this dichotomy. ICSD3 employs a diagnosis of chronic insomnia disorder,[7,8] which includes all insomnias of at least 3 months' duration, regardless of presumed etiology. Insomnia of less than 3 months' duration is classified as short-term insomnia. A

Table 61-1 Sleep Disorder Diagnoses and Codes

Proposed ICSD3 Diagnoses	Proposed ICD-9CM Diagnosis and Codes	Proposed ICD-10CM Diagnosis and Codes
Insomnia Disorders		
Chronic insomnia disorder	307.42 Persistent insomnia—nonorganic origin	F51.01 Other insomnia not due to physiologic or substance
Short-term insomnia disorder	307.41 Transient insomnia	F51.02
Sleep-Related Breathing Disorders		
Obstructive Sleep Apnea		
Obstructive sleep apnea (adult and pediatric)	327.23	G47.33
Central Sleep Apnea		
Central sleep apnea with Cheyne-Stokes breathing	786.04 Cheyne-Stokes respiration	R06.3
Central sleep apnea due to a medical disorder without Cheyne-Stokes breathing	327.27	G47.37
Central sleep apnea due to high-altitude periodic breathing	327.22 High-altitude periodic breathing	G47.32
Central sleep apnea due to drug or a substance	327.29 Other organic sleep apnea	G47.39
Primary central sleep apnea	327.21	G47.31
Primary central sleep apnea of infancy	770.81 Primary apnea of newborn	P28.3
Primary central sleep apnea of prematurity	770.82 Other apnea of newborn	P28.4
Treatment emergent central sleep apnea	327.29 Other organic sleep apnea	G47.39
Hypoventilation/Hypoxemia		
Obesity hypoventilation syndrome	278.03 Morbid obesity with alveolar hypoventilation	E66.2
Congenital central alveolar hypoventilation syndrome	327.25 Congenital central alveolar hypoventilation syndrome	G47.35
Late-onset central hypoventilation with hypothalamic abnormalities	327.26 Sleep-related hypoventilation/hypoxemia in conditions classifiable elsewhere	G47.36
Idiopathic central alveolar hypoventilation	327.24	G47.34
Sleep-related hypoventilation due to drug or substance	327.26 Sleep-related hypoventilation/hypoxemia in conditions classifiable elsewhere	G47.36
Sleep related hypoventilation due to medical or neurologic condition	327.26 Sleep-related hypoventilation/hypoxemia in conditions classifiable elsewhere	G47.36
Sleep-related hypoxemia	327.26 Sleep-related hypoventilation/hypoxemia in conditions classifiable elsewhere	G47.36

Continued

Table 61-1 Sleep Disorder Diagnoses and Codes—cont'd

Proposed ICSD3 Diagnoses	Proposed ICD-9CM Diagnosis and Codes	Proposed ICD-10CM Diagnosis and Codes
Hypersomnolence Disorders		
Narcolepsy— type 1 (with cataplexy and/or hypocretin deficiency)	347.01 Narcolepsy with cataplexy	G47.411
Narcolepsy—type 2 (without cataplexy or hypocretin deficiency)	347.00 Narcolepsy without cataplexy	G47.419
Idiopathic hypersomnia	327.11 Idiopathic hypersomnia with long sleep time	G47.11
Kleine-Levin syndrome (recurrent hypersomnia)	327.13 Recurrent hypersomnia	G47.13
Hypersomnia due to drug or substance	292.85 (drug induced) 291.82 (alcohol induced)	F11–19
Hypersomnia due to medical condition	327.14	G47.14
Hypersomnia associated with psychiatric disorder	327.15 Hypersomnia due to mental disorder	F51.13 Hypersomnia due to mental disorder
Insufficient sleep syndrome	307.44	F51.12
Circadian Rhythm Sleep-Wake Disorders		
Delayed sleep phase disorder	327.31	G47.21
Advanced sleep phase disorder	327.32	G47.22
Irregular sleep-wake rhythm disorder	327.33	G47.23
Nonentrained (free-running) disorder	327.34	G47.24
Shift work disorder	327.36	G47.26
Jet lag disorder	327.35	G47.25
Circadian rhythm sleep-wake disorder, not otherwise specified	327.30 (other)	G47.20 CRSWD, unspecified
Parasomnias		
Confusional arousals	327.41 Confusional arousals	G47.51
Sleepwalking	307.46 Sleepwalking	F51.3
Sleep terrors	307.46 Sleep terrors	F51.4
Sleep-related eating disorder	327.49 Other organic parasomnia	G47.59
REM sleep behavior disorder	327.42 REM sleep behavior disorder	G47.52
Recurrent isolated sleep paralysis	327.43 Recurrent isolated sleep paralysis	G47.53
Nightmare disorder	307.47 Other dysfunctions of sleep stages or arousal	F51.5 Nightmare disorder
Exploding head syndrome	327.49 Other organic parasomnia	G47.59
Sleep-related hallucinations	368.16 Psychophysical visual disturbances	H53.16
Sleep enuresis	788.36 Nocturnal enuresis	N39.44
Parasomnia due to a medical disorder	327.44	G47.54
Parasomnia not otherwise specified	327.40 Parasomnia NOS	G47.50

Table 61-1 Sleep Disorder Diagnoses and Codes—cont'd

Proposed ICSD3 Diagnoses	Proposed ICD-9CM Diagnosis and Codes	Proposed ICD-10CM Diagnosis and Codes
Sleep-Related Movement Disorders		
Restless legs syndrome	333.94	G25.81
Periodic limb movement disorder	327.51	G47.61
Sleep-related bruxism	327.53 Sleep-related bruxism	G47.63
Sleep-related leg cramps	327.52 Sleep-related leg cramps	G47.62
Rhythmic movement disorder	327.59 Other sleep-related movement disorder	G47.69
Benign sleep myoclonus of infancy	327.59 Other sleep-related movement disorder	G47.69
Propriospinal myoclonus at sleep onset	327.59 Other sleep-related movement disorder	G47.69
Sleep-related movement disorder due to drug or substance	327.59 Other sleep-related movement disorder	G47.69
Sleep-related movement disorder due to medical condition	327.59 Other sleep-related movement disorder	G47.69
Sleep-related movement disorder not otherwise specified	327.59 Other sleep-related movement disorder	G47.69

diagnosis of other insomnia may be employed when a patient presents with insomnia complaints but does not meet full criteria for either chronic or short-term insomnia.

In ICSD3, insomnia is defined as persistent sleep difficulty despite adequate opportunity and circumstances for sleep, which is accompanied by daytime consequences that are attributable to the sleep disturbance. Complaints may include difficulty with sleep initiation, sleep maintenance, or early awakening. In addition, resistance to going to bed on an appropriate schedule or difficulty sleeping without the intervention of a parent or caregiver constitutes an insomnia complaint (see Chapters 80 to 88). The disturbance may be reported by the patient or by the patient's parent or caregiver. Although any one of these complaints meets the first criterion for an insomnia diagnosis, it is not uncommon for patients to present with two or more of these symptoms (e.g., difficulty initiating and maintaining sleep). In the case of children or cognitively impaired adults, it is often the parent or caregiver who reports the problem. Symptoms of bedtime resistance or requirement of parent or caregiver intervention applies primarily to these groups. An ICSD3 insomnia diagnosis also requires associated daytime consequences (e.g., fatigue, impaired concentration, mood disturbance, or other occupational, social, or academic impairment) and adequate opportunity and environmental circumstances for sleep. A duration of at least 3 months and the presence of symptoms at least three times per week are necessary to establish a diagnosis of chronic insomnia. When the symptoms have been present for less than 3 months, a diagnosis of short-term insomnia disorder is applied. Insomnia symptoms occur frequently in conjunction with many medical and psychiatric disorders. A diagnosis of chronic insomnia disorder should be invoked only when the insomnia component is the focus of independent clinical assessment and treatment.

ICSD3 discusses previously identified subtypes within the context of chronic insomnia disorder. Although these clinical subtypes are no longer considered independent diagnoses, there may be characteristics of these subtypes that are clinically relevant. Major subtypes of what was previously termed *primary insomnia* include (1) psychophysiologic insomnia, characterized by a heightened level of arousal with learned sleep-preventing associations and an excessive concern with the inability to sleep; (2) paradoxical insomnia (formerly known as *sleep state misperception*), which is a complaint of severe insomnia that occurs without evidence of objective sleep disturbance and without daytime impairment to the extent that would be suggested by the amount of sleep disturbance reported; (3) idiopathic insomnia, a long-standing form of insomnia that appears to date from childhood and has an insidious onset; and (4) behavioral insomnia of childhood,[9] including limit-setting sleep disorder and sleep-onset association disorder. The former is a resistance or refusal to go to sleep, which is the result of insufficient limit-setting on the part of care providers. Sleep-onset association disorder occurs when there is reliance on inappropriate sleep associations, such as rocking, watching television, holding a bottle or other object, or specific environmental conditions such as a lighted room or an alternative place to sleep. Previously identified secondary, or comorbid, insomnias include (1) insomnia due to medical condition, which has been applied when a medical or neurologic disorder is believed to give rise to the insomnia; (2) insomnia due to drug or substance, when excessive use, dependence on, or withdrawal from a substance such as alcohol, a recreational drug, or caffeine is causative; and (3) insomnia due to mental disorder, which is employed when an underlying mental disorder is the major etiologic factor.

Patients often present with multiple comorbidities that may contribute to an insomnia complaint. It is often difficult

to ascertain the cause-and-effect relationship between comorbidities and the insomnia. When an ICSD3 diagnosis of chronic insomnia is employed, clinicians are encouraged to list all pertinent comorbidities along with the diagnosis.

Short sleep and excessive time in bed are listed as normal variants within the insomnia section. A short sleeper is a person with a routine pattern of obtaining 6 hours or less of sleep in a 24-hour day without sleep complaints or identifiable daytime consequences. In children, this sleep length can be 3 hours or less than the norm for the age group.

SLEEP-RELATED BREATHING DISORDERS

The disorders in this group are characterized by disordered respiration during sleep (see Chapters 107 to 122). ICSD3 includes four major categories of sleep-related breathing disorders: (1) obstructive sleep apneas (OSAs), (2) central sleep apnea (CSA) syndromes, (3) sleep-related hypoventilation disorders, and (4) sleep-related hypoxemia disorder.

The OSAs are disorders in which obstruction of the airway during sleep results in reduced or absent airflow despite adequate respiratory effort. OSA, adult,[10,11] is characterized by repetitive episodes of cessation of breathing (apneas), reduced breathing (hypopneas), or arousal associated with increased airway resistance and respiratory effort (respiratory effort related arousal). The term *upper airway resistance syndrome* is no longer employed because the underlying pathophysiology and potential consequences are essentially those of OSA. Heavy snoring is reported in most of these patients. Apneic and hypopneic events are often associated with reduced blood oxygen saturation. Diagnosis requires the presence of five or more events (apnea, hypopneas, or respiratory effort related arousals) coupled with at least one sign or symptom (e.g., snoring, observed pauses, excessive sleepiness, insomnia) or medical or psychiatric complications. A predominantly obstructive event frequency of greater than 15/hour meets diagnostic criteria regardless of the presence or absence of symptoms.

OSA, pediatric,[12] is characterized by features similar to those seen in the adult, but cortical arousals may not occur, possibly because of a higher arousal threshold. This may give rise to a pattern of obstructive hypoventilation, which may require CO_2 monitoring for detection. Signs or symptoms are required for a diagnosis of pediatric OSA. These must be coupled with at least one obstructive event per hour of sleep or a pattern of hypoventilation, as evidenced by $Paco_2$ greater than 50 mm Hg for more than 25% of sleep time.

CSA syndromes[13,14] include those in which airflow is diminished or absent in an intermittent or cyclical fashion as a result of reduced or absent respiratory effort. ICSD3 includes nine CSA syndromes. All of the adult forms of these disorders, with the exception of high-altitude periodic breathing, require PSG demonstration of a central apnea or hypopnea index greater than 5/hour. Adult presentations also require associated signs or symptoms (e.g., sleepiness, sleep disturbance, awakening short of breath, snoring, or witnessed apnea), although presence of a medical condition such as congestive heart failure, stroke, or renal failure precludes this requirement in CSA with Cheyne-Stokes breathing. CSA with Cheyne Stokes breathing[15] is characterized by recurrent central apneas or hypopneas alternating with a respiratory phase in which tidal volume waxes and wanes in a crescendo-decrescendo pattern. This pattern is characteristically seen in non–rapid eye movement (NREM) sleep. CSA due to a medical condition without Cheyne-Stokes breathing is generally a result of brainstem lesions of varying etiologies. CSA due to high-altitude periodic breathing[16,17] is seen following recent ascent to altitude, typically greater than 2500 meters, although some individuals may experience the symptoms at lower altitudes. The condition may be diagnosed on the basis of ascent to altitude and symptoms alone, although PSG, if performed, shows a central index greater than 5/hour. CSA due to drug or substance[18] is most commonly associated with long-term opioid use. The substance causes a respiratory depression by acting on the μ receptors of the ventral medulla. Primary CSA is a disorder of unknown cause characterized by recurrent episodes of cessation of breathing during sleep without associated ventilatory effort. A complaint of excessive daytime sleepiness, insomnia, or difficulty breathing during sleep is reported. The patient must not be hypercapnic (Pco_2 >45 mm Hg).

Primary sleep apnea of infancy (if conceptional age of the child is 37 weeks or greater) or prematurity (when conceptional age is less than 37 weeks) is a disorder of respiratory control caused by developmental issues (immaturity of brainstem respiratory centers) or other medical disorders. Diagnosis requires observation of an episode of apnea or cyanosis or detection of apnea or desaturations by monitoring. Recurrent, prolonged (>20 seconds) central apnea or periodic breathing for more than 5% of total sleep time must be demonstrated.

ICSD3 includes a new CSA diagnosis: treatment emergent central sleep apnea.[19] This disorder, which has been referred to in the literature as complex sleep apnea, is characterized by predominantly obstructive apnea on baseline PSG, with resolution of obstruction and emergence or persistence of predominantly central apnea during administration of positive airway pressure without backup rate. The term *complex sleep apnea* has also been used to describe emergence of CSA with Cheyne-Stokes breathing or due to drug or substance in the context of treated OSA. However, a diagnosis of treatment emergent CSA should not be applied when another etiology for the CSA is established. In such cases, clinicians should diagnose both OSA and CSA due to Cheyne-Stokes breathing or substance.

Sleep-related hypoventilation disorders[20] and sleep-related hypoxemia disorder, previously subsumed under a single heading, are now distinct diagnoses. The hypoventilation disorders comprise six disorders associated with hypoventilation during sleep. Hypoventilation must be established by demonstration of elevated Pco_2 (as defined in the most recent version of the AASM scoring manual) by blood gas or, more commonly, by proxy measures such as end-tidal or transcutaneous CO_2.

Obesity hypoventilation[21,22] requires demonstration of daytime hypercapnia, whereas other sleep-related hypoventilation disorders require only sleep-related hypoventilation and may or not be associated with daytime hypercapnia. Congenital central alveolar hypoventilation syndrome[23] is a failure of automatic central control of breathing associated with mutation of the *PHOX2B* gene. The hypoventilation begins in infancy and worsens during sleep. Idiopathic central alveolar hypoventilation refers to sleep-related hypoventilation that is not attributable to another disorder. Late-onset central hypoventilation with hypothalamic dysfunction[24] is newly added to ICSD3. It is characterized by onset of symptoms

after the first several years of life. In addition to hypoventilation, symptoms may include obesity, endocrine abnormalities of hypothalamic origin, emotional and behavioral disturbances, and neural tumors. Sleep-related hypoventilation due to a medical condition[25] may result from pulmonary airway or parenchymal disease, extrinsic factors such as chest wall disorder, or neuromuscular disease. Sleep-related hypoventilation may also be caused by substances such as opioid or other respiratory depressants.

Sustained declines in Po_2 (Sao_2 <88% [90% for children] for ≥5 minutes), in the absence of demonstrated elevation of Pco_2, is diagnosed as sleep-related hypoxemia disorder.

Snoring is an isolated symptom included in this section and is identified when a respiratory sound, typically associated with inspiration, is disturbing to the patient, a bed partner, or others. This term applies when the snoring occurs without evidence of upper airway obstructive events or sleep-wake complaint such as insomnia or excessive sleepiness. Not only can snoring lead to impaired health, but it may also be a cause of social embarrassment and can disturb the sleep of a bed partner.

CENTRAL DISORDERS OF HYPERSOMNOLENCE

Central disorders of hypersomnolence are characterized by a primary complaint of daytime sleepiness that is not attributable to another sleep disorder[26,27] (see Chapters 89 to 91 and 100). Most of these conditions are caused by central nervous system abnormalities or the effects of substances or other disorders on the central nervous system. One exception to this is insufficient sleep syndrome, which results from behaviorally induced sleep deprivation. Daytime sleepiness is defined as the inability to stay alert and awake during the major waking episodes of the day, typically resulting in unintended lapses into sleep. Other sleep disorders may be present, and they must first be effectively treated to establish a diagnosis of a hypersomnia disorder. The disorders previously named narcolepsy with cataplexy and narcolepsy without cataplexy[28-32] are now termed *narcolepsy type 1* and *type 2*. Narcolepsy type 1 is diagnosed on the basis of complaint of excessive sleepiness and the presence of definite cataplexy plus multiple sleep latency test (MSLT) findings (mean sleep latency ≤8 minutes and evidence of two or more sleep-onset REM periods [SOREMPs]. More recent evidence suggests that a SOREMP on the PSG before the MSLT is highly specific for narcolepsy type 1.[33] Therefore a nocturnal PSG SOREMP (within 15 minutes of sleep onset) may substitute for one of the MSLT SOREMPs in making the diagnosis. Alternatively, narcolepsy type 1 can be diagnosed when subjective sleepiness and hypocretin deficiency are present, even in the absence of cataplexy. Some patients with hypocretin deficiency may not manifest cataplexy, at least at the time of initial diagnosis. It is for this is reason that the term *narcolepsy with cataplexy* has been rendered obsolete. Narcolepsy type 2 criteria include subjective sleepiness and the MSLT findings described previously for narcolepsy type 1. Cataplexy is absent, and hypocretin levels, if obtained, must not meet type 1 criteria.

Idiopathic hypersomnia[34] is the diagnosis given to those with subjective sleepiness complaints that are not explained by other sleep disorders, medical illness, or psychiatric illness. An MSLT latency of 8 minutes or less is required with fewer than two SOREMPs (including a SOREMP on the preceding night's PSG, if present) for diagnosis. It is recognized that some patients with legitimate sleepiness problems may not demonstrate mean latencies of 8 minutes or less. Clinical judgment is required in such cases. Particular care must be exercised to exclude insufficient sleep as a cause or contributing factor before a diagnosis of idiopathic hypersomnia is established.

Kleine-Levin syndrome[35,36] is a form of recurrent hypersomnia that persists for days to several weeks and is associated with one or more of the following: eating disorder (most commonly hyperphagia), cognitive dysfunction, perceptual disturbance, or uninhibited behavior, often sexual. In some women, recurrent hypersomnia may manifest as a menses-related phenomenon and is now termed *menstrual-related Kleine-Levin syndrome*.

Insufficient sleep syndrome occurs in patients who maintain a habitually short sleep episode relative to age-appropriate norms. These patients typically sleep significantly longer when schedules allow (i.e., weekends or vacation). Extended sleep resolves the sleepiness. Consideration must be given to the possibility of long sleep requirement in patients presenting with otherwise unexplained sleepiness and ostensibly normal sleep times.

Hypersomnia due to medical condition is excessive sleepiness that is caused by a medical or neurologic disorder.[37,38] Cataplexy or other diagnostic features of narcolepsy are not present. Hypersomnia due to drug or substance is diagnosed when the complaint is believed to be secondary to current or past use of drugs. Hypersomnia associated with a psychiatric disorder is excessive sleepiness that is temporally associated with a psychiatric diagnosis, although, as the name implies, cause-and-effect relationships between true hypersomnolence and psychiatric disorders are not well established.

Long sleeper, a normal variant within the hypersomnolence conditions, applies when a person sleeps more in the 24-hour day than the typical person. Sleep is normal in architecture and quality. Usually, sleep lengths of 10 hours or greater qualify for this diagnosis. Symptoms of excessive sleepiness occur if the person does not get that amount of sleep.

CIRCADIAN RHYTHM SLEEP-WAKE DISORDERS

The circadian rhythm sleep-wake disorders[39,40] share a common underlying chronophysiologic basis. The major feature of these disorders is an alteration of the circadian clock or a persistent or recurrent misalignment between the patient's sleep pattern and the pattern that is desired or required by societal demands or the environment (see Chapters 32 to 40 and 75). Because of the misalignment between circadian sleep-wake propensity and behavioral schedules, these individuals often experience symptoms of both insomnia and excessive sleepiness.

Delayed sleep-wake phase type,[41-44] which is more commonly seen in adolescents and young adults, is characterized by a delay in the phase of the major sleep period in relation to the desired sleep time and wake time, whereas advanced sleep phase type,[44] which is more commonly seen in older adults, is characterized by an advance in the phase of the major sleep period in relation to the desired sleep time and wake-up time. In both of these disorders, sleep, when allowed to occur with the circadian sleep propensity, is normal in quality and duration, but at a later (delayed sleep phase) or earlier (advanced sleep phase) time of day than desired. The irregular sleep-wake type, a disorder that involves the lack of a clearly defined circadian rhythm of sleep and wakefulness, is most

often seen in institutionalized older adults or chronically mentally ill patients and is associated with a lack of synchronizing agents such as light, physical activity, and social schedules. The non-24-hour type or free-running sleep-wake rhythm[44] is a result of lack of entrainment to the 24-hour light-dark cycle. The most common case of this condition is total blindness, although in some individuals an unusually long circadian period (which lies outside the range of entertainment) or an abnormal response to light may be causative. These patients generally exhibit cycles of poor sleep alternating with better sleep as their clock moves in and out of phase with the light-dark cycle.

Jet lag disorder[45,46] is related to an abrupt desynchronization between circadian sleep-wake propensity and the environmental day-night cycle as a result of transmeridian travel across two or more time zones. The severity of the disorder is influenced by the number of time zones crossed and the direction of travel, with eastward travel usually being more disruptive. Shift work type[45,47,48] is characterized by complaints of insomnia or excessive sleepiness that occur as a result of work hours that overlap the usual sleep period.

In addition to the subjective reports of insomnia or sleepiness that are attributable to the circadian disturbance, sleep logs that demonstrate the sleep-wake schedule alteration are required for most of these disorders. Although objective measures are not required, increasing emphasis is placed on the use of actigraphy and biomarkers such as dim-light melatonin onset to provide more accurate diagnosis and treatment guidance. Other sleep disorders (e.g., chronic insomnia disorder) may mimic or overlap with circadian rhythm sleep-wake disorders and must be identified and addressed as part of the overall therapeutic approach.

PARASOMNIAS

The parasomnias[49-51] are undesirable behaviors or experiences that accompany sleep (see Chapters 101 to 105). The motor activity of parasomnias is typically more complex than that observed with the sleep-related movement disorders. Experiences may be cognitive-emotional (as in nightmares or sleep terrors) or sensory (e.g., sleep-related hallucinations or exploding head syndrome). The parasomnias can arise during NREM sleep, REM sleep, or sleep-wake transitions. Many parasomnias represent a sleep-wake state disassociation between either NREM and wake (disorders of arousal from NREM) or REM and wake (REM sleep behavior disorder and recurrent sleep paralysis). The parasomnias often occur in conjunction with other sleep disorders such as OSA. It is not uncommon for several parasomnias to occur in one patient. NREM parasomnias[52,53] include the disorders of arousal from slow wave NREM sleep (confusional arousal, sleepwalking, and sleep terrors) and the closely related sleep-related eating disorder. Confusional arousals are characterized by mental confusion or confusional behavior that occurs during or after arousal from sleep. These arousals are common in children and can occur not only from nocturnal sleep but also from daytime naps. Sleepwalking[54-56] is a series of complex behaviors that arise from sudden arousals from slow wave sleep and result in locomotion during a state of impaired consciousness. Sleep terrors[57,58] also occur from slow wave sleep and are associated with a cry or piercing scream accompanied by autonomic system activation and behavioral manifestation of intense fear. Individuals may be difficult to arouse from the episode and when aroused can be confused and subsequently amnestic for the episode. These two disorders, sleepwalking and sleep terrors, often coexist and may result in a potentially dangerous state of terror and walking or running. Sleep-related eating disorder[59-61] involves recurrent eating and drinking episodes during partial arousals from sleep. The eating behavior is uncontrollable, typically involves ingestion of unusual or inedible substances, and may be associated with potential injury (e.g., from cooking) or other adverse health consequences. Often the patient has limited or no awareness during the episode and impaired recall of the behavior.

Several parasomnias are typically associated with the REM sleep stage. Pathophysiologic mechanisms related to REM sleep may underlie these disorders. REM sleep behavior disorder (RBD)[62-66] involves dream enactment behaviors that occur in REM sleep and may result in injury or sleep disruption. The behaviors are often violent with dream enactment that is action filled. The disorder can occur in narcolepsy and in many patients with Parkinson disease. The delayed emergence of neurodegenerative disorders, primarily synucleinopathies, occurs in a significant percentage of patients with idiopathic RBD, especially in men older than 50 years. Recurrent isolated sleep paralysis can occur at sleep onset or on awakening and is characterized by a frightening inability to perform voluntary movements. Ventilation is usually unaffected. Hallucinatory experiences often accompany the paralysis. Nightmare disorder[67-69] is characterized by recurrent anxiety-laden, often terrifying dreams that occur most commonly in REM sleep and result in an awakening with intense anxiety, fear, or other negative feelings. The disorder is most commonly encountered in the clinical setting in conjunction with posttraumatic stress disorder. The diagnosis should be invoked in this context only when the nightmare component is the focus of independent clinical assessment and treatment.

Sleep enuresis[70-72] is recurrent involuntary voiding that occurs during sleep. Enuresis is considered primary in a child who has never been dry for 6 months or longer; otherwise, it is called secondary. Exploding head syndrome is characterized by a loud imagined noise or sense of a violent explosion that occurs in the head as the patient is falling asleep or during waking in the night.

Sleep-related hallucinations are hallucinatory experiences that occur at sleep onset or on awakening. They may be difficult to distinguish from vivid dreams or nightmares but usually are complex images that occur when the patient is clearly awake. Sleeptalking, an isolated symptom, may arise during NREM or REM sleep and can be idiopathic or associated with other disorders such as REM sleep behavior disorder or sleep-related eating disorder.

SLEEP-RELATED MOVEMENT DISORDERS

The sleep-related movement disorders[73-77] are characterized by relatively simple, usually stereotyped movements that disturb sleep (see Chapter 106). Disorders such as periodic limb movement disorder and restless legs syndrome (RLS) are classified in this section.

RLS[78-83] consists of a complaint of a strong, nearly irresistible urge to move the legs, often accompanied by uncomfortable or painful symptoms. The sensations are worse at rest and

occur more frequently in the evening or during the night. Walking or moving the legs relieves the sensation temporarily. The disorder is commonly associated with repetitive limb movements during sleep, but a separate diagnosis of periodic limb movement disorder is not used when the movement occurs in the context of RLS, narcolepsy, untreated OSA, or RBD.

Periodic limb movement disorder[84,85] is an independent disorder of repetitive, highly stereotyped limb movements that occur during sleep. The movements must give rise to sleep disturbance or daytime sleepiness to meet criteria. Sleep-related leg cramps[86,87] are painful sensations associated with sudden intense muscle contractions, usually of the calves or small muscles of the feet, which occur during the sleep period and can lead to disrupted sleep. Relief is usually obtained by stretching the affected muscle.

Sleep-related bruxism[88,89] is characterized by clenching of the teeth during any stage of sleep and can result in arousals. Often the activity is severe or frequent enough to result in symptoms of temporomandibular joint pain or wearing down of the teeth. Sleep-related rhythmic movement disorder[90] is a stereotyped, repetitive rhythmic motor behavior that occurs during drowsiness or light sleep and results in large movements of the head, body, or limbs. Typically seen in children, the disorder can also be seen in adults. Head and limb injuries can result from violent movements. Rhythmic movement disorder can also occur during full wakefulness and alertness, particularly in individuals who are mentally retarded. Benign sleep myoclonus of infancy[91] is a disorder of myoclonic jerks that occur during sleep in infants. It typically occurs from birth to age 6 months, is not associated with known adverse consequences, and resolves spontaneously. Propriospinal myoclonus at sleep onset[92] is a disorder of recurrent sudden muscular jerks in the transition from wakefulness to sleep. The disorder may be associated with severe sleep-onset insomnia.

Isolated symptoms and normal variants within this section include sleep starts, excessive fragmentary myoclonus, and hypnagogic foot tremor and alternating leg muscle activation. Sleep starts (hypnic jerks) are sudden brief contractions of the body that occur at sleep onset. These movements are often associated with a sensation of falling, a sensory flash, or a sleep-onset dream. Hypnagogic foot tremor and alternating leg muscle activation occur at the transition between wake and sleep or during light NREM sleep. These conditions are listed together because they may represent somewhat different manifestations of a single disturbance. Hypnagogic foot tremor consists of rhythmic movement of the feet or toes, whereas alternating leg muscle activation consists of a PSG pattern of repetitive, transient activation of one anterior tibialis muscle group, alternating with activation of the contralateral tibialis. Excessive fragmentary myoclonus is manifest as small muscle twitches in the fingers, toes, or the corner of the mouth that do not cause actual movements across a joint. The myoclonus is often a finding during PSG that may be asymptomatic or associated with daytime sleepiness or fatigue.

OTHER SLEEP DISORDERS

The diagnosis of other sleep disorder is employed when a sleep disorder cannot be classified elsewhere. Disorders that demonstrate clear features of other, more specific categories (e.g., circadian, parasomnia, or movement disorder), but do not meet criteria for a specific diagnosis, should be classified within those respective categories as "unspecified."

ICSD2 included a diagnosis of environmental sleep disorder within this section. Significant controversy exists regarding this condition. Sleep disturbance that is purely a function of environmental disturbance such as a physical stimulus (e.g., noise or light) or environmental danger does not technically meet criteria for insomnia. When a complaint such as this is encountered in the clinical setting and is accompanied by adverse consequences, a diagnosis of other sleep disorder may be employed.

SLEEP-RELATED MEDICAL AND NEUROLOGIC DISORDERS

The ICSD3 lists six disorders that, although not sleep disorders, per se, are commonly encountered in association with sleep or may have unique presentations during sleep.

Fatal familial insomnia[93-95] is a progressive disorder characterized by difficulty in falling asleep and maintaining sleep that develops into enacted dreams or stupor. Autonomic hyperactivity with pyrexia, excessive salivation, and hyperhidrosis leads to cardiac and respiratory failure. The disease is caused by a prion, and it leads eventually to death. Sleep-related epilepsy[96-98] is the diagnosis employed when epilepsy occurs during sleep. Several epilepsy types are associated with sleep, including nocturnal frontal lobe epilepsy, benign epilepsy of childhood with centrotemporal spikes, benign epilepsy with occipital paroxysms, and juvenile myoclonic epilepsy. Sleep-related headaches[99,100] are cephalgias that occur during sleep or on awakening from sleep. Some are uniquely associated with sleep. Migraine headaches, chronic paroxysmal hemicrania, hypnic headaches, and cluster headaches can all occur during sleep. Sleep-related laryngospasm[101,102] is a disorder in which patients report choking and difficulty breathing at night typically associated with pronounced fear or panic. Its etiology is not well established, although it is encountered in multisystem atrophy and may be associated with OSA or gastroesophageal reflux. Sleep-related gastroesophageal reflux[103,104] is characterized by regurgitation of stomach contents into the esophagus during sleep. Shortness of breath or heartburn can result, but occasionally the disorder is asymptomatic. Sleep-related myocardial ischemia is due to reduction of blood flow to the myocardium during the sleep period.

CURRENT AND FUTURE CLASSIFICATION CONSIDERATIONS

Insomnia

From the outset of modern-day sleep disorder classification systems, insomnia has been compartmentalized into primary insomnia and insomnia due to a medical disorder, mental illness, or substance use, with several additional subtypes of primary insomnia identified. The consolidation of insomnia diagnoses in ICSD3 is not intended to suggest that there may not be important differences among certain subtypes of insomnia. However, from a clinical perspective, differentiation among a number of these subtypes has proved rather unreliable,[105,106] and after treatment of comorbidities is accounted for, therapeutic approaches do not differ significantly among the various subtypes. Additional research is necessary to determine whether significant, identifiable

differences among insomnias do exist and to what extent these differences might alter therapeutic approaches.

Previous classification systems such as ICSD2 included a complaint of "nonrestorative sleep"[107,108] in the list of possible insomnia symptoms. However, the nature of this complaint is ambiguous. Although many insomnia patients describe their sleep as "nonrestorative," majority of these patients have a presenting complaint of sleep initiation or maintenance resulting in extended periods of nocturnal wakefulness. Although a small percentage of patients have an isolated complaint of nonrestorative sleep, it is not clear that this complaint should be categorized as insomnia. Many other disturbances of sleep (e.g., sleep apnea) result in sleep that can fairly be categorized as "nonrestorative." The limited studies of isolated nonrestorative sleep suggest that, although sufferers share many symptoms in common with insomnia patients, they differ in certain respects. Therefore the evaluation and classification of patients with this complaint are left to the judgment of the clinician.

Sleep-Related Breathing Disorders

Treatment emergent central sleep apnea is a new diagnosis in ICSD3. Some controversy has existed regarding the validity of this diagnosis as well as the conditions under which it should be applied.[19] The diagnosis is established on the basis of PSG performed during positive airway pressure titration for obstructive sleep apnea. However, it has been recognized that some central sleep apnea that is observed during PAP titration resolves spontaneously over time. This may create the potential for unnecessary application of more complex and expensive treatment modalities such as auto servo ventilation. Further long-term follow-up is necessary to address these issues.

The criteria for pediatric obstructive sleep apnea have remain unchanged. However, limited information exists about the appropriate thresholds for treatment intervention in this population.

Central Disorders of Hypersomnolence

Destruction of hypocretin neurons, possibly on an autoimmune basis, is the presumptive etiology of narcolepsy type 1.[31] However, 5% to 10% of patients with narcolepsy and cataplexy (type 1) have normal hypocretin levels suggesting either a downstream problem with hypocretin (e.g., at the receptor level) or an alternative pathophysiologic mechanism. It is unclear whether narcolepsy type 2 shares some common pathophysiologic aspects with narcolepsy type 1 or is a disorder based on an entirely different pathophysiology. Idiopathic hypersomnia is still poorly understood because there is no clear pathophysiologic mechanism identified. Diagnosis can be challenging in that some patients may present with a convincing history of debilitating sleepiness, yet fail to meet the MSLT requirement for the diagnosis. Further work is needed to define the pathophysiology of narcolepsy type 2 and idiopathic hypersomnia as well as the genetic basis of these disorders.

Circadian Rhythm Sleep-Wake Disorders

Much has been learned regarding the biologic basis for circadian sleep-wake disorders. However, use of biologic markers that will enhance diagnostic precision is still quite limited, in part owing to issues of availability and reimbursement. ICSD3 encourages greater use of these markers for diagnosis in the hope that they will become a diagnostic standard and more widely available.

Parasomnia

Definitions of REM sleep without atonia, necessary to establish the diagnosis of RBD, have been somewhat complex and variable in recent years. Users should consult the AASM scoring manual for the most recent definitions. Substantial numbers of patients present with dream enactment behaviors in the context of OSA or antidepressant use. It remains unclear whether these patients carry the same unfavorable prognosis with respect to development of synucleinopathies as those with idiopathic RBD.

CLINICAL PEARLS

- ICSD3 incorporates all chronic insomnia within a single diagnosis of chronic insomnia disorder.
- Treatment emergent central sleep apnea, previously referred to as complex sleep apnea, has been added to ICSD3. This excludes central apneas of known etiology, such as substance-induced central apnea and Cheyne-Stokes breathing.
- Narcolepsy is now classified as type 1 and type 2. This allows more accurate description of hypocretin-deficient cases (type 1) without cataplexy.

SUMMARY

The classification of sleep disorders allows accurate diagnosis, improved communication among physicians, and the standardization of data for research purposes. The ICSD3 identifies six major groups of disorders, including insomnia, sleep-related breathing disorders, hypersomnolence disorders, circadian rhythm sleep-wake disorders, parasomnias, and sleep-related movement disorders. New sleep disorders have been recognized and previous sleep disorders have been clarified with a better understanding of their diagnostic and epidemiologic features. The ICSD3 increases the refinement of sleep disorder diagnoses because of recent advances in sleep research. Referral to the ICSD3 will help clinicians establish a rational differential diagnosis when evaluating patients.

Selected Readings

American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.

Berry BR, Gamaldo CE, Harding SM, et al. For the American Academy of Sleep Medicine. *AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, version 2.0.3. Darien, IL: American Academy of Sleep Medicine; 2014.

Bornemann MA, Mahowald MW, Schenck CH. Parasomnias: clinical features and forensic implications. *Chest* 2006;**130**(2):605–10.

Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. *Lancet* 2014;**383**(9918):736–47.

Khan MT, Franco RA. Complex sleep apnea syndrome. *Sleep Disord* 2014;**2014**:798487.

Malhotra S, Kushida CA. Primary hypersomnias of central origin. *Continuum (N Y)* 2013;**19**(1 Sleep Disorders):67–85.

Morin CM, Benca R. Chronic insomnia. *Lancet* 2012;**379**(9821):1129–41.

Nishino S, Mignot E. Narcolepsy and cataplexy. *Handb Clin Neurol* 2011;**99**:783–814.

Zee PC, Attarian H, Videnovic A. Circadian rhythm abnormalities. *Continuum (N Y)* 2013;**19**(1 Sleep Disorders):132–47.

A complete reference list can be found online at ExpertConsult.com.

Chapter 62

Epidemiology of Sleep Medicine

Amy W. Amara; Mary Halsey Maddox

Chapter Highlights

- Sleep medicine is a young field of medicine, and as such, the epidemiology of sleep medicine is in the early stages of study. Because of this relatively short history and the evolving classifications of disorders within the field, there is still much to be learned about the epidemiology of sleep medicine.
- Sleep disorders are common and include sleep-disordered breathing, hypersomnia, circadian rhythm disorders, insomnia, restless legs syndrome, and parasomnias, all of which have documented adverse outcomes for patients and many of which have adverse societal outcomes.
- Sleep-disordered breathing causes significant health and economic burdens and affects all ages and ethnicities, although it is most common in older men.
- Parasomnias are divided into rapid eye movement and non–rapid eye movement parasomnias. The non–rapid eye movement parasomnias are more common in children, but lifetime prevalence is high for both.
- Insomnia is common, adversely affects quality of life, and is associated with other health conditions. It is important to exclude other sleep disorders, such as circadian rhythm disorders and restless legs syndrome, when examining the epidemiology of insomnia.

Hippocrates said, "It is far more important to know what person the disease has than what disease the person has." Epidemiology includes not only the number of new and existing cases of certain illnesses and diseases occurring within a population (incidence and prevalence) but also the impact and burden those illnesses and diseases have on society and the people who live in it. The epidemiology of sleep medicine is in its youth and is challenging to define because of the evolution of diagnostic criteria over time, as well as differences in study methodologies.[1] As such, this chapter highlights the larger epidemiologic studies and attempts to summarize the data available, recognizing that, within the study of epidemiology of sleep medicine, there are gaps in knowledge, particularly in the areas of impact and economic burden.

This chapter is divided into sections (Sleep-Disordered Breathing, Hypersomnia, Circadian Rhythm Disorders, Insomnia, Restless Legs Syndrome, and Parasomnias) and generally follows the *International Classification of Sleep Disorders*, third edition (ICSD3) for inclusion criteria of each. The epidemiology of each seeks to focus on the people of the disease rather than the disease itself, as Hippocrates so eloquently stated.

SLEEP-DISORDERED BREATHING

Although disruptions of breathing in sleep had been reported previously (and attributed to hypercapnea),[2] airway obstruction was first recognized as the cause of sleep apnea between 1964 and 1965. At that time, Jung and Kuhlo, who are credited with performing the first tracheostomy as treatment for this disorder, and Gastant and colleagues simultaneously described sleep-related airway obstruction.[3-5] In the 1970s, recognition of obstructive sleep apnea (OSA) expanded in conjunction with enhancement of polysomnography (PSG) techniques and development of diagnostic parameters.[3] Before the 1980s, tracheostomy and weight loss were the only treatments for OSA. In 1981, continuous positive airway pressure was introduced[6] and revolutionized the treatment of OSA. With expansion of the field of sleep medicine, health care providers increasingly recognize sleep-disordered breathing and the adverse health effects and comorbidities of untreated OSA.

Snoring

Although the definition varies somewhat from study to study, habitual snoring in adults is most commonly defined as snoring for three or more nights per week.[7] Most of the studies evaluating snoring prevalence rely on self-report by the subject or spouse based on questionnaires, although some use a two-stage evaluation of questionnaire followed by portable or center-based PSG. In most cases, reports of snoring prevalence do not distinguish between snorers with apnea and those without apnea. The prevalence of habitual snoring ranges from 14% to 84%,[8,9] depending on the population studied. One large questionnaire-based study evaluated 22,389 blood donors in New Zealand and found self-reported snoring in 33%.[10] In the Sleep Heart Health Study (SHHS), snoring prevalence among 13,194 participants ranged from 32% to

52% in men and from 19% to 29% in women, depending on race and ethnicity.[11] Snoring prevalence is consistently higher in males than females and is influenced by body mass index (BMI), age, smoking status, ethnicity, and the presence or absence of a bed partner.

Obstructive Sleep Apnea

The prevalence of OSA varies by study based on how OSA is defined and the population in which it is measured. The ICSD3[12] defines OSA as either (1) 5 or more predominantly obstructive respiratory events per hour measured by PSG or out-of-center sleep testing (OCST) plus a comorbid medical condition, sleep complaint (sleepiness, insomnia, nonrestorative sleep), or witnessed apnea; or (2) 15 or more predominantly obstructive respiratory events per hour by PSG or OCST in the absence of other symptoms or comorbid conditions. This current definition has not been applied to most of the previously performed epidemiologic studies, and the challenges in interpreting prevalence in the setting of such different methodologies is discussed by Davies and Stradling.[13] Despite these challenges, large population-based studies provide very useful information about OSA epidemiology.[14] In the Wisconsin Sleep Cohort Study, the obstructive sleep apnea syndrome (OSAS) (defined in this study as an apnea hypopnea index [AHI] of 5 or greater in combination with the presence of excessive daytime sleepiness [EDS]) was estimated to affect 2% of women and 4% of men between the ages of 30 and 60 years.[15] The estimation of sleep-disordered breathing (SDB) (AHI ≥ 5 with or without EDS) was 9% in women and 24% in men.[15] In the SHHS,[16] 5615 subjects between 40 and 98 years of age were evaluated with questionnaires and OCST. In this sample, 18% had an AHI of 15 or greater (25% of men and 11% of women), and 29% of participants had an AHI between 5 and 14 (33% of men and 26% of women). In addition to sex, age and race also influenced the results.[16] Duran and colleagues evaluated a large cohort of 2148 subjects in Spain in two stages with questionnaire and home-based nocturnal oxygen monitoring in the first phase followed by PSG in a subset of the group. In the first phase, 35% were habitual snorers (46% of men and 25% of women). PSG monitoring showed an AHI of 10 or greater in 19% of men and 14.9% of women.[17] Mehra and colleagues investigated the prevalence of sleep-disordered breathing among 2911 men older than 65 years, finding that 26.4% of these older male participants had a respiratory disturbance index (RDI) of 15 or greater. Age and BMI influenced these outcomes.[18] Using home-based single-channel nasal airflow monitoring in a subset of the Australian Busselton Health Study cohort, Simpson and colleagues reported that 9.1% of 793 subjects had an RDI of 15 or greater (12.4% of men and 5.7% of women).[19] In addition to these prevalence studies, OSA incidence was documented in the Cleveland Family study, which demonstrated a 5-year incidence of 16% for mild to moderate OSA (AHI ≥ 10) and an incidence of 7.5% for moderately severe OSA (AHI ≥ 15) within 5 years following a normal PSG (AHI < 5).[20] The incidence was influenced by body habitus, sex, age, and cholesterol concentration.[20]

In summary, mild OSA with more than 5 events per hour likely affects approximately 1 in 5 adults, with moderate to severe OSA affecting at least 1 in 15 adults.[13,21,22] These may be an underestimation because most of the epidemiologic studies evaluate only those with symptoms, and many groups, including those with diabetes,[23] heart failure,[24,25] Parkinson disease,[26] and others, are less likely to have the typical symptoms of snoring and daytime sleepiness.

Sleep-Disordered Breathing and Sex

A consistent finding among epidemiologic studies investigating snoring and sleep apnea is the increased prevalence among men compared with women. In the SHHS, which reported snoring frequency based on ethnicity among 13,194 individuals older than 40 years, habitual snoring was reported by 32% to 52% of the men evaluated, but only by 19% to 29% of the women.[11] In a study of the Bussleton cohort, which followed 967 nonsnoring subjects from 1981 to 1994, 21% of men and 8% of women reported development of habitual snoring by the conclusion of the study.[27] The Cardiovascular Health Study evaluated participants at baseline for sleep complaints. Of the 5888 subjects, 34% of men and 20% of women self-reported snoring.[28] In a study of 4648 subjects between 20 and 69 years of age in Northern Sweden, 14.6% of men and 6.7% of women reported problematic snoring.[29] This higher prevalence among men than women is also reported in investigations of sleep apnea risk. Wheaton and colleagues evaluated data collected from 2005 to 2008 in the National Health and Nutrition Examination Survey (NHANES). In this population, 6% of men and 3.1% of women had a prior diagnosis of sleep apnea, and habitual snoring (five or more nights per week) was reported by 37.2% of men and 22.4% of women.[30] Bixler and colleagues performed a two-phase study in which they interviewed 12,291 women and 4364 men in Pennsylvania for sleep apnea risk and then randomly selected 1000 women and 741 men for PSG, with oversampling of those with more risk factors. In this sample, 7.2% of men and 2.2% of women had an AHI of 10 or greater, and 3.9% of men and 1.2% of women had an AHI of 10 or greater plus daytime symptoms. The prevalence was lower in premenopausal women and postmenopausal women on hormone replacement therapy compared with the prevalence in postmenopausal women not on hormone replacement therapy.[31] In a subset of the SHHS in which 5615 subjects underwent OCST, men had 2.7 times higher odds of having an AHI of 15 or greater compared with women.[16] These differences between men and women were also evident in the Wisconsin Sleep Cohort and the Busselton Health Study as described earlier in this paragraph.[15,19] Although the cause of these differences in prevalence is not completely defined, male gender appears to be a risk factor for development of OSA compared with premenopausal women.[31] However, because women with OSA tend to have fewer of the symptoms typically associated with sleep-disordered breathing, prevalence among women is likely underestimated.[32,33]

Sleep-Disordered Breathing and Race and Ethnicity

Race and ethnicity may also influence the prevalence of SDB. Many of the earlier epidemiologic studies on sleep apnea focused on Western populations, but in more recent years, more prevalence information is available for other groups. In the SHHS, O'Connor and colleagues evaluated snoring prevalence based on ethnicity.[11] After adjustment for age, BMI, and presence of a bed partner, the odds of frequent snoring were 2.25 times higher among Hispanic women and 1.55 times higher among black women compared with white women. Hispanic men had 2.3 times the odds of frequent

snoring than white men.[11] Redline and colleagues performed OCST in 14,440 Hispanic or Latino participants who live in the United States. In this population, the age-adjusted prevalence of sleep apnea was 25.8% for AHI of 5 or greater, 9.8% for AHI of 15 or greater, and 3.9% for AHI of 30 or greater.[34] The odds of having SDB were higher among persons who were male, obese, and older.[34] In the Northern Manhattan Study, Hispanic subjects had a 3.6 times higher odds of frequent snoring compared with white subjects, and there was no significant difference in self-reported snoring between black and white subjects.[35] However, a PSG-based study comparing severity of SDB among white and black subjects showed a higher AHI in African American males younger than 39 years or between 50 and 59 years when compared to white males.[36] Similarly, Redline and colleagues found that African American subjects 25 years or younger had 1.88 times higher odds of having SDB compared with whites.[37] Ancoli-Israel and colleagues showed no difference in prevalence of SDB (RDI ≥ 15) between whites and African Americans older than 65 years, but African Americans were more likely to have more severe sleep apnea (RDI ≥ 30).[38] Among studies performed in Asian populations, there is a broad range of suspected sleep apnea prevalence as discussed in a meta-analysis by Mirrakhimov and colleagues.[39] This large disparate prevalence is related to differences in study population based on sample size, sex, age, body habitus, and study design.[39] In a comparison of prevalence of OSA (AHI ≥ 5) among individuals based on OCST, Leong and colleagues found that obese South Asians had higher prevalence of OSA (85%) compared with obese white Europeans (66%).[40] Among Chinese men and women evaluated in two studies with questionnaires followed by PSG in a subset of subjects, 23% of men and 14.8% of women reported habitual snoring. The estimated prevalence of sleep apnea was 8.8% of Chinese men and 3.7% of Chinese women with an AHI of 5 or greater and 4.1% of men and 2.1% of women with an AHI of 5 or greater plus EDS.[41,42] In India, estimated population prevalence of OSA (AHI ≥ 5) is 9.3% to 13.7% and of OSAS (AHI ≥ 5 plus EDS) is 2.8% to 3.57%.[43,44] Yamagishi and colleagues reported higher prevalence of OSA in Americans and Hispanics than in Japanese, but this difference was attenuated after controlling for BMI.[45] In Thailand, the estimated prevalence of OSA and OSAS is 11.4% and 4.4%, respectively.[46] Two studies reported the prevalence of habitual snoring in Nigeria ranging from 14% to 18.5%.[8,47] The differences in prevalence of sleep apnea based on ethnicity or race may be in part related to body habitus[40,45] and differences in craniofacial features,[48] but there are differences independent of these factors.[38] Based on available data, the most consistent findings seem to be that Hispanics have higher prevalence of SDB than whites, and African Americans have more severe SDB that is present at a younger age than among whites. The prevalence among Asians varies based on the ethnicity within this group.

Sleep-Disordered Breathing and Age

Studies evaluating epidemiology of sleep apnea have documented a clear effect of age on prevalence of SDB. In the study by Duran and colleagues, which included white men and women between the ages of 30 and 70 years, there was a 2.2 times higher odds of having an AHI of 5 or greater for each 10-year increase in age.[17] Enright and colleagues assessed habitual snoring in subjects older than 65 years and showed that snoring actually decreased with aging, with 41% of men and 20% of women between the ages of 65 and 69 years reporting loud snoring, and only 17% of men and 10% of women older than 80 years reporting loud snoring.[49] A pattern that might reconcile these two observations was noted in the San Marino study, in which snoring was reported to increase with age until between 60 and 65 years, but then plateaued at older ages.[50] Bixler and colleagues studied 4364 men by telephone survey followed by PSG in a subset (741 subjects) and found that OSA increased with age until age 60 years but then decreased with older age.[51] A similar pattern was noted among women.[31] These studies together indicate that sleep apnea prevalence increases with age into the 60s but then plateaus.

Central Sleep Apnea

Central sleep apnea is the absence of airflow for at least 10 seconds with concomitant absence of respiratory effort. According the ICSD3, central sleep apnea can be classified as due to Cheyne-Stokes breathing; due to a medical condition without Cheyne-Stokes breathing; due to high-altitude periodic breathing; due to a medication or substance; or within a category of primary central sleep apnea.[12] Cheyne-Stokes breathing is most common in subjects older than 60 years and is most often associated with heart failure or stroke.[12] In heart failure patients, prevalence of SDB (AHI ≥ 5) is estimated to affect 70% to 80% of patients,[52] with moderate to severe central sleep apnea (AHI ≥ 15) in 21% to 40%.[53-56] The prevalence of central sleep apnea due to other conditions or due to opioids is not known, although healthy persons commonly develop central sleep apnea due to high altitude.[12]

Health and Economic Impact of Sleep-Disordered Breathing

Among adults, SDB results in increased health care use and increased economic burden. A retrospective evaluation of 1,867,876 veterans showed that those with a diagnosis of OSA had more hospitalizations and more emergency department visits than those without a diagnosis of OSA. Additionally, those with OSA had higher odds of having heart failure, hypertension, lung disease, obesity, stroke, depression, diabetes, and other health problems.[57] Similarly, Smith and colleagues found that patients diagnosed with OSA used 23% to 50% more healthcare resources in the 5 years prior to diagnosis compared to age-matched control subjects.[58] The OSA patients were also at higher risk for development of heart disease, lung disease, and depression.[58] These health consequences and other associated symptoms of sleep apnea result in decreased work performance and increased risk for work- and leisure-related injuries, further increasing the economic impact of this common disorder.[59] In the year 2000, the estimated cost of sleep apnea–related motor vehicle collisions was $15.9 billion and 1400 lives.[60] Treatment of sleep apnea can reduce some of this burden.[59,60] In a study of 22,275 Union Pacific Railroad employees, treatment of sleep apnea resulted in a $4.9 billion differential cost savings compared with untreated sleep apnea.[61] The effects and dependence of SDB on other health conditions, such as hypertension, cardiovascular disease, heart failure, cognition, lung disease, obesity, neurologic disorders, and metabolic dysfunction, are explored in more extensive detail in other chapters in this book (see Chapters 93, 94, 96, 97, 117–118, and 126–129).

Pediatric Sleep-Disordered Breathing

Osler described pediatric SDB in reference to pediatric neurocognitive impairment in 1892 in *The Principles and Practice of Medicine,* but it was not identified as clinically important in pediatrics until 1976 when Guilleminault and colleagues described "Sleep Apnea in Eight Children."[62,63] Consequently, the primary base of information about SDB in children stems from the past 35 to 40 years, suggesting that there is much still to be discovered. Over the past 3 to 4 decades, however, SDB in pediatrics has changed from a disorder observed in a small case series to a disease with a significant mental, physical, social, and economic impact.

Pediatric Obstructive Sleep Apnea and Primary Snoring

Most of the epidemiologic data for SDB in children report a prevalence of 4% to 11%, with the prevalence of OSA ranging from 1.2% to 5.7%.[64-66] Most studies evaluating prevalence are based on questionnaires, although two large studies have performed more objective evaluation with PSG in combination with questionnaires. Li and colleagues administered questionnaires and then performed PSG in Chinese school-aged children. The prevalence of symptomatic OSA (defined as AHI > 5) was 4.8% for the entire sample, 5.7% in boys and 3.8% in girls; the prevalence of asymptomatic OSA (AHI > 5) was 9.1% and 5.7% in boys and girls, respectively.[67] In the United States, Bixler and colleagues had a similar protocol with healthy school-aged children that showed a prevalence of 1.2% for an AHI of greater than 5, 25% for an AHI between 1 and 5, and 15.5% for primary snoring.[65] In a meta-analysis of pediatric SDB, Lumeng and Chervin, in their meta-analysis of pediatric SDB reported a prevalence of 4% to 11%, with a breakdown of percentages based on how individual studies framed the questions about primary snoring and apnea. Primary snoring makes up the higher end of the former percentages, ranging from 2.5% in Turkey to 6.2% in Sweden for those who snore "always."[66,68,69] For those who snore "often," the prevalence is 3.2% in Iceland, 14.8% in Spain, and 34.5% in Italy.[66,70-72] OSA makes up the lower end of the spectrum, but prevalence depends on the criteria used for diagnosis. Overall, however, the prevalence of OSA based on PSG ranges from 0.1% in Singapore for an AHI of greater than 1 to 13% in Italy for an oxygen desaturation index of 5 or greater.[66,72,73] Parent reported apnea ranges from 0.2% to 18.6%.[66,72,74] Certainly there is room for more standardized trials involving consistent definitions of OSA, primary snoring, and SDB with larger numbers of subjects to identify the true prevalence of pediatric SDB. Based on available studies, the prevalence of primary snoring appears to range from 1.5% to 27.6% and the range for OSA is 1% to 5%.[75]

Pediatric Sleep-Disordered Breathing and Age

Over time the distribution of SDB by age has changed within the pediatric population as well. In the 1990s, the prevalence of primary snoring was cited at 6% to 12%.[69,76-78] These older studies reported the overall prevalence of primary snoring at 10% of children 2 to 8 years of age with a decreasing prevalence at about 9 years of age.[79] More recent studies suggest a bimodal distribution of sleep apnea with a second peak in adolescence. Although the prevalence of SDB in adolescence is not definitive, it is clear that SDB increases with obesity in this age group. Kohler and colleagues found a 3.5-fold increase in AHI relative to increase in BMI Z-score in pediatric patients older than 12 years.[80] The Childhood Adenotonsillectomy Study (CHAT) found that children younger than 9 years with OSA had better resolution of symptoms if they were not overweight or obese regardless of treatment group (adenotonsillectomy or watchful waiting).[81] Of course, obesity is a problem worldwide and has increased in the pediatric population over the past 30 years, which certainly contributes to the increasing prevalence of SDB and the bimodal peak. Furthermore, studies have begun to investigate racial, social, and health differences among pediatric patients with SDB. These reports suggest that African American ethnicity, prematurity, and low socioeconomic status result in worse SDB in pediatric patients.[37,82-85]

Health and Economic Burden of Pediatric Sleep-Disordered Breathing

In the pediatric population, health and societal implications of SDB are reported with increasing frequency. Perhaps one of the greatest concerns is the neurocognitive impairment seen in children with SDB. Beebe (2006) and Owens (2009) both provide excellent reviews of the association between neurocognitive impairment and pediatric SDB.[86,87] SDB, inclusive of primary snoring and OSA, clearly results in varying degrees of neurocognitive impairment in children. The more recent Childhood Adenotonsillectomy Study study demonstrated that children who were treated for their SDB had more improvement in secondary behavioral scores and quality of life than those in the "watchful waiting" arm, although treatment of SDB did not result in any improvement in executive function and attention scores on neuropsychological testing.[81] Similar to other studies, Bourke and colleagues found cognitive and academic impairment in Australian children, but this did not correlate with the degree of SDB.[88] Landau and colleagues observed a small cohort of Israeli preschoolers in whom treatment of SDB resulted in resolution of the neurocognitive problems and "catch-up" to controls after 1 year.[89] On the other hand, Perfect and colleagues found as part of the Tucson Children's Assessment of Sleep Apnea Study that there is concern for lifetime impairment in those with history of SDB versus controls who have never had SDB.[90] Most studies, however, suggest that treatment of SDB does improve neurocognitive performance.[75] For an extensive review of the studies of neurocognitive function and SDB, please see the Technical Report of the American Academy of Pediatrics Recommendation on the Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome.[75]

In addition to neurocognitive impairment, pediatric SDB also results in physical health problems. Historically, when SDB was underrecognized or when diagnosis was delayed, toddlers had failure to thrive, and some pediatric patients presented with cor pulmonale. With earlier recognition of the medical complications of pediatric SDB, these outcomes are now rarely seen, but there is mounting evidence for cardiac, inflammatory, and endocrine effects of SDB. The prevalence of such complications is not yet known, but cardiovascular complications include right and left ventricular changes, blood pressure changes, changes in brain natriuretic peptide, changes in cerebral blood flow, and autonomic dysregulation.[75] Inflammatory consequences and endocrine and growth changes need to be more systematically studied, but early research suggests

that there is an association between SDB in pediatric patients and changes in C-reactive protein and insulin levels.[75] For a thorough review of the studies that have delved into these medical complications of pediatric SDB, please see the Technical Report of the American Academy of Pediatrics Recommendation on the Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome.[75] Certainly more longitudinal studies are needed to further characterize and delineate the prevalence and long-term consequences of these metabolic, cardiovascular, and inflammatory complications of SDB in children.

Economically, there have been few studies that have addressed the burden of cost of SDB in pediatrics. Tarasiuk and colleagues looked at a cohort in Israel and found that pediatric patients with SDB used health care resources 215% more than healthy matched controls.[91] Other studies have documented the increased cost in pediatric patients with sleep problems or specific cohorts with other medical problems such as sickle cell disease, but the increased cost has not been delineated for SDB.[92,93] More research is needed, but it stands to reason that children with decreased neurocognitive capabilities and increased medical problems not only will use more health care and social resources but also will not achieve their full potential, a cost that cannot be truly measured.

Other Sleep-Disordered Breathing in Pediatrics

Other SDB in children includes apnea of infancy, congenital central hypoventilation syndrome, central apnea related to obesity, SDB related to neuromuscular disease, and central sleep apnea related to Chiari malformation, but there are no large studies evaluating the prevalence and cost of these types of SDB.[12]

HYPERSOMNIA

The definition of hypersomnia has not been consistent in population-based studies. Some studies define it as perceived insufficient sleep, whereas others define it as too many or too few hours slept compared with the "ideal" number of hours. Still others define hypersomnia as the feeling of "sleepiness" or the ability and urge to sleep during the day. The previous edition of this book reviewed these difficulties defining hypersomnia and includes extensive tables and references, which estimate a prevalence between 0.3% and more than 30%, depending on how hypersomnia is defined.[14] Ohayon and colleagues also reviewed the literature and found that there was little consistency in study methods reporting either excess of sleep or increased propensity to sleep during wakefulness.[94] As such, hypersomnia is acknowledged as a worldwide problem, but one that needs more exploration. For the purposes of this review, prevalence data are divided by diagnosis, limited to the central disorders of hypersomnia as defined by the ICSD3, excluding hypersomnia due to a medical problem, because that topic is too broad for this context and depends on the specific medical problem.[12]

Narcolepsy

At the time of the last edition of this book, there were approximately 30 studies evaluating the prevalence of narcolepsy in the general population. Clear ethnic differences were seen in a handful of studies, but overall, the prevalence seemed to range from 0.025% to 0.05%, with Japan as an outlier, having the highest prevalence at 0.16%. These studies had significant differences in methodology, including patient questionnaires, population reviews, chart review, physician interview, self-report, and HLA typing.[14] More recently, Wijnans and colleagues combined seven patient databases in six European nations and compared the incidence of narcolepsy with or without cataplexy before, during, and after the H1N1 pandemic and reported pooled incidence of about 1 per 100,000 person-years.[95] These results suggested a higher incidence than the 0.6 per 100,000 reported by Longstreth and colleagues in the United States based on multistage screening of patients.[96] Similar findings were reported by Silber and associates from Olmstead County in Minnesota at an incidence of 1.37 per 100,000 person-years for narcolepsy with or without cataplexy and an incidence of 0.74 per 100,000 person-years for narcolepsy with cataplexy.[97] To date, there have been no published epidemiologic studies using the ICSD3 definitions of narcolepsy type 1 and type 2.[12]

After the 2009 H1N1 pandemic in China and the introduction of the ASO3 adjuvant H1N1 vaccine (Pandemrix) in Europe, narcolepsy with cataplexy appeared to have an escalating incidence in certain European countries and parts of China in which H1N1 influenza was epidemic, starting in Sweden and Finland.[98,99] China also noted a sharp increased incidence of narcolepsy after the H1N1 pandemic.[100] This acute increase in diagnoses of narcolepsy in 2009 prompted larger studies of national patient registries across Europe to examine the incidence of narcolepsy and what was previously defined as narcolepsy with cataplexy. In Sweden, the incidence in those who received the Pandemrix vaccine was found to be increased three-fold in those younger than 20 years, two-fold in those 21 to 30 years, and unchanged in those older than 40 years.[101] A small increase in narcolepsy among unvaccinated individuals was also seen; however, it is unknown whether that is due to population bias, increased recognition, or perhaps a true increase related to H1N1 infection.[100,101]

Studies show conflicting results regarding whether narcolepsy preferentially affects men or women, suggesting that more research is needed in this area. There are clear increases in the incidence of narcolepsy in people 10 to 30 years of age compared with other age groups, even excluding those affected by the H1N1-related increase reported previously.[96,97,102] It is thought that there is an autoimmune component because there is a known genetic association with HLA DQB1*0602 as well as an increased relative risk in first-degree relatives.[103]

The burden of narcolepsy is significant in terms of morbidity, health care costs, and societal costs. A higher rate of motor vehicle crashes is reported in narcolepsy, and drowsy driving is a known risk factor for increased collisions. Furthermore, the Burden of Narcolepsy Disease (BOND) study suggested narcoleptic patiens have two to three times the health care use and costs compared with age-matched controls in the United States, higher rates of short-term disability, and more missed work days, suggesting implications for long-term productivity.[104] A Danish study by Jennum and colleagues reported increased morbidity from endocrine, other sleep-related, neurologic, musculoskeletal, ophthalmic, and respiratory disorders in persons with narcolepsy.[105] Ohayon and associates also reported a mortality rate increase of 1.5 in narcoleptic patients across age groups.[106]

Idiopathic Hypersomnia

The prevalence of idiopathic hypersomnia is not known, although it is estimated to be less than that of narcolepsy based on referrals to sleep centers.[12,94,107,108] Further research is needed in this area to determine its prevalence.

Kleine-Levin Syndrome

Kleine-Levin syndrome is known to most commonly present in the second decade, although there are outliers, and it is far more common in males.[109,110] Prevalence is estimated at 1 to 2 per million based on retrospective studies. Additionally, there may be increased risk within families with an affected member.[12,109]

Behaviorally Induced Insufficient Sleep

Behaviorally induced insufficient sleep has been the focus in only two studies estimating prevalence. Komada and colleagues estimated prevalence at 7.1% with a 7:3 male preponderance and Pallesen and associates reported a prevalence of 10.4% of behaviorally induced insufficient sleep in Norwegian high school students, although 22.3% of all the students reported excessive sleepiness.[111,112] However, more studies are needed because cultural pressures and norms most likely play a role in incidence and prevalence.[12]

Other Hypersomnia

Hypersomnia due to a medication or substance depends on the substance used and whether the patient is intoxicated or withdrawing. Prevalence data also depend on which substance is used. However, hypersomnia from stimulant withdrawal occurs most frequently in teens and young adults.[12] Similarly, hypersomnia associated with a psychiatric disorder also depends on the disorder being considered and medications used, but it is known to occur in more than 50% of people with seasonal affective disorder.[12]

Clearly more studies are warranted in determining the epidemiology of hypersomnia, particularly because most of those affected present in the second and third decades, making hypersomnia a lifelong disease and one that affects quality of life significantly.[104,105,113]

CIRCADIAN RHYTHM DISORDERS

Circadian rhythm disorders (CRDs) encompass several disorders as defined by the ICSD3. Epidemiologic data are divided according to each disorder; however, true prevalence is not known for most disorders or for CRD as a whole.

Delayed Sleep Phase

Delayed sleep phase (DSP) is probably the most common of the CRDs and may account for 5% to 10% of patients referred to sleep clinics.[12,114] The true prevalence of DSP is not known; however, it has been estimated to be as high as 7% to 16% in adolescents[12] and as low as 0.1% to 0.2% in the general population.[114,115] Adolescence is the peak age for DSP, with the circadian clock "lateness" peaking at 20 years for most people.[116] Several studies examining prevalence among teenagers have reported rates ranging from 1.1% in Australia to 3.3% in Norway to 8.4% in a different Norwegian study.[117-119] A study out of San Diego found a prevalence of less than 1% in adults ages 40 to 64 years.[120] Two studies have shown that male gender predisposes to eveningness, but one additional study suggested females are predisposed to eveningness.[121-123] No studies have been done to examine racial and ethnic differences among those with DSP.[12] More studies are needed to determine the true prevalence and effect of race and ethnicity in DSP.[124] DSP has a familial pattern in up to 40% of patients, and there is some evidence that DSP is associated with human *hPer3,* arylalkylamine *N*-acetyltransferase, human leukocyte antigen, and *CLOCK* polymorphisms, although it has not been validated in all studies.[12,125,126]

The clinical significance of DSP has primarily been identified as a problem in teens and young adults. Saxvig and colleagues found that grades were negatively affected and that smoking, alcohol use, anxiety, and depression were increased with DSP.[117] Also in teens, DSP is a risk for hypersomnia because of early school start times and, as such, is a risk for the consequences of hypersomnia.[127] Please see the Hypersomnia section of this chapter for further description of the consequences of hypersomnia.

Advanced Sleep Phase

Advanced sleep phase (ASP) is rare and is thought to be found in 1% of the population, although this may be an underestimation because patients may not be sufficiently distressed to seek help. With the exception of familial cases, there are only four reported cases in the literature, and there is not a known gender predilection.[12] Indeed, a study of 10,000 people in Norway found no cases of ASP by strict ICSD criteria.[128] Older age and neurodevelopmental disorders seem to predispose people to ASP.[12,124]

Irregular Sleep-Wake Rhythm Disorder

Irregular sleep-wake rhythm disorder is also quite rare, and there are a paucity of data to support prevalence numbers. However, neurologic disease, mental disability, and dementia all are risk factors, and the prevalence increases with age.[129] There is not enough information to comment on gender or ethnic influences.[12,124]

Non-24-Hour Sleep-Wake Rhythm Disorder

Non-24-hour sleep-wake rhythm disorder is found in more than half of completely blind individuals, and the disorder is demonstrated in 5% to 15% of other types of blindness.[12,114] In blind individuals, onset can be at any age, and there are no data to suggest increased risk based on sex. In sighted individuals, non-24-hour sleep-wake rhythm disorder is rare and associated with psychiatric disorders (25%), traumatic brain injury, and male gender, although it is unclear whether male gender is truly a risk.[12,124]

Shift Work Disorder

Shift work disorder has garnered more interest over the past few years with the recognition that increased morbidity is associated with shift work in general. The prevalence in the general population is estimated to be 1% to 4% and 10% to 33% in shift workers.[12,130-132] Shift work disorder results in accidents, decreased alertness, decreased quality of life, and morbidity.[133] The specific societal and personal consequences are addressed in the chapter in this text about shift work disorder.

Jet lag disorder and circadian sleep-wake disorder not otherwise specified do not have sufficient studies documenting epidemiologic data to comment.

INSOMNIA

The ICSD3 has reclassified insomnia, abandoning previous insomnia categories of primary and secondary insomnia for a broader classification, including chronic insomnia disorder, short-term insomnia disorder, and other insomnia disorder.[12] The criteria for diagnosis of insomnia disorders include a difficulty initiating or maintaining sleep or undesired early awakening despite adequate opportunity for sleep that results in daytime, functional, or social impairment or distress.[12] The prevalence estimates of insomnia vary based on the diagnostic criteria used. In fact, the America Insomnia Survey evaluated 10,094 participants by different diagnostic criteria and found a range of insomnia prevalence within that group from 3.9% based on International Classification of Diseases, 10th revision (ICD-10) criteria to 22.1% based on *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition (DSM-IV) criteria.[134] The prevalence range also varies depending on whether subjects are evaluated for dissatisfaction with their sleep, any symptoms of insomnia (sometimes restricted to a certain frequency or severity of these symptoms), or actually qualifying for an insomnia diagnosis.[135] For example, in two separate Canadian cohorts evaluated by telephone survey, 19.8% to 25.3% reported sleep dissatisfaction, 29.9% to 40.2% reported at least one insomnia symptom, and 9.5% to 13.4% met criteria for insomnia based on a combination of DSM-IV and ICD-10 diagnostic criteria.[136,137] This difference is also reflected in a study of 12,778 French individuals, which reported at least one sleep problem more than three times weekly in 29% of subjects, but a prevalence of 19% for those with daytime consequences of the sleep complaint.[138] Bixler and colleagues studied a group of 1741 subjects in Central Pennsylvania with PSG and questions about insomnia and poor sleep. This study showed a prevalence of sleep difficulty in 22.4% and chronic insomnia (symptoms longer than 1 year) in 7.5%.[139] The subjects who did not have chronic insomnia at baseline were evaluated 7.5 years later with the same questions about insomnia and poor sleep. Incidence of chronic insomnia in this group was 9.3% (12.9% in women and 6.2% in men).[140] Within this cohort, incident poor sleep in those with baseline normal sleep was 18.4%. Those who had poor sleep at baseline had a remission rate of 44%, whereas 39% remained poor sleepers and 17% transitioned to chronic insomnia.[141] Regarding acute or short-term insomnia, Ellis and colleagues reported prevalence of 9.5% in the United States and 7.9% in the United Kingdom.[142] Based on these studies and others, the prevalence range of insomnia diagnosis is likely 4% to 22%, with some symptoms of insomnia affecting 20% to 45% of adults.[134,135,143-146]

The prevalence of insomnia appears to be increasing over time. A study performed in England assessing insomnia prevalence over 15 years reported prevalence of 3.1% in 1993 and 5.8% in 2007.[147] Based on the National Health and Nutrition Examination Survey 1999–2010, use of prescription medications for insomnia have also increased over time, with 2% of studied individuals using such medications in the 1999–2000 period and 3.5% in the 2009–2010 period.[148] Notably, the estimated prevalence of insomnia may be underestimated in certain populations because many of those with symptoms do not seek help from a physician.[136,149]

Pediatric Insomnia

Similar to adults, insomnia prevalence rates among adolescents vary based on study design and definitions of insomnia. Pediatric insomnia includes behavioral insomnia of childhood, psychophysiologic insomnia, and sleep-onset association disorder, and all can have overlap with other sleep disorders. As such, prevalence ranges vary. If simply insomnia symptoms are reported, prevalence ranges from 25% to 35%, whereas the prevalence rates are lower when defined by insomnia diagnostic criteria (DSM-IV or ICSD22), ranging from 4% to 14%.[150-154] In a study of 10,220 adolescents in Norway, prevalence of insomnia was 23.8% based on DSM-IV criteria, 18.5% based on DSM-5 criteria, and 13.6% based on the quantitative criteria for insomnia, which requires a 6-month duration of symptoms.[155] The presence of insomnia among adolescents can have significant influence on safety, cognition, and mood and may predict future sleep habits.[156-158] Studies consistently show a higher prevalence of insomnia among adolescent girls than boys, but this difference appears to emerge only after onset of menses.[150,155,156]

Insomnia is also common among children and preadolescents, affecting up to 41% of those aged 2 to 14 years.[159] Most reports, however, place pediatric insomnia prevalence at 10% to 20%. An evaluation of 5- to 12-year-olds by Calhoun and colleagues found a prevalence of insomnia symptoms in 19.3%, with the highest prevalence among preteen girls (age 11 to 12 years).[158] Similar prevalence was found in a study of 4989 Australian preschoolers (age 4 to 5 years), among whom 19.8% had mild sleep problems and 13.8% had moderate to severe sleep problems. Those who reported difficulty falling asleep were more likely to have worse health-related quality of life and diagnosis of attention deficit disorder.[160] The Sleep in America poll of 2004 surveyed parents of children from infancy to 10 years of age, and sleep problems were reported among 6.3% of infants, 10.5% of toddlers, 10.2% of preschoolers, and 10.8% of school-aged children. Those with reported sleep problems were more likely to have a later bedtime and to have a parent present at sleep onset.[161] An important consideration in evaluation of childhood sleep complaints is that there can be considerable impact of parental perception of sleep problems based on cultural and ethnic differences.[162,163]

During infancy, the prevalence of sleep problems is approximately 10%, with most concerns being related to nocturnal awakenings and short sleep duration.[164] Interestingly, sleep difficulty at this age appears to predict sleep problems during the early childhood years. For example, infants who do not self-soothe are more likely to have difficulty with sleep onset at age 2 years.[165] Other longitudinal studies show that sleep problems at 6 to 12 months of age predict sleep trouble at 3 to 4 years of age.[164,166,167] These changes may not persist to older ages because another longitudinal study showed no predictive value of infant sleep problems on sleep outcomes at age 6 years.[168] Regardless, insomnia is common in the pediatric population and can have a significant effect on health, cognition, and safety among patients and their parents.

Insomnia Demographics

The prevalence of insomnia is consistently reported to be higher in women than men.[135,139,169,170] A meta-analysis performed by Zhang and colleagues found a risk ratio of 1.4 for men versus women. The higher prevalence among female adults is present at all age groups but increases with age, with a 1.28 higher risk in women aged 15 to 30 years and a 1.73 risk ratio among those older than 65 years.[171] Aging is often identified as a risk factor for insomnia, but some studies actually show higher rates of incident and persistent insomnia in middle age or in younger age groups.[135,141,172,173] Insomnia in older adults may be associated with more nocturnal awakenings and early morning awakening than in younger adults.[170] The association of age with insomnia varies based on the population studied and may depend on the presence of associated health conditions (which could increase with aging) and the increased use of technology among younger age groups.[173] Race and ethnicity, education, and socioeconomic status may also influence the prevalence of insomnia.[135,170] The influence of these demographic features and health comorbidities on insomnia is further explored in Chapters 81 and 84.

One key thing to remember about insomnia is that it must result in daytime, functional, or social impairment or distress. As such, the effect on quality of life and the cost and burden of insomnia are difficult to quantify, but it stands to reason that there is a societal cost to insomnia related to all the problems related to sleep deprivation, as well as the known medical morbidities related to sleep disorders.

RESTLESS LEGS SYNDROME

Restless legs syndrome (RLS) is characterized by an urge to move the legs, often associated with an uncomfortable sensation, which occurs during periods of rest, is relieved by movement, occurs predominantly in the evening, and causes distress.[12] RLS can be divided into primary and secondary RLS, which is RLS due to precipitating factors, such as iron deficiency, pregnancy, renal failure, neuropathy, and use of certain medications. Primary RLS tends to have younger onset of symptoms, and those with primary RLS are more likely to have an affected family member. Comparatively, secondary RLS usually develops at an older age.[174] Studies addressing the prevalence of RLS may differ in the diagnostic criteria employed and as to whether diagnosis was based on questionnaire report of symptoms, physician diagnosis of RLS, or frequency and severity criteria.[175] For example, Allen and colleagues studied a cohort of subjects in Western Europe, where prevalence of RLS symptoms among 10,564 respondents was 7.6%, but the prevalence decreased to 3.5% when patients underwent physician interview for diagnosis.[176] Similarly, in the United States the prevalence, based on the four diagnostic criteria of RLS by questionnaire screening, was 7.3%.[177] Further, more detailed evaluation to exclude secondary RLS resulted in an adjusted prevalence of primary RLS of 2.4%. The prevalence of moderately to severely distressing symptoms at least twice per week decreased further to 1.5%.[177] Using International Restless Legs Syndrome Study Group criteria, prevalence rates range from approximately 5% to 15%.[174,175,178-181] The prevalence of distressing symptoms occurring at least twice weekly is lower, from approximately 1% to 10%.[174-177,182] Regardless of which criteria are used to establish prevalence, RLS is common and is likely underdiagnosed.[174]

RLS is consistently reported to be approximately twice as high in women as in men.[175,181] There is a higher prevalence of RLS with increasing age.[183] Regarding ethnicity, the prevalence of RLS has been reported to be lower in Asian populations in some studies, but other studies in Asian populations show comparable prevalence rates.[184,185]

The health and societal effect of RLS are significant. RLS can lead to insomnia,[186] daytime sleepiness,[187] increased health care costs, and increased indirect cost related to lost productivity.[188] Additionally, RLS has been associated in some studies with hypertension, cardiovascular disease,[186,189,190] gastrointestinal disease,[191] mood disorders,[182,186] stroke,[190] migraine,[192] multiple sclerosis,[193] and, in men, increased mortality.[194] More studies are needed to determine whether these medical comorbidities are a result of RLS or may actually predict incident RLS.[190,195] RLS has a clear relationship with periodic limb movements of sleep, which are present in 85% to 95% of patients with RLS.[183]

RLS prevalence in school-aged children is estimated to be 2% to 4%.[196] RLS can also be present in young children and infants, although its prevalence in this age group is not known.[197] Prevalence of RLS in adolescents ranges from 1% to 2.8%.[198] Pediatric RLS is associated with attention deficit hyperactivity disorder, depression, anxiety, insomnia, and daytime sleepiness.[197,199-201] As with adults, iron deficiency can be a cause of secondary RLS in the pediatric population.[196,197]

PARASOMNIAS

Parasomnias are divided in the ICSD3 into non–rapid eye movement (NREM) parasomnias, rapid eye movement (REM)-related parasomnias, other parasomnias, and normal variants.[12] Although some observations can be made about parasomnias in general, it is more pertinent to define the epidemiology by specific disorders. Epidemiologic studies of parasomnias must be, in large part, acknowledged as rough estimates because most studies are population based, involve varying methodologies, and may involve recall bias because many of the NREM parasomnias are more prevalent in childhood. Furthermore, parasomnias may be underestimated in cases in which there is no bed partner or self-report is required. Overall prevalence of parasomnias has been reported as high as 88% in preschoolers and 73% in children aged 3 to 13 years, with lifetime prevalence ranging from 4% to 67% in adults.[202-204] Prevalence data are summarized in Table 62-1.

NREM Parasomnias

NREM parasomnias, according to the ICSD3, include somnambulism (sleepwalking), confusional arousals, sleep terrors, and sleep-related eating disorder.[12] Confusional arousals, somnambulism, and sleep terrors tend to initially present in childhood and adolescence; however, somnambulism and confusional arousals may present at any age. All three disorders may persist throughout life, although the incidence of all three decreases with age.[205] These arousal disorders can be triggered by other sleep disorders (e.g., OSA, RLS) or environmental stimuli, and sleep deprivation and stress are known primers for the occurrence of arousal parasomnias.[12]

Table 62-1 Prevalence of Parasomnias

	Adult (%)	Pediatric (%)	Lifetime (%)
NREM Parasomnias			
Confusional arousals	1.8–6.9	ID	18.50
Sleep terrors	2.2–2.7	1–14.7	10
Somnambulism	0.6–3.9	3.5–14.5	22–30
Sleep-related eating	0.4–2.2	ID	4.50
REM Parasomnias			
REM sleep behavior disorder	0.38–2.1	ID	ID
Nightmare disorder	2–8	65–80.5	66.20
Recurrent isolated sleep paralysis	4.7–41	ID	6.20
Other Parasomnias			
Sleep-related hallucinations	23	ID	ID
Enuresis	2–3	Age dependent	ID
Normal Variants			
Somniloquy	6.3–17.7	Age dependent	66.80

ID, Insufficient data.

Confusional Arousals

Even though confusional arousals have been described for years, both in literature as well as in cultural viewpoints and behaviors, very few epidemiologic studies have evaluated this parasomnia.[206] Ohayon and colleagues performed a telephone-based study on 13,057 subjects older than 15 years in the United Kingdom, Germany, and Italy and found a prevalence of 2.9%, with 1.9% of subjects having confusional arousals at least once a month.[206] In the United Kingdom alone, Ohayon and colleagues reported a prevalence of 4.2% in the almost 5000 people evaluated with the Sleep-EVAL system.[205] Bjorvatn and colleagues also performed a telephone based study on 1000 adults in Norway and found a lifetime prevalence of 18.5%, current prevalence (at least once in the last 3 months) of 6.9%, and current prevalence (and occurring at least once a week) of 1.8%.[204] Although no gender difference has been reported, confusional arousals have been reported more by night and shift workers.[206] In pediatric patients, confusional arousals are sometimes confused with somniloquy by the parents, and prevalence data are not well documented.

Sleep Terrors

The prevalence of sleep terrors varies by age, ranging from 2.2% to 2.7% in adults, with a lifetime prevalence of 10% and a much higher prevalence in pediatric populations.[12,204,205] Pediatric epidemiologic reports differ based on the age of subjects studied but range from 1% to 14.7% in children aged 3 to 13 years and older to 36.9% in 18-month-olds.[12,203,207] Night terrors generally abate by age 65 years and occur in less than 1% of subjects older than 65 years.[12]

Somnambulism

Somnambulism also has varying prevalence data. Lifetime prevalence based on a telephone survey by Bjorvatn and colleagues was 22.4%, current (in the last 3 months) was 1.7%, and current (at least once a week) was 0.6%.[204] Within a European population, Ohayon and colleagues reported the prevalence of somnambulism, defined as occurring frequently and being perceived as problematic by the patient, as 2%.[205] Bixler and colleagues reported a prevalence of 2.5% in the United States in the late 1970s.[208] More recently, Ohayon and colleagues reported a lifetime prevalence of nocturnal wandering of almost 30%.[209] Hublin and colleagues reported similar prevalence data in the Finnish twin cohort of 3.9% of male adults and 3% of female adults.[210] Lifetime data for "never" was similar to other studies at 73% to 74%.[210] Compiling these larger studies of somnambulism reveals a lifetime prevalence of 25% to 30% of sleepwalking or nocturnal wandering and a current prevalence of 0.6% to 3.9%, depending on the definition of "current." In pediatrics, the prevalence rates are higher at 3.5% to 14.5%.[202,203,211] There is a clear genetic predisposition, and the same things that can precipitate and prime for confusional arousals can precipitate and prime for somnambulism. Additionally, Z-drugs (nonbenzodiazepine hypnotics) may also precipitate events.[12,210]

Sleep-Related Eating Disorder

Few epidemiologic studies have been performed on sleep-related eating disorder, although a telephone survey by Bjorvatn and colleagues reported a lifetime prevalence of 4.5%, current prevalence (in the last 3 months) of 2.2%, and current prevalence (at least once per week) of 0.4%.[204] In a self-report questionnaire, prevalence in college students was 4.6%, 8.7% in outpatients with an eating disorder, and 16.7% in inpatients with an eating disorder.[212] More studies are needed, but sleep-related eating disorder generally presents in the third decade, tends to be more common in women, and can be associated with sedative-hypnotic use.[12]

REM Parasomnias

REM parasomnias include REM sleep behavior disorder (RBD), nightmare disorder, and recurrent isolated sleep paralysis (RISP). Because REM behavior disorder and nightmares

arise from REM sleep, they most often occur in early morning hours and are associated with recall of dream content. As with other sleep disorders, the diagnostic criteria include the presence of distress related to the condition.

REM Behavior Disorder

RBD is diagnosed by the presence of repeated vocalizations or complex movements that occur during REM sleep, with documented REM sleep without atonia (RWA) on PSG.[12,213] The prevalence of RBD in the general population is estimated to be 0.38% to 2.1%.[214-216] Kang and colleagues additionally reported that the prevalence of subclinical RBD (RWA without history of sleep-related injury) is 4.95%. There is a male predominance of RBD, and it is more common in older individuals, typically with an onset older than 50 years.[217,218] The recognition of RBD is important because it can be associated with injury to the patient or bed partner.[217,218] Patients with RBD tend to have increased muscle twitches and periodic limb movements as well.[217]

There is a significant association between RBD and neurodegenerative diseases, such as Parkinson disease, dementia with Lewy bodies, multiple systems atrophy, and mild cognitive impairment.[218-222] In fact, up to 81% to 90% of those with RBD ultimately develop neurodegenerative disease, with risk increasing over time: 33.1% at 5 years, 75.7% at 10 years, and 90.9% at 14 years.[222,223] Additionally, in those who have synucleinopathy (Parkinson disease, dementia with Lewy bodies, multiple systems atrophy), there is an increased prevalence of RBD compared with the general population and those with other neurodegenerative diseases, such as tauopathies (Alzheimer disease, progressive supranuclear palsy, corticobasal degeneration).[220] In multiple systems atrophy, the prevalence of RBD is particularly high, affecting 90% to 100% of patients.[224-226] This strong association calls into question whether *idiopathic* RBD truly exists. In younger patients with onset of RBD symptoms before age 50 years, RBD is often associated with narcolepsy, and about one third of patients with narcolepsy with cataplexy have RBD.[227-230] There may be a genetic predisposition to RBD because Dauvilliers and colleagues found increased odds of having dream enactment behavior in family members of patients with confirmed idiopathic RBD.[231] In addition to chronic RBD, there can also be acute episodes that are related to medications, such as selective serotonin reuptake inhibitors, alcohol withdrawal, or drug abuse.[12] RBD has been described in the pediatric population as well, with one study reporting RBD in 32.3% of narcoleptic patients younger than 19 years.[230] Additionally, Lloyd and colleagues performed a chart review within their sleep center and found 15 children (age range, 3 to 17 years) with RBD or RWA.[232]

Nightmare Disorder

Nightmare disorder diagnosis requires repeated episodes of dysphoric, often threatening dreams that cause significant functional impairment and from which a person awakens with rapid orientation, dream content recall, and a feeling of distress.[12] The prevalence studies of nightmares are difficult to compare because some studies query "bad dreams" rather than establishing a diagnosis of nightmare disorder. Frequent nightmares that cause distress have been reported to affect 2% to 8% of the general population, although lifetime prevalence is higher.[12,153,204,233-237] In a cross-sectional study of 1000 randomly selected adults, Bjorvatn and colleagues reported the lifetime prevalence of nightmares as 66.2% (72% in women and 61% in men) and the prevalence of nightmares in the preceding 3 months as 19.4%. In addition, 2.8% of subjects reported current nightmares occurring at least once per week.[204] Most epidemiologic studies of nightmares in adults show a higher prevalence among women than men, although this gender difference does not seem apparent in children and older adults.[204,235,238] Nightmare disorder is more common among patients with psychiatric diagnoses and those who have undergone traumatic events.[235,239-241] In fact, a study by Swart and colleagues showed that 29.9% of subjects undergoing outpatient psychiatric treatment reported nightmares at least once per week that resulted in distress.[240] Nightmares can also be associated with insomnia, with a prevalence of approximately 18%.[233,242,243]

The prevalence of occasional nightmares in the pediatric population is likely higher than that among adults. In young children, a longitudinal study (with parental report at 29, 41, and 50 months of age and at 5 and 6 years of age) showed a 65% to 69% prevalence of nightmares "sometimes," whereas there was only a 1.3% to 3.9% prevalence of nightmares "often."[244] Other studies have reported prevalence up to 80.5% in children between the ages of 4 and 12 years.[245] Frequent nightmares in children can be associated with insomnia, hyperactivity, poor academic performance, and mood disturbances.[246] At the opposite end of the age spectrum, nightmares seem to decrease in frequency and prevalence with aging.[241]

Recurrent Isolated Sleep Paralysis

RISP has been described by many names within many cultural contexts, making a true epidemiologic study difficult. Most studies have evaluated small, specific populations, reporting varying prevalence estimates of 4.7% to 41%.[247,248] Ohayon and colleagues performed a larger telephone survey with the Sleep-EVAL system of 8085 subjects in Germany and Italy and reported a lifetime prevalence of 6.2%.[247] Several studies have documented the association of RISP with anxiety and panic disorder, and the association of sleep paralysis with narcolepsy is also well documented.[12,247,249] Bell and colleagues have also reported an increase in the prevalence of RISP in the African American population compared with reports of the population as a whole.[248] RISP usually presents in the second decade and is thought to affect men and women equally.[12]

Other Parasomnias

Other parasomnias include exploding head syndrome, sleep-related hallucinations, sleep enuresis, parasomnia due to a medical disorder, parasomnia due to a medication or substance, and parasomnia, unspecified. There are insufficient epidemiologic data to comment on any of these other than sleep-related hallucinations and sleep enuresis.

Sleep-Related Hallucinations

Sleep-related hallucinations are divided into hypnagogic and hypnopompic hallucinations. Using a telephone survey with the Sleep-EVAL system in the United Kingdom, Germany, and Italy, Ohayon and colleagues reported an overall prevalence of any type of hallucination at 38.7%, but when divided into subgroups, prevalence was 18% for hypnagogic and 4.9%

for hypnopompic hallucinations.[250] In this population, hallucinations were found more commonly in women and in the younger age groups.[250] In a smaller study of medical students in Spain, sleep-related hallucinations occurred with greater frequency in those with insomnia.[170] There is, of course, a well-documented higher prevalence of sleep-related hallucinations in narcolepsy.[12] More studies are needed to determine the frequency of sleep-related hallucinations in the general populations outside of Western Europe.

Sleep Enuresis

Sleep enuresis has not been extensively studied in adults, but existing data suggest a prevalence between 2% and 3%. However, nocturnal enuresis has been reported to be as low as 0.2% in a military population and as high as 6% in the United Kingdom.[251-253] It occurs with much higher frequency in pediatric patients, with prevalence estimates of 15% to 25% of 5-year-olds and decreasing with age until adulthood.[202,203] Enuresis has a strong familial component and is associated with SDB in adults and children. This disorder can result in significant psychosocial ramifications and expense.[12]

Normal Variants

Somniloquy

Somniloquy (sleeptalking) is considered a normal variant and occurs in children and adults. In a cross-sectional study in adults, Bjorvatn and colleagues reported a lifetime prevalence of 66.8%, current (occurring once in the last 3 months) of 17.7%, and current (occurring at least once a week) of 6.3%.[204] In children, the prevalence numbers are higher; for example, Laberge and colleagues report a prevalence of 55.5% in 3- to 13-year-olds.[203]

Despite the work already completed on the epidemiology of parasomnias, more is needed. Most of the previously mentioned studies have primarily been done in the Western European population or in the United States. Additionally, recall bias continues to affect epidemiology of parasomnias. Nonetheless, the presence of parasomnias is significant; in the cross-sectional study by Bjorvatn, only 9.8% of subjects did not have a parasomnia, making parasomnias prevalent indeed.

CLINICAL PEARLS

- The epidemiology of sleep medicine depends on the population studied and methodologies used.
- Sleep disorders have significant health, economic, and safety implications.
- Most people have at least one sleep disorder over their lifetime.

SUMMARY

The epidemiology of sleep medicine is a complex topic, and although many excellent studies have been published, more work is needed to define the true prevalence and incidence of each of the sleep disorders, as well as their health and societal implications. The epidemiologic outcomes differ depending on the population studied, the research methodologies employed, and the diagnostic criteria used. As the definitions of sleep disorders evolve (e.g., with the recent publication of the ICSD3), we hope that study methodology will be standardized and the burdens of the diseases of sleep medicine will be better identified, allowing for targeted treatments. This chapter summarizes the available data on this broad topic.

Selected Readings

Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999–2010. *Sleep* 2014;**37**(2):343–9.

Bixler EO, Vgontzas AN, Lin HM, et al. Sleep disordered breathing in children in a general population sample: prevalence and risk factors. *Sleep* 2009;**32**(6):731–6.

Bjorvatn B, Gronli J, Pallesen S. Prevalence of different parasomnias in the general population. *Sleep Med* 2010;**11**(10):1031–4.

Black J, Reaven NL, Funk SE, et al. The Burden of Narcolepsy Disease (BOND) study: health-care utilization and cost findings. *Sleep Med* 2014;**15**(5):522–9.

Iranzo A, Fernandez-Arcos A, Tolosa E, et al. Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: study in 174 patients. *PLoS ONE* 2014;**9**(2):e89741.

Lumeng JC, Chervin RD. Epidemiology of pediatric obstructive sleep apnea. *Proc Am Thorac Soc* 2008;**5**(2):242–52.

O'Connor GT, Lind BK, Lee ET, et al. Variation in symptoms of sleep-disordered breathing with race and ethnicity: the Sleep Heart Health Study. *Sleep* 2003;**26**(1):74–9.

Ohayon MM, Guilleminault C, Chokroverty S. Sleep epidemiology 30 years later: where are we? *Sleep Med* 2010;**11**(10):961–2.

Parthasarathy S, Vasquez MM, Halonen M, et al. Persistent insomnia is associated with mortality risk. *Am J Med* 2015;**128**:268–75.

Petit D, Pennestri MH, Paquet J, et al. Childhood sleepwalking and sleep terrors: a longitudinal study of prevalence and familial aggregation. *JAMA Pediatrics* 2015;**169**:653–8.

Roth T, Coulouvrat C, Hajak G, et al. Prevalence and perceived health associated with insomnia based on DSM-IV-TR; International Statistical Classification of Diseases and Related Health Problems, tenth revision; and Research Diagnostic Criteria/International Classification of Sleep Disorders, second edition criteria: results from the America Insomnia Survey. *Biol Psychiatry* 2011;**69**(6):592–600.

Schenck CH, Boeve BF, Mahowald MW. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep Med* 2013;**14**(8):744–8.

Wijnans L, Lecomte C, de Vries C, et al. The incidence of narcolepsy in Europe: before, during, and after the influenza A(H1N1)pdm09 pandemic and vaccination campaigns. *Vaccine* 2013;**31**(8):1246–54.

Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* 2002;**165**(9):1217–39.

Zhang B, Wing YK. Sex differences in insomnia: a meta-analysis. *Sleep* 2006;**29**(1):85–93.

A complete reference list can be found online at ExpertConsult.com.

Chapter
63

Sleep Medicine, Public Policy, and Public Health

Raghu Pishka Upender

Chapter Highlights

- This chapter emphasizes the growing interface between sleep medicine and public health by surveying the various dimensions of sleep known to affect health outcomes. These dimensions include sleep duration, sleep timing, sleep efficiency, sleep quality, alertness, and performance.
- This chapter also surveys how the evolving knowledge of sleep physiology and sleep disorders is shaping public policy in various safety-sensitive occupations, such as the transportation and the health care industries. Brief historical backgrounds behind many of the regulations are also discussed.
- The chapter ends by discussing how the sleep medicine community can shape public policy and improve sleep health and safety by partnering with governmental regulatory agencies and industry stakeholders.

SLEEP MEDICINE AND PUBLIC HEALTH

Sleep is an essential biologic function that is thought to have evolved to help organisms cope with light and dark cycles that occur naturally on Earth. The physiologic importance of sleep is underscored by the fact that humans spend nearly one third of their lives sleeping. Although the function of sleep remains to be fully elucidated, current evidence suggests key roles in development, neurocognitive performance, mood regulation, and metabolic homeostasis. Accumulating evidence suggests that even partial sleep deprivation has far-reaching consequences on health and well-being. *Sleep health*, a term often encountered in the media, is an apt term emphasizing the strong connection between sleep and health.

Public health interest in sleep has been steadily increasing over the past two decades. This interest is driven not only by the growing knowledge of the physiologic importance of sleep but also by the recognition that sleep deprivation and sleep disorders are pervasive among the general population. In a health survey conducted by the Centers for Disease Control and Prevention in the United States, approximately 29% of U.S. adults reported sleeping less than 7 hours per night.[1] The American Institute of Medicine estimates that 50 to 70 million people have chronic sleep and wakefulness disorders.[2]

Poor sleep in its various manifestations is associated with adverse health outcomes and a huge economic burden at both the individual and societal levels. The costs of sleep loss and sleep disorders are estimated to be billions of dollars. For example, the economic burden of obstructive sleep apnea alone is estimated to be between $65 and $165 million in both direct and indirect costs related to comorbid medical conditions, hospitalization, accidents, and productivity loss.[3]

Initiatives to increase public awareness of sleep health are underway in many countries. In the United States, for example, the government-sponsored public health initiative, *Healthy People 2020,* now includes a dedicated section on sleep health to promote public awareness of the ill effects of sleep loss and sleep disorders.[4] The U.S. Army has adopted a program called *Performance Triad* that includes sleep as one of the three pillars of health and performance alongside nutrition and physical activity.[5] Through efforts like this, sleep health literacy is improving, but the chasm between knowledge and health behavior remains formidable. To fully realize the benefits of sleep, public health initiatives will need to spotlight the various dimensions of sleep—sleep duration, sleep timing, sleep efficiency, subjective sleep satisfaction, and daytime alertness—that have been associated with health outcomes.[6]

Sleep Duration

The amount of sleep required for optimal physiologic function varies across age and individuals. For example, newborns spend 16 hours or more sleeping, much of it in rapid eye movement (REM) sleep. By age 2 years, sleep duration declines to about 11 to 12 hours. School-aged children sleep about 10 hours, whereas adolescents sleep 9 to 10 hours each night.[7] Adults sleep about 8 hours per night when unfettered by lifestyle demands. Total sleep time declines with age at a rate of 10 minutes per decade, and sleep efficiency declines at a rate of 3% per decade. Components of sleep such as slow wave and REM sleep also decrease, but at different rates.[8]

Modern lifestyles with long working and commuting hours are increasingly encroaching on traditional sleep time and creating a society that is chronically sleep deprived. By some reports, average sleep duration in America has declined by 20% over the past century.[9] The advent of the electric light bulb and cheap artificial light has had dramatic effects on the 24-hour sleep-wake patterns of humans. Light exposure activates brainstem arousal systems and attenuates sleep signal by suppressing melatonin. As a result, the peak circadian wake signal that normally occurs at the end of the day is delayed and allows individuals to remain awake well into the night.[10] Early start times at school and work prevent compensatory sleep in the morning. Thus sleep duration is squeezed at both

ends. A 2010 National Health Survey found that 30% of all employed U.S. adults (40.6 million workers) reported averaging less than 6 hours sleep per night.[11] Even more alarming is the declining sleep duration of school-aged children, adolescents, and young adults. More than one fourth of high school and college students were found to be sleep deprived.[12] The 2006 *Sleep in America Poll* by the National Sleep Foundation found that sleep duration declined from 8.4 to 6.9 hours per night among 6th and 12th graders, even though physiologic sleep need does not decline significantly across this age span.[13] The problem is global as indicated by a recent systematic review, which found that sleep duration declined by 0.75 minute nightly per year and 1 hour per night over the study period of 1905 to 2008.[14]

It should be noted that not all studies have shown declining sleep duration. For example, a 2003 face-to-face interview of nearly 2000 British subjects aged 16 to 93 years found that self-reported sleep duration was not significantly different compared with sleep surveys conducted in 1969.[15,16] Another study using time use diaries from eight surveys across a 31-year period from 1975 to 2006 indicated that the odds of short sleep had not changed for part-time workers, retired workers, homemakers, or the unemployed and that the odds actually decreased for students, who represented less than 5% of the participants. Full-time workers were the only group that showed an increase in the odds ratio of short sleep of 1.19 (95% confidence interval, 1.00, 1.42; $P = .05$) over 31 years. Long work hours were much more common in those sleeping less than 6 hours, suggesting a possible cause-and-effect relationship.[17] Another study examining data from 15 countries over a period spanning between the 1960s and 2000s found that the average sleep duration of adults actually increased in seven countries: Bulgaria, Poland, Canada, France, Britain, Korea, and the Netherlands (range, 0.1 to 1.7 minutes per night each year) and had decreased in six countries: Japan, Russia, Finland, Germany, Belgium, and Austria (range, 0.1 to 0.6 minute per night each year). The findings were inconsistent in the United States and Sweden.[18,19] Some of the conflicting findings noted in these studies could be related to methodologic factors. It is important to note that most epidemiologic studies evaluating sleep duration employ subjective measures and are susceptible to recall and response bias. Subjects may confuse sleep time with time in bed. Therefore future public health surveillance of sleep practices should include more objective measures of sleep duration. Low-cost activity monitors and social networking platforms may enable collection of more objective measurement of sleep duration.

A large body of literature has linked sleep duration with health outcomes. One of the earliest associations between sleep duration and mortality risk was found in a prospective study of more than 1 million subjects who were followed for more than 2 years. Those who reported sleeping 7 hours per night had a lower death rate than those who reported either more or less sleep than this.[20] These findings were corroborated in a study of 7000 subjects in Alameda County, California who were followed for 9 years.[21] The analysis indicated that those sleeping less than 6 hours or more than 9 hours per night had 1.6 times the total age-adjusted death rate of those sleeping 7 to 8 hours per night. This U-shaped relationship between mortality rates and sleep duration has also been found in a number of studies around the world and holds true across the adult life span.[22] Studies have suggested that short sleep can lead to higher mortality rates even after controlling for comorbidities.[23] Sleep loss is likely to cause higher mortality rates through various adverse effects on physiology. Studies have found higher likelihood of hypertension,[24,25] atherosclerosis,[26] dyslipidemia,[27] and diabetes[28] in short sleepers. A number of studies in both adult and pediatric populations have shown strong associations between obesity and sleep duration, potentially mediated through alterations in hormones regulating appetite.[29] The rising rates of childhood obesity and diabetes are considered to be partly related to chronic partial sleep loss.

Sleep Timing

Traditional sleep-wake schedules are changing to meet the needs of 24/7 modern global economy. About 20 million U.S. workers (17.7% of the workforce) are estimated to work in shifts that at least partly fall outside the traditional 6 AM to 6 PM schdule.[30] As many as 4.3% of workers work primarily during the night. This trend has increased over the past 50 years with globalization of the economy. The rates of nontraditional work hours vary across industry and are the highest among protective service, food service, transportation, and health care industries. Nontraditional schedules can have significant effects on circadian processes that originated to cope with light-dark phases of the environment. The human circadian system is responsible for providing alerting signal during the light phase and sleep signal (melatonin) during the dark phase, thus optimizing sleep-wake functions. Sleep occurring during the light phase is often fragmented and short in duration because the sleep-promoting signal is absent or suboptimal. Studies have shown that night shift workers average about 30 to 60 minutes less sleep on average then daytime workers. Conversely, maintenance of alertness during the dark phase is extremely difficult for most people, especially in early morning hours when the circadian alerting signal reaches its nadir.

Adverse effects of altering the timing of sleep are readily apparent to most people traveling across time zones and are responsible for the phenomenon known as *jet lag*. Rapid travel across multiple time zones leaves circadian rhythms out of sync with the destination's light-dark cycles because circadian rhythms are slower to adapt and can only do so on an average rate of 1 hour per day. As a result, circadian signals conflict with environmental and social cues in the new location and lead to an unpleasant symptom complex that includes daytime fatigue, irritability, poor concentration, digestive problems, and excessive sleepiness and nocturnal insomnia. Less intense symptoms may be experienced by people who delay bed and wake-up time by 1 to 2 hours on weekends. The "social jet lag" could partially explain "Monday morning blues."

Circadian misalignment is a perpetual problem for shift workers because it is practically impossible for people to maintain consistent sleep-wake schedules that are out of phase with environmental light-dark cycles. Even small amounts of ambient light can drive circadian phase to shift toward the light or active phase of the rhythm. Sociologic factors force most shift workers to revert to a more traditional sleep-wake schedule on days off in order to spend time with family and friends and attend to business affairs. This creates a situation in which shift workers are always functioning and sleeping during adverse circadian phases. Ability to cope with

the physiologic challenges created by circadian misalignment varies across individuals and may be inheritable. Approximately 20% to 30% of shift workers experience what has come to be known as shift work disorder.[31] Symptoms are persistent and not dissimilar to jet lag and include fatigue, insomnia, impaired concentration, low mood, and memory. Workers experiencing shift worker disorder have higher rates of depression and anxiety and are more prone to accidents. Higher prevalence of gastric ulcers, heart disease, ischemic stroke, obesity, and metabolic syndrome has been noted in the literature. There has been recent interest in the connection between shift work and increased risk for cancers, especially breast cancer.[32] Melatonin is thought to play a role in tumor surveillance, and its suppression by nocturnal light exposure has been hypothesized as a potential mechanism of tumorigenesis in shift workers. Some recent studies have questioned whether the increased risk for cancer is real.[33,34]

Sleep Efficiency and Sleep Quality

Sleep efficiency is defined as total sleep time divided by time spent in bed trying to sleep and tends to decline with age. Although some of the decline is related to loss of gamma-aminobutyric acid-ergic neurons in the ventral preoptic area, multiple factors—medical, psychological, and sociologic—converge to compromise sleep efficiency with advancing age. It is important to recognize these factors because some are amenable to treatment.

One of the most prevalent sleep problems is insomnia, a condition characterized by difficulty with sleep onset and sleep maintenance. Chronic insomnia affects approximately 30 million Americans and causes much distress to its sufferers. A number of studies have indicated a striking association between insomnia and depression. Some postulate that insomnia may be an early marker for the onset of depression. Although the pathophysiologic relationship remains to be clarified, there may be overlap of neural pathways for anxiety, arousal, and circadian disturbance.[35] The close association of insomnia and depression also raises the tantalizing possibility that treating insomnia may prevent some cases of depression, but limited data are available.[36]

Sleep-disordered breathing is another common sleep disorder that is associated with poor sleep efficiency. The prevalence of obstructive sleep apnea (OSA), as defined by an apnea-hypopnea index greater than 5 and excessive sleepiness, is 9% in women and 24% in men in the adult population.[37] The prevalence of OSA is rising owing to the obesity epidemic and aging demography. Multiple factors are likely to be responsible for the increasing prevalence of sleep apnea in older individuals. These include age-related weight gain, reduced pharyngeal dilator tone, and decline in the sensorimotor responsiveness to hypoxia, hypercapnia, and respiratory load. The cessation of apneic and hypopneic events is often contingent on somatic or cortical arousal, and therefore sleep fragmentation is intrinsic to the pathophysiology of sleep apnea. Sleep fragmentation is very responsive to positive airway pressure therapy, which provides an opportunity to markedly improve sleep efficiency and quality.

Sleep efficiency is also reduced in patients suffering from restless legs syndrome, a neurologic condition characterized by an irresistible urge to move the legs and nocturnal limb movements that affects approximately 5% of the general population.[38]

A number of studies have shown an association between sleep efficiency, subjective sleep quality, and various health outcomes.[6] For example, a Japanese population-based cohort study of self-reported sleep parameters found that woman who reported poor awakening state experienced a higher mortality rate (relative risk, 1.97) compared with those who awakened normally.[39]

Alertness and Performance

The value of sleep for most individuals is its ability to restore alertness and improve performance during the waking period. When sleep is suboptimal, people feel non-restored and experience a number of neurocognitive deficits that include poor memory, decreased concentration, slowed reaction times, and impaired judgment. Elegant studies using psychomotor vigilance tests have shown dose-dependent slowing of response times with increasing sleep restriction.[40] Mean number of lapses on psychomotor vigilance tests increased with increasing sleep restriction. Performance precipitously dropped with sleep durations of less than 5 hours per night. Several nights of recovery sleep were required for performance to be restored to baseline levels.

Beyond their effects on daytime performance, decreased levels of alertness during the day have been linked to poor health. In a study of 5888 individuals, daytime sleepiness was the only sleep disturbance symptom that was associated with mortality, cardiovascular disease, and congestive heart failure.[41] These associations persisted in women after adjustment for age and other factors. In another study of community dwelling older adults (>65 years old), mortality rate was accelerated by 1.73 times in those who napped most of the time and made two or more errors on cognitive tests.[42]

The importance of alertness and performance is nowhere more relevant than it is in occupational safety. The next section describes the interconnectedness of alertness and performance with fatigue and the evolving public policies that aim to mitigate fatigue risk in safety-sensitive industries.

SLEEP MEDICINE AND PUBLIC POLICY

Fatigue related to work demands has long been recognized as a factor contributing to accidents, especially in industries that operate 24/7. With globalization of the economy in the past century, the number of industries and workers operating around the clock has increased. Commensurately, fatigue-related accidents have become more common and have created a greater threat to the environment and public safety. For example, the Three Mile Island nuclear reactor disaster of 1979 resulted from coalescence of human error and mechanical factors that allowed a large amount of nuclear reactor coolant to escape into the environment. Fatigue was implicated in the accident, not surprisingly because the accident occurred at 4 AM. Another fatigue-related accident occurred in 1989 when the oil tanker Exxon Valdez struck a reef off the coast of Alaska and spilled 11 to 32 million gallons of crude oil. It was at the time the largest and most devastating human-caused environmental disaster. Although multiple factors played a role in the accident, crew fatigue was identified as a major factor. Investigators found the crew to be understaffed and overworked. Fatigue has also been implicated in the Chernobyl nuclear and Challenger space shuttle disasters. Other examples of preventable fatigue-related accidents abound in the transportation,

health care, and emergency response industries. In response to these accidents, a variety of governmental regulations have been established in the interest of promoting public safety. Brief histories of these regulations in the transportation and health care industries are discussed next.

Transportation Industry

Railroad

At the end of nineteenth century after a period of dramatic expansion of railroads in America, it became apparent that train accidents, including those related to worker fatigue, were resulting in unacceptable loss of life and economic damages. Railroad workers were permitted and in many cases required to work extremely long hours, especially during harvest time. There was a public outcry for more regulatory oversight, which led to the first public policy attempting to address fatigue-related accidents, the "Hours of Service Act of 1907" (45 USC Sect. 61; 1907). The law as originally adopted prevented workers from working more than 16 consecutive hours in a 24-hour period. The act also established a minimum of 10 consecutive hours of rest after a 16-hour shift, and a minimum of 8 hours of rest after an aggregate of 16 hours of work in a 24-hour period. Although it is difficult to assess the direct impact of the law owing to poor reporting standards, the available data indicate reductions in injuries and fatalities during the 10 years following the enactment of the law despite a rise in passenger and freight traffic.[43] The law was subsequently changed to allow only 14 hours in 1969, and 12 hours in 1971. In response to several fatal rail accidents in 2002 and 2008, Congress passed the Rail Safety Improvement Act of 2008, which enabled the Federal Railroad Administration (FRA), a member of the U.S. Department of Transportation (USDOT), to promulgate new safety regulations governing different areas related to railroad safety, including hours of service requirements. This law provides statutory limits on the total on-duty and "limbo" time (time spent traveling to duty assignment) for rail and signal employees to 276 hours per month; limits total allowable shift time for employees to 12 consecutive hours; increases uninterrupted off-duty hours from 8 to 10 hours in a 24-hour period; requires 2 consecutive days off after 6 consecutive days worked and 3 consecutive days off after 7 consecutive days worked; and reduces allowable limbo time to 30 hours per month.[44] It is important to note that of all transport modes regulated by USDOT, railroad hours-of-service standards are the only ones locked into statute rather than being adjustable by administrative regulations.

Aviation

Pilot fatigue has long been recognized as a factor contributing to pilot error that could exacerbate the dangers inherent to flying an aircraft. In 1931, the U.S. Commerce Department set a monthly flight-time limit of 110 hours as a compromise between the 140 hours wanted by airline operators and the 85 hours advocated by the Airline Pilot Association. In 1938, the Civil Aeronautic Board issued domestic flight-time rules, limiting flight time to 8 hours in a 24-hour period.[45] These flight-time and duty hour regulations have evolved over the years, driven by public safety concerns, widely publicized airplane crashes, and evolving understanding of human fatigue in operational environment. The Federal Aviation Administration (FAA), the U.S. regulatory body overseeing aviation safety, completed a major overhaul of regulations in 2011. Key components of the 2011 rules include varying flight and duty requirements based on what time the pilot's day begins. Flight duty period limits range from 9 to 14 hours for a pilot depending on when the pilot's day begins and the number of flight segments he or she is expected to fly. Flight time is limited to 8 to 9 hours depending on the start time of pilot's entire flight duty period. Pilots are required to have a 10-hour minimum rest period that includes 8 hours of uninterrupted sleep opportunity before the flight duty period, an increase of 2 hours of rest over the previous rules. To address cumulative fatigue, the 2011 rule includes weekly, monthly, and annual limits of flight and duty hours. For example, pilots are required to have 30 consecutive hours off every week. Pilots and airlines are expected to take joint responsibility when considering a pilot's fitness for duty, including fatigue resulting from preduty activities such as commuting. Pilots are required to affirmatively state their fitness for duty, and the airlines are required to remove a pilot from duty when he or she is fatigued or unfit to fly.[46] Airlines are also required to implement a comprehensive fatigue risk management plan. The 2011 rules only apply to passenger pilots and exclude pilots who fly only cargo, even though they fly the same types of aircrafts on the same routes and are susceptible to the same levels of fatigue.

Trucking

As in other transportation modalities, concerns about unsafe commercial driver scheduling practices led to governmental regulations in the United States in the mid-1930s. The first scientific study addressing driver fatigue was conducted in 1938 by the U.S. Public Health Service and supported the need for regulatory limits on hours of service (HOS) to ensure highway safety.

In 1989, the USDOT sponsored a field study, the Driver Fatigue and Alertness Study, to determine the relationships among HOS regulations, driver fatigue, and frequency of serious accidents involving commercial motor vehicles. The study was completed in 1996 and found that time of day was the strongest and most consistent factor influencing driver fatigue and alertness. Drowsiness as determined by video recording of drivers' faces was markedly greater during night driving than during daytime driving. Time of day was a much better predictor of decreased driving performance than hours of driving (*time on task*) or the cumulative number of trips made. The study also found that drivers spent an average of 5.2 hours in bed, which is about 2 hours less time than their reported "ideal" daily amount of sleep. Drivers with night start time spent the least time in bed, about 4.4 hours. Not surprisingly, there was a negative correlation between the length of the principal sleep period and amount of drowsiness during the next driving trip; that is, more sleep was associated with less drowsiness. There was a tendency for drivers to rate themselves as more alert than the performance tests indicated.[47]

Analysis of crash data has revealed that crash risk is statistically similar for the first 6 hours of driving and then increases nonlinearly after 6 hours. The 11th hour has a crash risk more than three times greater than the first hour. Multiday driving schedules are also associated with statistically significant increases in crash risk, comparable in magnitude to driving time.[48]

The previously referenced study and other similar studies have led to modern hours of service regulations that at least

attempt to consider for human circadian physiology. The 2011 Federal Motor Carrier Safety Administration (FMCA) regulations, for example, incorporate a rest period of at least 34 consecutive hours that include two periods between 1 AM and 5 AM, the window of circadian trough. This provision gives drivers who routinely work nights and put in very long workweeks an opportunity to overcome the chronic fatigue that can build up when working nights. In addition, the regulations reduced the maximal workweek from 82 to 70 hours. Previous HOS rules that include an 11-hour driving limit with a mandatory 30-minute break in a 14-hour driving window every 24 hours were maintained.

Marine

Fatigue is a major concern in the marine transport industry, and its operational challenges are unlike those found in other modes of transportation. Crewmembers spend extended time away from home, often working as many as 3 to 6 months at a time in the open seas. They often work under harsh conditions on moving vessels that are subject to unpredictable weather. The absence of clear separation between work and recreation contributes to increased levels of stress, especially when working with shipmates from different countries. Cramped quarters, environmental noise, vibration, heat, and bad weather compromise sleep quality.[49] Twenty-four-hour operations constrain sleep duration and require some crew members to sleep during the adverse phase of their circadian rhythm. These factors contribute to levels of fatigue that are much greater than those seen in other industries.[50] A study relating fatigue to marine casualties found that 33% of personnel injuries and 16% of critical vessel casualties had crew fatigue as a causal or contributing factor.[51]

The first set of international standards for minimum competence and safety for seafarers was drafted in 1978 by the International Marine Organization, an agency of the United Nations. The resulting Standards of Training, Certification, and Watchkeeping for Seafarers (STCW) included a weekly rest hour minimum that was subsequently increased to the current 77 hours of rest per week. In the United States, the Coast Guard regulates inland ("brown water") and coastal ("blue water") waterway operations and sets minimal hours of rest regulations based on STCW. Currently, a minimum of 10 hours per 24-hour period and 77 hours per 7-day period of rest are required. Rest hours can be divided into no more than two periods in any 24-hour period, and one of the periods must be at least 6 hours in length.[52] The inland waterway industry has adopted a square watch consisting of a 6-hour watch period alternating with 6-hour rest period such that each crew member observes two watch and two rest periods per 24 hours. Although this arrangement limits shift length and thus mitigates time-on-task related fatigue risk, it necessitates some crewmembers to work and sleep in the adverse circadian phase. For example, crews working the second watch typically work from midnight to 6 AM and then again from noon to 6 PM. Both of these work periods include circadian temperature minimums that can be associated with increased sleepiness, especially in sedentary work environments such as the wheelhouse. The U.S. Coast Guard has developed the Crews Endurance Management Systems (CEMS), a set of tools and practices to help maritime operators manage their productivity and safety levels in their work environment.[53] CEMS outlines a number of strategies to mitigate fatigue risk, including the use of artificial light to shift the circadian temperature trough into off-duty period. Sleep quarter improvements (noise and light abatement), sleep hygiene, exercise, nutrition, and stress management are also emphasized in the CEMS manual. Although the strategies outlined in CEMS are laudable because they address fatigue risks in a more comprehensive manner than duty and rest hour regulations, the true effect of such efforts on sleep health and operational safety remains to be determined.

Other fatigue risk management systems employ mathematical models that attempt to predict the risk for fatigue-related accidents based on factors known to influence alertness and performance[54] (see also Chapter 73). These factors include homeostatic sleep drive, circadian variations in alertness, wake inertia, task type, and time on task. These models have the potential to identify at-risk individuals so that corrective measures can be undertaken. Although these novel approaches improve current duty hour limitations, caution must be exercised when implementing them in the operational environment. For fatigue models to be applicable, the fatigue scores derived from such modeling must have sufficient positive and negative predictive values for a given operational setting. In other words, the fatigue score must be accurate enough to help management decide which individual should or should not work. Low accuracy models that falsely indicate high fatigue have the potential to disrupt operations and infringe on labor's income and are unlikely to be successfully implemented. Furthermore, many of the available models do not yet incorporate individual differences in susceptibility to fatigue.

Health Care

The care of the sick necessitates around-the-clock operations whereby continuity of care is paramount. Work shifts tend to be long and thus increase the risk for fatigue-related accidents. This problem is in no place more acute than in the training of physicians and surgeons. In the early year of modern medicine, physician education included brief periods of intense training during which "resident" physicians cared for patients 24 hours a day, 7 days a week. By the latter half of twentieth century, "residencies" became multiyear experiences that incorporated new learning modalities.[55] Duty hours remained long and were thought to be critical to physician training. By the 1970s, congruent with accumulating evidence on the effects of sleep deprivation, it became apparent that postcall residents made more errors.[56] As early as the 1980s, some internal medicine and pediatric training programs attempted to balance service and educational needs with personal needs of the residents. The 1984 medication error by residents working 36 hours under insufficient supervision caused the death of Libby Zion in New York and sparked a national debate that continues to this day. It took nearly three decades to arrive at current resident duty hours as promulgated by the Accreditation Council for Graduate Medical Education. The 2011 standards eliminate overnight call responsibilities for first-year residents by limiting the maximal duty period to 16 hours per 24-hour period and 80 hours per 7-day period. One off-duty day per 7 days is also stipulated in the new standards.

A number of studies have looked at error rates before and after implementation of changes in resident duty hours. A 2004 study compared the traditional resident work schedules (>24-hour shifts and long workweeks) with newer schedules

(shorter, <24-hour shifts and fewer weekly work hours). Interns made 35.9% more serious medical errors during the traditional schedule than during the modified schedule (136.0 vs. 100.1 per 1000 patient-days; $P < .001$), including 56.6% more non-intercepted serious errors ($P < .001$). Interns also made 5.6 times more serious diagnostic errors during the traditional schedule as during the intervention schedule (18.6 vs. 3.3 per 1000 patient-days; $P < .001$).[57]

Resident duty hour limits are not without unintended adverse effects. Some studies have argued that the increased number of handoffs necessitated by duty hour limits disrupts continuity of care and leads to delayed diagnoses, lengthier hospital stays, and increased number of preventable complications.[58] Training programs remain concerned that trainees may not gain adequate exposure to case mix and leave the already protracted years of postgraduate training with insufficient experience. Ultimately, new training models will have to address these concerns while balancing the educational, health, and safety needs of residents.

Hazardous Workplace

In addition to the industries discussed earlier, there are a large number of occupations that operate around the clock and have the potential to affect public safety and health. Workers in these occupations include firefighters, paramedics, police, disaster recovery workers, military personnel, utilities operators, and operators of plants producing hazardous or toxic chemicals. Examples of fatigue-related accidents leading to loss of life, environmental disaster, and economic damages are abundant. The 1986 Space Shuttle Challenger disaster occurred partly because of a sequence of ill-fated decisions made by sleep-deprived individuals. The legacy of the environmental damage caused by the grounding of Exxon Valdez in 1989 by overworked and understaffed crew can still be felt today in Prince William Sound, Alaska.

ROLE OF SLEEP MEDICINE COMMUNITY

Progress in recognizing fatigue as an important factor for health and safety is evident in a number of industries, as described previously. However, much work remains to be done. Sleep clinicians, scientists, and advocates play an important role in socializing the evolving fund of knowledge in sleep medicine and science. In addition to raising awareness and educating the public about sleep's role in health and safety, sleep professionals need to play an active role in developing and promoting industry-specific regulations and best practices that mitigate fatigue-related accidents. These fatigue countermeasures have to be practical and must address operational and economic needs of the stakeholders while promoting health and safety.

Before advocating a policy that has the potential for disrupting an industry, sleep professionals must gain reasonable understanding of the operational constraints of a given industry. For example, the inland waterway industry has long adopted a split sleep schedule whereby work and rest periods alternate every 6 hours in what has become known as *square shift*. On the front watch, a pilot operates the boat from 6 AM to noon and then again from 6 PM to midnight. On the back watch, another pilot operates the boat from noon to 6 PM and again from midnight to 6 AM. This schedule allows two pilots to operate a towboat around the clock with neither of them enduring fatigue-inducing 12-hour shifts. The downside to this schedule is that it limits sleep opportunities to short intervals and requires some crewmembers to sleep during the adverse phase of the circadian rhythms. Based on what is currently known about human sleep physiology, this type of work and rest schedule is expected to cause chronic sleep deprivation, performance deficits, and greater risk for accidents. A policy that argues for a longer rest periods without understanding the economic forces that shape the current square-shift practice will bring strong resistance from the industry stakeholders, especially when unequivocal evidence linking the current work-rest schedule with operational safety risk is lacking. Profit margins in the transportation industry are often thin, and employing an additional pilot to expand the rest periods can render the barge industry unable to compete with other modes of transportation.

Sleep professionals are more likely to be successful in bringing about meaningful changes within an industry when they partner with industry advocates. American Waterways Organization and industry leaders have supported fatigue research to clarify the effect of square shift on crew members' sleep quality and duration.[59] Preliminary findings indicate that both front watch and back wheelhouse crew report similar time in bed and sleep duration. One of the sleep periods is longer and is described as anchor sleep, whereas the second shorter sleep period is described as a nap. When the sleep periods are combined, the total sleep duration is same on the boat as at home. Pilots slept on average 6.6 hours even though they were in bed for about 8.1 hours (sleep efficiency, 81.4%). Sleep quality on the other hand is worse for back watch crew, possibly because their sleep periods (6 AM to noon and 6 PM to midnight) occur during adverse circadian phases. Policies that aim to improve sleep efficiency and sleep quality are more likely to be accepted by waterway operators and ultimately reduce fatigue risk in the industry than policies that mandate changes to long observed scheduling practices.

Sleep advocates can also help shape policies that affect public health and workplace safety by engaging governmental regulatory agencies. Familiarity with these agencies and the industries they administer is required to effectively navigate a complex regulatory ecosystem that has various stakeholders vying to promote their own interests. In the United States, the National Transport Safety Board (NTSB) is responsible for investigating major accidents in all modes of transportation and making safety recommendations; however, it lacks the authority to enforce them. The USDOT and its member agencies (i.e., the FAA, FRA, FMCA, and Pipeline and Hazardous Materials Safety Administration) are responsible for creating and enforcing regulations that govern their respective industries.

The regulatory agencies must consider economic impact, cost-benefit analysis, and industry input when formulating new regulations or policy changes. They are often subject to political pressures. To give an example, in late 2013, the FAA decided to modify rules governing untreated OSA, a disqualifying condition for airmen and traffic controllers when untreated. The proposed modification further mandated that all airmen with a body mass index greater than 40 be evaluated by a board-certified sleep specialist and treated for OSA in order to be medically certified. The policy modification also indicated that the body mass index threshold will be gradually lowered until every airman with OSA is treated."[60] Reaction

from industry was swift. In a letter to the FAA, the Aircraft Owners and Pilots Association argued that the "policy inappropriately bypasses the rulemaking process; overlooks potentially more effective and efficient solutions; provides no clear safety benefit; and imposes unjustified costs on the user community.… In 2011 the FAA identified 124,973 airmen who are considered obese, making them potential candidates for testing under an expanded policy … the potential cost to pilots is between $99 million and $374 million for testing alone."[61] As a result of effective lobbying efforts, Congress took action within a matter of days of learning of the FAA's intent and released legislation that would require the FAA to go through the rulemaking process before making policy changes on sleep apnea. The rulemaking process is a public forum through which industry can influence policymaking. The revised guidelines related to OSA require "aviation medical examiners to consider all of the OSA risk factors and make a recommendation regarding an OSA evaluation." Airmen are allowed to see any physician whether or not they are qualified to diagnose and treat OSA. "Evaluations do not require a laboratory sleep study or even a home study if the certifying physician does not feel the pilot requires it."[62] This modification clearly dilutes the intent of the initial policy modification. Without formal sleep testing and clearly defined screening criteria, at-risk pilots may go undiagnosed and untreated by inadequately qualified physicians. Furthermore, financial burden and regulatory scrutiny combine to disincentivize patients and nonsleep specialists to underestimate symptoms, signs, and the full impact of sleep disorders. This example is typical of how a well-intended policy that promotes sleep health and safety becomes less effective because of socioeconomic factors. It is difficult to create policies that may be perceived as an undue burden on individuals when it affects their right to earn their livelihood, especially in an industry like aviation, for which the safety record is strong.

Sleep advocates can help empower industry stakeholders to implement sleep science–based fatigue risk management programs. Even profit-motivated corporations will embrace health and safety promoting measures when those measures make economic sense. Berger and colleagues, with the support of Schneider National Incorporated (SNI), created a sleep apnea diagnosis and treatment program that eliminated or lowered many of the barriers to treating sleep apnea.[63] Education about sleep apnea was provided through multiple channels to increase awareness and reduce driver anxiety and misconceptions about the condition. Occupational health professionals, safety officers, and training engineers were trained to monitor for symptoms of sleep apnea. Drivers referred to the program incurred no out-of-pocket expenses, thus eliminating one of the major financial obstacles. Evaluations were performed at locations near major SNI operating centers and allowed for a 2-day turnaround from diagnosis to treatment of apnea, thus minimizing scheduling constraints and time away from work. SNI also explicitly identified what documentation is required of their contracted USDOT physicians to ensure uniform high-standard reporting. Among 348 drivers diagnosed with sleep-disordered breathing and who were treated, medical costs and accident rates declined by 57.8% and 73%, respectively. The driver retention rate of continuous positive airway pressure (CPAP)-treated individuals was 2.29 times greater than the total company driver population.[63]

Success of the SNI sleep apnea program has encouraged other transportation companies to pursue similar programs. To realize the full health and safety benefits, however, programs need to go beyond screening and treating sleep apnea in order to ensure long-term patient engagement and CPAP adherence. A major challenge to the development of such comprehensive sleep apnea monitoring programs is providing care to a geographically distributed workforce. A small study involving towboat wheelhouse crew members showed that centralized care coordination, remote adherence monitoring, and telemedicine tools can provide sleep apnea care that is acceptable to patients and achieves high CPAP adherence rates.[64]

The most important aspect of any fatigue management program is the presence of a corporate safety culture that puts the safety of the employee and the public ahead of corporate profits. Sleep professionals can play an important role in promoting such culture by engaging and educating industry leaders, governmental agencies, and the public at large. They can draw inspiration from past successful public health campaigns that brought about meaningful behavioral changes and societal benefits. Efforts of Ignaz Semmelweiss and those that followed helped establish hand washing practices that have saved countless number of lives and much suffering over the past 150 years. The many lives saved by reductions in cigarette smoking and widespread use of seatbelts would not have occurred were it not for the efforts of pulmonologists and emergency medicine physicians. Similarly, sleep medicine has an immense potential to improve public health and safety, especially because sleepiness and sleep disorders have far-reaching influences on health and safety.

CLINICAL PEARL

Sleep medicine has the potential to have an immense positive influence on the public because of the many ways in which sleep and sleep loss affect public health and safety. The promise of sleep medicine will only be realized when there is greater awareness of the importance of sleep. Sleep professionals are in a unique position to increase this awareness through education and advocacy. Sleep professionals have many opportunities to help transform sleep science knowledge to practical solutions that improve public health and safety through partnership with industry and governmental agencies.

SUMMARY

Sleep is an essential biologic function whose physiologic importance has become clear only in the past several decades. Despite this growing knowledge and awareness of sleep's role in health, insufficient sleep is all too common in the modern world. The American Institute of Medicine estimates that 50 to 70 million people suffer from chronic sleep loss. Many factors are driving this worrisome trend toward short sleep. Economic forces that drive long work hours, long commute times, and ever-expanding 24/7 operations commingle with modern lifestyles that overconsume digital entertainment and social media to create a culture that champions "work hard and play hard" and leaves little room for sleep. The availability of inexpensive artificial light has had a dramatic effect on the

natural 24-hour sleep-wake patterns in humans. Light exposure interferes with circadian rhythms by delaying the normally occurring peak alertness and allows people to extend wake activities well into the night at the expense of sleep. Early start times common to many occupations necessitate early arousal, and hence the sleep duration is squeezed at both ends of the sleep period. It is estimated that 30% of all employed U.S. adults (40.6 million workers) average less than 6 hours sleep per night. The sleep deficiency is not unique to employed adults. Sleep trends in school-aged children and adolescents indicate a decrease in sleep time of more than 1 hour over the past 100 years. Insufficient sleep has been associated with a number of chronic medical conditions, such as hypertension, diabetes, and cardiovascular disease, as well as all-cause mortality. When sleep disorders are weighed in, the resulting sleep loss is probably the single most important public health issue of our time. This chapter describes current trends in sleep practices across the age spectrum; their potential effect on health, performance, and public safety; and the role sleep professionals can play in promoting sleep health and public safety.

Selected Readings

Hale L, Kryger MH, et al. *Sleep Health* 2015;1(1):1–75.

Howard S, Gaba DM, Rosekind MR, Zarcone VP. The risk and implications of excessive daytime sleepiness in resident physicians. *Acad Med* 2002; **77**:1019–25.

Institute of Medicine. *Sleep disorders and sleep deprivation: an unmet public health problem*. Washington, DC: The National Academies Press; 2006. <http://www.iom.edu/reports/2006/sleep-disorders-and-sleep-deprivation-an-unmet-public-health-problem>; [Accessed April 28, 2015].

Institute of Medicine. *Resident duty hours enhancing sleep, supervision, and safety*. Washington, DC: The National Academies Press; 2009. <http://www.nap.edu/catalog.php?record_id=12508>; [Accessed April 28, 2015].

McNicholas WT, Krieger J, Levy P, et al. Public health and medicolegal implications of sleep apnoea. *Eur Respir J* 2002;**20**:1594–609.

Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep* 2004;**27**(7):1255–73.

Strohl KP, Brown DB, Collop N, et al. An official American Thoracic Society clinical practice guideline: sleep apnea, sleepiness, and driving risk in noncommercial drivers. *Am J Respir Crit Care Med* 2013;**187**(11): 1259–66.

A complete reference list can be found online at ExpertConsult.com.

Chapter
64

Sleep and Athletic Performance

Scott J. Kutscher

Chapter Highlights

- Sleep is increasingly being recognized as part of the foundation of optimal athletic performance. Sleep influences on athletic performance include circadian and homeostatic factors.
- Sleep disorders may affect the performance and health of athletes. Although specialized populations of athletes may be at increased risk for certain sleep disorders, exercise itself may be therapeutic for a variety of sleep disorders.
- Rest is a vital component of recovery from training and competition, whereas improper rest may hasten fatigue and predispose to injury.
- Properly managing sleep and identifying problem sleepers and potential sleep disorders requires cooperation of a multidisciplinary team of athletes, coaches, trainers, and physicians. Scheduling of multiple factors, including travel and practice, need to be addressed for optimizing sleep in the athlete.

SLEEP AND ATHLETIC PERFORMANCE

Sleep is increasingly regarded as a major contributor in variation of athletic performance, with emerging evidence suggesting a strong and often intertwined influence of both circadian and homeostatic factors, outlined in Figure 64-1.

Circadian Variations in Performance

Twenty-four-hour variation has been demonstrated in a wide array of athletic performance measures. Peak performance correlates well to variations in core body temperature, with highest results occurring close to the maximal body temperature (Tmax), and poorer performance near the body temperature trough (Tmin). Physiologic studies have confirmed a time of day effect, with peak output typically occurring in the evening (16:00 to 18:00) for isometric power output of leg and back muscles.[1-4] Similar evening preference has been substantiated in elbow flexor muscles.[5,6] There is an evening increase in VO_2max during aerobic energy production,[7] as well as during anaerobic exercises such as stair running and broad jump.[8]

Of course, demonstrating clinical significance of these effects is important if one is to consider a therapeutic intervention. For athletic performance, the most obvious place to look is the place athletes and fans turn to in order to measure results: the scoreboard. To this effect, multiple attempts have been made to establish clinical significance of circadian performance variations.

Major League Baseball (MLB) has been closely studied in this regard, perhaps owing to the frequency and set-up of games. With the same teams typically playing a series of consecutive games, each team serves as its own control. In an evaluation of matchups between teams on the East and West coasts, home teams were found to be at an advantage after cross-country eastward flights and during day games,[9] suggesting effect of both travel and time of day. Another study evaluating 10 years of MLB games demonstrated that the team with a circadian advantage (i.e., the team that did not have to travel across one or more time zones) had a higher winning percentage (52%); and at a 3-hour time zone difference, this actually outperformed winning percentage based on home field advantage alone (60.6% vs. 53.7%)[10]—indicating a team can be at a disadvantage playing at their own stadium, if they had to travel across the country to be there.

The same effect seems to appear in professional football, despite the fact that teams only play one game per week, indicating that circadian effect over effects of fatigue and travel are at play. Through 40 years' worth of games between teams on the East and West coasts, a distinct pattern emerges, with even winning percentages when compared against the point spread in games starting before 8 PM (49% winning percentage for West Coast teams vs. East Coast teams). However, after 8 PM, West Coast teams have a distinct advantage, outperforming the point spread in a significant majority (66%) of games.[11] These findings suggest a circadian advantage for these teams, aligned with their peaks in circadian performance, and bring the obligatory disclaimer that gambling on professional sports is illegal in most places, including the location of this author's residence.

Sleep Deprivation

Another aspect to consider is the effect of sleep deprivation on performance. Studies of acute total sleep deprivation have shown mixed results—with little change in muscle strength and endurance noted even after 60 hours of total sleep deprivation.[12] Sleep deprivation of 24 hours did not demonstrate a significant difference in performance of weight lifters, despite an increase in subjective sleepiness and mood symptoms.[13] Other efforts contradict these studies, demonstrating decrease in maximal bench press,[14] distance covered on self-paced treadmill run,[15] and time to exhaustion on stationary cycling test,[16] after periods of acute sleep deprivation. These tests are

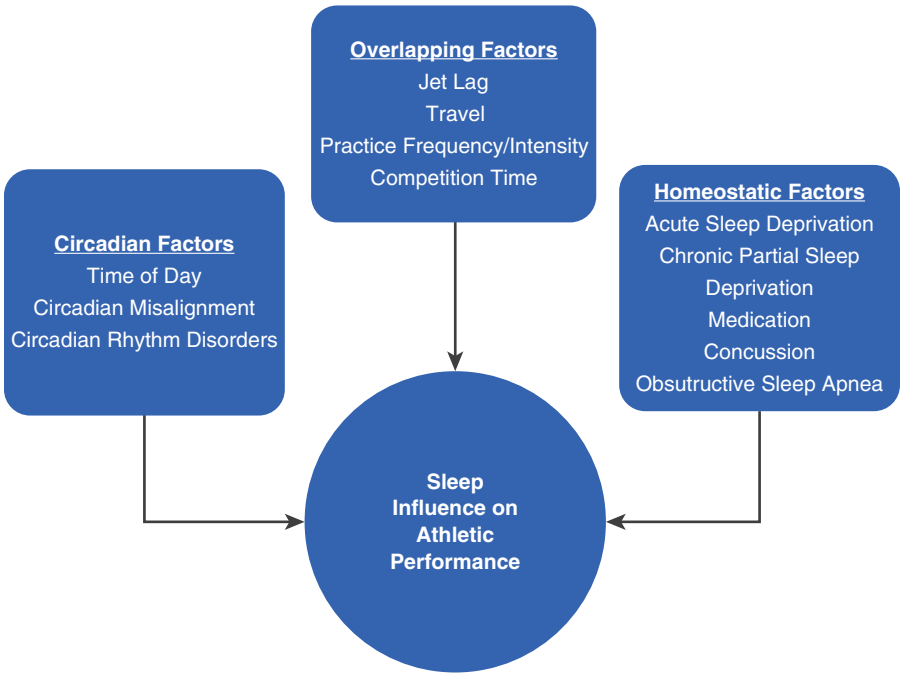


Figure 64-1 The interplay of circadian, homeostatic, and overlapping factors influencing athletic performance.

routinely limited by small sample sizes, and differences in results may simply be attributable to a high rate of individual variability in sensitivity to sleep deprivation.[17]

Of greater clinical concern—assuming athletes are not intentionally sleep depriving themselves for 60 hours—is the role of chronic partial sleep deprivation. Athletes with demanding schedules that include travel, practice, and variable game times may be particularly susceptible to a state of chronic partial sleep deprivation, to say nothing of amateur athletes who may have additional obligations of work or school. This state of insufficient sleep has been confirmed in elite swimmers, who demonstrated significantly less time in bed and time asleep on nights before training days than nights before rest days.[18] As mentioned later in the chapter and also elsewhere in this text, insufficient sleep has its own set of sequelae on health and daytime functioning (see Chapter 5).

Studies that evaluate strength and stamina in a controlled setting are, by definition, removing some of the aspects of athletics, such as judgment, reaction time, and sustained vigilance, that may be particularly susceptible to the effects of chronic sleep deprivation. Batters in MLB, for whom attention and reaction time are vital to success, have been shown to have steadily declining plate discipline during the course of a season. This is despite of potential improvement because of practice and repetition, and this erosion of the ability to tell balls from strikes may indicate a fatigue effect.[19] A study of men's professional basketball further supports this idea, demonstrating significant improvement in team performance with increasing number of rest days between games, with peak performance occurring with 3 days off between games.[20]

Another way of considering sleep deprivation is to measure the effects of sleep extension. One study of male collegiate basketball players examined performance on metrics such as sprint time and shooting accuracy and found a significant improvement after a 5- to 7-week period of sleep extension to 10 hours, compared with each athlete's own baseline sleep.[21]

Finding time for such sleep extension over so many weeks can be difficult for *any* individual. When taken as a whole, studies to date suggest sleep deprivation may have an insidious effect on performance that can be difficult to overcome without clear planning.

SLEEP DISORDERS

Having established a relationship between performance and sleep, it is also important to remember that athletic performance may be acutely sensitive to sleep disorders. Although some athletes, as summarized in Table 64-1, may be considered at higher risk for certain disorders, exercise itself may be therapeutic for others.

Obstructive Sleep Apnea

This author has had a patient—a former professional wrestler—relate anecdotally that he found larger wrestlers would lose consciousness faster when subjected to a chokehold, that is, external pressure to constrict the airway. Although our Independent Review Board might raise some concerns if we tried to confirm his theory with a controlled trial, it certainly is in line with much of what we know about the risk factors contributing to airway narrowing and upper airway collapse.

With elevated body mass index (BMI) and enlarged neck circumference each contributing to obstructive sleep apnea (OSA), it stands to reason that athletes for whom largeness is a performance advantage may be at higher risk for developing OSA. This cohort would include athletes such as linemen in football, weight lifters, sumo wrestlers, and, yes, professional wrestlers.

Unfortunately, population studies in these specialized groups are largely lacking. One study of active professional football players found OSA (apnea-hypopnea index [AHI] ≥10) in 14 of 52 subjects tested by polysomnography. Most

Table 64-1 Suggested Risks of Sleep Disorders Specific to Athlete Populations

Insomnia and Poor Sleep Quality
- Elevated scores on the Pittsburgh Sleep Quality Index
- Excessive worry preceding competition
- Injuries, pain, medications
- Acute concussion, postconcussive syndrome

Hypersomnia and Excessive Daytime Sleepiness
- Insufficient sleep due to schedule demands
- Injuries, pain, medications
- Acute concussion, postconcussive syndrome

Obstructive Sleep Apnea
- Elevated body mass index ≥40 in specialized sport populations
  - Sustained or increased concern in retired athletes
- Concussion

Restless Legs Syndrome
- Training or competition in long distance running
  - Should be carefully differentiated from cramping or other mimics of restless legs syndrome

Circadian Rhythm Disorders
- Advanced sleep phase suggested by early morning preference on Morningness-Eveningness Questionnaire
- Shift work characterized by early, late, or variable competition start times

of these positive cases were offensive and defensive linemen, with elevated BMI higher than 40 being the most significant driver of risk.[22] A larger study of 137 active National Football League (NFL) players using only unattended nasal pressure transducer found that 19% had a respiratory disturbance index of 5 or higher, but only 4.4% had a respiratory disturbance index of 15 or higher.[23]

The seeming driver for the focus on sleep apnea in football players is a 1994 study by the Centers for Disease Control and Prevention that found retired NFL linemen had a 52% higher rate of cardiovascular mortality than the general population and were three times more likely than other position players to die of heart disease.[24] Indeed, the only study of OSA in retired NFL players confirmed this discrepancy, with linemen being more likely to have an AHI of 10 or greater (61% vs. 46%) and obesity (83% vs. 52%) than other position players.[25]

Besides the obvious long-term health risks, there remains the question of whether OSA has an influence on performance. One study attempting to answer this question found that amateur golfers with a diagnosis of OSA (AHI ≥15) could improve their handicaps after continuous positive airway pressure therapy,[26] a finding that may do more to motivate excellent continuous positive airway pressure compliance than any other single intervention. There is also the question of whether exercise itself has an effect on OSA. To this end, a randomized control trial of sedentary adults found that a 12-week exercise regimen significantly lowered AHI compared with stretching, without a similar significant change in BMI to otherwise explain this reduction.[27]

Despite the potential for higher incidences of OSA and cardiovascular disease among certain populations of active and retired athletes, as well as theoretical limitations to performance, outreach and awareness remain barriers to care. Although attempts have been made to provide educational resources for the NFL,[28] no such global outreach has been attempted for other sports, either amateur or professional. Overall diagnosis and treatment rates remain unknown, a fact that larger epidemiologic studies could ameliorate.

Circadian Rhythm Disorders

Given the circadian variation of athletic performance, it is possible that one can be at an advantage or disadvantage depending on how one's own circadian rhythms align with practice and competition time. This is especially true in the major professional leagues, where the financial incentives of television often dictate game times. Given the potential for high variability of game times in some sports, one may reasonably classify some athletes as shift workers.

Few studies address individual circadian variability, which remains largely theoretical. Some effect has been demonstrated in MLB players. Those who scored as advanced on a Morningness-Eveningness Questionnaire (MEQ) had higher batting averages during day games before 2 PM (0.267 vs. 0.259), but the higher average shifted to players with a delayed-type circadian rhythm during evening games (0.261 vs. 0.252).[29] Somewhat counterintuitively, one survey of elite adolescent athletes revealed a preponderance of morning-type circadian rhythms,[30] although it is unclear whether these athletes are truly phase advanced or whether they simply grow accustomed to the early wake-up and practice times often required of elite athletes and adjust their answers accordingly.

Although jet lag is not always directly mentioned, it should be considered a potential confounder of performance in any athlete who has to travel across multiple time zones to compete. Training performance can be affected by eastward or westward travel, and effects can be seen after traveling across as few as two time zones.[31] Performance decrement may be worse after westward travel; however, it has been found to recover to baseline before full reentrainment after flights crossing up to 8 hours.[32]

It is unclear whether studies of travel on, for example, Australian netball players, are generalizable to a population other than Australian netball players. There may be little to no effect in performance demonstrated, even after 4 to 6 hours of transmeridian travel.[33,34] Although circadian studies of performance often site travel, true circadian effect directly attributable to jet lag can be difficult to isolate from confounders like perception of jet lag symptoms, decrements in cortisol level, distance traveled, sleep deprivation, and mood.[35]

On the other hand, with the chronobiotic effect of exercise, typically manifesting as a phase delay when administered in the early evening,[36] it may be considered therapeutic for jet lag. Although these studies, in general, are limited by numerous confounding variables, exercise has been shown to accelerate reentrainment to an advanced circadian rhythm after forced desynchronization under isolation and constant dim light.[37]

Restless Legs Syndrome

Despite the fact that restless legs syndrome (RLS) is a disorder of movement, there is precious little knowledge on its relationship to athletic performance. The suspected pathway

of dopamine dysfunction may implicate motor systems under control of the basal ganglia, such as in Parkinson disease. However, RLS subjects did no worse than controls on multiple tests of motor function.[38]

The fact that RLS is improved by rest suggests a possible role for exercise in therapy. A positive effect was confirmed by a small randomized trial of RLS patients, with a significant decline in RLS severity scale after 12 weeks of aerobic exercise,[39] as well as by one randomized trial in hemodialysis patients with uremia showing improvement in International Restless Legs Syndrome Rating Study Group severity scores after 6 months of graduated resistance training.[40] However, another small study surveying marathon runners actually showed a higher prevalence of RLS than in the general population.[41]

Insomnia and Poor Sleep Quality

It should not be surprising that athletes would have a high prevalence of insomnia. If using the Spielman model of insomnia, it is rather easy to identify predisposing (e.g., hyperawareness), precipitating (e.g., competition time), and perpetuating (e.g., irregular training times, travel, medications) factors. One survey found that 64% of elite athletes questioned admitted to worse sleep on the nights before competition in the preceding 12 months. Most (82%) of these complaints were of sleep-onset insomnia, with preoccupying thoughts (83.5%) and nervousness (43.8%) the most common reasons cited.[42] Incidence of chronic insomnia is unknown, although multiple factors could contribute, including psychosocial factors, travel and jet lag, variable sleep times, and medications.

The mechanism underlying the relationship between exercise and sleep quality is unclear, but there have been well-established positive effects in a number of populations. Exercise has been shown to improve sleep latency and sleep efficiency in patients with chronic primary insomnia.[43,44] Adolescents with increased physical activity report fewer symptoms of insomnia and better sleep quality, along with objectively measured improvements in total sleep time and wake after sleep onset.[45] Subjective sleep latency and total sleep times were improved with moderate-intensity exercise in sedentary older adults.[46] In middle-aged women, greater recent exercise was associated with improved sleep efficiency and sleep quality.[47] Exercise is a proposed modality to improve sleep quality and quality of life in people with diabetes,[48] heart failure,[49] breast cancer,[50] pregnancy,[51] and menopause,[52] among others.

Results of the 2013 National Sleep Foundation Sleep in America Poll—focused on sleep and exercise, surveying 1000 U.S. adults aged 23 to 60 years—support findings in the scientific literature. Despite similar sleep habits, respondents who exercise were more likely to report a good night's sleep (83%, 77%, 76%, and 56% for vigorous exercisers, moderate exercisers, light exercisers, and nonexercisers, respectively) and to feel their overall sleep needs were being met. Nonexercisers were also significantly more likely to report being in poor health, although the assumptions from this correlation are clearly bidirectional (perhaps they cannot exercise *because of* their poor health).

Surprisingly, the results of the survey did not demonstrate significant difference in sleep quality based on timing of exercise, with those performing vigorous or moderate exercise within 4 hours of bedtime reporting similar overall sleep quality to those whose exercise was more than 4 hours before sleep.[53] This runs contrary to the long-held belief that exercise close to bedtime can decrease sleep quality and suggests that if future studies are confirmatory of these survey results, a transition away from this recommendation—and focus on encouragement of exercise at any time—may be in order.

INJURIES

Injury prevention and management are as important as any goal in the practice and preparation athletes undergo before competition. Just as sleep is an integral part of athletic performance, it should be considered integral to maintenance of that performance.

Fatigue-Related Injuries

The concept of fatigue can have several definitions in regard to athletics. At its most precise, fatigue is the decline in force over sustained muscle activity. However, more generally and more frequently, fatigue refers to physical and mental exhaustion due to prolonged exertion.[54] Fatigue-related injuries are common, seen in up to 10% of athletes surveyed. Moreover, in one study, these injuries were significantly associated with both subjective perception of inadequate rest before injury, as well as self-reported sleep times of 6 hours or less.[55] In another study, sleep was consistently listed as the single most important modality for recovery in a survey of 890 male and female elite team athletes, regardless of gender, sport, or level of competition.[56]

Despite the subjective support of the role of sleep in fatigue-related injuries, there are no objective data on the mechanism linking sleep to athletic injuries. Nonetheless, the link between sleep and fatigue-related errors or accidents is so well established—both under laboratory conditions such as seen with psychomotor vigilance testing, as well as safety-conscious fields like military and transportation—that the conditions in athletes are largely extrapolated under the assumption of similar effects.

Sport-Related Concussion

Consensus definition of concussion is a biomechanically induced alteration of brain function, with primary pathophysiology involving a functional, rather than anatomic, disruption. This is distinguished from traumatic brain injury (TBI), which may have evidence of both. Indeed, some argue that concussion represents a distinct subset of mild TBI (mTBI), whereas others use the terms interchangeably.[57] This new working definition of concussion in sport should be noted for what it lacks: requirements for loss of consciousness or even direct contact with the head. Epidemiologic studies demonstrate variable rates of concussions dependent on sport, although there is evidence that incidence is increasing across all sports, to a large degree likely related to increasing awareness.[58]

The consensus statement on concussion in sport, updated in 2012, includes sleep among the modifying factors for concussion.[59] Despite this designation, the relationship between sleep and sport-related concussion (SRC) remains poorly understood.

Increasingly, diagnosis of SRC, as well as decisions for return to play, have incorporated neurocognitive screening

tools. These tools are useful for comparison of multiple domains of cognitive performance, as well as subjective symptoms, in the postconcussion state to personal or group-averaged baselines. Sleep may play an important role in interpretation of these tests. High school and collegiate athletes with sleep times of 7 hours or less on the night before testing have lower scores and higher total symptoms on baseline, preconcussion testing.[60] Lower subjective sleep quality is also correlated to higher number of symptomatic complaints on baseline neurocognitive screening.[61] These findings are consistent with effects of sleep deprivation seen with other neurocognitive tests, such as the psychomotor vigilance test, and suggest that sleep time and sleep quality need to be considered when interpreting concussion screening tools.

As symptoms of SRC, sleep disturbances are among the most common complaints and can include symptoms of hyperarousal, hypersomnia, or both, with sleep symptoms present in up to 30% to 70% of cases.[62] These symptoms may persist as part of the postconcussion syndrome.[63] However, fatigue at the time of injury was not considered predictive of a protracted recovery in one study of high school athletes.[64] Another study examining polysomnography of concussed athletes and controls found no difference in sleep parameters, despite an increase in subjective sleep complaints in the SRC group. This same study did find significant difference in waking electroencephalogram patterns, with the SRC group displaying increased delta and decreased alpha activity.[65]

It should be noted that these findings differ from overall mTBI research, which has shown increased sleep latency, shorter rapid eye movement (REM) latencies, increased N1, and decreased REM sleep.[66,67] In addition to frequent symptoms of insomnia, symptoms of excessive daytime sleepiness may be the result of comorbid sleep disorders such as OSA, periodic limb movement disorder, and narcolepsy.[68,69]

This discrepancy may be partly explained by differing mechanisms of injury. However, this field of research in general is limited to date by small sample sizes and long lead times between injury and study. Furthermore, survey questions following concussion are often vague, including only general terms like *fatigue*, which may be interpreted in multiple ways. Given that cognitive rest encompasses the definitive protocol for management of SRC, it would seem at least somewhat probable for sleep to play a part. Of particular interest is the possible role of sleep disorders in chronic traumatic encephalopathy, a neuropathologically distinct degenerative disorder attributable to repetitive head trauma.[70] Given the significant burden of disease, a better understanding of sleep in SRC is vital.

MANAGING SLEEP IN ATHLETICS

Along with exercise and nutrition, sleep should be considered a pillar of any training regimen, with goals to improve performance and aid in recovery and injury prevention. Managing sleep in the athlete can take on many forms. Summarized in Figure 64-2, this should include broad oversight of scheduling, including practice and travel, to optimize sleep time and circadian alignment, as well as understanding of individual variability, including consideration of adequate sleep satisfaction, mood and stress, and screening for sleep disorders. One should always be cognizant that outside constraints—particularly time and financial pressure on travel, training, and competition times—may require secondary solutions (e.g., if

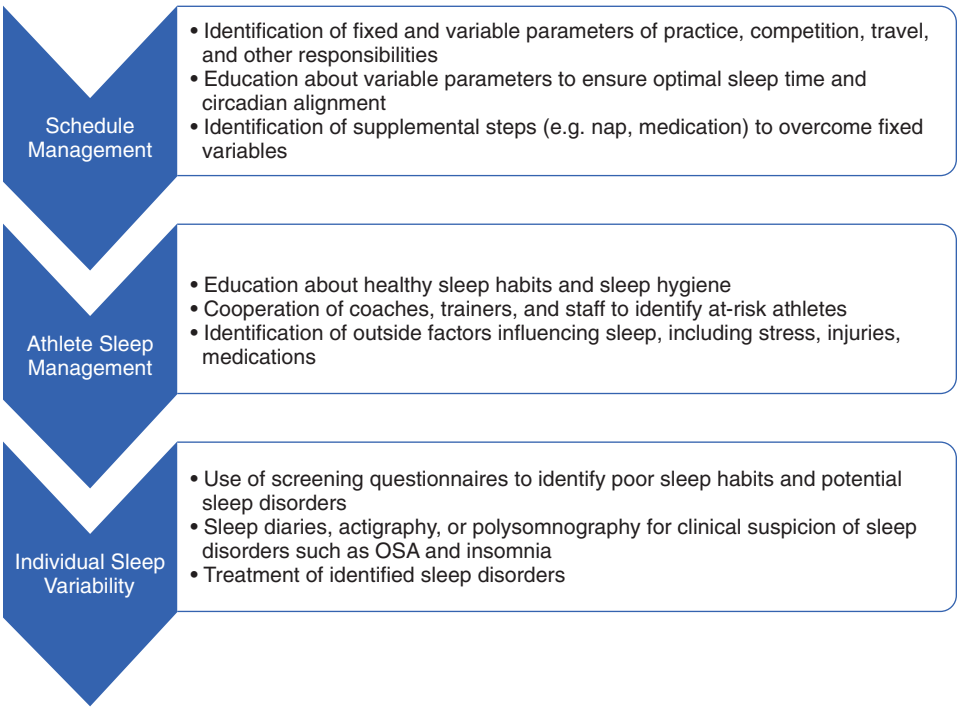


Figure 64-2 Sleep management in athletics includes broad goals generalizable to all athletes as well as highly individualized treatment plans. OSA, Obstructive sleep apnea.

a competition starts at 6 AM, one cannot simply demand a later start time).

Training in Elite Athletes

One concern about sleep management in elite athletes is whether their specific sleep requirements are different than those of the general population, for whom sleep advice is both widely available and well established. Elite Olympic athletes do report poorer sleep quality than controls when measured by Pittsburgh Sleep Quality Index (PSQI),[71] whereas a survey of college athletes showed elevated PSQI and Epworth Sleepiness Scale scores, although these could have been attributed to environmental and social factors.[72] The data are limited, but given the intensity of training often required, it is reasonable to consider the standard to at least be elite performers in other fields such as military, transportation, medicine, and business.

Sport-Specific Considerations

Understanding of sport-specific requirements is vital for adequate sleep opportunity. Each sport represents a unique combination of variables to address for sleep management. Factors to consider include level of competition; seasonal versus year-round sports; intensity of training regimen during off-season, preseason, competition, and postseason times; frequency of competition; frequency of travel and rest days; and time of competition. Athletes who train in the early morning or have late evening competitions may be at increased risk for insufficient sleep and more reliant on supplemental naps. Amateur athletes may have work or school requirements in addition to competition and training, further compromising sleep habits.

Screening for Sleep Disturbance

A required formal sleep evaluation for all athletes would be both costly and impractical. Awareness and education are key; the coaching, training, and ancillary support staffs should be enlisted to identify athletes at risk. Off-season routine physical examinations may be appropriate times to administer sleep screening questionnaires, such as the PSQI or Epworth Sleepiness Scale, the latter of which has been inversely correlated with service time in MLB.[73]

Actigraphy, either alone or in combination with sleep diaries, may be particularly useful in tracking sleep and ensuring adequate time in bed, especially during periods of travel or high-intensity training, when there is high risk for insufficient sleep. Polysomnography is useful for athletes who screen at high risk for OSA. Although less useful as a screening test itself, primarily because of low incidence of findings in this typically healthy population, a treatable sleep disorder should not be missed or ignored.

Managing Sleep Deprivation

Much of the theory behind mitigating sleep deprivation in athletes is derived from our understanding of fatigue, particularly the physical and neurocognitive consequences of chronic partial sleep deprivation, in other specialized patient populations and from laboratory studies of healthy adults.

With the small margins of error inherent in elite athletics, limiting effects of sleep deprivation is vital. A major contributor to sleep deprivation in this population is overscheduling. One should view the goals as similar to legally mandated requirements in medical and transportation fields, with proper time for rest and recovery allotted between practices and competitions. Naps can be useful adjuncts to an anchor sleep period, particularly during times of heavy travel or training.

Making time for sleep is of major importance but is only useful insomuch as an athlete is able to actually sleep during that time. It would be unreasonable to expect an athlete to fall asleep mere minutes after competition or at times at odds with their circadian rhythms. Symptoms of insomnia should be routinely addressed, with behavioral interventions considered first-line therapy.

Managing Circadian Effects

Although it is important to understand time-of-day variability in performance, the consequences remain vague. There is no evidence to date demonstrating that training at one specific time of day (i.e., close to Tmax) actually improves results compared with other times. Certainly, given the predominance and health consequences of a sedentary lifestyle, for many people *any* activity, regardless of relationship to circadian time, should be encouraged, so long as it does not predispose to an insufficient sleep pattern.

For more seasoned athletes, either amateur or professional, training is encouraged at or near time of competition. Although high-output training (i.e., muscle and endurance exercise) should be scheduled to coincide with peak circadian times, it does not preclude less intense activity such as stretching at off times.

Travel and Jet Lag

Fatigue from travel and jet lag should be properly differentiated; the latter consists mostly of a constellation of symptoms resulting from acute circadian disruption, whereas the former can result in cumulative fatigue over time.[74] Travel should be scheduled to allow for continued adequate sleep durations whenever possible. Transmeridian travel across three or more time zones may require circadian adjustment before departure, at the rate of one time zone per day. However, athletes may be better served by simply staying on their home time throughout shorter trips of less than 3 days.[75] Again, complex scheduling may limit the practice of pretravel time zone adjustment.

Medication and Supportive Therapy

Medication effects are largely unstudied in athletes, and principles are derived from general patient populations. Supplemental melatonin may be useful as a sedative or chronobiotic, particularly in the reentrainment period during or after transmeridian travel. Nonbenzodiazepine hypnotics may likewise serve as temporary aides to improve sleep onset and maintenance, with a consistent goal of limited, short-term therapy. Effect of long-term use of these medications in this population is unknown, although anecdotal concern for overuse is present.

Bright-light therapy can be considered for circadian entrainment and may be beneficial both during periods of travel and for athletes who must train at times at odds with their own circadian clocks. Caffeine is a safe supplement whose stimulant properties, mediated by antagonism of andesine receptors, can be useful in addressing fatigue. Prescription stimulants are almost universally banned in elite competition and should not be started without justifiable cause.

CLINICAL PEARL

Peak athletic performance generally occurs in the late afternoon, concomitant with maximal core body temperature. Performance is degraded by chronic partial sleep deprivation, although foundation for this belief is largely based on research in nonathlete populations. Inadequate rest may hinder recovery and exacerbate injury. Sleep disorders should be considered in specific athlete populations. Concussion may also influence sleep, and sleep complaints are common following concussion in athletes. Properly addressing the sleep needs of athletes requires understanding the complexity of variables influencing circadian and homeostatic factors and cooperation of a multidisciplinary team of trainers, coaches, and physicians.

SUMMARY

Athletes, coaches, and trainers are increasingly looking to sleep in an effort to achieve and maintain peak athletic performance. There exists circadian variation in a number of performance variables, correlated to variations in core body temperature, as well as homeostatic drivers of performance, with athletic schedules often predisposing to risk for chronic partial sleep deprivation. Although certain populations of athletes may be at increased risk for sleep disorders such as insomnia and OSA, exercise itself has demonstrable efficacy in treatment of a number of sleep disorders. Poor sleep may inhibit recovery from training and predispose to injury, and sleep symptoms are a common complaint in SRC. Athletes of all levels should consider sleep as a core component of training and recovery, and for some, a comprehensive analysis of sleep habits, training, competition, and travel schedules, as well as medication and supplement use, may be indicated to promote optimal sleep patterns.

Selected Readings

Forbes-Robertson S, Dudley E, Vadgama P, et al. Circadian disruption and remedial interventions: effects and interventions for jet lag for athletic peak performance. *Sports Med* 2012;**42**(3):185–208.

Kline CE, Crowley EP, Ewing GB, et al. The effect of exercise training on obstructive sleep apnea and sleep quality: a randomized controlled trial. *Sleep* 2011;**34**(12):1631–40.

Mah CD, Mah KE, Kezirian EJ, Dement WC. The effects of sleep extension on the athletic performance of collegiate basketball players. *Sleep* 2011; **34**(7):943–50.

McCrory P, Meeuwisse WH, Aubry M, et al. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport, Zurich, November 2012. *J Athl Train* 2013;**48**(4):554–75.

Samuels C, James L, Lawson D, Meeuwisse W. The athlete sleep screening questionnaire: a new tool for assessing and managing sleep in elite athletes. *Br J Sports Med* 2015. doi:10.1136/bjsports-2014-094332; [Epub ahead of print].

Samuels CH. Jet lag and travel fatigue: a comprehensive management plan for sport medicine physicians and high-performance support teams. *Clin J Sport Med* 2012;**22**(3):268–73.

Sufrinko A, Pearce K, Elbin RJ, et al. The effect of preinjury sleep difficulties on neurocognitive impairment and symptoms after sport-related concussion. *Am J Sports Med* 2015;**43**:830–8.

A complete reference list can be found online at ExpertConsult.com.

Section 9

Legal Topics in Sleep Medicine

Chapter 65

Sleep Forensics: Criminal Culpability for Sleep-Related Violence

Michel A. Cramer-Bornemann; Mark W. Mahowald

Chapter Highlights

- Sleep forensics is the application of neuroscience to somnology and sleep medicine to investigate unusual, irrational, or bizarre human sleep-related behavior associated with alleged criminal activity. This investigation is typically used to form the basis of an expert opinion for use in a criminal trial regarding a defendant's state of mind.
- Consciousness is not an all-or-none state but occurs on a spectrum. In addition, sleep and wakefulness are not mutually exclusive states of consciousness. Wakefulness, NREM sleep, and REM sleep may occur simultaneously or oscillate rapidly. This phenomenon is key to understanding the forensic implications of violent parasomnias.
- Consciousness can be dissociated from behavior. Neurophysiologic mechanisms can account for violent or other asocial behaviors associated with sleep.
- In a criminal proceeding, the offer of a clinical diagnosis alone is often insufficient to secure a conviction. Sleep medicine specialists have a role in legal proceedings to describe and address aspects of a defendant's consciousness and culpability.

To blame a person is to express moral criticism, and if the person's action does not deserve criticism, blaming him is a kind of falsehood, and is, to the extent the person is injured by being blamed, unjust to him.

Sanford Kadish, 2000[1]

Philosophers from time immemorial have grappled with the mind–body dilemma. Recent advances in neuroscience put us now on the verge of solving how our brains affect our minds and behavior.

THE DEVELOPMENT OF SLEEP FORENSICS

Sleep forensics is formally defined as the application of the principles and tools of neuroscience as applied to somnology and sleep medicine. These principles and tools have been widely accepted under international peer review as means for investigating unusual, irrational, or bizarre sleep-related behavior associated with alleged criminal activity. The investigation undergoes further examination in a courtroom pursuant to rules of criminal law.

The best application of sleep forensics involves an adaptable conceptual approach. An adaptable approach applies current neuroscientific concepts of consciousness and sleep-wake state dissociation to sleep medicine. This dynamic method is preferred to the approach set forth in the U.S. Model Penal Code (MPC), which uses static definitions and clinical disorder markers from which criminal behavior might be extrapolated.

Therefore a medical expert called on to investigate criminal allegations will need to do more than just evaluate for a possible sleep disorder. Ultimately the expert's determination of

the defendant's state of consciousness will prove pivotal. This requires an understanding of the neuroscience of consciousness, an awareness of relevant neuroscientific models for types of potential behaviors that may arise from sleep, as well as a determination concerning the appropriate application of consensus-driven clinical guidelines to assist in determining purported acts of violence arising from sleep.

The medical expert should also recognize the sleep specialist's primary role when interfacing with lawyers, judges, and law enforcement. The sleep specialist can help facilitate the discourse concerning advances in cognitive neuroscience and help develop the framework for further research, particularly in parasomnias.

EVOLUTION OF LEGAL THOUGHT ON CRIMINAL MENTAL STATES

General

Influenced by Sir Edward Coke, Chief Justice of the King's Bench (1613), Anglo-American law defines criminal offenses whereby a person must be in a certain mental state, called the *mens rea* (guilty mind), necessary to have committed a crime. Persons possessed of *mens rea* cannot be convicted absent a corresponding criminal act, called the *actus reus* (guilty act). Traditionally, intention is found within *mens rea,* and the physical part of the offense resides within *actus reus*. Proof of both is essential to secure a conviction.

Recognition that an impaired mental state might mitigate criminal punishment appears to date back to at least 1772 BC as recorded in the Code of Hammurabi. The Roman Empire also appeared to recognize a person's altered mental status to find defendants not guilty due to *non-compos mentis*, meaning without mastery of mind.[2]

In 1843 Sir Nicolas Tindal established what has become known as the *M'Naghten Rule*. These rules still provide the conceptual legal framework for excusing a person's criminal act committed while the defendant was suffering from a defect of reason or disease of the mind. For the defense to operate, the actor must be shown to be in a state of mind such that he did "not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong."[3,4]

From a neuroscientific perspective, the criminal act or *actus reus* component of the crime is of less interest than the essential *mens rea* element. Implementation of the M'Naghten Rule thereby becomes a watershed moment concerning the influence of neuroscience on criminal law because it is the state of the mind—or perhaps, more accurately, the brain—to which inquiry is focused.

Sleep

With respect to criminal culpability, the inquiry surrounding sleep is whether a person in a sleep state possesses sufficient *mens rea* to support a conviction for the actor's behavior. The first appearance of the "sleepwalking defense" in an American court of law came in *Massachusetts v Tirrell*[5] in 1846. In this landmark case, Rufus Choate, a skilled orator and U.S. senator, successfully employed the "insanity of sleep" defense in the murder trial of Albert Tirrell. The evidence in that case proved that Tirrell brutally killed the victim with a razor, almost severing her head from her body, set the horrifically bloody crime scene ablaze, and then attempted to flee the country.[6] Choate, an innovative legal tactician influenced by the advent of the M'Naghten Rule in the United Kingdom, argued in part that Tirrell, a sleepwalker, murdered the victim in an unconscious sleepwalking state and was able to convince the jury to acquit the Tirrell on this basis.

Later, in the mid to late 1800s there were no plausible medical explanations to account for sleepwalking, let alone account for complex violent actions that apparently arose during sleep. Still, courts were willing to adopt and apply defenses to deadly crimes committed by persons in a sleep state by pleas of a temporary "defect of reason" or "disease of mind." See, for example, *HMS Advocate v Fraser* (1878)[7] and *Fain v Commonwealth* (1879).[8]

Until physiologic aspects of sleep could be objectively measured and verified using validated neuroscientific instruments, defending criminal behavior arising from sleep often meant associating such behavior with other, better-understood medical or psychiatric conditions such as insanity or automatism. For example, courts apply the insanity defense to excuse criminal actions resulting from a diseased mind incapable of knowing right from wrong. Where indicated, courts might also withhold criminal punishment for acts caused by a defendant's involuntary bodily movements exhibited even while in a conscious or sane state. Thus a defense of automatism may be appropriate for acts arising from epileptic seizures, fugue states, and limbic psychotic trigger reactions, whereas an insanity plea would be appropriate if the defendant acted in the throes of fulminant delusional paranoid schizophrenia.

The legal community's perspective toward sleep began to shift in 1968 with Roger Broughton's seminal publication characterizing the relationship among somnambulism, nightmares, confusional states of arousal, and rapid eye movement (REM) sleep.[9] By creating a clear demarcation between sleep disorders and other medical or psychiatric conditions, this appears to be the first scientific sleep-related publication with direct legal implications, as demonstrated by the 1992 Canadian criminal case of *Regina v Parks* (1992)[10] in which Broughton served as an expert witness on behalf of the defense.

The defendant in the *Parks* case claimed that while sleepwalking in the early morning hours he drove to the house of his wife's parents and, provoked to attack by his in-laws' physical contact, killed his mother-in-law with a kitchen knife and left his father-in-law seriously injured.[11] The defendant defended his actions on the basis of automatistic sleepwalking rather than insanity. Expert witnesses for the defense testified that sleepwalking is not a neurologic, psychiatric, or other illness but rather is a sleep disorder very common in children and also found in adults.

The jury acquitted the defendant on the basis of automatism, which is a complete acquittal, as opposed to finding the defendant not guilty by reason of insanity, which typically leads to some institutional incarceration. The Canadian Supreme Court took up the case to decide the single legal issue of whether sleepwalking should be classified as noninsane automatism or insane automatism arising from a "disease of the mind," giving rise to a special verdict of not guilty by reason of insanity. Based on the unchallenged expert testimony that sleepwalking is a sleep disorder rather than a mental defect, the court rejected the characterization of sleepwalking as a mental health disorder.

EVOLUTION OF CONSCIOUSNESS THOUGHT

Criminal law presumes that most human behavior is voluntary and that individuals are consciously aware of their acts. All criminal liability is based on a voluntary act, or an omission to engage in a voluntary act that the defendant would otherwise have been capable of performing. Voluntariness is the first step in establishing *mens rea.* If the state proves *mens rea*, then the state will assess liability according to four levels of culpability: purpose, knowledge, recklessness, and negligence. In criminal law, the level of culpability determines the category of homicide (murder, manslaughter, or negligent homicide), and the category directly influences the severity of punishment.

Because voluntariness is absolutely fundamental to *mens rea,* it is surprising that the MPC offers examples of involuntary acts in lieu of explicitly defining the term *voluntary acts*. One example of an involuntary act is bodily movement during unconsciousness or sleep. Thus the MPC equates sleep with unconsciousness and deems bodily movements performed in a sleep state to be involuntary and presumably excused from criminal punishment. The other three examples of involuntary acts in the MPC include reflex convulsion; bodily movement that is not otherwise a product of the effort or determination of the actor, either conscious or habitual; and conduct during hypnosis or hypnotic suggestion.[12,13]

Waking Consciousness

A comprehensive review of the neuroscience of consciousness is well beyond the scope of this chapter. However, consciousness involves awareness of our environment, awareness of our bodies, and introspection (self-awareness), and it can only fully occur when we are awake.

To neuroscientists, *consciousness* is a term that has varied meanings, although its definition in the legal realm has held steadfast. In science, for example, consciousness may be used to indicate whether an individual is in a conscious state, as in whether it has been altered, reduced, or even lost. On the other hand, consciousness may be a trait or an attribute of a psychological process, as in the ability to think, see, and feel consciously. With trait consciousness, further distinctions may be made between conscious representations, which are usually phenomenal, and required conscious access.

Unfortunately a direct objective marker for the neural basis of state and trait consciousness that is independent of a person's external expressions or behavior has yet to be determined. Nevertheless it is believed that the neuronal processes that mediate access to consciousness take place in a network of frontoparietal cortical regions of the brain. These networks play an important role in attentional and behavioral selection of incoming and stored information. Because the frontoparietal cortical regions govern behavioral selection, it is not surprising that this region becomes active when patients in a vegetative state recover or becomes further activated when healthy subjects perform demanding perceptual tasks.

Sleep is regularly and actively induced by a shift in neuronal activity and neurotransmitter balance in brainstem nuclei. Functional neuroimaging studies performed on sleeping subjects reveal that during both REM and non–rapid eye movement (NREM) sleep the prefrontal and parietal cortical regions become deactivated in comparison to the resting wakeful state.[14-16] The most active regions during the resting wakeful state include the left dorsolateral and medial prefrontal areas, the inferior parietal cortex, and the posterior cingulate and precuneus.[15] In REM sleep, despite overall increases in cerebral blood flow and energy demands, relatively low regional cerebral blood flow persists in prefrontal and parietal cortex.[17] Because consciousness can only fully occur when we are awake, the frontal cortex would appear to be indispensable to consciousness.

Another important region is the dorsal thalamus, the "gateway to the cortex," and its accompanying "guardian of the gateway," the reticular complex (part of which is often called the *perigeniculate nucleus*).[18-21] Francis Crick believed that the input and output gating of the reticular complex were topographically arranged to approximate a map of the entire cortex. In his searchlight hypothesis, the reticular complex thereby was able to heat up the warmer parts of the thalamus and cool down the cooler parts so that "attention" would remain focused on the most active thalamocortical regions.[22] Although the function of the thalamic reticular complex remains incompletely understood, it is essential for consciousness, whereas cerebellar circuits, in contrast, are not.

Consciousness as a Continuum

The modern concept of consciousness was perhaps first established by influential American scientific psychologist and pragmatist William James (1842–1910). Consciousness has been subdivided into nine distinct components (Table 65-1), all of which are seamlessly integrated into our own personal conscious experience.

From the notion that consciousness is graded and not dichotomous, J. Alan Hobson developed the AIM concept. This concept creates a four-dimensional "mind space" in time that is transformed by three variables: activation (A), input-output gating (I), and neuromodulation ratio (M) (as measured by the aminergic-to-cholinergic ratio). All of these determine changes in the state of consciousness, which in turn govern the oscillation from wake to sleep.

These variables account for the physiologic properties within each state.[23] The interaction of these variables can cause temporary nonhomeostatic conditions that, in turn, might trigger or cause undesirable behavior without consciousness or memory. Such behaviors might include violent outbursts related to REM sleep behavior disorder.[24]

The recurrent pattern of state-determining parameters is amazingly consistent. However, there are numerous clinical

Table 65-1 The Nine Components of Consciousness

Component	Function
Perception	Representation of input data
Attention	Selection of input data
Memory	Retrieval of stored representations
Orientation	Representation of time, place, and person
Thought	Reflection on representations
Narrative	Linguistic symbolization of representations
Instinct	Innate propensities to act
Intention	Representation of goals
Volition	Decisions to act

and experimental examples of dissociation of state components. Such dissociation can be explained as simultaneous mixtures of clinical and neurophysiologic elements of the three states of being—wake, NREM sleep, and REM sleep. These fall into three categories, as reviewed by Mahowald and Schenck,[25] and include neuroanatomic lesions or stimulation, pharmacologic mechanisms, and sleep deprivation. Neuroanatomic lesions or stimulation include hypothalamic, thalamic, and brainstem manipulation or stimulation inducing state dissociation. Pharmacologic mechanisms include manipulation of the cholinergic or glutamate neurotransmitter systems resulting in a variety of state dissociations. A consequence of general anesthesia is "cognitive unbinding," which further explains state dissociation.[26] As for sleep deprivation, recent studies by Montplaisir and colleagues suggest that sleepwalking results from a dysfunction of the mechanism responsible for sustaining stable slow-wave sleep and that sleepwalkers are particularly at risk when exposed to increased homeostatic sleep pressure.[27,28]

State dissociations are the consequence of timing or switching errors in the normal process of the dynamic reorganization of the central nervous system as it moves from one state (or mode) to another. Elements of one state persist or are recruited erroneously into another state, often with fascinating and dramatic consequences.

Objective support for state dissociation is provided by depth electrode electroencephalographic studies demonstrating areas of wakefulness and sleep occurring simultaneously in humans.[29-32] This concept helps to explain such phenomena as sleep inertia, waking hallucinations, narcolepsy, REM sleep behavior disorder, lucid dreaming, out-of-body experiences, near-death experiences, repressed or recovered memories of childhood sexual abuse, alien abductions, and disorders of arousal (Figure 65-1).[33-39] The concept of state dissociation supports the notion that consciousness occurs on a continuum.

Fixed Action Patterns and Central Pattern Generators: A Neuroethologic Approach to Behavior

Ethology is the study of whole patterns of animal behavior under natural conditions in a manner that highlights the functions and the evolutionary process of those patterns. With an ever-increasing physiologic approach through the application of refined and elegant laboratory research techniques to animal behavior, neurobiology and ethology have coalesced to develop neuroethology.[40]

An important behavior type in ethology is the fixed action pattern (FAP). This is an instinctive indivisible behavioral sequence that when initiated will run to full completion. FAPs are invariant and are produced by a neural network known as the *innate releasing mechanism* in response to an external stimulus known as a *sign stimulus*.

FAPs are ubiquitous in the animal kingdom and are seen from invertebrates to higher primates. Movements resulting in FAPs may be initiated by central pattern generators (CPGs): "Movements are generated by dedicated network of nerve cells that contain the information that is necessary to activate the different motor neurons in the appropriate sequence and intensity to generate motor patterns. Such networks are referred to as *Central Pattern Generators*."[41]

Tassinari and coworkers recognized that motor events related to certain epileptic seizures and parasomnias share very similar features. This suggests a stereotyped inborn FAP, perhaps initiated by CPGs.[42] Tassinari recognized CPGs as genetically determined neuronal aggregates in the mesencephalon, pons, and spinal cord that, from an evolutionary perspective, were linked with innate primal behavior essential for survival (e.g., feeding, locomotion, reproduction).

In higher primates, CPGs are inhibited by the influence of neocortical control. Many of the CPGs are located in the brainstem and in close proximity to processes that govern the wake, NREM sleep, and REM sleep transitions. Despite diurnal neocortical inhibition, Tassinari provides a neuroethologic model whereby both epilepsy and sleep can lead to a temporary loss of control of the neomammalian cortex that is provided a pathway through a common arousal platform initiated by CPGs, which in turn triggers these FAPs (Figure 65-2), resulting in the abrupt onset of bizarre motor or

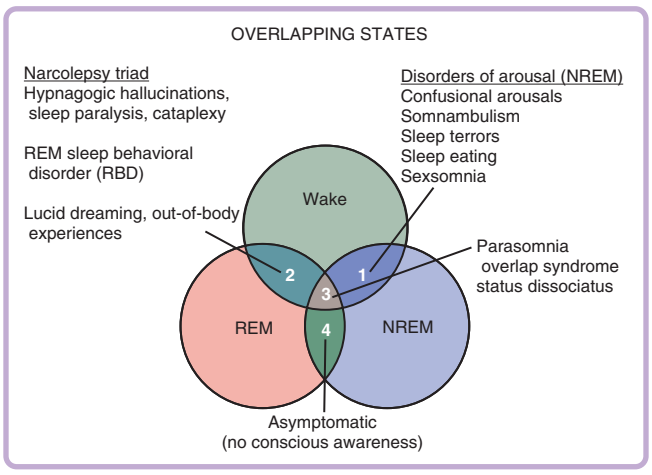


Figure 65-1 Areas of overlap among states of being. (Modified from Mahowald MW, Schenck CH. Dissociated states of wakefulness and sleep. *Neurology* 1992;42:44–52).

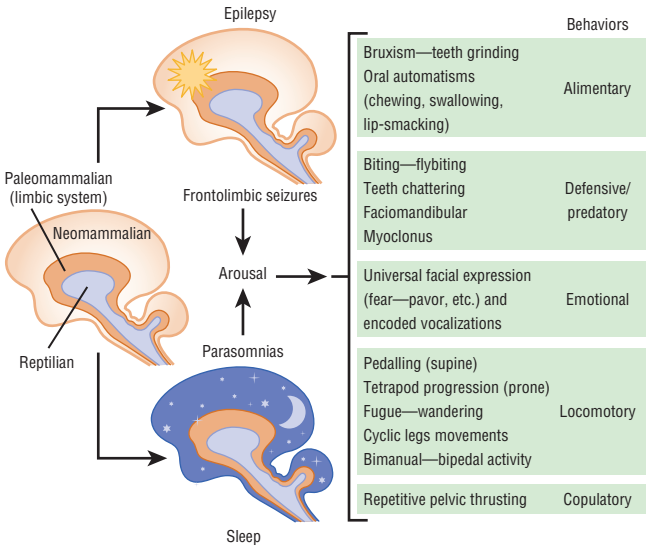


Figure 65-2 The emergence of innate primal behavior facilitated through central pattern generators from the arousal platform. (Modified from Tassinari CA, Rubboli G, Gardella E, et al. Central pattern generators for a common semiology in fronto-limbic seizures and in parasomnias: a neuroethologic approach. *Neurol Sci* 2005;26:S225–32).

emotional expressions that are uncharacteristic of awake neocortical-mediated diurnal behavior.

Tassinari's concept of the role of CPGs and FAPs provides a physiologic explanation for parasomnias. This concept is particularly useful in sleep forensics because parasomnias and epileptic seizures tend to have patterned stereotyped actions without conscious awareness. When addressing criminal allegations and their potential association with sleep-related conditions, the sleep medicine specialist can use behavior pattern recognition, applying neuroethologic concepts that indicate process fractionation, and neurobehavioral investigative techniques. Such an approach could be particularly beneficial and would be consistent with the direction of current mainstream neuroscience.

Dreaming Consciousness

It is obvious with sleep onset that sensory input is largely lost and our ability to interact with the external environment is curtailed. The conscious state paradigm outlined by J. Alan Hobson recognizes that all nine components of consciousness change to varying degrees as the brain changes state and does so in a repetitive and stereotyped manner over the sleep-wake cycle. Furthermore, consciousness is graded, and the state changes appear to be of such dramatic magnitude that strong inferences can be made about the major physiologic underpinnings of consciousness.[23] Sleep physicians on a daily basis appreciate, by interpreting polysomnographic studies, the state-determined uniformity of physiologic events that are consistent from patient to patient. To analyze the transitions in process, from wake to sleep or from NREM to REM, and isolate its individual components in order to deduce its underlying state, including its associated degree of consciousness, is a method called *process fractionation* (Figure 65-3). The application of the conscious state paradigm has led Hobson to declare three important principles.

First, consciousness rides on the crest of the brain activation process. Therefore, even a small perturbation in activation level leads to lapses in waking vigilance. Second, the brain remains highly active and capable of processing information even though consciousness may be largely deactivated. Functional imaging studies reveal that the brain remains about 80% active even when consciousness has largely subsided. Lastly,

	Wake	NREM	REM	Causal Hypothesis
Sensation & Perception	Vivid, externally generated	Dull or absent	Vivid, internally generated	Presynaptic inhibition, Blockade of sensory input
Thought	Logical and progressive	Logical and perseverative	Non-logical and bizarre	Loss of attention memory and volition leads to failure of sequencing and rule inconstancy; analogy replaces analysis
Attention	Intact, vigilant	Lost	Lost	Decreased aminergic modulation causes a decrease in signal to noise ratio
Orientation	Intact	Unstable	Unstable	Internally inconsistent orienting signals are generated by cholinergic system
Emotion	Inhibited	Weak	Episodically Strong	Cholinergic hyperstimulation of amygdala and related temporal lobe structures
Instinct	Inhibited	Weak	Episodically Strong	Cholinergic hyperstimulation of hypothalamus and limbic forebrain triggers CPG/FAP axis
Aminergic Inhibition (-)				Aminergic Inhibition (-)
Cholinergic Excitation (+)				Cholinergic Excitation (+)

Figure 65-3 Process fractionation: contrasts in components of consciousness between states. CPG, Central pattern generator; FAP, fixed action pattern. (Modified from Hobson JA. States of consciousness: normal and abnormal variation. In: Zelazo PD, Moscovitch M, Thompson V, editors. *The Cambridge handbook of consciousness*. New York: Cambridge University Press; 2007. p. 69.)

most brain activity is *not* associated with consciousness. In relation to its evanescence, consciousness "is a very poor judge of its own causation and of information processing by the brain."[23]

There have been seismic shifts in cognitive neuroscience that the legal system has yet to appreciate and incorporate into the legal arena. Rather confusingly, the terms *conscious* and *unconscious* are still used in the lexicon of neuroscience, but the ideas and principles behind these terms have been substantially altered and continue to be refined, with one such example being Tononi's information integration theory of consciousness.[43,44]

Advances in neuroscience since the 1980s support the existence of a continuum of conscious and unconscious processes, and neuroscience has largely dispensed with Freudian-influenced psychoanalytic concepts and theories. The boundaries between conscious and unconscious, as between wake and sleep, are permeable, dynamic, and interactive. As such, there is no valid scientific support for the sharp dichotomy between consciousness and unconsciousness currently held by the MPC and the legal community. It is this model of state dissociation that assists in the explanation of unusual, irrational, or bizarre human behavior in sleep forensics.

COMPLEX BEHAVIOR ARISING FROM SLEEP

Increasingly, experts in sleep medicine are being called on by attorneys to assist in reviewing legal cases, most of which involve allegations of criminal behavior. The conventional path is to make an evaluation for parasomnias to explain a wide variety of sleep-related violent behavior. The typical defense strategy is to clothe the behavior as a parasomnia symptom and thus completely exonerate the perpetrator's actions. Case requests to review bizarre nocturnal activities as mere rage reactions attributed to pharmaceutical agents such as benzodiazepines and, particularly, nonbenzodiazepine drugs are not uncommon.

Incidents of violent sleep-related behavior have been reviewed in the context of automatic behavior in general, with many well-documented cases resulting from a wide variety of disorders. Conditions associated with violence during the sleep period fall into two major categories: neurologic and psychiatric (Box 65-1). Those actions arising from a primary neurologic condition can be explained by applying conceptual approaches based on models of evanescent consciousness, the overlapping physiology of clinical disorders, and the platform of CPGs supported by semiotic neuroethology.

CLINICAL GUIDELINES TO ASSIST IN DETERMINING PURPORTED VIOLENCE ARISING FROM SLEEP

Legal implications of automatic behavior have been discussed and debated in the medical and legal literature.[45-49] The identification of a specific underlying organic or psychiatric sleep and violence condition does not establish causality for any given deed.

To assist in determining the existence of an underlying sleep disorder in a specific violent act, practitioners should follow guidelines based on peer-reviewed international clinical experience. Several clinical guidelines have been proposed[50-53] that identify certain features occurring as a result of a sleep disorder:

- There should be reason by history to suspect a bona fide sleep disorder. Similar episodes, with benign or morbid outcome, should have occurred previously.
- There has to be some degree of interaction with the environment. This behavior cannot be entirely passive in nature.
- The duration of the action is usually brief (seconds), although action of longer duration (minutes) does not necessarily exclude a sleep disorder or a sleep-related behavior. The action is usually abrupt, immediate, impulsive, and senseless—without apparent motivation. Although ostensibly purposeful, it is completely inappropriate to the total situation, out of (waking) character for the individual, and without evidence of premeditation.
- The victim is someone who merely happened to be present, usually in close proximity, and who may have been the stimulus for the arousal. Sleepwalkers rarely, if ever, seek out victims.[53,54]
- Immediately following return of consciousness, there is perplexity or horror, and there is no attempt to escape, conceal, or cover up the action. There is evidence of lack of awareness on the part of the sleepwalker during the event. There is usually some degree of amnesia for the event, but this amnesia need not be complete.
- In the case of sleep terrors, sleepwalking, or sleep inertia, the act may occur on awakening (rarely immediately on falling asleep) and usually at least 1 hour after sleep onset. It occurs on attempts to awaken the subject. The action has been potentiated by sedative-hypnotics or by prior sleep deprivation.
- Lastly, the violent behavior cannot be better explained by another mental disorder, medical condition, medication, or substance use. Ultimately, to attribute a violent behavior with criminal implications to parasomnia is a diagnosis of exclusion with the explicit understanding that other conditions are often more statistically likely. Note that this final guideline is also in accord with the diagnostic criteria for parasomnias in the *International Classification of Sleep Disorders*, 3rd edition.

Box 65-1 CONDITIONS ASSOCIATED WITH AUTOMATIC BEHAVIOR ARISING FROM THE SLEEP PERIOD

Primary Sleep Disorders (Neurologic Conditions)

Disorders of arousal
- Confusional arousal
- Sleep-related abnormal sexual behavior (sexsomnia)
- Sleep terror
- Somnambulism

REM sleep behavior disorder
Nocturnal seizures
Compelling hypnagogic hallucinations
Somniloquy

Psychiatric Conditions

Dissociative states (may arise exclusively from sleep)
Posttraumatic stress disorder
Malingering
Munchausen syndrome by proxy
Psychopathy

The guidelines for determining the role of a sleep disorder in violence are not meant to be perceived as a rigid rule nor as a set of necessary criteria. They merely provide direction to gauge whether an argument in favor of a sleep disorder could be sustained in the formulation of a possible criminal defense. The strength of the argument should consider current neuroscientific models of consciousness and behavior as supported by the medical expert's specialized clinical experience.

THE ROLE OF THE SLEEP MEDICINE SPECIALIST

To address the problem of junk science in the courtroom, some professional societies have developed guidelines for expert witness qualifications and testimony. The American Academy of Sleep Medicine's stance on expert witness testimony is to accept opinions held by the American Medical Association in its 2004 Report of the Council on Ethical and Judicial Affairs.[55] Similarly, influenced by both the American Academy of Neurology and the AMA, the following guidelines should serve as a compass.

Expert Witness Qualifications

Expert witnesses should have a current, valid, unrestricted medical license. Expert witnesses testifying about sleep medicine should be diplomats of the American Board of Sleep Medicine or should have passed the American Board of Internal Medicine specialty examination in sleep medicine. Membership in the Sleep Research Society is strongly encouraged. An expert witness in sleep medicine must be a recognized resource within the sleep medicine community and should have been actively involved in clinical practice in a manner consistent with the requirement of the criminal case at the time of the event. Given the essential position of *mens rea* in criminal law and the pivotal role of levels of consciousness, an expert witness should have significant direct experience in either neurology or neuroscience.

Guidelines for Expert Testimony

Expert testimony must be impartial. The ultimate test for accuracy and impartiality is a willingness to prepare testimony that could be presented unchanged for use by either the plaintiff or the defendant. Fees should relate to time and effort and should not be contingent on the outcome of the claim. Fees should not exceed 20% of the practitioner's annual income. The practitioner should be willing to submit such testimony for peer review. To establish consistency, the expert witness should make records from his or her previous expert witness testimony available to the attorneys and expert witnesses of both parties. The expert witness must not become a partisan or advocate in the legal proceeding.

It is not the role of the medical expert to win the case for a client, although it is not uncommon to use irrelevant disingenuous technicalities in an attempt to deceive so as to attain an advantage to secure the decision. Instead, the salient ethical decision for those who assume this mantle of medical expert witness is to recognize and value the privileged position given within our society as an educator inside the legal system by promoting current published peer-reviewed science and minimizing bias while rendering an opinion. The goal of the expert witness is not to simply ascertain or promote an argument addressing "reasonable doubt" in any given case because this is best deferred to legal parlance most often provided during counsel's closing statements. Instead, the role of the expert witness is therefore to attempt to succinctly and clearly communicate scientifically valid information without bias within the context of the case to the jury, who in turn determines culpability based on this information. The weight of the decisions of either guilt or innocence should never rest in the hands of medical experts, whose task is to contribute to the due process of an efficient and functional legal system by ensuring that the jury is educated and well informed.[56]

CLINICAL PEARLS

In a court of law, the undisciplined use of scientific technical data is a real concern, especially given the public misperception that science is a field that deals with absolute certainties when in actuality it is a field that reflects probabilities of occurrence. In many ways, the legal community has misrepresented the nature of science for many years and continues to attempt to do to so in an admittedly adversarial environment.

One problem with presenting results of scientific research is a condition called *brain overclaim syndrome*.[57] In general, the public has an overfascination with new developments in science and often assumes statements of scientific evidence are true, even when the statements cannot be conceptually or empirically sustained. Studies have shown, for example, that the results of a simple experiment in cognitive psychology are more positively evaluated and considered important if a brain scan is included in the report of the results.[58] The limitations of neuroscience data in the courtroom must be appreciated.[59]

In a sleep-related example, a review has clearly established that polysomnography performed after the fact is of absolutely no value in determining whether the accused was sleepwalking at the time of the criminal activity. On very infrequent occasions, polysomnography may be considered to assess potential mitigating influences, such as obstructive sleep apnea. Even sleepwalking during a formal sleep study would only indicate that the person was a sleepwalker, not that sleepwalking was involved at the time of the crime.[60]

SUMMARY

Advances in neuroscience are increasing our understanding of how the brain enables action, including everything from simple movement, to thought, to the diurnal and nocturnal variability of sleep-wake processing.

The societal and cultural implications of these scientific advances have yet to be understood or even conceived.[61] However, the legal community is aware of the implications of this new neuroscience because science directly challenges the law's currently held constructs of consciousness as defined by *mens rea* and the voluntary act requirements. To study these problems, the John D. and Catherine T. MacArthur Foundation has established the Law and Neuroscience Project in 2007 (www.lawandneuroscienceproject.org) comprising 40 neuroscientists, legal specialists, and philosophers.[62] Two important concepts to be incorporated into the legal community are that consciousness is not all-or-none but occurs on a spectrum and that consciousness can be dissociated from behavior.

Sleep forensics does more than provide medical expert testimony in individual legal cases. The growth of cognitive neuroscience will continue to change our understanding of what it means to be human, and as a result the law will have to change in conformity with it.

The conceptual approach to sleep forensics encourages further research to define and characterize mixed states of wake and sleep and the parasomnias. Understanding all of these is beneficial in understanding the spectrum of complex human behavior. Close collaboration among basic neuroscientists, sleep medicine clinicians, and the legal community will facilitate the development of a commonly shared concept of consciousness and culpability.

Selected Readings

Casartelli L, Chiamulera C. Opportunities, threats and limitations of neuroscience data in forensic psychiatric evaluation. *Curr Opin Psychiatry* 2013;**26**(5):468–73.

Cramer Bornemann MA. Sexsomnia: a medico-legal case-based approach in analyzing potential sleep-related abnormal sexual behaviors. In: Kothare SV, Ivanenko A, editors. *Parasomnias*. New York: Springer Science+Business Media; 2013. p. 431–61.

Goldrich MS. Report of the Council on Ethical and Judicial Affairs. *CEJA Report 12-A-04* American Medical Association; 2004.

Ingravallo F, Poli F, Gilmore EV, et al. Sleep related violence and sexual behavior in sleep: a systematic review of medical-legal case reports. *J Clin Sleep Med* 2014;**10**:927–35.

Mashour GA. Cognitive unbinding: a neuroscientific paradigm of general anesthesia and related states of unconsciousness. *Neurosci Biobehav Rev* 2013;**37**:2751–9.

Popat S, Winslade W. While you were sleepwalking: science and neurobiology of sleep disorders and the enigma of legal responsibility of violence during parasomnia. *Neuroethics* 2015;**8**(2):203–14.

Pressman MR. Disorders of arousal from sleep and violent behavior: the role of physical contact and proximity. *Sleep* 2007;**30**(8):1039–47.

Pressman MR. Factors that predispose, prime, and precipitate NREM parasomnias in adults: clinical and forensics implications. *Sleep Med Rev* 2007;**11**:5–30.

Pressman MR, Caudill DS. Alcohol-induced blackout as a criminal defense or mitigating factor: an evidence-based review and admissibility as scientific evidence. *J Forens Sci* 2013;**4**:932–40.

Pressman MR, Mahowald MW, Schenck CH, et al. Alcohol-induced sleepwalking or confusional arousal as a defense to criminal behavior: a review of scientific evidence, methods and forensic implications. *J Sleep Res* 2007;**16**:198–212.

Schenck CH, Arnulf I, Mahowald MW. Sleep and sex: what can go wrong? A review of the literature on sleep related disorders and abnormal sexual behaviors and experiences. *Sleep* 2007;**30**:683–702.

Tassinari CA, Rubboli G, Gardella E, et al. Central pattern generators for a common semiology in fronto-limbic seizures and in parasomnias: a neuroethologic approach. *Neurol Sci* 2005;**26**:s225–32.

Tononi G. Consciousness as integrated information: a provisional manifesto. *Biol Bull* 2008;**215**:216–42.

Zadra A, Desautels A, Petit D, et al. Somnambulism: clinical aspects and pathophysiological hypotheses. *Lancet Neurol* 2013;**12**:285–94.

A complete reference list can be found online at ExpertConsult.com.

Legal Obligations of Persons Who Have Sleep Disorders or Who Treat or Hire Them

Chapter 66

Daniel B. Brown

Chapter Highlights

- Patients have a right to refuse treatment for sleep-related breathing disorders, but they may be liable if the foreseeable results of their sleep-disordered actions cause injury to others.
- Drivers who experience a "sudden blackout" while behind the wheel are excused from liability for damages arising from crashes caused as a result of their unexpected unconsciousness or seizure. However, this legal defense fails if the driver knew that his or her "blackout" or sleep episode was imminent or otherwise foreseeable.
- At least two states, New Jersey and Arkansas, factor fatigue into a criminal finding of reckless driving by drowsy drivers.
- Employers will be found liable for the negligence of their employees if the accident causing the injury occurred during the course of the employee's employment. Thus a trucking company will likely be found liable for injuries caused by an employee truck driver who falls asleep at the wheel while on the job.
- Employers are not likely to be found liable for injuries caused by their employees' sleep-induced actions that occur outside the scope of the employee's employment, even if the employee's fatigue is caused directly by the employer's overscheduling the employee's work hours.
- Whether a sleep disorder is a "disability" for purposes of the Americans with Disabilities Act (ADA) depends in large part on whether the discriminatory actions taken by an employer against a sleep-disordered individual occurred before or after passage of the ADA Amendments Act of 2008 (ADAAA). Sleep disorders are much more likely to be considered disabilities eligible for ADA protection after the effective date of the ADAAA.

Sleep disorders such as obstructive sleep apnea (OSA) have been shown to fragment sleep and deprive the sufferer from restful and regenerative slumber.[1] In addition to comorbidities such as cardiopulmonary diseases, untreated OSA can lead to daytime hypersomnolence, which, in turn, can directly and adversely affect an OSA sufferer's overall daytime performance on the job or on the road.[2]

Legal systems examine a person's adherence to reasonable or statutorily defined standards of conduct and whether variance from such standards causes injury to others. Civil legal remedies generally award monetary penalties to injured parties as compensation for their damages incurred at the hands of a negligent party.[3] Criminal law metes out fines, imprisonment, or other punishments for conduct determined by state or federal statute to be injurious against the public at large.[3]

Legal proceedings under civil law focus on whether the wrongdoer had a duty of care to the injured party and whether the actor's breach of that duty caused the injured party's damages.[3] Criminal court actions focus on whether the actor's conduct violated each element of an activity that a particular statute considers to be criminal.[4]

Because there exists a causal connection between a person's sleep-disordered fatigue and such person's impaired daytime activity, the law will examine whether the sleep-disordered patient or those treating or employing the patient owe any special duty of care. That duty might arise to protect the patient or those with whom the patient may come into contact. This chapter briefly discusses the legal duties owed by sleep-disordered patients, the health care professionals who diagnose and treat them, and those persons or entities who employ potentially fatigue-prone persons under current U.S. law.

SURVEY OF SLEEP APNEA IN THE LAW

What ultimately was called *sleep apnea* was described by several European groups in the 1960s, and the first description in a major journal appeared in 1973[5] (see Chapter 1). It was not until the 1980 Social Security benefits case of *Parks v. Harris* that the disorder was first seen in a reported decision in the United States.[6] Parks, who had recently been diagnosed with sleep apnea, sued to overturn his benefits denial based on a vocational expert's testimony that Parks suffered from uncontrolled somnolence due to a sleep disorder.[6] This early case did not consider disease treatment as a factor to defeat the disability claim because continuous positive airway pressure (CPAP) treatment for OSA was not available until 1981, or a year after the *Parks* decision.[7]

Legal notice of OSA was slow to develop following *Parks*. For example, in the 9 years after the *Parks* case, references to sleep apnea in all reported U.S. appellate decisions appeared in only 17 cases. However, recognition of OSA in U.S. jurisprudence grew quickly after 1980 and fairly exploded after 2005. References to OSA in reported U.S. cases grew more than 4000% in the 25 years from 1980 to 2005, with 772 cases making some reference to OSA during this period. The references to OSA in U.S. case law grew an astounding 5196 additional cases in the 8-year period between January 1, 2006 and July 1, 2014, mostly in cases associated with patient disability. To be clear, most of the combined 5985 cases since *Parks* contain only passing references to OSA. For example, OSA often appears only as one of many conditions that make up a defendant's general health profile and has no bearing on the outcome of the case. For example, only 215 of these cases include any mention of malpractice, and hardly any those address a health care practitioner's malpractice with regard to a sleep disorder.

Legal Obligations of Persons with Obstructive Sleep Apnea

A long-established general rule states that patients may choose to ignore their medical conditions and refuse treatment without facing legal consequences.[8] That right of refusal does not give those patients the right to ignore the risks to public safety caused by their decisions.[8] In California, for example, there is a statute requiring those with certain infectious diseases to take precautions to avoid the willful spread of their diseases through public contact.[9] Similarly, although an OSA patient may choose not to treat his or her condition with CPAP, oral therapy, or surgery, there can still be legal consequences to driving on the road while fatigued or performing other safety-sensitive activities.

A 1925 Connecticut Supreme Court case addressed the then novel question of a driver's legal duty when possessed by sleep or other unconscious episode. In *Bushnell v. Bushnell,*[10] Mr. and Mrs. Bushnell drove from Connecticut to Rhode Island to drop their son off at college. On the return trip, Mr. Bushnell dozed off at the wheel and crashed into a tree in a single-car accident. Mrs. Bushnell, who had been sleeping while Mr. Bushnell drove, was injured in the accident.

Mrs. Bushnell sued her husband in negligence for failing to operate the car in a reasonable manner. Mr. Bushnell demurred, arguing in essence that he was to be excused from his duty to maintain control of the car while asleep because sleep, like any other unforeseeable blackout condition, comes about without warning. In essence, Mr. Bushnell argued that he had no duty to adjust his driving conduct that day because the instant of sleep onset cannot be determined in advance.

The Court challenged Mr. Bushnell's assertion that he had no warning of sleep onset. The court received medical evidence indicating that unlike a sudden blackout, sleep displays routine and recognizable precursor conditions. Based on early twentieth century medical knowledge, these conditions include sensations of well-being, fatigue, and dulling of the senses. On the basis of this medical evidence, the Court ruled that Mr. Bushnell knew, or should have known, that sleep was overtaking his driving and that he should have pulled off the road. Because his sleep episode was foreseeable, the Court found Mr. Bushnell liable for the cost of his wife's injuries. This ruling sets out the rule that an unforeseeable loss of consciousness, such as a suddenly unexpected seizure or blackout, would excuse the driver's duty to exercise due care in driving.

The "sudden blackout" rule is an important legal protection for drivers who suffer from a sudden and unforeseen onset of sleep or seizure disorder. This awareness may exist as a result of disease or past experience of a tendency to either fall asleep or lose consciousness while driving. If a patient knew that he or she suffered from frequent seizures or narcoleptic episodes several times a day, it would be negligent for that individual to get behind the wheel of a car even if the seizure events were unexpected.

An example of the endurance of the sudden blackout rule is seen in a 2006 Vermont case, *State v. Valyou*.[11] In *Valyou,* the defendant dozed off multiple times on the way to work yet continued his drive, eventually colliding with another vehicle after falling asleep.[11] The Vermont Supreme Court recited the Bushnell rule that falling asleep at the wheel does not, in and of itself, constitute gross negligence.[11] "On the other hand, when a driver is on sufficient notice as to the danger of falling asleep but nevertheless continues to drive, the driver's subsequent failure to stay awake may be grossly negligent."[11] The Vermont Supreme Court held the defendant liable because he remained at the wheel despite his knowledge that he was at high risk for falling asleep and injuring others.

CRIMINAL LIABILITY FOR DROWSY DRIVING

Some states have reviewed their negligent homicide or reckless driving laws and amended these to recognize erratic driving behavior caused by sleep deprivation. The first state to do so was New Jersey through enactment of Maggie's Law in 2003.[12] This law arose following the death of college student Margaret "Maggie" McConnell. McConnell was struck by a driver who had not slept for 30 hours leading up to the crash and had smoked crack cocaine hours before the accident.[13] Unlike drunkenness, no New Jersey statute required the State to consider drowsy driving as a condition or factor in the driver's reckless operation of the car. The judge refused to admit evidence of the driver's sleep deprivation to establish reckless behavior, and the driver received only a $200 penalty for causing the accident.[13]

Maggie's Law is an evidentiary rule establishing that proof of driving after 24 hours of sleeplessness "*shall* give rise to an inference that the defendant was driving recklessly" in order to convict a defendant for vehicular homicide[12] (*emphasis added*). The law also establishes that falling asleep while driving may infer recklessness without regard to sleeplessness.[12]

As finally adopted, Maggie's Law does not criminalize drowsy driving. Fatigued driving provides only inferential evidence that a defendant was driving recklessly. This is to be contrasted with intoxication, the presence of which under Maggie's Law "*shall* give rise to an inference that the defendant was driving recklessly."[12] Thus, unlike proof of intoxication, evidence of drowsy driving in New Jersey will not, by itself, automatically lead to a conviction of reckless driving.

Unlike inebriation, proof of reckless fatigue under Maggie's Law is itself difficult under the law's definition of sleeplessness. Conviction under Maggie's Law requires proof that the defendant was driving "after having been without sleep for a period in excess of 24 consecutive hours." Under this language, evidence that the driver took a 10-minute nap during the

relevant 24-hour period can defeat a prosecutor's offer of the inference of reckless driving due to sleeplessness. Despite its shortcomings, Maggie's Law does open a path for states to consider drowsy driving as a factor in proving reckless driving in vehicular homicide cases.

In 2013, Arkansas became the second state to expand its negligent homicide statute to recognize fatigue as a factor in proving criminal vehicular homicide. The Arkansas law provides that a defendant shall be guilty of a class B felony for negligently causing the death of another person as a result of operating a vehicle, aircraft, or watercraft in a state of intoxication, while passing a stopped school bus, or while fatigued.[14] Fatigue is defined as "having been without sleep for a period of twenty-four consecutive hours," or "having been without sleep for a period of twenty-four consecutive hours and in the state of being asleep."[14] Unlike New Jersey's statute, Arkansas creates no distinction between driving fatigued and driving while intoxicated for purposes of proving negligent homicide.

Several states are now considering drowsy driving statutes similar to those found in New Jersey and Arkansas. High-profile events, such as the sleep-deprived Wal-Mart truck driver who crashed into comedian Tracy Morgan's limousine in 2014, frequently raise public awareness of the issue and draw significant attention of lawmakers.

DUTIES OF PHYSICIANS TO SLEEP DISORDERED PATIENTS

On establishment of the physician-patient relationship, the physician owes the patient the duty of reasonable care present in the community when treating the patient.[15] Very few reported cases can be found that specifically hold a health care provider liable for malpractice surrounding diagnosis or treatment of sleep disorders. A rare example is the 1993 Louisiana case of *Cornett v. State, W.O. Moss Hospital,*[16] which stands for the principle that a physician owes a the duty of reasonable care to treat the patient's conditions and warn the patient of potentially fatal risks that may be associated with untreated sleep disorders such as OSA.

In the Cornett case, the hospital treating Mr. Cornett was found liable for his death after he suffered from cardiopulmonary arrest. Mr. Cornett had complained to the hospital's physicians of sleep apnea symptoms, including his falling asleep at the wheel on three separate occasions. Nonetheless, the hospital physicians focused for months on Mr. Cornett's other medical conditions, ordering endocrinology tests for diabetes and acromaly.

Although the hospital physicians knew that OSA is a potentially fatal condition, Mr. Cornett was never tested or treated for OSA or warned of the risks posed by untreated sleep apnea. At trial, expert witnesses testified that Mr. Cornett's death was likely caused by his untreated sleep apnea. The hospital's own medical expert testified that OSA is an emergency condition. On these facts, the hospital and its physicians were unable to avoid liability for malpractice following the patient's death.[16]

Within the duty of care that a physician owes to patients is the duty to obtain all pertinent information that may be relevant to a patient's health care. In *Feitzinger v. Simon,*[17] the patient, Mr. Feitzinger, was brought in for a routine hernia operation. The anesthesiologist neglected to inquire about Mr. Feitzinger's sleep apnea, and he did not learn about his history of OSA and his current use of CPAP. As a result, CPAP was not recommended for use during the surgery or during recuperation.

Mr. Feitzinger developed pneumonia and died of cardiac arrest 3 days after the hernia operation. His estate sued for malpractice, claiming that if the anesthesiologist had taken Mr. Feitzinger's complete medical history, the anesthesiologist would have known of the patient's OSA and would have recommended CPAP as part of the patient's procedure and recovery. According to expert testimony, CPAP use could have prevented the patient's pneumonia and eventual death. The court determined that the anesthesiologist's failure to check for a sleep-related breathing disorder established sufficient cause to take the trial to a jury.[17]

The principle of informed consent requires a physician who orders surgical treatment for OSA to inform the patient of surgical risks and treatment alternatives such as CPAP.[18] Thus, in *Russell v. Brown,*[18] the plaintiff visited an otolaryngologist complaining of snoring and recurrent tonsillitis. The physician recommended a tonsillectomy for the tonsillitis and the surgical uvulopharyngoplasty procedure for snoring after diagnosing the patient with mild sleep apnea.

The patient suffered complications from the surgery and brought suit against the physician, claiming that the physician never informed the patient of the risks of surgery or the availability of nonsurgical alternatives to his sleep apnea such as CPAP or laser surgery. The jury nonetheless found in favor of the physician.[18] Medical experts testified at trial that the physician's actions reasonably fell within the accepted standard of care. In addition, the patient had signed a broad consent form that undercut his claim of invalid consent. Although the testimony supported the jury verdict in this case, the judge noted the established rule in informed consent cases that patients must be informed of alternative methods of treatment and the risks and benefits of such treatment.[18]

PHYSICIAN'S DUTIES WITH REGARD TO THIRD PARTIES

In the usual case, a physician is not liable for damages to third parties caused by the negligent or criminal acts of his or her patient.[19] In legal terms, it is atypical for a physician to owe a legal duty to unknown third parties.[19] Whether liability for the acts of a physician's sleep-disordered patient redounds to the sleep physician depends on the facts of the case and the physician's documented warning to patients of the potential adverse effects of failure to adhere to disease treatment.

Under a seminal California case, if a physician knew or should have known that his or her patient was likely to cause serious bodily harm to others, such as a psychiatric patient who tells his psychologist of his plans to kill a woman he was stalking, then a duty arises for the physician to take reasonable steps to prevent the patient causing injury to others.[20] However, this duty rarely attaches in circumstances in which the physician merely dispenses medication or diagnoses a potentially debilitating condition.[21] Although some contrary law exists,[22] courts usually reason that it is beyond a physician's control whether a patient takes the medication prescribed for the patient's condition[23] or, in the case of sleep apnea, whether the patient complies or fails to comply with the patient's CPAP therapy.

If a physician prescribes treatment whose use or nonuse could cause an impaired condition, then the physician has a duty to warn the patient of the risks stemming from use or misuse of the treatment.[24] In *Gooden v. Tips*,[25] a physician prescribed Quaalude tablets for his patient but neglected to warn her of the dangers of driving under the influence of the medication. After taking the pills, the patient injured third parties while driving. The injured third parties sued the physician for damages.

The court ruled that the physician was liable to the injured third parties not because the physician had a duty to prevent his patient from driving but because the physician had the duty to warn the patient not to drive, which he failed to do.[25] A physician treating a patient for OSA or another sleep disorder would have a similar duty to warn the patient about the risks of driving while drowsy or under the influence of medication that impairs performance.

The duty to warn applies only when the physician knows or should have known of the patient's condition that could give rise to injuries. In the case of *Calwell v. Hassan*,[26] the defendant physician examined his patient, Sharon Rylant, who complained of disordered sleep and fatigue. The physician ruled out narcolepsy and prescribed Elavil for Ms. Rylant's excessive daytime sleepiness. Ms. Rylant visited the physician for the next 3 years, but she never complained of drowsy driving during the years following her first visit. Ms. Rylant fell asleep while driving sometime later and ran into Calwell, who was bicycling along the street. Calwell sued the physician for failing to warn Rylant not to drive.

The court determined that the physician was not liable for Calwell's injuries.[26] The court focused on the years-long gap period during which Rylant continued to visit the defendant physician but during which she never complained of hypersomnolence. On these facts, the court determined that the defendant did not breach a duty of care by neglecting to warn Rylant of the dangers that hypersomnolence posed to her driving.[26]

A physician may discharge his or her duty to warn by complying with state public safety disclosure laws. These laws, which are discussed in more detail in Chapter 67, either require or recommend physicians to inform governmental public health or motor vehicle officials of the identity of patients who present a danger to the public by continuing to drive.

EMPLOYER'S DUTIES WITH REGARD TO EMPLOYEES WITH SLEEP DISORDER

In the United States, the legal doctrine of *respondeat superior* provides that an employer may be held vicariously liable for the acts of an employee that are performed as part of the employee's duties.[27] Because many employers, such as trucking companies, taxi cab companies, or mail delivery services, employ drivers, those employers will be vicariously liable if the driver falls asleep at the wheel and injures a third party in a crash while performing his or her job.[28]

If the accident occurs within the scope of employment, then the trucking company will often try to defend by denying that their driver acted negligently. Thus, when Norman Munnal fell asleep at the wheel of a tractor-trailer and killed a woman when he drifted into oncoming traffic, his employer invoked Ohio's version of the "sudden blackout" doctrine by blaming the accident on Munnal's "sudden unconsciousness."[29]

As discussed, the sudden blackout defense fails if the defendant knew that loss of consciousness was likely to occur and thus was foreseeable.[29] In the *Munnal* case, Munnal testified that he had a propensity to fall asleep at unpredictable times and that he had fallen asleep at the wheel at least once before.[29] Munnal's fiancé testified that he slept poorly and only slept an average of 3 hours a night despite Munnal's testimony that he slept 8 hours. Munnal was also diagnosed with severe OSA following a sleep test ordered after the accident.[29]

Although there was no evidence that Munnal was aware he had sleep apnea before the test, or that his fiancé had shared her concerns about the quality of his sleep, the court found sufficient evidence that Munnal was aware of his excessive sleepiness. As a result of this knowledge and an expert's testimony that Munnal probably fell asleep rather than suffering from a sudden blackout, the court found Munnal negligent for failing to operate the truck in a safe manner. Munnal's employer was held vicariously liable because Munnal had been operating the rig within the scope of his employment.

EMPLOYER'S OVERSCHEDULING LIABILITY

Some employees who fall asleep at the wheel after the end of an overly long work day have sought to hold their employers liable for scheduling excessive work time. Courts generally refuse to hold employers liable for the postwork actions of their employees, even if the employer contributed to the employee's alleged fatigue.[30]

The case of *Black v. William Insulation Co.*[31] is representative. In that case Black fell asleep while driving and crossed over into oncoming traffic. The resulting collision killed Black. Black's widow sued her husband's employer, alleging that the trucking company negligently required Black to commute long distances and work long hours and that the employer had failed to provide proper training or safeguards to prevent such an accident from occurring.[31]

The court determined that the employer owed no duty to Black and that it was Black's personal responsibility to ensure that he was capable of safely driving to work.[31] Some of Black's personal decisions, such as working a second job and completing long commutes, were considered more significant contributions to his sleepiness than his working conditions.[31]

In *Barclay v. Briscoe*,[32] Sgt. Barclay suffered catastrophic injuries after a head-on collision with Richardson, who had fallen asleep on his way home following a 22-hour work shift. Barclay sued Richardson's estate and his employer. The court stated that it was the employee's responsibility to get to or from work and that absent special circumstances, an employer was not liable for the actions of employees traveling to and from work.[32] The court did not determine the 22-hour shift to be a special circumstance, especially because Richardson elected to complete such a long shift rather than having it forced on him by the company.[32]

EMPLOYER'S DUTY TO ACCOMMODATE AN EMPLOYEE'S SLEEP DISORDER UNDER THE AMERICAN WITH DISABILITIES ACT

The federal Americans with Disabilities Act (ADA) works in conjunction with various state job protection statutes to protect

employees with disabilities from discrimination by their employer.[33] The ADA prohibits employers from firing, failing to promote, or failing to provide "reasonable accommodations" to employees who suffer from a protected disability.[33]

A protected disability is one that limits one or more "major life activities" as defined by the court system.[34] The court seldom adds new major life activities, but it has long held that sleeping and breathing are major life activities.[35] As discussed later, few cases of OSA or other sleep disorders rose to the level of a qualified disability under the law before adoption of the ADA amendments in 2008.

To establish an ADA violation, an employee must show that (1) he or she is disabled; (2) he or she is otherwise qualified to perform the essential functions of the job with or without reasonable accommodation; and (3) the employer took an adverse job action against the employee because of the disability or failed to make a reasonable accommodation.[36]

Before passing of the American with Disabilities Act Amendments Act (ADAAA) in 2008,[37] few courts found that a diagnosis of OSA affected either sleep or breathing sufficiently to trigger ADA protections. By passing the ADAAA, Congress lowered the plaintiff's burden so that he or she needs only to show a "degree of functional limitation," which is a lower standard than the "substantially limits" standard applied before the ADAAA. The ADAAA applies to fact patterns occurring on or after January 1, 2009.

Before the ADAAA, allegations of OSA or sleeping difficulties without proof of an impairment to a major life activity were found insufficient to establish a substantial limitation in the major life activity of sleeping.[37] One court concluded that an individual's inability to get to sleep did not adversely affect the activity of sleep because difficulty sleeping was deemed extremely widespread and because the plaintiff in question did not provide evidence that his difficulties were any worse than difficulties suffered by a large number of other adults.[38] Successful use of oral appliances or positive airway pressure therapy to treat symptoms would also remove OSA as a disability for ADA purposes altogether because the courts found that successful treatment negated disability symptoms.[39]

There were rare circumstances when a court would find OSA severe enough to trigger pre-ADAAA protections,[40] particularly in situations in which therapy such as CPAP failed to relieve fatigue caused by severe sleep apnea.[41] The 2009 Pennsylvania case, *Peter v. Lincoln Technical Institute, Inc.*, is illustrative.[41] There, the plaintiff testified to waking up five to six times a night and falling asleep during the day, even while working and driving to work. The patient's disorder was unresponsive to a variety of therapies, including CPAP, tonsil surgery, oral medication, and pure oxygen therapy. The extreme and untreatable aspects of this particular patient's sleep apnea led to a disability determination.

The ADAAA was designed to refocus ADA cases on determining whether employers are discriminating against their employees, rather than extensively analyzing whether the patient is truly disabled under the legal definition. The plaintiff still possesses a burden of proof to show that he or she is disabled compared with the general population, yet the burden may be met by merely providing physician testimony of a sleeping disorder that interferes with sleep and leads to falling asleep on the job. In addition, successful CPAP treatment no longer acts as a mitigating measure to prevent OSA suffers from claiming a disability under the ADA.

The 2013 case of *Orne v. Christie*[42] showcases several of the changes seen in ADA decisions since the ADAAA was passed. While working as Primary Counsel to the Virginia State Corporation Commission's Bureau of Financial Institutions, Orne was informed he had been sleeping on the job, which he attributed to having difficulty sleeping at night. Orne went to a sleep disorders specialist, was diagnosed with OSA, and was prescribed CPAP. This treatment proved effective in treating Orne's symptoms.

Orne made his employer aware of the diagnosis and treatment plan, yet, after a few months, he was told that he would have to accept a demotion with a pay cut or face termination. The court found that Orne's sleep apnea, and the corresponding sleepiness he experienced during the day, was sufficient to qualify him as disabled under the new ADAAA standards. In addition, the successful CPAP treatment had no impact on the presence of a disability.[42]

SUMMARY

OSA and other types of sleep-related breathing disorders may impair an individual's ability to sleep. This can result in drowsiness during the day and can impair one's ability to work or drive. Although patients have a right to refuse treatment for sleep-related breathing disorders, they may be liable if that decision harms others. For instance, new criminal negligence laws in certain states address impaired driving due to fatigue.

Physicians have obligations when treating a patient with a sleep-related breathing disorder. Physicians who have created a physician-patient relationship must inform patients of the risks of driving while suffering from a sleep-related breathing disorder that may result in severe sleepiness. Employers will be liable for the negligence of their employees, but do not owe any duty to employees for actions after they leave work, even if the employer scheduled the employee to work extremely long hours.

Finally, a 2008 update to the ADA has loosened the definition of a disability for purposes of ADA protections. As a result, sleep-related breathing disorders are more likely to qualify as a disability under the ADAAA language.

Selected Readings

Colvin LJ, Collop N. Commercial motor vehicle driver obstructive sleep apnea screening and treatment in the United States: an update and recommendation. *J Clin Sleep Med* 2015;[Epub ahead of print].

Fouladpour N, Jesudoss R, Bolden N, et al. Perioperative complications in obstructive sleep apnea patients undergoing surgery: a review of the legal literature. *Anesth Analg* 2015;[In press].

Idzikowski C, Rumbold J. Sleep in a legal context: the role of the expert witness. *Med Sci Law* 2015;**55**(3):176–82.

Masullo A, Feola A, Marino V, et al. Sleep disorders and driving licence: the current Italian legislation and medico-legal issues. *Clin Ter* 2014; **165**(5):e368–72.

Rumbold J. Automatism and driving offences. *J Forensic Leg Med* 2013;**20**(7):825–9.

Svider PF, Pashkova AA, Folbe AJ, et al. Obstructive sleep apnea: strategies for minimizing liability and enhancing patient safety. *Otolaryngol Head Neck Surg* 2013;**149**(6):947–53.

A complete reference list can be found online at ExpertConsult.com.

Chapter 67

Legal Aspects of Fatigue- and Safety-Sensitive Professions

Daniel B. Brown; Jeffrey Masor

Chapter Highlights

- Lack of sleep can cause mistakes in the workplace. These mistakes can have tragic consequences.
- Various industries, including maritime, aerospace, medical, military, nuclear, and shipping, have implemented regulations concerning fatigue and sleep to reduce workplace accidents.
- Drowsy driving regulations have been implemented in some states, potentially resulting in legal liability when driving without adequate sleep.

Lack of sleep can inhibit performance and cause mistakes in the workplace. Over the past few decades, high-profile incidents resulting from a lack of sleep have created a public awareness of the dangers of sleep deprivation. In response, Congress, as well as many administrative agencies and professional groups, have weighed the harmful risks of lack of sleep and have crafted safety regulations to protect its members in the workplace and the community as a whole.

HISTORY

Lack of sleep can cause mistakes in the workplace that have tragic repercussions. Historically, many accidents have been attributed to lack of sleep. A partial list follows:

- In 1979, a nuclear reactor located on Three Mile Island in Pennsylvania experienced a partial meltdown due to human error.[1] Shift workers were suffering from lack of sleep and did not notice that a stuck valve was causing the reactor to lose coolant.[2] Eventually, the core overheated and was damaged.[3]
- In January 1986, Space Shuttle Challenger exploded moments after launch and resulted in the deaths of all astronauts onboard.[4] This incident has been partially attributed to lack of sleep because key managers had had only 2 hours of sleep and had been on duty since 1 AM, which resulted in poor decision making.[5]
- In April 1986, the Chernobyl nuclear plant experienced a serious meltdown.[6] This disaster was a result of human error in conditions where employees were fatigued after working more than 13 hours.[7] The aftermath of this event resulted in the most horrific nuclear accident in human history.[8]
- In 1989, the Exxon Valdez ran aground and spilled 10.8 million gallons of crude oil into the ocean.[9] This disaster occurred after the third mate stayed awake for more than 18 hours and, being greatly fatigued, did not properly account for the ship's position.[10] This oil spill dealt a great deal of damage to the surrounding wildlife, and the area's environment has still not fully recovered.[11]
- In 1999, American Airlines Flight 1420 crashed during a severe thunderstorm.[12] While the main cause was due to the weather, both pilots for the aircraft were near the end of their 14-hour duty shift and displayed poor judgment by trying to land in an area with such dangerous conditions.[13]
- In 2013, an engineer operating a Metro-North train fell asleep at the controls, which resulted in the train derailing from the track.[14] Although the engineer had no previous record of a sleep disorder, later investigation found that he had sleep apnea.[15]

LEGAL OBLIGATIONS REGARDING FATIGUE IN NON-SAFETY-SENSITIVE INDUSTRIES

Businesses engaging in retail, entertainment, lodging, and other activities not generally considered to be safety sensitive typically operate without regulatory limits on employees' hours of service. Employers in these fields are free to schedule their employees for as many hours of service as their employee desires and as the employer is willing to pay. An employer's legal risk in these industries for their employees' impaired job performance due to fatigue is far more limited than the risk faced by employers in safety-sensitive industries such as transportation.

For example, courts routinely find employers who over-schedule their employees to be free from liability for injuries caused by fatigued employees at the end of their lengthy shift. The case of *Barclay v. Briscoe*, 47 A.3d 560, 427 Md. 270 (Md., 2012), is instructive. In Barclay, a longshoreman fell asleep at the wheel while traveling home after working a 22-hour shift at his job site located at the Port of Baltimore. The longshoreman crashed head-on into a morning commuter, causing catastrophic injuries to the commuter and the longshoreman's own death. The commuter brought suit against the longshoreman's employer for primary negligence in failing to protect the motoring public from an employee driving after an unreasonably long shift.

Following rules laid down by most American state courts, the Maryland Court of Appeals denied the injured commuter's request for damages. The Court found that the employer Ports did no more than establish the work schedule for the job. It did nothing to affirmatively control whether Richardson drove home in a fatigued state. As such, the Court held that an employer has no duty to a third party who might be injured by a commuting employee based solely on the fact that an employee's fatigue was a foreseeable consequence of the employment.[16] In that regard the Court flatly refused "to fashion some type of judicially-imposed maximum working hours standard across all industries."

INDUSTRY REGULATIONS

Unlike employers in non-safety-sensitive industries, employers whose workers' tasks have the direct ability to affect public safety are subject to legal obligations for the fatigue-impaired actions of their employees. These legal obligations arise in the form of government and industry regulation addressing hours of service and fatigue management. The remainder of this chapter discusses a variety of these regulatory schemes.

Maritime

U.S. Maritime law, codified in 46 U.S.C. § 8104, has set rules to reduce accidents due to fatigue. These rules set limits on how many hours an individual may work per day, while leaving an exception for emergencies.

As stated in § 8104, for vessels leaving port, an officer in charge of deck watch must have been off duty for at least 6 hours of the previous 12 hours.[17] When serving on certain kinds of vessels under 100 gross tons, mariners may not work for more than 9 of 24 hours when in port and more than 12 of 24 hours when at sea.[18] An exception waives these requirements during an emergency.[19] On certain larger vessels over 100 gross tons, mariners on the boat must be split into at least three separate watches, where each watch rotates duty shifts.[20] This allows those not on duty to rest and recuperate. Those on deck or in the engine department may not be required to work for more than 8 hours a day.[21] Again, exceptions can apply in case of an emergency.[22]

Specific rules may apply to certain kinds of vessels. For instance, on towing vessels, mariners may not work for more than 12 hours per day, except in an emergency.[23] Furthermore, on a tanker, a mariner may work up to 15 hours of every 24-hour period or up to 36 hours of a 72-hour period, except for emergencies.[24]

National Aeronautics and Space Administration

The National Aeronautics and Space Administration (NASA) has internal regulations over employees and subcontractors that limit work hours "to ensure personnel safety, protect high value assets and to maintain the quality of on-the-job performance."[25] For missions involving suborbital and certain orbital projects, personnel may work up to 72 hours per 7-day period, unless they get a waiver.[26] Only the Director of suborbital and special orbital projects may issue waivers.[27] Personnel may also only work up to 13 days consecutively without a break.[28] NASA has also set a minimum time off between work periods, between 8 and 24 hours, depending on the length of the previous work period.[29]

Medical Residents

New York was the first state to implement residency work duration limits.[30] Support for these regulations grew as a result of the death of a patient due to improper medical care, believed to have been a result of overworked physicians and residents.[31] New York's work duration limits on medical residents, codified in 10 CRR-NY § 405.4, states that work hours for medical residents with inpatient care responsibilities must not exceed 80 hours per week over a 4-week period.[32] In addition, residents must not be scheduled to work for more than 24 consecutive hours.[33] On-call duties for surgical residents are exempt from these requirements if (1) such shifts have infrequent interruptions limited to patients that the resident has a continuing responsibility, (2) this responsibility occurs no more than every third night, (3) a continuous assignment that includes a night shift on-call is followed by a nonworking period of at least 16 hours, and (4) additional policies are implemented by the hospital to relieve residents after an unusually active on-call period.[34]

In 2003, after New York implemented these work hour restrictions, the Accreditation Council for Graduate Medical Education (ACGME) created a new accreditation requirement, mirroring New York's law, which applies to medical residents in all accredited institutions.[35] In 2011, ACGME issued a revised rule that further restricts resident work hours.[36] Currently, ACGME requires that duty hours be limited to 80 hours per week, averaged over a 4-week period, inclusive of all in-house call activities and moonlighting.[37] A review committee may grant programs a limited exception, up to a maximum of 88 hours per week.[38] Residents must be scheduled for a minimum of 1 free day a week, averaged over 4 weeks.[39] At-home calling cannot be assigned on these free days.[40] In addition, duty periods of residents are restricted to a certain duration per day, with minimum time off requirements between scheduled duty periods.[41]

United States Military

Unlike many other regulations that mandate compliance, regulations issued by the U.S. Army tend to act more as a guidance; however, that is to be expected given the uncertain day-to-day conditions in the field. The Army's training regulations recommend that each trainee receive 7 hours of sleep per night.[42] The Army's *Combat and Operational Stress Control Manual for Leaders and Soldiers in the Field* describes the factors considered by military personnel when scheduling time for sleep.[43] The Army recommends the best time for sleep as between 11 PM and 7 AM because of the body's natural circadian rhythms.[44] It also recommends 7 to 8 hours of sleep per 24-hour period and cautions that any reduction in this amount will degrade performance.[45] Although continuous sleep is preferred, sleep may be divided between two or more shorter periods to obtain the full 7 to 8 hours of sleep.[46] The manual also prioritizes the need to sleep based on the task being performed.[47] Leaders making critical decisions have a top priority for sleep.[48] Soldiers on guard duty, performing tedious tasks, or analyzing information have second priority for sleep.[49] Finally, soldiers performing duties that require only physical labor have third priority for sleep.[50]

Nuclear Power Plants

In light of several high-profile nuclear incidents that were at least partially caused by operator fatigue, the U.S. Nuclear Regulatory Commission has mandated limits on the work hours of nuclear power plant operators. As codified in 10 C.F.R. § 26.205, employees may work up to 16 hours per 24-hour period, 26 hours per 48-hour period, and 72 hours per 7-day period.[51] Exceptions may be granted in limited circumstances.[52] Employees must also be provided rest breaks after each work period.[53] These rest break lengths vary depending on how long the employee had been working.[54] For instance, employees must receive at least a 10-hour break after working for at least 10 hours.[55]

In addition, employees must receive a minimum number of days off, depending on their work schedule.[56] For example, employees working 10-hour shifts must have, on average, at least 2 days off of work per week.[57] In the event of an unscheduled preparedness drill, the duration of an employee's unscheduled participation in this drill is not calculated into the hours worked that day.[58]

Railroads

With passage of the Rail Safety Improvement Act of 2008, Congress directed certain railroads to develop and implement a railroad safety risk reduction program.[59] By statute, the program must include a fatigue management plan designed to reduce fatigue experienced by railroad employees and to reduce the likelihood of accidents, incidents, injuries, and fatalities caused by fatigue.[60] This fatigue management plan must consider physiologic and human factors that affect fatigue and promote appropriate fatigue countermeasures to address safety concerns, such as scheduling practices, napping policies, and avoidance of abrupt changes in rest cycles for employees.[61]

Existing regulations address hours of service limitations for railroad employees as follows. These regulations, codified in 49 U.S.C. § 21103, provide that employees may spend up to 276 hours per month working for their carrier.[62] In addition, employees are limited to no more than 12 consecutive hours of work.[63] Each employee must receive at least 10 consecutive hours of off-duty time per 24 hours.[64] Employees must not work more than 6 or 7 consecutive days, depending on how much time the employee had off before that work duration.[65] Exceptions may apply in case of emergency; however, even then, employees are limited to 4 additional hours per 24-hour period.[66]

Trucking

Drowsiness among truck drivers is a highly prevalent risk in the trucking industry. For instance, a Federal Motor Carrier Safety Administration–sponsored study found that 45% of U.S. truck drivers sometimes or often had difficulty staying awake while driving.[67] In addition, a number of high-profile accidents have involved truck drivers who fell asleep at the wheel,[68] and a 5-day study found that truck drivers tended to average 4.78 hours of sleep per day with only 5.18 hours in bed per day.[69] As a result, regulations limit the number of hours a truck driver may be on the road.

Trucking regulations cover commercial motor vehicles (CMVs). A CMV vehicle is one that either (1) weighs 10,001 pounds or more; (2) is transporting hazardous materials such that it requires a placard; (3) is designed or used to transport 16 or more passengers, including the driver, not for compensation; or (4) is designed or used to transport 9 or more passengers, including the driver, for compensation.[70]

Solo truck drivers that transport property may drive for a maximum of 11 hours after 10 consecutive hours off duty.[71] At the same time, a driver cannot be on duty for more than 14 consecutive hours.[72] Drivers may not drive more than 60 hours per 7-day period or 70 hours per 8-day period.[73] However, truck drivers who reach this hour limit may rest for 34 consecutive hours that include two nights from 1 AM to 5 AM, and then they may resume driving.[74] This "restart" may be used up to once per 168-hour period.[75] Truck drivers are also required to take a 30-minute break during the first 8 hours of a shift.[76]

The U.S. Department of Transportation (DOT) is also working toward requiring drivers to undergo sleep apnea testing. Although the DOT gained the congressional authority to require this testing in October 2013, a lengthy, formal rule-making proceeding must first be completed before this requirement may take effect.[77] A formal rule-making proceeding is a process whereby the administrative agency researches the costs and benefits of a new rule and allows for public comment from the industry.[78] Some applauded this decision, arguing that testing and treatment of obstructive sleep apnea could cost the industry more than $1 billion; the government should take its time to ensure that proper analysis is completed and that suitable regulation language is chosen.[79] Others see the rule-making process as an unnecessary delay in a much-needed rule.[80] These groups worry that the rule-making process will take several years and that new sleep testing requirements may never be implemented.[81]

Airlines

Airline pilots are restricted from working excessive hours by the Federal Aviation Administration (FAA). Pilots of a one- or two-pilot crew may fly for up to 500 hours per calendar quarter, 800 hours per two consecutive calendar quarters, and 1400 hours per calendar year.[82] For a 24-hour period, work duration limits vary depending on whether it is a one- or two-pilot crew.[83] For a one-pilot crew, the total flight time cannot exceed 8 hours.[84] For a two-pilot crew, the total flight time cannot exceed 10 hours.[85] After a duty period, a varying amount of rest is required depending on whether the destination crosses time zones and whether the flight was extended because of unforeseen factors, such as bad weather.[86] For instance, a multiple–time zone flight that had no unforeseen extensions of flight time would require that the pilot rest for at least 14 hours.[87]

Currently, the FAA requires that all pilots diagnosed with OSA be treated in order to be fit to fly an aircraft.[88] Like the DOT, the FAA is also working toward requiring sleep testing.[89] Although Congress gave the FAA the authority to implement this new testing requirement, Congress also required that the FAA complete a formal rule-making process.[90] The FAA initially planned to limit this requirement to pilots with a body mass index of more than 40.[91] Over time, the FAA intended to reduce the requirements for mandatory sleep testing until all pilots had been tested.[92] However, this rule has undergone several revisions in its quest to become a new mandatory requirement.[93] For instance, on March 28, 2014, the FAA proposed rule had been changed to require

pilots to undergo testing based on a recommendation by an Aviation Medical Examiner who considers all obstructive sleep apnea risk factors.[94]

DROWSY DRIVING

Drowsy driving can greatly increase the risk for a collision. For instance, a study found that the risk for a car crash nearly tripled when drivers had less than 5 hours of sleep in a 24-hour period.[95] Drowsiness has also been compared with alcohol consumption. In one study, individuals who had been awake for 24 hours performed like an individual with a blood alcohol concentration of 0.10%.[96] For comparison, the legal blood alcohol concentration limit in the U.S. is 0.08%.[97]

In addition to regulating professionals' sleep, some states have amended laws to factor sleep deprivation into their definition of criminally reckless driving. The first state to do so was New Jersey, which enacted legislation referred to as "Maggie's Law" as part of the state's criminal code in 1997, following the death of college student Margaret "Maggie" McConnell.[98] A driver, who had not slept for 30 hours before the crash and had smoked crack cocaine hours before the accident, hit McConnell.[99] During litigation, the judge refused to admit evidence of the driver's sleep deprivation to establish reckless behavior, and the driver received only a $200 penalty for causing the accident.[100]

Maggie's Law is an evidentiary rule establishing that proof of driving after 24 hours of sleeplessness "*may* give rise to an inference that the defendant was driving recklessly" *(emphasis added)*.[101] The law also establishes that falling asleep while driving may infer recklessness without regard to sleeplessness.[102] Although it may prove difficult to establish that a driver was asleep at the wheel at the time of the collision, or had been awake for the entirety of the 24-hour statutory period, Maggie's Law establishes that drowsy driving, if proven, can uphold a criminal conviction in the State of New Jersey.

In 2013, Arkansas became the second state to pass a drowsy driving amendment to its negligent homicide statute, making it a class B felony to negligently cause the death of another person as a result of operating a vehicle, aircraft, or watercraft while fatigued.[103] *Fatigued* is defined as "having been without sleep for a period of twenty-four (24) consecutive hours," or "having been without sleep for a period of twenty-four (24) consecutive hours and in the state of being asleep."[104]

Several states, including Massachusetts, Tennessee, New York, and Oregon, have attempted to pass statutes similar to those found in New Jersey and Arkansas that would criminalize drowsy driving.[105] Although these attempts have proved unsuccessful thus far, high-profile events, such as the sleep-deprived Walmart truck driver who crashed into comedian Tracy Morgan's limousine in 2014,[106] frequently raise public awareness of the issue and draw significant attention of lawmakers.

CLINICAL PEARLS

- As a general matter, in situations after work hours, employers do not have a legal duty to protect the motoring public from the drowsy driving of employees whose fatigue may have been caused by the employer overscheduling the employee's work shifts.
- On the other hand, employers are liable under law for the negligent acts of their employees, such as vehicular accidents caused by impaired driving because of fatigue, if the accident occurred during the employee's course of employment.
- To protect the public, the U.S. government imposes hours of service regulations on many safety-sensitive industries, such as commercial trucking, nuclear power, and commercial airlines.

SUMMARY

Lack of sleep can cause mistakes in the workplace and while driving, leading to tragic consequences. Historically, many high-profile accidents have been at least partially attributed to lack of sleep. Various industries, including maritime, aerospace, medical, military, nuclear, and shipping, have implemented regulations concerning fatigue and sleep in order to reduce workplace accidents. For instance, the ACGME requires that medical residents limit their duty hours in order for a medical institution to maintain its accreditation. Some states have also implemented drowsy driving regulations, which result in criminal liability for a driver who causes a collision after remaining conscious for at least 24 hours.

Selected Readings

Mansfield D, Kryger M. Regulating danger on the highways: hours of service regulations. *Sleep Health* 2015. [In press].

National Sleep Foundation. Available at: <www.DrowsyDriving.org>.

Rasleara TG, Gertlerb J, DiFioreb A. Work schedules, sleep, fatigue, and accidents in the US railroad industry. *Fatigue Biomed Health Behav* 2013;**1**(1–2):99–115.

Williamson A, Friswell R. Fatigue in the workplace: causes and countermeasures. *Fatigue Biomed Health Behav* 2013;**1**(1–2):81–98.

A complete reference list can be found online at ExpertConsult.com.

Chapter
68

Sleep Medicine Clinical Practice and Compliance—United States

Daniel B. Brown

Chapter Highlights

- The delivery of sleep medicine and the treatment of sleep disorders in the United States are highly fragmented. Government and most commercial insurance payers cover sleep medicine, sleep testing, and sleep therapy under separate reimbursement categories with separate requirements for licensure, accreditation, personnel certification, and coverage conditions. Legal restrictions on patient self-referrals can limit, in certain cases, the ability of treating physicians to dispense continuous positive airway pressure (CPAP) or other items of durable medical equipment to treat their own patients' sleep disease.
- In-laboratory and home sleep testing for obstructive sleep apnea performed by physicians and stand-alone sleep centers are reimbursed by Medicare, Medicaid, TRICARE, and most commercial insurance payers. Polysomnography and home sleep testing (HST) for many other sleep disorders, such as insomnia, restless legs syndrome, and narcolepsy, are not typically covered by government or private insurance.
- The government views polysomnography, HST, and CPAP treatment as an activity ripe for fraud and abuse. Accordingly, government health care programs condition reimbursement for sleep medicine services to a wide array of technical requirements, including, but not limited to, a face-to-face examination by a treating physician whose notes must include certain specified notations, testing by laboratories whose programs are accredited by select agencies and who use only certified sleep technicians, test interpretations by physicians who hold certain board certifications, and CPAP dispensing only by durable medical equipment suppliers who, in the case of home sleep tests, are not affiliated with the provider of the home sleep test.

Like many chronic conditions, detection and treatment of obstructive sleep apnea (OSA) and other sleep disorders triggers a complex range of clinical activity. Patient care plans for sleep-disordered patients include medical examinations, overnight testing, ongoing respiratory or dental appliance treatment with durable medical equipment, monitoring, and follow-up. This collection of care services touches a wide variety of legal, regulatory, and reimbursement rules, few of which operate in concert to promote seamless delivery of clinical sleep medicine.

This chapter addresses some of the legal, regulatory, and reimbursement roadblocks sleep practitioners in the United States face in offering their services to their sleep-disordered patients.

STATE LICENSING OF SLEEP SERVICES

State Licensing for Sleep Laboratories

Many states allocate health care resources within their borders through the Certificate of Need (CON) process.[1] Obtaining a CON for the operation of a health care facility is a lengthy, expensive, and often adversarial process. Fortunately, sleep laboratories will typically fall below the threshold for CON review and escape review requirements in CON states.[2]

Separate from CON review are state laws requiring health care facilities to obtain licensure before operation. Although diagnostic testing performed as part of a hospital or physician practice is almost always exempt from state health care facility licensure, freestanding sleep laboratories may fall within a state's regulatory jurisdiction as a "health care clinic" or other health care facility.[3] Freestanding laboratories are sleep laboratories operated separately from physician practices or hospitals.[4]

At least three states now require freestanding sleep laboratories to obtain health facility licensure before operating. These states are Florida,[5] New Jersey,[6] and Alabama.[7] Licensure requires completing an application and disclosing the laboratory's ownership structure, medical supervision, and affiliations with other health care facilities.[8] Interestingly, these states require facility licensure even if the entity performs only home or portable sleep tests in the absence of any physical health care structure visited by patients. Other states may join in regulating the licensing of freestanding sleep laboratories.

Sleep Physician Certification

Physician certification in the specialty of sleep medicine is a fairly recent phenomenon. The American Board of Sleep Medicine (ABSM) began issuing diplomate status in Sleep Medicine to physicians and PhDs in the late 1970s. After 2007, the American Board of Medical Specialties (ABMS) began issuing subspecialty board certification to physicians in

sleep medicine as part of the boards of internal medicine, family medicine, anesthesiology, otolaryngology, pediatrics, and psychiatry and neurology.[9]

As discussed later, government and commercial payers often require that the physicians who interpret polysomnograms or home sleep tests hold ABSM, ABMS, or certain other credentials as a condition for reimbursement. However, one state—Oklahoma—has adopted legislation that prohibits non–board-certified physicians from interpreting sleep tests without regard to the source of payment for the test.

Believing that public safety requires specific regulation of sleep disorders testing, the Oklahoma legislature in 2009 adopted the Oklahoma Sleep Diagnostic Testing Regulation Act.[10] The act establishes minimum medical standards of care in the field of sleep medicine. It declares that it is illegal in Oklahoma for a licensed physician to interpret a sleep test unless the physician is board-certified in sleep medicine by the ABSM or ABMS. Alternatively, the interpreting physician must have completed a 1-year sleep medicine fellowship accredited by the Accreditation Council for Graduate Medical Education or received a Certification of Special Qualifications or a Certification of Added Qualifications in Sleep Medicine issued by the American Osteopathic Association. Otherwise, it is illegal for a medical doctor in Oklahoma to interpret the results of a diagnostic sleep test performed in Oklahoma.[10]

Oklahoma also establishes minimum requirements for sleep testing facilities and their employees.[10] Under the Act, (1) sleep diagnostic testing facilities in Oklahoma must be fully or provisionally certified or accredited as a sleep laboratory by the American Academy of Sleep Medicine (AASM), the Accreditation Commission for Health Care (ACHC), or The Joint Commission; (2) Oklahoma sleep laboratories must be supervised by board-certified sleep physicians; and (3) home sleep testing must be performed by technicians who are supervised by board-certified sleep physicians.[10] In this regard, the Oklahoma law is the most intrusive in the nation.

State Licensing for Nonphysician Polysomnography Technicians

Nonphysician persons who perform and score sleep tests have specialized skills and training. These individuals have enhanced their professional skills and standing by earning certification through organizations such as the Board of Registered Polysomnographic Technologists[11] and the American Association of Sleep Technologists.[12] These credentialing entities generally recognize three levels of sleep technician expertise: trainee, technician, and technologist.[13]

Sleep laboratories are encouraged to use technicians holding Board of Registered Polysomnographic Technologists or American Association of Sleep Technologists credentials. Some payers require the participation of certified technicians as a condition to payment. For example, Medicare will not consider a sleep test performed by an independent diagnostic testing facility to be medically necessary unless the technician attending the overnight test or handling the home sleep test is properly credentialed.

Quite apart from bodies that credential sleep technicians are a handful of state medical boards that impose licensure requirements on sleep technicians operating in their state. For example, the states of Idaho,[14] Tennessee,[15] New York,[16] and California,[17] among several others, have recognized the allied health profession of polysomnography (PSG) and adopted mandatory licensure requirements for these nonphysician sleep test personnel. In New York, for example, licensure of PSG technologists requires applicants to be at least 18 years old, complete educational requirements that provide knowledge of PSG technology, pass an examination, and prove "good moral character."[18] Sleep laboratories that fail to use licensed sleep technicians, when required, face potential state law criminal or civil penalties.[19]

Because CPAP titration performed during PSG borders on respiratory therapy, some state respiratory therapy boards have brought disciplinary actions against some sleep laboratories for the unlicensed practice of respiratory therapy.[20] States that have adopted PSG technology laws typically exempt licensed PSG technicians from any requirement to hold a respiratory therapy license.[20]

State Licensing for Dispensing Continuous Positive Airway Pressure

Because CPAP is a prescriptive device under the U.S. Food and Drug Administration rules,[21] CPAP is considered a medical device for most state law regulatory purposes. CPAP is also an item of durable medical equipment (DME) for purposes of state law[22] and for insurance reimbursement purposes.[23]

Roughly half of states, including Florida,[24] Tennessee,[25] and Ohio,[26] require persons or entities that dispense DME, like CPAP, to obtain a state license before delivery of DME items. Almost all states that license DME suppliers exempt physicians or other licensed health care practitioners, who dispense devices from their practices to their patients, from licensure.

Notwithstanding licensure of the CPAP supplier, most states consider CPAP titration and education to be the practice of respiratory therapy to be performed only by persons holding respiratory therapy licenses issued by the applicable state.[27] States that have licensed or regulated the activities of PSG technicians have almost universally exempted PSG technicians from also obtaining a respiratory therapy license.[28]

REIMBUREMENT FOR SLEEP MEDICINE SERVICES

Reimbursement drives the delivery of health care services in America. The provision of sleep medicine and the types of testing and treatment depend, in large part, on the coverage conditions that health insurance plans impose on the providers of sleep disorders services.

Medicare Coverage for Sleep Testing

The Medicare Part B Program covers in-laboratory and home sleep testing performed by physicians or stand-alone independent diagnostic testing facilities as "other diagnostic tests" payable under the under Medicare's Physician Fee Schedule.[29] Medicare reimburses hospitals for hospital sleep tests under Medicare's Outpatient Prospective Payment System.[30] Services under this system are classified into groups called ambulatory payment classifications (APCs), with payment rates established for all activities assigned to each classification group.[31] For example, home sleep tests are payable under APC group 213,[32] and attended PSGs are in the APC group 209.[32]

Medicare coverage for all diagnostic tests requires a physician's order and proper physician supervision.[33] The Centers for Medicare and Medicaid Services (CMS) has delegated to

Medicare Administrative Contractors (MACs) the authority to specify special medical necessity coverage conditions for activities in the MAC's jurisdiction.[34] Some MACs, such as are Cahaba Government Benefits Administrators, LLC, have not adopted local coverage determinations (LCDs) describing special conditions for sleep test coverage in Cahaba's oversight area of Alabama, Georgia, and Tennessee. Thus sleep laboratories seeking Medicare reimbursement for sleep tests in those states need not hold any sleep laboratory accreditations or, other than independent diagnostic testing facility sleep laboratories, engage only certified technicians.

On the other hand, the Part B MAC for Florida, First Coast Service Options, Inc., and many other MACs have published LCDs for sleep testing in their jurisdictions. First Coast's LCD, L29949, is quite restrictive.[35] For example, in Florida, a sleep test will be medically necessary and payable by Medicare only if (1) the laboratory is accredited by the AASM, ACHC, or The Joint Commission; (2) the physician interpreting the test is board-certified or meets certain other criteria; and, among other items (3) the technicians performing the test, even the home sleep test, holds certain certifications or evidence of training.[35] First Coast further requires that each patient have a face-to-face examination with a physician before testing.[35]

The notes of the physician visit covered under the First Coast Local Coverage Determination must contain a sleep history and physical examination, an Epworth sleepiness scale, notations of the patient's body mass index and neck circumference, and a focused cardiopulmonary and upper airway evaluation.[35] First Coast takes the position that sleep tests for Medicare patients whose initial examination notes do not show the measurement of neck size circumference, body mass index, or Epworth scores are not medically necessary and are ineligible for Medicare coverage, even if all other indications for OSA are present and made clear in the examination notes.[35]

Medicare Coverage for CPAP Therapy

Because Medicare categorizes CPAP devices, CPAP masks, oral appliances, and items of CPAP resupply as DME, Medicare coverage for such items derives from a Medicare fee schedule and system separate from Medicare coverage available for sleep examinations and sleep testing.[36] This bifurcation means that Medicare reimbursement for OSA treatment devices follows its own separate rules and coverage conditions.

For example, CPAP suppliers who enroll in the Medicare system must adhere to a complex set of 30 or so supplier standards, including minimum space, insurance coverage and hours of operation requirements, possession of accreditation from selected accrediting bodies, and the posting of a $50,000 surety bond to offset losses due to improper practices.[37] Medicare enrollment and payments are governed by regional DME Medicare Administrative Contractors, each of which has published a similar LCD indicating the conditions of coverage for CPAP reimbursement in the contractor's jurisdiction.[38] According to these LCDs, Medicare will reimburse a DME supplier for CPAP to treat OSA only in the following circumstances:

1. The beneficiary has a face-to-face clinical evaluation by the treating physician before the sleep test to assess the beneficiary for OSA.
2. The beneficiary has a sleep test showing minimum criteria for OSA.
3. The sleep test is performed by a test provider who meets Medicare coverage for sleep testing.
4. The sleep test is interpreted by a boarded sleep physician or a physician on staff of a sleep center accredited by the AASM, ACHC, or The Joint Commission.
5. The beneficiary shows continued use of the device and that he or she benefits from the device within 90 days of initiation by subjective and objective evidence.[38]

In 2008, CMS stated that it believes that the sleep test provider has a self-interest in the result of that test if the provider, or its affiliate, also supplies the CPAP device. To CMS, affiliations between the test provider and the CPAP supplier provided an incentive to test more frequently or less frequently than is medically necessary and to interpret a test result with a bias that favors self-interest.[39]

In an effort to curtail these perceived abuses, CMS adopted a special payment prohibition in 2008 that prohibits Medicare from paying a DME supplier for CPAP if the CPAP supplier is directly or indirectly affiliated with the provider of the sleep test from which a diagnosis of OSA is obtained.[40] *Affiliation* means a relationship among parties by compensation arrangement or ownership.[41] This special payment prohibition only applies if the underlying sleep test is a home sleep test. In other words, Medicare will reimburse a DME supplier for CPAP even if the DME supplier is affiliated with the provider of the sleep test, as long as full, in-laboratory PSG was used to diagnose the patient's OSA.[40]

Medicaid Coverage

Medicaid is a state-run health care program subsidized with federal funds but operated under terms and conditions adopted by individual states for their citizens. A comprehensive 50-state survey of Medicaid coverage for sleep testing is beyond the scope of this chapter. However, the coverage conditions for sleep tests under the State of Washington's "Apple Health" Medicaid program are illustrative.

Apple Health will reimburse participating providers only for the performance of full, in-laboratory PSG or a Multiple Sleep Latency Test.[42] Home sleep testing is not a covered service under Apple Health at this time.[42] The PSG and Multiple Sleep Latency Test are covered only if ordered by the client's physician and performed in an agency-designated center of excellence that is an independent diagnostic testing facility, sleep laboratory, or outpatient hospital.[42] Providers obtain eligible center of excellence status if they hold sleep laboratory accreditation from the AASM and at least one staff physician is board-certified in sleep medicine.[43]

Commercial Insurance Coverage of Sleep Testing

Almost all commercial health insurance plans cover testing for OSA. The conditions for coverage vary widely according to the terms of the individual plan. Although commercial payers cover the attended, in-laboratory PSG as the gold standard method of detecting OSA, these payers are now promoting the less expensive home sleep test for their members.[44] Large health plans do this by requiring plan members to obtain prior authorization for the more expensive in-laboratory PSG.[45] Absent a showing that an in-laboratory PSG is medically necessary, the payers will cover only the less expensive home sleep test for purposes of detecting OSA.[45]

FRAUD AND ABUSE LAWS

Health care fraud and abuse laws can be separated into two separate conceptual categories. One category deals with prohibited self-referrals. These laws prohibit physicians from referring their patients to entities in which the physicians or their families have an ownership or other financial interest. On the federal level, the major regulation in this area is the Stark Law.

The other broad category is anti-kickback laws. These prohibit any person (not just doctors) from paying or receiving money, or other items of value, for the referral of a health care service.[46] The federal anti-kickback law extends beyond payments for referrals alone.[46] This law also covers payments, gifts, or other valuable items to anyone who merely recommends or even arranges health care services reimbursed by Medicare, Medicaid, or any other federal health care program.[46]

The Stark Self-Referral Law

Absent an exception, the federal Stark Law prohibits a physician (or an immediate family member of such physician) from making a referral for a designated health service to an entity in which the physician has a direct or indirect ownership or compensation arrangement—if the service is reimbursed by Medicare or Medicaid.[47]

Penalties for violating the Stark Law include denial of payment, refunds of amounts collected in violation of the statute, and a civil money penalty of up to $15,000 for each bill or claim for a service a person knows or should know is for a service for which payment may not be made.[48] If the physician or the entity engages in a circumvention arrangement that the physician or entity knows, or should know, has a principal purpose of indirectly evading the Stark Law, the civil money penalty jumps to $100,000 for each such arrangement or scheme.[49]

Only referrals for designated health services are prohibited. Overnight PSG or home sleep testing, generally characterized as a type of electroencephalogram, does not fall within any of the categories of designated health services.[50] Consequently, a physician's referral of a Medicare or Medicaid sleep test falls *outside* the Stark Law prohibition, unless the sleep test is performed in a hospital setting. Inpatient and outpatient hospital services are designated health services, and a physician's referral of a Medicare or Medicaid sleep test performed as an inpatient or outpatient hospital service would be considered a referral of a designated health service. However, a violation would occur only if the referring physician had an ownership or compensation arrangement with the hospital that did not meet a Stark Law exception.

Self-Referrals for Durable Medical Equipment

Unlike referrals for nonhospital sleep tests, referrals for DME *are* referrals for designated health services under the Stark Law.[51] Because CPAP and oral appliances are items of DME, the entirety of the Stark Law applies to these items of OSA treatment. This means that, absent an exception, a physician may not refer a patient to a DME supplier in which the physician or a member of the physician's family has an investment or compensation interest, if the DME supplier requests payment of the item from a government health care program.

The Stark Law exempts referrals of certain items of DHS that are ancillary to the physician's in-office procedures, such as imaging or prescription drug services. This in-office ancillary services exception permits physicians to refer items of DHS to be furnished by the physician's own practice if certain conditions are met.[52]

Unfortunately, items of DME such as CPAP or oral appliances are not included in the in-office ancillary services exception.[52] This means that a physician may not furnish CPAP to his or her own OSA patient from the physician's office if the patient is a Medicare or Medicaid beneficiary, unless another Stark Law exception exists, such as the rural provider or personally performed exceptions.

State Self-Referral Laws

Many states have their own laws restricting self-referrals. Some states, such as Michigan, adopt language very close to the federal Stark Law.[53] Therefore an exception under the federal law is likely an automatic exception under the state law.[53] Other states, such as Georgia, use their own definitions and exceptions.[54]

Importantly, these states punish self-referrals regardless of the reimbursement of the item or service by a federal or state health care program.[54] Referrals of even private-pay patients to entities owned by the physician or in which the physician has a compensation arrangement not covered by an exception may be unlawful under these laws.

Federal Anti-Kickback Statute

The federal Anti-Kickback Statute makes it unlawful for anyone to knowingly and willfully solicit or receive any payment in return for referring an individual to another person or entity for the furnishing, or arranging for the furnishing, of any item or service that may be paid in whole or in part by any federally funded health care program.[55]

Violations of the law require the solicitation, offer, payment, or acceptance of illegal remuneration.[55] Remuneration includes the transfer of anything of value in cash or in kind (e.g., goods), whether made directly or indirectly, and whether made overtly or covertly.[55,56]

The statute is a two-way street. Soliciting for and accepting payments in return for referrals is as bad as paying the kickback itself. A violation of the Anti-Kickback Statute constitutes a felony punishable by a maximum fine of $25,000, imprisonment for up to 5 years, or both.[55] Conviction will also lead to automatic exclusion from Medicare, Medicaid, and other federally funded health care programs.[57]

Civil monetary penalties may be applied to violations of the Anti-Kickback Statute in the amount of $50,000 for each act that violates the statute, plus three times the amount of remuneration unlawfully transferred.[58] Significantly, the imposition of civil monetary penalties requires proof by only a preponderance of the evidence, not proof beyond a reasonable doubt, which is required for criminal penalties.[59] Consequently, it is easier for the government to establish violations in a civil proceeding than in a criminal proceeding. The addition of civil monetary penalties increases the risk associated with practices that implicate the statute but are not protected by statutory exceptions or regulatory safe harbors, discussed later.

Anti-Kickback Exceptions and Safe Harbors

The Anti-Kickback Statue has the breadth to capture almost every health care transaction in the United States.[55] At risk

are not only envelopes stuffed with cash but also a wide array of negotiated business practices—sales commissions, below-market rent, distributions arising from joint ventures with suppliers, expensive gifts, medical director fees, and certain equipment rental arrangements. It is these arrangements that the government or its contractors may review as part of its audit of sleep test services.[60]

Responding to industry concerns, Congress has included several exceptions in the law and approved the promulgation of specific "safe harbor" payment practices.[61] Examples of safe harbor arrangements common in sleep medicine include joint ventures among referring physicians and sleep test providers, medical director and other personal services agreements, and space lease arrangements. Compliance with all aspects of the applicable safe harbor protects the actor from prosecution under the Anti-Kickback Statute. However, failure to meet each of the elements of the applicable safe harbor does not automatically mean that the activity is illegal. The activity may be acceptable in the eyes of the government, depending on a variety of factors.

State Anti-Kickback Laws

Many states have enacted their own anti-kickback laws. Almost all state physician-licensing boards prohibit paying or sharing fees for referrals. Some, like California and Florida, criminalize payments or in-kind exchanges to refer patients.[62,63]

The Florida Patient Brokering Act makes it illegal for any person, including any health care provider or health care facility, to (1) offer or pay any commission, bonus, rebate, kickback, or bribe, directly or indirectly, in cash or in kind, or engage in any split-fee arrangement, in any form whatsoever, to induce the referral of patients or patronage from a health care provider or health care facility; (2) solicit or receive any commission, bonus, rebate, kickback, or bribe, directly or indirectly, in cash or in kind, or engage in any split-fee arrangement, in any form whatsoever, in return for referring patients or patronage to a health care provider or health care facility; or (3) aid, abet, advise, or otherwise participate in such conduct.[63] The act specifically covers the actions of attorneys and other advisors and participants who counsel persons in the participation of such arrangements.[63]

Federal False Claims Act

Presenting a claim to CMS for payment for an item or service performed in violation of the Stark Law or Anti-Kickback Statute may constitute a false claim under the federal False Claims Act. Enacted during the Civil War to deter war profiteers,[64] the False Claims Act permits private "whistleblowers" to bring actions against health care companies for filing fraudulent claims with the government.[65] Penalties include repayment of the fraudulent claim and a mandatory civil penalty of at least $5500 and no more than $11,000 per claim, which amounts can be tripled.[66] The whistleblower gets to keep a percentage of the damages and penalties, which could be as much as 30% in some circumstances.[66] Because the penalties could reach upward of $30,000 per sleep study performed, potential recoveries under the law add up quickly.

The risk to sleep laboratory operators under false claims whistleblower statutes has increased in recent years. A wave of states have adopted state False Claims Acts that mirror the federal False Claims Act in many respects.[67]

Sleep laboratories and sleep physicians participating in government reimbursement programs are well advised to adopt compliance plans, or take other audit and monitoring actions, to ensure compliance with fraud and abuse laws.

SUMMARY

Although sleep testing and CPAP devices may be reimbursed under Medicare, Medicaid, and private payers, many wide-reaching prohibitions on reimbursement fraud and abuse have been established. On the federal level, the Anti-Kickback Statute prohibits medical practices from offering or receiving any direct or indirect remuneration to encourage the referral of patients. In addition, the Stark Law prohibits physicians from referring designated health services to an entity owned by themselves, an immediate family member, or to someone with whom the physician has a financial relationship, unless an exception applies. Furthermore, the False Claims Act prohibits medical providers from fraudulently billing for medical services from government payers. States may also have their own fraud and abuse legislation. To provide these services, some states require licensing for technologists performing sleep testing. Oklahoma has even more stringent rules that require physicians to possess specialized credentials to interpret sleep tests.

Selected Readings

Centers for Medicare and Medicaid Services. *Medicare fraud and abuse: prevention, detection, and reporting*. August 2014. http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/Fraud_and_Abuse.pdf. Accessed January 10, 2016.

Office of Inspector General. *A roadmap for new physicians: fraud and abuse laws*. http://oig.hhs.gov/compliance/physician-education/01laws.asp. Accessed January 10, 2016.

Staman J. *Health care fraud and abuse laws affecting Medicare and Medicaid: an overview. (Congressional. Report No. RS22743)*. Washington DC: Library of Congressional Research Service, September 8, 2014. https://www.fas.org/sgp/crs/misc/RS22743.pdf. Accessed January 10, 2016.

A complete reference list can be found online at ExpertConsult.com.

Sleep Medicine Clinical Practice and Compliance—Europe

Thomas Penzel

Chapter 69

Chapter Highlights

- European sleep medicine is coordinated through the European Sleep Research Society (ESRS). The ESRS is composed, in part, of the Assembly of National Sleep Societies with delegates from all European Sleep societies. The levels of sleep medicine and sleep medicine delivery differ in European countries according to the health care systems in the various countries.
- The ESRS has published European guidelines for certification of sleep physicians, psychologists, scientists, and sleep technologists. Examinations for European somnologists started in 2012.
- European sleep center certification guidelines have been published, and a network of research centers has been created. Harmonized guidelines for sleep medicine center accreditation in Europe are being considered by various European national sleep societies with support from the ESRS.

Sleep medicine research in Europe has a long tradition with a strong focus on basic science. A variety of medical and educational institutions across Europe long engaged in sleep research resulted in pioneering insights in the past century. In 1972, a small group of individual sleep researchers and clinicians sought a forum to exchange and advance scientific ideas relating to sleep research and founded the European Sleep Research Society (ESRS or Society) in Switzerland.[1] Since then the Society has become the pre-eminent aggregator and distributor of sleep research in Europe. Its goals are to promote sleep research in Europe, to improve care for patients with sleep disorders, and to disseminate information regarding sleep research. The ESRS is a founding member of the World Federation of Sleep Research Societies. In 1992 the ESRS founded the *Journal of Sleep Research,* and the Society holds biannual conferences in different cities all over Europe.

Unlike sleep research, the development and delivery of sleep medicine for European patients suffering from sleep disorders have evolved independently along national lines. This is because the health care systems in Europe differ much among the countries. In the beginning European sleep medicine was closely related to clinical research, which the ESRS exchanged among pioneering sleep medicine groups in various European countries.

Today there are large national sleep societies with a few thousand members in those European countries that have large populations. Many of these national sleep societies outnumber the membership of the ESRS. However, many European countries with smaller populations do not have reasonably sized national sleep societies and do not have the infrastructure to run professional societies. For these countries the ESRS is of utmost importance to provide support for national sleep medicine needs and services.

In 1994 the ESRS set up a clinical committee to exchange national experiences regarding sleep medicine services in Europe. The committee also sought to coordinate activities among the different insurance and reimbursement schemes across Europe. Additional goals of the committee included the development of standard practice papers and support for the educational exchange of sleep medicine clinicians across Europe.

The Assembly of National Sleep Societies (ANSS or Assembly) grew out of this committee's work. The Assembly is a membership organization comprising approximately 30 national European sleep societies. The association is devoted to the clinical needs of sleep physicians and is made up of national delegates who exchange ideas and concepts to serve sleep-disordered patients and promote sleep medicine in Europe. The ANSS exists under the auspices of the ESRS.

In 2004 the ESRS board met with the national society presidents as representatives of the ANSS for the first time. The Assembly agreed that a number of unifying policies and procedures addressing the delivery of sleep medicine in Europe would be desirable. Acting through task forces, the ANSS and ESRS developed uniform standards for European sleep center accreditation in 2006,[2] certification of sleep professionals in 2009,[3] clinical procedures for adults in accredited sleep centers,[4] and clinical education standards in 2014.[5] Each of these standards and procedures has been published in the *Journal of Sleep Research*.

To address the patchwork standards of European countries that did not adopt ESRS protocols, the ESRS established the Sleep Medicine Committee (SMC) of the ESRS in 2010. The SMC is charged with adopting and promoting (1) standards of practice papers and guidelines for clinical service, (2) certification of sleep medicine professionals (physicians,

psychologists, and other scientists, technicians), and (3) accreditation of sleep medicine centers.

The SMC provides certification and accreditation to sleep professionals and sleep centers in countries that do not have access to national certification and accreditation or that desire to supplement national accreditation with the ESRS emblem. Working with national delegates at the ANSS, the SMC works on educational courses together with other groups of the ESRS to provide educational opportunities. The SMC acts as an educational clearinghouse to help standardize sleep medicine education throughout Europe.

STANDARDS

Sleep Technicians and Technologists

As elsewhere, performance of overnight polysomnography (PSG) in Europe contemplates the participation by nonphysician technicians and technologists to set up and attend the overnight test. The involvement of trained sleep technicians is part of the ESRS sleep center accreditation standards. Several European countries recognize the separate allied health profession of PSG Technician or Technologist. Several European countries have established their own sleep technician societies, and the independent European Society of Sleep technologists (ESST) was formed in 1996. The ESST meets regularly together with the biannual ESRS conferences and works closely together with the ESRS in terms of education and certification.

Sleep Center Accreditation

The ESRS is also working toward an accreditation of European sleep centers. In 2006, the Steering Committee of the ESRS published European Guidelines for the Accreditation of Sleep Medicine Centres in the *Journal of Sleep Research*. Implementation of an accreditation of European sleep centers is still in progress.

The establishment of institutional sleep medicine programs in Europe is less prevalent than in the United States.[6] This is because many sleep centers in Europe are linked to individuals active in the field who, as yet, have not sold or otherwise had their centers incorporated into institutional health care systems or facilities. Some scholars view this as a loss because the implementation of academic sleep centers is needed to cover the increasing challenges brought up by the importance of sleep medicine as part of medical health care.[7,8]

CERTIFICATION OF SLEEP MEDICINE PROFESSIONALS

The ESRS has inaugurated a sleep medicine committee to serve the needs of the various European national sleep societies and to coordinate sleep medicine standards in Europe. The larger national sleep societies have established national certifications for sleep medicine for sleep professionals in their respective countries. Depending on the national medical education system, these examinations are implemented by the national sleep society or by a chamber of physicians, a university, or some other educational institute.

National recognition of these certifications varies largely across Europe. At this time only Germany and Hungary recognize a national subspecialty in sleep medicine. The German Chamber of Physicians has established the subspecialty in sleep medicine under the primary specialties of pneumology (pulmonology), neurology, psychiatry, pediatrics, and ear-nose-throat medicine.

The lack of universal certification for sleep medicine across Europe has led to a European initiative for all countries to harmonize certifications under the authority of the ESRS. A European certification of sleep professionals started in 2012 based guidelines for certification requirements published by an ESRS task force in 2009.[3] In general it is recognized that certification of sleep physicians improves health care in patients with sleep disorders.[9]

The task force guidelines specify four different certifications: one for physicians, one for psychologists, one for scientists, and one for sleep technologists, depending on the previous education of the applicant. All certifications are called *European somnologist* with mention of the specific type as an attribute.

The basic requirement for certification includes a full-time clinical training period in a sleep center for 12 months. During this time the applicant should have evaluated at least 100 patients, with cases of sleep-disordered breathing, insomnia, hypersomnia, movement disorder, and circadian disorder. Certification requires experience in clinical interviewing, use of diagnostic criteria and classification systems, use of sleep diaries, questionnaires and rating scales, psychometric evaluation, and physiological monitoring. Assessment with actigraphy is required.

PSG experience needs to include hook-up, nighttime surveillance, PSG scoring, interpretation, and reporting. Experience with the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test as well as other tests must be proved. In relation to treatment, applicants need to show skills in patient education, treatment delivery, and experience with treatment of patients related to their original professional discipline.

The certification programs require experience in a variety of sleep disorder treatments, such as pharmacotherapy, continuous positive airway pressure (CPAP), cognitive behavioral therapy, and health behavior. Some of these experiences may be obtained at approved sleep medicine courses.

All certifications require applicants to sit for a written examination. The first examinations in 2012 and 2013 were held according to a grandparenting rule. These "grandparents" had to prove at least 10 years of practice and strong engagement in sleep medicine in the past. The first regular examinations for physicians, psychologists, and scientists began in 2014. Sleep technologists had their first grandparenting examination in 2014 and will have their regular technologist examinations in 2016. In addition to taking their examinations, they have to visually score a number of cases.

Somewhat parallel to the ESRS push for commonly recognized European sleep specialty certifications is the contemporaneous effort of the European Respiratory Society (ERS) to create European certifications for respiratory sleep physicians.[10] Although the ESRS focuses on the needs of the national sleep societies to have their sleep expert members show proof of qualifications, the ERS approach has been to create a curriculum, then courses, then certificates to practitioners who successfully completed a final examination after the courses.

The ERS approach is part of a larger framework of certifications for Europeans in respiratory medicine launched in

2009.[11] The framework is a mission to harmonize education in respiratory medicine for European specialists (HERMES) because medical education in Europe is primarily a national duty. To date, European medical associations have been slow to harmonize medical education. The ESRS and the ERS are working together to align their parallel activities in terms of educational content and recognition of educational and examination modules.

ACCREDITATION OF SLEEP MEDICINE CENTERS

Like physician certification, accreditation of sleep centers in Europe is currently a patchwork of rules. The larger national sleep societies in Europe, such as Great Britain and Germany, have established an accreditation procedure for their sleep centers. Depending on the national health care system of these particular European countries, the accreditations are recognized by health insurance, by health care officials, or for quality of care. However, sleep center accreditation standards have not been adopted by most European countries.

Although the ESRS has initiated a European somnologist certification, it is not expected that the ESRS will implement a common European sleep center accreditation. In 2009, the *Journal of Sleep Research* published European guidelines for the accreditation of sleep medicine centers.[2] Currently it is envisaged to check national accreditations that are in place in some countries against the published recommendations and endorse these national accreditations if applicable. For countries in which there are no national accreditations, the ESRS will help the national sleep societies to create a national accreditation system or will organize site visits with sleep center accreditation as preferred by the host country.

It is expected that the current accreditation recommendations[2] will be updated to reflect changes in recording and scoring sleep. In addition to the technical update the revisions will include a reflection on sleep medicine services with varying degrees of specialization ranging from the family physician level through sleep medicine services in university level facilities.[12]

MANAGEMENT OF PATIENTS

The management of patients with sleep disorders differs greatly among European countries. This depends much on the health care systems in place and the activity of the local sleep societies. The management of sleep-related breathing disorders had been investigated across Europe.[13] Home sleep testing is prevalent in several European countries for managing patients with sleep-disordered breathing. Some countries require level III home sleep testing, whereas other countries accept level IV home sleep testing as sufficient to diagnose sleep breathing disease and to initiate treatment. Still other countries require cardiorespiratory PSG for the diagnosis of obstructive sleep apnea and the prescription of CPAP therapy.

EDUCATION

The ESRS supports education for researchers and physicians and specialized education for sleep physicians and sleep scientists. In addition to accreditation and certification, the ESRS sleep medicine committee supports the development of educational material for future somnologists in Europe.

A first step will be establishment of a base set of knowledge and skills. A handful of European universities have implemented master classes in sleep science or sleep medicine for a number of selected students. It is believed that a common curriculum for sleep medicine can be built from these initial experiences.

In 2014 the *Journal of Sleep Research* published the *Catalogue of Knowledge and Skills for Sleep Medicine*.[5] The catalogue is intended to describe a standardized curriculum for sleep medicine education across Europe. The ESRS Board and its SMC compiled the catalogue based on textbooks, standard of practice publications, systematic reviews, and professional experience. The compilation was later validated by an online survey completed by 110 delegates specializing in sleep medicine from different European countries.

The catalogue is intended to be a basis for sleep medicine education, for sleep medicine courses, and for sleep medicine examinations, serving not only physicians with a medical specialty degree but also PhD and MSc health professionals such as clinical psychologists and scientists, technologists, and nurses, all of whom may be involved professionally in sleep medicine. The treatise comprises 10 chapters covering sleep physiology, pathology, diagnostic and treatment procedures, as well as selected societal and organizational aspects of European sleep medicine. A European textbook on sleep medicine was published following this outline of chapters.[15] Required levels of knowledge and skills are defined, as is a proposed workload of 60 points according to the European Credit Transfer System. In the future, the catalogue will be revised in accordance with advances in the field of sleep medicine.

COMPLIANCE WITH REGULATIONS AND REIMBURSEMENT

Government regulations for the delivery and reimbursement of sleep medicine in Europe differ as much as the health care systems in European countries differ. Although many countries recognize the need to diagnose, treat, and follow patients with sleep disorders as a matter of good medicine and—with regard to hypersomnolent drivers—public safety, there is little movement to align the disparate regulatory schemes at this time.

A 2007 study indicates that most European countries recognize the public safety risks presented by drivers with excessive daytime sleepiness caused, in part, by sleep-disordered breathing.[14] However, a review of the licensure requirements of 25 different European countries as part of the study showed that less than half of the countries referenced obstructive sleep apnea in their licensure regulations. Still, at the time, seven countries required a physician's medical certificate indicating treatment compliance to process applications for drivers with sleep breathing disorders, and in two countries it was up to the patient to decide (on the physician's advice) whether to drive again. France required a Maintenance of Wakefulness Test for symptomatic commercial drivers to pass fitness-to-drive standards.[14] Still, driving license regulations are different throughout European countries, and efforts are being made to harmonize regulations. To proceed with this, on July 1st, 2014, an update of the European directive on driving was published, which now mentions obstructive sleep apnea and daytime sleepiness specifically. The directive has to be adopted by national laws in Europe until Dec 15th, 2015, in order to become effective.

HEALTH INSURANCE COVERAGE FOR SLEEP MEDICINE: THE GERMAN EXPERIENCE

The insurance and regulatory aspects of covering and delivering sleep medicine in Germany provide one example of the European experience. Everybody in Germany has health insurance and expects that all medical care is covered by the insurance. In Germany there are a little less than 200 insurances available to citizens. Very few patients are not insured.

There are two types of insurances: a general basic insurance for 85% of the population that is open to anybody and a so-called private insurance for 15% of the population. Although the medical care is the same under each insurance program, small differences exist. Patients with private insurance have a wider choice to select their physician and better chances to obtain single-bed rooms if admitted to a hospital. Physician reimbursement is higher for private patients than for general plan patients for the same service.

German citizens with private insurance pay monthly premiums that vary according to age and risk. Thus premiums are relatively low for younger insured persons and higher for older persons. Therefore a person must earn a high salary to be eligible for private health insurance. The cost of health insurance under the general insurance plan is simply a percentage of the insured person's income.

In general, health insurance in Germany covers all costs for a diagnosis of sleep apnea and the treatment of sleep apnea and all replacement and service items. A patient with complaints of irregular snoring or observed apneas will go to the family physician to get guidance for his or her problem. The patient cannot go to a sleep center directly. Sleep centers are only allowed to admit referred and diagnosed patients.

If the patient's family physician suspects sleep apnea, the physician will send the patient to a pneumologist who has a license for home sleep testing. To receive reimbursement for home sleep testing from the health insurance, the pneumologist has to get a license, which requires attending a 5-day course about basics of sleep medicine with an emphasis on sleep-disordered breathing and passing a 60-minute examination at the end of the course.

If the home sleep test is positive in terms of sleep apnea, the pneumologist will refer the patient to a sleep center for CPAP titration with attended PSG. The sleep center has to be accredited by either the German Sleep Society or by another institution to obtain health insurance reimbursement. The patient then returns after 6 months of treatment to the pneumologist for another home sleep test as a treatment follow-up. Thereafter are no planned follow-ups unless the patient experiences new complaints.

If the health insurance company determines that a diagnosis was wrong or a referral at any step was not justified, the insurance company can ask a control body to review the patient's medical records. These review bodies, called medical service of insurances, typically consist of special physicians employed by the control body. They will check the case and will see whether all diagnostic steps were done according to guidelines. They will review the pneumologist for credentials and licensure as well as the sleep test data to determine the quality of the recording signals and whether the diagnostic and therapeutic decisions are justified.

The criteria used by the reviewing physicians derive from evidence-based literature and are compiled in a reviewers' guide for sleep-disordered breathing for the medical insurance services. If the reviewing body denies reimbursement for the claim, the pneumologist cannot appeal to the patient or to the insurance company, but only to the independent control body. The physician is not prohibited from asking the patient to pay privately for the services provided, but as a cultural matter, German patients do not expect to have to pay privately for health care services. Accordingly, providers almost never ask for patients to cover services from their own pockets. If the prescription was found to be erroneous or a procedure was found to be not necessary, then the physician, not the hospital or institution, is liable and will be charged for the prescription. To cover this, German physicians have a professional insurance. Hospitals may cover this claim, if agreed on in the contract with the employed physician.

CLINICAL PEARLS

- Sleep medicine education in Europe is in the process of being coined in curricula that are based on evidence and worldwide knowledge on sleep physiology, sleep disorders, and treatment.
- Certification of sleep professionals and accreditation of sleep centers is advancing in all countries, and large efforts are taken to achieve the same consensus on patient service and quality of care independent of the health care system and more dependent on the underlying pathologies.
- Management of sleep disorders is developed through academic institutions that are in the process of installing sleep medicine programs and aligning these programs to achieve comparable goals to finally optimize patient care.

SUMMARY

Europe has a long tradition of sleep research. Sleep medicine evolved in parallel with other countries worldwide. The ESRS covers both sleep science and sleep medicine. Sleep medicine is more a national issue because health care systems are very different across the European countries. Some countries have large national sleep societies with sleep center accreditation and sleep expert certification. Other countries lean much on the ESRS. The Society tries to set consensus rules for sleep expert certification, for sleep center accreditation, and provides education for sleep experts and technologists. Publications on these issues are compiled and published in the *Journal of Sleep Research*. The implementation of these steps is coordinated among the European countries and is recently coordinated with similar activities initiated by the ERS. The management of patients with sleep disorders differs much among European countries and is strongly linked to reimbursement schemes.

Selected Readings

Bassetti C, Dogas Z, Peigneux P. *Sleep Medicine Textbook*. Regensburg: European Sleep Research Society; 2014.

Zee PC, Badr MS, Kushida C, et al. Strategic opportunities in sleep and circadian research: report of the joint task force of the sleep research society and American academy of sleep medicine. *Sleep* 2014;**37**:219–27.

A complete reference list can be found online at ExpertConsult.com.

Section 10

Occupational Sleep Medicine

Chapter 70

Introduction

Gregory Belenky; Torbjörn Åkerstedt; Nancy J. Wesensten

Occupational sleep medicine draws on clinical, experimental, and field research to sustain human performance. Broadly, the goal of occupational sleep medicine is to sustain performance, and attendant productivity, safety, health, and wellbeing, in the workplace and in the operational environment. An operational environment is a workplace in which human performance is critical to system output; if the human fails, the system fails.

Sleep loss (time awake), adverse circadian rhythm phase (time of day), and workload (time on task/task difficulty) interact to degrade performance and increase self-reported fatigue and sleepiness. With increasing fatigue, performance degrades, productivity decreases, and the risk of error, incident, and accident increases. Occupational sleep medicine aims to mitigate these adverse effects.

Occupational sleep medicine applies to all operational environments. Examples of operational environments include military operations, maritime operations, medicine, land transportation, aviation, security work, energy generation, resource extraction, financial markets, industrial production, the information media, and intelligence-gathering operations.

Occupational sleep medicine applies to any system involving human performance and 24/7 operations. Occupational sleep medicine will likely develop as a subset of sleep medicine with ties to occupational medicine and industrial and organizational psychology. Occupational sleep medicine could be the basis for enterprise-wide systems of fatigue risk management.

What follows in this introduction are brief overviews of the chapters making up the section on occupational sleep medicine.

CHAPTER 71: PERFORMANCE DEFICITS DURING SLEEP LOSS AND THEIR OPERATIONAL CONSEQUENCES

Does the pattern of sleep-loss–induced neurobehavioral (alertness, response speed) deficits seen during laboratory studies tell us anything about how workplace performance is likely to be affected by insufficient sleep? Under controlled laboratory conditions, the duration of lapses in sustained attention increase as sleep loss accumulates—an effect that is exacerbated during the circadian alertness trough. As the ability to sustain attention degrades, time on task becomes more salient, and breaks (time off task) restore performance. However, this effect is temporary, and its efficacy (duration and magnitude) decreases with increasing sleep loss.

Under operational conditions, accidents occur when lapses in sustained attention temporally align with the requirement for attention to avoid critical errors. Accident risk is thus a function of duration and frequency of attention lapses (which themselves are a function of sleep-wake and circadian factors) and exposure to critical events. Although the ability to predict a specific individual's risk has yet to be achieved, predicting general risk or likelihood based on sleep-wake history and time of day is possible.

CHAPTER 72: SLEEP AND PERFORMANCE PREDICTION MODELING

Currently available mathematical sleep and performance prediction models are based on a conceptual two-process model of sleep regulation. According to such models, performance

(and in the original conceptual model, sleep) is governed by two processes: a homeostatic process that is determined by sleep-wake amounts and a relatively (but not entirely) static circadian (24-hour) process. Such models can be used to quantify the effect of a given sleep-wake-work schedule on mental effectiveness. Importantly, the *relative* effect of several sleep-wake-work schedules can be quantified. The ability to quantify effectiveness associated with any sleep-wake-work schedule allows for proactive risk management and mitigation in the form of alternative scheduling, shifting the timing of critical events outside the window of reduced effectiveness, and implementation of countermeasures (e.g., naps). Application of predictive models as components of comprehensive fatigue risk management systems (FRMS) is gaining momentum in transportation sectors, most notably commercial aviation.

CHAPTER 73: FATIGUE RISK MANAGEMENT SYSTEMS

Fatigue risk management is an emerging applied arm of occupational sleep medicine. As indicated by its name, fatigue risk management involves actively managing risk (which in turn implies acceptance of some amount of risk in the same way that risk associated with poor weather or other factors would be considered and managed). In contrast, prescriptive hours-of-service rules specify shift duration, between-shift intervals, and within-shift breaks, thus imposing a priori boundaries that are designed to eliminate risk. However, because such hours-of-service (HOS) rules are not based in the physiology that drives performance (i.e., the human circadian rhythm and homeostatic drive for sleep), they are overly restrictive in some aspects (e.g., restricting consecutive hours worked) and potentially unsafe in other aspects (e.g., allowing for 23-hour days and other schedules that are incompatible with circadian physiology).

An FRMS based on sleep-wake and circadian principles affords an alternative to prescriptive HOS. FRMS is adapted to the operational environment and is iteratively reviewed and revised to meet operational demands. Unlike prescriptive HOS rules, in which operators are incentivized to work to HOS limits, FRMS may incentivize operators to sleep (e.g., in-flight napping) to extend duty hours (e.g., ultra–long-range flights). This unexpected consequence of FRMS implementation results from shifting the locus of responsibility for safety away from the regulator and toward employers and employees.

CHAPTER 74: DROWSINESS IN TRANSPORTATION WORKERS

Driving while drowsy places the driver, other motorists, and pedestrians at risk for injury. Professional drivers may be at particular risk (perhaps owing to the previously indicated issues with HOS regulations that allow for long, nighttime driving bouts), and it is likely that the problem of drowsy driving in professional drivers is underestimated. Individuals driving to work during the circadian alertness trough and from work after extended duty hours are also at risk for drowsiness-induced automobile collisions. Depressant drugs, sleep-disordered breathing, and narcolepsy increase these risks. Degree of impairment (and thus risk) can be quantified by objective measures including the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test. Effective treatment of sleep-disordered breathing and other sleep disorders reduces accident risk.

Because drowsy driving in professional drivers poses serious risks, some have suggested that a provision be included to screen for sleep disorders in this population. However, to date no transnational standards for such evaluation exist, nor is there agreement regarding what constitutes effective treatment and corresponding return-to-duty (or remain-on-duty) standards.

In lieu of regulatory requirements, physicians can routinely evaluate patients for risk for drowsy driving by clinical interviews that include queries related to sleep disorder symptoms as well as questions regarding sleep hygiene. Public service campaigns regarding the dangers of driving drowsy, magnification of risk associated with alcohol use, and the value of countermeasures (e.g., avoiding driving during the circadian low, tactical use of napping and caffeine) can be of value.

CHAPTER 75: SHIFT WORK, SHIFT-WORK DISORDER, AND JET LAG

In industrial economies, shift work and travel across time zones are common. Both displace sleep and wake from their usual temporal alignments with the 24-hour alertness rhythm. The circadian alertness rhythm will eventually (over days) resynchronize to a new time zone, but in most shift workers, the circadian alertness rhythm does not resynchronize to night-shift work. The resulting insomnia during the daytime sleep period and excessive sleepiness during the nighttime work period, when sufficiently severe, are diagnosed as shift work sleep disorder (SWSD). In addition to a thorough sleep history, tools including the Epworth Sleepiness Scale, the Insomnia Severity Index, and the Pittsburgh Sleep Quality Index can be used to diagnose SWSD with or without polysomnography. Distinguishing fatigue from sleepiness is critical to differentiate depression from SWSD.

In lieu of organization-level adjustments in shift timing and duration, treatments to mitigate SWSD include environmental manipulations, such as nocturnal bright light and daytime sleep in darkness, and pharmacologic tools, such as stimulants to maintain alertness while on shift at night and sleep-inducing medications to maintain and extend sleep off-shift during the day. Jet lag is characterized by excessive daytime sleepiness and nocturnal insomnia at the new local time. In contrast to SWSD, jet lag is self-limited because the circadian alertness rhythm gradually resynchronizes to new local time as a function of consistent daylight exposure. Thus bright light exposure can facilitate resynchronization if timed properly. Naps and caffeine are safe and effective countermeasures that can be implemented to reduce excessive daytime sleepiness.

CHAPTER 76: SLEEP PROBLEMS IN FIRST RESPONDERS AND THE MILITARY

First responders such as police and firefighters are required to work extended hours and engage in shift work, both of which result in sleep loss and circadian desynchrony. In these occupational groups, impaired judgment may result in more severe and enduring adverse consequences than in other occupational groups. The same is true for deployed military personnel. In police forces, organizational (e.g., direction and speed

of shift rotation, staffing), social (e.g., overtime, second jobs), and individual (e.g., overall health) factors promote fatigue. High prevalence of sleep disorders further increases morbidity and mortality. Primary care physicians may serve as the first line of defense by screening for sleep disorders and educating police officers regarding sleep hygiene, weight management, and other controllable factors.

CHAPTER 77: SLEEP, OCCUPATIONAL STRESS, AND BURNOUT

Occupational stress plays a major role in the etiology of insomnia. Results from both cross-sectional and prospective studies support this connection. The inability to stop thinking about work in the evening and ruminating over anticipated next-day stress appears to be the main factor causing insomnia. High job demands coupled with low job control, lower socioeconomic status, and low social support contribute to work-related insomnia. Long-term exposure to occupational stress may result in extreme fatigue, impaired cognition, and depressed mood; associated sleep disturbance is characterized by sleep fragmentation, increased sleep latency, and reduced slow wave sleep. A blunted cortisol response (dexamethasone suppression test) and an exaggerated cortisol awakening response indicate involvement of the hypothalamic-pituitary-adrenal axis. Neural markers include reduced volume in cortical and subcortical areas. It is tempting to see these effects as the culmination of a chain of events that begin with occupational stress, pass through stress-induced insomnia and sleep disturbance, and eventually lead to extreme fatigue and inability to work, frequently labeled "occupational burnout." Occupational burnout is associated with increased sickness absence and should be considered in patients with unusually high absentee rates. Occupational burnout seems at least partly reversible.

CHAPTER 78: OPTIMIZING SHIFT SCHEDULING

Scheduling of shift timing and duration is a key organization-level factor that determines worker on-duty alertness and off-duty ability to obtain sleep. In principle, a schedule that minimizes circadian misalignment with the sleep-wake schedule, minimizes build-up of sleep loss over a single shift and across a shift cycle, and permits sufficient recovery during days off will best maintain on-duty alertness and off-duty sleep quality. In practice, although evidence indicates that day work is superior to night work, there is little evidence to support rotating over permanent night-shift work or fast-rotating over slow-rotating shifts. In contrast, evidence strongly supports clockwise (progressively later) shift rotation over counterclockwise (progressively earlier) shift rotation. Early-morning starts (e.g., 3:00 AM) and long shifts are associated with reduced sleep amounts, and duration of time off between shift rotations affects amount of time available for sleep recovery. Under conditions in which workers sleep onsite, workers have few (or no) social obligations, and light exposure can be controlled (e.g., oil rig workers), numerous consecutive long-duration (12-hour) shifts can be worked with few problems.

Evidence supports the use of countermeasures, including caffeine, preduty napping, and on-duty napping. Arguments that the possibility for postnap sleep inertia should preclude on-shift napping are largely unsubstantiated.

CHAPTER 79: OBSTRUCTIVE SLEEP APNEA IN THE WORKPLACE

Obstructive sleep apnea (OSA)-induced sleep fragmentation causes excessive daytime sleepiness and impaired mental performance. In addition to workplace safety risks, OSA exacts a toll in terms of absenteeism and lost productivity. Managing OSA in the workplace consists of a tiered approach that starts with screening for both nonmodifiable and modifiable risk factors. Nonmodifiable screening risk factors for OSA include older age, male gender, postmenopausal status in females, and ethnicity. The most significant modifiable screening risk factor is obesity or adiposity, which can be objectively determined by body mass index and neck circumference. Screening also consists of self-report tools aimed at daytime sleepiness and related symptoms and functional performance tests for detecting impairment. Workers exceeding cutoffs are referred to a sleep disorders medicine specialist for diagnosis (ranging from portable monitoring to full polysomnography), treatment (positive airway pressure plus prescription stimulants to treat residual sleepiness; weight loss management), and compliance monitoring. Currently only the U.S. Federal Aviation Administration lists untreated OSA as a disqualifying condition of employment. Requirements that medical examinations for operator licenses (e.g., commercial motor vehicle operator licenses) be obtained from a certified medical examiner (CME), combined with sleep disorders education for CMEs, will serve to increase rates of OSA detection and treatment.

GROWTH IN OCCUPATIONAL SLEEP MEDICINE

Sleep medicine focuses mainly on diagnosis and treatment of specific sleep disorders (e.g., sleep apnea) in the individual patient. Complementing the clinical practice of sleep medicine, occupational sleep medicine is implemented at the group level not only to sustain workplace productivity and safety but also to maintain overall worker health across a career.

Occupational sleep medicine is advancing in several areas, including fatigue risk management, prevention of drowsy driving, mitigation of SWSD, behavioral and pharmacologic interventions in special populations (e.g., first responders), and even genetics. In particular, fatigue risk management efforts have benefited directly from growth in occupational sleep medicine through wider awareness of the effect of sleep-wake and circadian factors on safety. In turn, occupational sleep medicine specialists benefit from increased opportunities to provide consultation to industries and occupational groups on screening, diagnosis, and treatment of sleep disorders and implementation of organizational fatigue risk management.

Selected Readings

Fang MZ, Ohman-Strickland P, Kelly-McNeil K, et al. Sleep interruption associated with house staff work schedules alters circadian gene expression. *Sleep Med* 2015;**16**(11):1388–94.

Merkus SL, Holte KA, Huysmans MA, et al. Nonstandard working schedules and health: the systematic search for a comprehensive model. *BMC Public Health* 2015;**15**(1):1084.

Chapter 71

Performance Deficits During Sleep Loss and Their Operational Consequences

Hans P.A. Van Dongen; Thomas J. Balkin; Steven R. Hursh

Chapter Highlights

- Sleep loss induces sleepiness and exerts profound negative effects on cognitive performance, increasing the risk for errors and accidents.
- Three operationally relevant phenomena associated with the impact of sleep loss on performance are presented, including sleep loss-induced performance instability, interaction of sleepiness with time on task, and sleepiness-induced alteration of regional brain activity.
- Current knowledge regarding the impact of sleep loss on performance is summarized, and issues pertaining to the real-world application of this knowledge in operational environments are discussed.

Although difficult to estimate, it is likely that sleep loss-induced performance deficits cost the world economy hundreds of billions of dollars per year in accidents, direct health care costs, and lost operational efficiency and productivity. This can be inferred from the fact that sleep insufficiency is experienced at least occasionally by everyone and is experienced chronically by a considerable proportion of the adult population.

According to the 2004–2006 National Health Interview Survey,[1] approximately 21% of adults in the United States typically obtain 6 or fewer hours of sleep per 24 hours—considerably less than the 7 to 8 hours of sleep that is generally recommended.[2] If those with sleep disorders (e.g., sleep apnea, insomnia) are considered along with those who are chronically sleep-restricted for other reasons (e.g., shift work, lifestyle), it is estimated that as many as 70 million Americans experience chronic sleep loss[2] and are therefore likely to experience correspondingly impaired performance on a daily basis.

In this chapter, current knowledge regarding the impact of sleep loss on performance is summarized, and issues pertaining to the application of this knowledge to real-world, operational environments are discussed.

THE NATURE OF SLEEP LOSS-INDUCED PERFORMANCE DEFICITS

Cognitive performance is not solely a function of (and is therefore not simply a direct reflection of) sleepiness level (drive to sleep). Nonetheless, level of sleepiness, as determined by duration of prior sleep, time since awakening, and circadian timing (see Chapter 37), affects performance on a variety of tasks in a predictable manner. For example, performance on a psychomotor vigilance test (PVT) declines in a dose-dependent fashion with decreasing amounts of nighttime sleep across multiple days of sleep restriction.[3,4] However, sleepiness-induced declines in performance are not the result of a wholesale shift (general slowing) in the entire distribution of reaction times (RTs), but rather reflect increased trial-to-trial variability in RTs,[5,6] with an increasing proportion of relatively long RTs intermixed with "normal" RTs (i.e., RTs within the range typical of the well-rested state).[6,7]

It has been posited that the increased performance variability that characterizes sleep loss is most likely a manifestation of a reduced level of stability in the physiologic processes by which wakefulness is maintained—specifically, that it is caused by intermittent intrusion of sleep into wakefulness.[6,8] An implication of this hypothesis is that performance deficits due to sleep loss should be generic; that is, sleep loss should affect all facets of cognitive performance. However, evidence from cognitive-behavioral and neuroimaging studies does not uniformly support this hypothesis.[9,10] Recent evidence regarding the organization of sleep-wake processes suggests that the effects of sleepiness are greatest in those neuronal pathways most intensively used during task performance. Accordingly, the deleterious effects of sleepiness have been found to be most salient for those specific cognitive processes mediated by these same neural pathways, in a use-dependent manner.[10,11]

Although it would be interesting from a scientific standpoint, and useful as a tool to increase the accuracy with which task performance could be predicted in operational environments, the relative extent to which any two cognitive processes (e.g., information encoding, working memory) are differentially affected by sleep loss cannot be specified, much less quantified. This is because (1) multiple cognitive processes are involved in any given performance task, with each individual (and typically unobservable) cognitive process being affected by sleep loss to an unknown extent,[12] and (2) there is no scale against which different cognitive processes can be commonly measured and thus compared (i.e., the old "apples vs. oranges" conundrum).

Certain cognitive task designs allow for the effects of sleep loss on a specific cognitive process to be assessed by contrasting distinct task conditions (e.g., performance on a working memory task with two, three, or four items to be held in memory),[13] but these require the assumption that the effects of sleep loss on all other relevant cognitive processes remain invariant under each tested condition. Although research in this area is gaining some traction,[9] the experimental and

logical challenges to achieving substantial scientific breakthroughs remain considerable. In part, this is because a test's sensitivity to sleep loss varies not only as a function of the sensitivity to sleep loss of the admixture of cognitive processes involved in the task performance, but also as a function of the parameters of the test itself. For example, manipulation of task duration, time pressure, and amount of feedback provided during testing can affect the sensitivity of a performance measure to sleep loss[14,15] (see Chapter 37).

Likewise, statistics such as effect size are useful for comparing the sensitivity of specific tests administered with a specific set of test parameters, under specific sleep loss conditions,[16,17] but they cannot be used as a basis for comparing the extent to which sleep loss generally affects one cognitive ability versus another. To complicate matters further, large, trait-like individual differences in vulnerability to sleep loss exist.[18,19] These uncertainties also exist in workplace settings, where a job typically involves a wide range of tasks with several individuals performing a given job.

These challenges complicate researchers' efforts to understand the neurocognitive underpinnings of performance deficits due to sleep loss. However, from a practical standpoint, application of the known principles by which circadian and homeostatic processes mediate alertness and performance can nevertheless be usefully applied in operational environments. That is, although the precise degree to which performance on a particular task will be degraded is not currently predictable for a specific individual, the general trends (both timing and extent) of sleep loss-induced performance deficits can be predicted and usefully applied in operational settings.

In this chapter, we discuss three phenomena that characterize sleepiness in operational settings. The first is concerned with the translation of performance instability into errors and accidents in operational settings. The second is the interaction of sleep loss with time-on-task effects. The third is sleepiness-induced change in the brain activation patterns associated with degraded cognitive performance.

PERFORMANCE INSTABILITY: EFFECT ON ERRORS AND ACCIDENTS

In large-scale correlational studies, work schedules that militate against adequate daily sleep (including extended work hours and shift work) have been linked to increased risk for human error and accidents,[20-22] resulting in reduced safety and productivity.[23-25] Yet, sleepy people do not make errors or cause accidents simply by virtue of the fact that they are sleep deprived, just as being fully alert does not guarantee error-free performance. In accident investigations, the unequivocal establishment of sleepiness as a causal factor is often impossible, even when the presence of sleepiness is itself undisputed (unless there is evidence that the accident was caused by frank sleep onset).

There are two reasons for this. First, sleepiness is rarely the sole reason that accidents occur; multiple, diverse factors ranging from personnel shortages to equipment failures and safety check overrides typically combine with human error to result in adverse outcomes. Second, as previously mentioned, human error because of sleepiness occurs against a backdrop of increased performance variability[6] and is therefore at least partly stochastic (i.e., random) in nature. In other words, although the likelihood or frequency of decremented performance increases with sleep loss (and is exacerbated during the descending phases of the circadian rhythm of alertness), actual task performance of a sleepy individual can vary from "impaired" to "normal" on a moment-to-moment basis.

Sleep-wake history and circadian rhythmicity interact to determine sleepiness-alertness and performance capability (see Chapter 37). Sleepiness varies as a function of time awake, with longer wakefulness inducing progressively increasing sleep drive. Sleepiness also varies as a function of time of day, with the biologic clock's alertness output declining during "nighttime," resulting in elevated drive to sleep (sleepiness). As a consequence, alertness level and performance capability are reduced when working extended hours involving sleep loss and when working at night or in the early morning.[26,27] Moreover, chronic sleep loss leads to cumulative degradation of performance across days, weeks,[3,4] and perhaps longer (definitive longer term studies have not yet been performed).

Not surprisingly, it has been determined that the circadian (i.e., time of day) variation in alertness and performance is associated with a circadian rhythm of accident rate and injuries.[20,28] There is less evidence for a relationship between accidents or injuries and changes in performance related specifically to duration of time awake. However, a relationship between duration of time awake and accident risk can be inferred from statistics on road crashes that were attributed to the driver having fallen asleep.[29,30] Although definitive evidence is lacking (and it should be noted that there is a possibility that investigators are more likely to attribute accidents to sleepiness when occurring at certain times of the night), it is reasonable to presume that the same interaction between sleep-wake history and circadian rhythm that increases sleepiness and decreases performance capability also increases the probability of driving errors and resulting traffic accidents.

Many occupationally relevant tasks, ranging from systems monitoring and threat detection to driving, are likely vulnerable to sleep loss, at least in part because these tasks require sustained attention.[31,32] In modern operational settings characterized by extensive automation, such tasks are common. Automation and other technologic innovations have broadly improved safety, but by shifting performance demands to sustained attention tasks, they may have simultaneously increased the likelihood of human error.[33] This is because the ability to sustain attention is negatively affected by sleep loss and time of day (and time on task, discussed later). The paradoxical result is that although serious accidents are increasingly rare, when they do occur such accidents can be especially devastating and costly.[34,35]

Because such catastrophic accidents are rare, and because their occurrence also typically depends on the chance convergence of random factors or events, it remains difficult to predict the risk for sleep loss-induced accidents. Even when considering incidents more broadly by including near-accidents ("near-misses"[36]) and other performance errors, a relationship between sleepiness and accident rate remains difficult to discern. Consideration of the stochastic nature of performance impairment due to sleep loss may shed some light on this issue.

Figure 71-1 illustrates how sleep loss-induced stochastic instability in performance impairment may lead to an accident. The PVT is a validated assay of sustained attention that is particularly sensitive to sleep loss.[16,37] As such, the series of

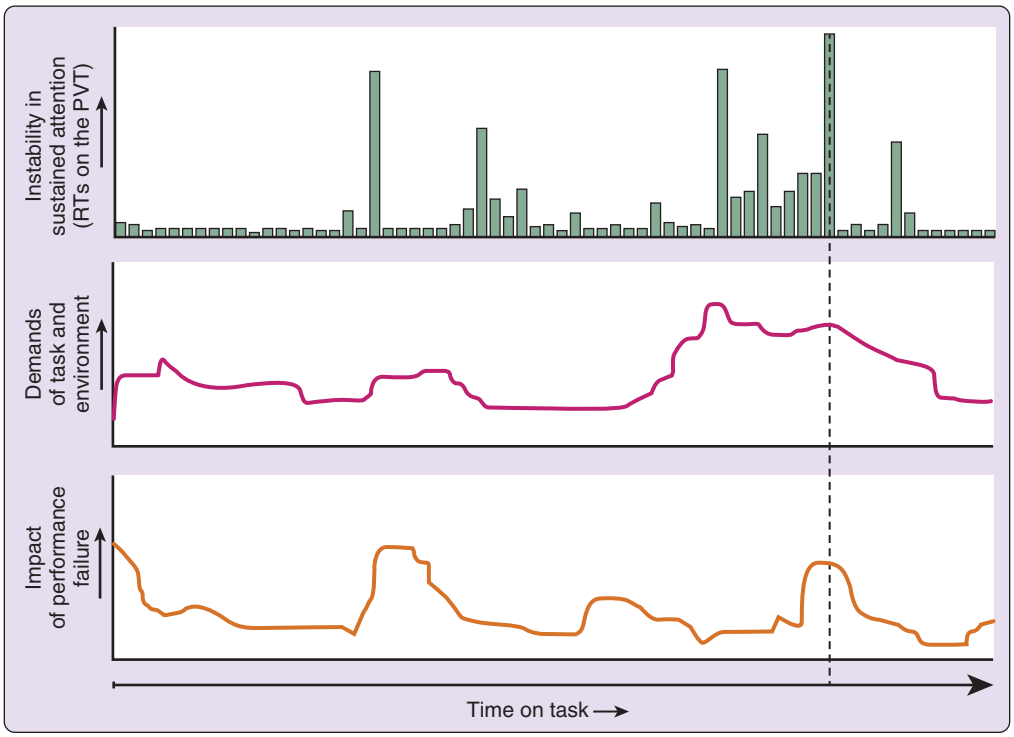


Figure 71-1 Schematic of a proposed mechanism by which sleep loss may contribute to accidents. The *top panel* depicts reaction times (RTs) across a 10-minute span of task performance on a psychomotor vigilance test (PVT) for an individual who continuously maintained wakefulness for 60 hours, as observed during an experiment published by Doran and colleagues.[6] This illustrates stochastic instability in sustained attention over a 10-minute interval. *Longer bars* represent slowed responses indicative of lapses of attention. The *middle panel* shows a hypothetical pattern of changing demands of the task at hand and the environment in which it is performed (*upward* corresponds to greater demands). The *bottom panel* displays a hypothetical level of impact that human error would have over the course of the task. In this view of how sleepiness contributes to accident causation, intervals of inattention when cognitive processing demands are high lead to human error, which in turn—if the impact of error is considerable—results in an accident. Thus, for sleep loss to actually lead to an accident, there must be temporal alignment of attentional lapses, substantial cognitive processing demands, and high impact of human error (in this case represented by the *dotted black line*). (From Doran SM, Van Dongen HPA, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol* 2001;139:253–67, with permission.)

RTs recorded in a PVT session can be considered a record of task inattentiveness. Assuming for the purpose of illustration that the same admixture of cognitive functions is required for a given task in an operational setting, long RTs on the PVT would reflect intervals of inattentiveness (lapses of attention) during the task at hand. When the demands for cognitive processing are high during such an interval of inattentiveness, human error would be likely. In addition if the negative impact of human error at that specific time is also high, then an accident might result.

For example, if the task is driving a car, and an interval of inattentiveness coincides with the approach to an intersection with a stop sign, then detection and processing of the stop sign might fail and the intersection would be crossed without braking—that is, human error. If at the same time another car enters the intersection, a collision could ensue. Yet, if no other car entered the intersection, or there had been no intersection, or if the interval of inattentiveness had occurred a little earlier or later, then the accident would not have occurred.

From this perspective, it is necessary for a period of inattentiveness and significant effect of error to temporally align in a manner that results in an accident. It follows that accident risk is proportional to both total time of inattentiveness (cumulative lapse time) and density of critical task events (i.e., prior risk or exposure[38]). Given information about the latter, it may be possible to predict accident risk by predicting cumulative lapse time. There is a strong correlation between the number of lapses of attention and their duration,[37] so a mathematical model that predicts PVT lapse counts[39] could potentially serve this purpose (see Chapter 72).

THE TIME-ON-TASK EFFECT

It is ironic that hours-of-service regulations intended to mitigate the effects of sleepiness on performance and risk in operational settings are typically focused exclusively on "time on duty" and fail to account for sleep-wake history and circadian rhythm. However, there are good reasons for this. Dealing effectively with sleep-wake and circadian factors would require regulation of the timing and duration of workers' sleep, which is nearly impossible to enforce from both logistical and socioethical standpoints. In addition, even when individuals are well rested, performance tends to deteriorate as a function of continuous work time[40] and is improved by rest breaks.[41] However, with the passing of work hours, time awake also accrues and the phase of the circadian rhythm of alertness

changes. Depending on whether these changes are toward increasing or reducing alertness, they may offset or amplify the performance-impairing effects of increasing work time. Therefore, prescriptive hours-of-service regulations that do not take all of these factors into account will never be more than partially effective.[42,43]

Performance degradation caused by continuous performance over time is a phenomenon commonly referred to as the time-on-task effect. This effect is especially pronounced for tasks requiring sustained attention, such as the PVT.[44,45] Such tasks are also especially sensitive to sleep loss.[14,45] The time-on-task effect has been conceptualized as arising from cognitive work-related depletion of cognitive resources over time—resources that, unlike those mediating sleepiness, require only rest (time off task) to effect recuperation.[45] Recent insights into the neurobiology of sleep and wakefulness suggest that the underlying mechanism may be use-dependent, local sleep-like states occurring increasingly frequently in neuronal pathways subserving the cognitive processes most intensely used while performing the task at hand.[11]

The time-on-task effect interacts with time awake and time of day. That is, when sleepiness is increased because of time awake or time of day, performance degrades faster during continuous task performance. This is illustrated in Figure 71-2. The figure depicts minute-by-minute mean response speed on the 10-minute PVT, which was administered every 2 hours across 40 hours of continuous wakefulness.[46] Time-on-task effects were evident before significant sleep loss, that is, from 0800 to about 2400 on the first day. Sleep deprivation exacerbated the time-on-task effect, especially during the night and early morning hours. For example, comparison of the performance change across the 10-minute test at 0800 on day 1 versus 0800 on day 2 suggests differences not only in overall performance but also in the rate at which performance declined across the 10 minutes.

Also noteworthy in Figure 71-2 is the extent to which performance on the PVT recovered from minute 10 on one trial to minute 1 on the next, despite the fact that there was no intervening sleep. This was evident even when overall performance was degraded most significantly (at about 0600 to 1000 on the second day). Thus mean speed on the PVT at minute 1 on the 0800 test was greater than mean speed at minute 10 on the 0600 test, despite the fact that average 10-minute performance declined across this time interval. In other words, the time-on-task effect is reversed by rest breaks (time off task), even during sleep deprivation.

One reasonable hypothesis is that time-on-task effects are at least partly motivational, that is, that the decrements across each 10-minute PVT bout represent declining within-bout motivation and that the recovery from the end of one PVT bout to the beginning of the next merely reflects some temporary restoration of motivation. However, it has been found that performance impairment from time on task may carry over to performance of another task performed immediately afterward,[47,48] strongly suggesting that declining motivation does not account for all of the time-on-task variance. That said, the time-on-task effect does not *necessarily* carry over between tasks performed back to back; in certain cases, performing a different task restores performance similarly to a

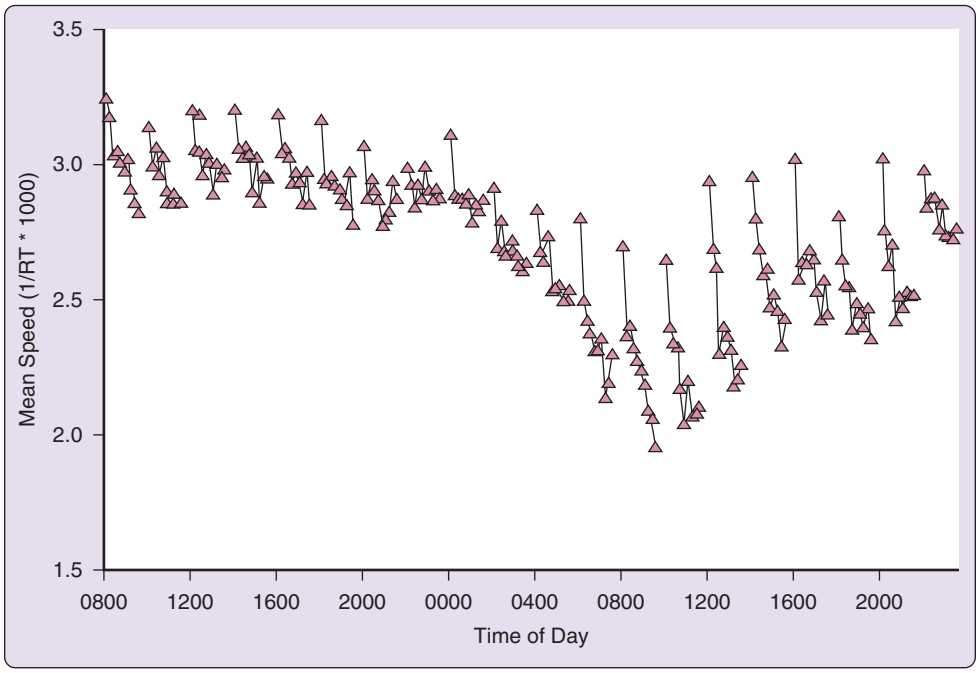


Figure 71-2 Time-on-task effect across a 10-minute psychomotor vigilance test (PVT) during 40 hours of total sleep deprivation. Each set of data points shows average speed (inverse of reaction time [RT]) in consecutive 1-minute bins (not drawn to scale on the time axis) during the 10-minute task duration. Note the changes in overall performance across the 40 hours of sleep deprivation due to the interaction of the effects of time awake and time of day. Furthermore, note the improvement from minute 10 of one PVT bout to minute 1 of the next PVT bout 2 hours later despite the absence of intervening sleep. (From Wesensten NJ, Belenky G, Thorne DR, et al. Modafinil vs. caffeine: effects on fatigue during sleep deprivation. *Aviat Space Environ Med* 2004;75:520–5, with permission.)

rest break. It has been hypothesized that time-on-task carryover effects reflect use of common neuronal pathways for cognitive processes, thus preventing rest-mediated recovery in these pathways and resulting in further degradation of performance (as discussed in the prior section).[45]

From a practical standpoint, the extent to which sleepiness from time awake and time of day interacts with time on task to produce performance deficits is an important issue. Sleep loss often occurs because of externally imposed (e.g., occupational) requirements for individuals to remain awake for extended periods, performing goal-directed tasks. Thus studies in which the effects of sleep loss are measured on tasks performed nearly continuously may best replicate real-world operational conditions. In one such study, subjects performed cognitively demanding work on a nearly continuous basis across 54 hours of sustained wakefulness, with occasional administration of subjective rating scales (including mood, fatigue, and sleepiness scales) and with short breaks for meals and personal hygiene.[49] The study did not include a comparison condition in which work was performed at a slower pace (i.e., with more frequent and/or longer breaks). Nonetheless, by assessing the relative rates of performance decline from this study with those of previously published sleep deprivation studies in which similar performance measures had been administered but for shorter periods of time, it was suggested that sustained cognitive work accelerates the rate at which performance declines during sleep deprivation.

Thus the rate and extent to which performance is impaired varies as a function of both time on task and sleepiness (a joint function of the homeostatic sleep pressure and the circadian rhythm of alertness).[50] The portion of the performance degradation that is due to time on task is reversed by simple rest (time off task), whereas the portion that is due to sleep loss is reversed only by recovery sleep.

FUNCTIONAL BRAIN IMAGING STUDIES OF SLEEP LOSS AND PERFORMANCE

Results from functional brain imaging studies have revealed reduced regional brain activity during sleep deprivation, with greatest reductions manifest in prefrontal cortex, inferior parietal and superior temporal cortex, and thalamus.[51] Based on such findings, specific sleep loss–induced deficits in various aspects of performance have been predicted. For example, noting that sleep deprivation results in reduced metabolic activity in regions of the prefrontal cortex known to mediate specific aspects of cognitive performance and perception, it has been predicted and confirmed that sleep deprivation causes deficits in executive mental functions such as risky decision making,[52] moral reasoning,[53] and humor appreciation[54] and that it results in reduced ability to differentiate odors.[55]

Although such results suggest a relationship between the absolute level of activity in specific brain regions and cognitive performance, the relationship between regional brain activation and performance is not always straightforward. For example, data from functional magnetic resonance imaging (fMRI) studies have revealed not only that sleepiness-related performance deficits are associated with region-specific reductions in activation but also that the ability to maintain performance near baseline levels on specific tasks during sleep deprivation is associated with relative activation of cortical regions that were not significantly activated during performance of the same task in the well-rested state.[56,57] Those individuals whose fMRI images revealed such "new" regional relative activations during sleep deprivation were better able to maintain task performance than those individuals whose fMRI images failed to reveal similar new activations. These findings suggest that individual differences in recruitment of resources from other brain regions may underlie task-specific individual differences in resilience to sleep loss.

However, results from subsequent fMRI studies have painted a more complex picture.[58,59] These results indicate that individuals possess varying degrees of redundant functional neuronal circuitry to process information in a given task context in the well-rested state.[10] The results also indicate that sleep deprivation reduces the number of functional neuronal circuits available.[10,59] These findings suggest an alternative explanation for task-specific individual differences in resilience during sleep loss. Namely, individuals with the greatest baseline redundancy in functional neuronal circuits for the task at hand are the ones who can tolerate the greatest reduction in available circuits during sleep deprivation and are therefore the most resilient to performance impairment from sleep loss.[10,60]

Paradoxically, the relationship between sleepiness and performance may also be investigated during the first several minutes after awakening. This immediate postawakening period is characterized by profound sleepiness and performance deficits, which rapidly dissipate over about 20 minutes of continuous wakefulness.[61,62] This phenomenon is referred to as "sleep inertia." Sleep inertia constitutes a state of sleepiness and performance deficits not necessarily associated with sleep loss. Studies during the immediate postawakening period characterize a sleepy, yet sleep-satiated, brain that is ascending toward alertness.[63]

It is informative to compare regional cerebral blood flow patterns, as measured with positron emission tomography, during sleep deprivation, non–rapid eye movement (NREM) sleep, and sleep inertia. Relative to well-rested wakefulness, sleep-deprived wakefulness is characterized by global reductions in brain activity, with the greatest reductions evident in heteromodal (primarily prefrontal) association cortices and thalamus.[51] A similar pattern is evident when comparing NREM sleep to well-rested wakefulness: global deactivation (albeit of a greater magnitude than during sleep-deprived wakefulness), with the greatest deactivations evident in anterior (prefrontal) cortical regions and centrencephalic regions, including the thalamus.[64] In contrast, the immediate postawakening, sleep inertia period is characterized by bidirectional changes in regional cerebral blood flow, with waxing activity in anterior cortical regions and waning activity in centrencephalic regions across the first 5 to 20 minutes of wakefulness[65] (Figure 71-3).

Because deactivated anterior/prefrontal cortex is the only common finding among these three states (sleep-deprived wakefulness, NREM sleep, and sleep inertia), it has been surmised that activity in the prefrontal cortices is a critical determinant of sleepiness and its attendant performance deficits. Taken together, findings from neuroimaging studies suggest that performance impairment on tasks affected by fluctuations in sleepiness vary as a function of regional brain activity and interregional connectivity patterns, especially those functional connections involving prefrontal cortices.

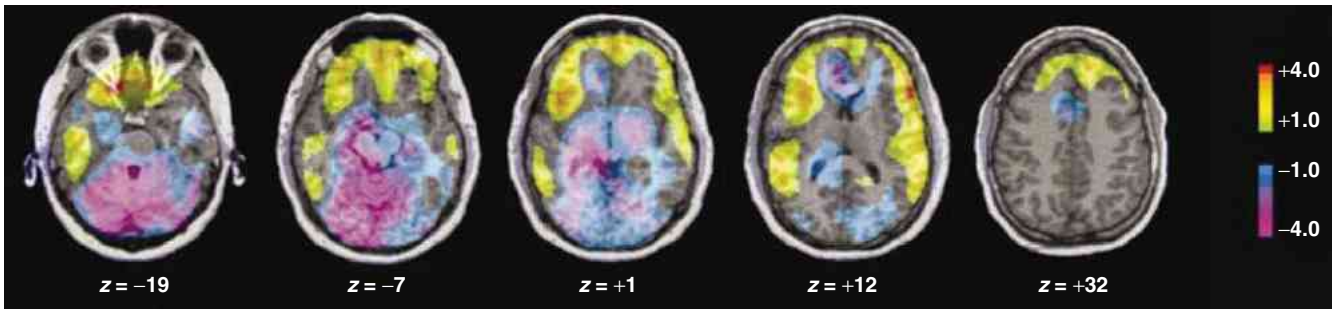


Figure 71-3 Brain map depicting changes in regional cerebral blood flow during the sleep inertia period between 5 and 20 minutes after awakening from stage 2 sleep. Color-coded values are Z-scores representing the significance level of changes in proportionally normalized regional cerebral blood flow in each voxel between scans acquired at 20 versus 5 minutes after awakening. The range of Z-scores is coded in the color table, with *red* designating Z-scores of greater than +4.0 and *purple* designating Z-scores of less than –4.0. Positive scores reflect increases in relative blood flow from 5 to 20 minutes after awakening; negative scores reflect concomitant, relative decreases during this period. (From Balkin TJ, Braun AR, Wesensten NJ, et al. The process of awakening: a PET study of regional brain activity patterns mediating the reestablishment of alertness and consciousness. *Brain* 2002;125:2308–19, with permission.)

CONCLUSION

Sleepiness contributes to hundreds of thousands of road accidents each year[66] and has been cited as a contributing factor in occupational disasters such as the meltdown of the Chernobyl nuclear reactor, the grounding of the Exxon Valdez oil tanker, and the decision to launch the ill-fated Challenger space shuttle. This chapter provides an overview of current thinking regarding the combined effects of sleep loss, the circadian rhythm of alertness, and time on task on performance. To the extent possible, the separate and combined effects of these factors should be considered when determining the risk for errors and accidents in operational environments and should be weighed against the costs and potential consequences of errors and accidents when determining and implementing fatigue management strategies.

Present-day hours-of-service regulations do not adequately account for all of these risk factors and are therefore of limited utility for promoting safety and productivity in operational environments. Recent developments that integrate mathematical performance prediction models into regulatory frameworks for hours of service can be used to address at least some of these issues (see Chapter 72).

In addition, indirect effects of sleep loss on performance and accident risk should be noted. Sleepiness tends to promote impulsivity and risk taking[67,68] and impairs self-monitoring of performance.[69,70] The extent to which these aspects of impaired judgment contribute to errors and accidents is unknown, although risk taking and sleep loss have been noted to be a potentially deadly combination in young male drivers.[71,72]

At the level of individuals, the consequences of sleepiness and the need to intervene remain difficult to gauge. A clinician seeing subjectively sleepy patients should discuss with them the risk for errors and accidents. The clinician may also try to estimate risk level by asking questions about the following:

- Type of job
- Nature of job (e.g., safety sensitive, mission critical)
- Level of exposure
- History of incidents or "near misses"
- Safety measures that could be put in place (e.g., rest breaks, ergonomic tools)
- Habitual sleep-wake-work schedules
- Tasks, hobbies, and circumstances that intrude on time for sleep
- Length of commute to and from work
- Possible consequences of adverse events

It is important to explain that a patient's own subjective evaluation of his or her impairment resulting from sleepiness is likely to be inaccurate, with a tendency to overestimate one's own capabilities.[4,73] Furthermore, because of the stochastic nature of human error, past performance or safety does not guarantee future performance or safety. Bringing these issues to the attention of patients may prompt them to make behavioral changes and thereby helps to reduce their risk for sleep loss-related errors and accidents.

CLINICAL PEARL

Interactions among multiple factors, such as recent and long-term sleep-wake history, circadian rhythm, time-on-task effects, and individual differences in vulnerability to sleep loss, lead to sleepiness-related performance deficits and contribute to errors and accidents. Such factors should be taken into consideration when devising operational work-rest schedules, implementing fatigue management strategies, and assessing patients who complain of sleepiness or fatigue.

SUMMARY

During the past decade, significant progress has been made toward elucidation of the physiologic basis of sleepiness and the neurobiologic and neurobehavioral correlates of sleepiness-induced performance deficits. The deleterious effects of sleep loss on various aspects of performance are well known and well documented. Nonetheless, inadequate sleep remains virtually ubiquitous and exacts an inestimable toll on society through increased errors and (sometimes catastrophic) accidents as well as reduced efficiency and productivity.

Sleepiness manifests as a function of the combined effects of the homeostatic pressure to initiate sleep and the circadian rhythm of alertness, with considerable trait-like individual differences in the ability to maintain performance under sleepiness-inducing conditions. Although it is not possible to compare the relative extent to which specific cognitive abilities or processes are decremented by sleepiness, it is known that

sleepiness is (for example) associated with an increased tendency toward risk taking and overestimation of one's own performance capabilities, deficits in judgment that can clearly contribute to the likelihood of errors and accidents.

Although sleep loss in real-world operational environments is typically characterized by continuous (or near-continuous) work, there is a dearth of studies in which the interaction between sleep loss and cognitive load or time on task has been addressed specifically or adequately. Nevertheless, the evidence that does exist provides compelling evidence that performance under operational conditions is the result of an interaction between sleepiness and cognitive workload. Future progress will be realized with an improved understanding and appreciation of the interactions between sleep loss and task-related variables, an improved ability to measure these effects, and implementation of alertness and performance management strategies that, unlike current hours-of-service regulations, effectively account for these effects.

Selected Readings

Balkin TJ, Braun AR, Wesensten NJ, et al. The process of awakening: a PET study of regional brain activity patterns mediating the reestablishment of alertness and consciousness. *Brain* 2002;**125**:2308–19.

Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res* 2003;**12**:1–12.

Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk TH, editor. *Sleep, sleepiness and performance*. Chichester, UK: John Wiley & Sons; 1991. p. 97–128.

Doran SM, Van Dongen HPA, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol* 2001;**139**:253–67.

Drummond SPA, Brown GG, Gillin JC, et al. Altered brain response to verbal learning following sleep deprivation. *Nature* 2000;**403**:655–7.

Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. *Semin Neurol* 2005;**25**:117–29.

Oonk M, Tucker AM, Belenky G, Van Dongen HPA. Excessive sleepiness: determinants, outcomes, and context. *Int J Sleep Wakefulness* 2008;**1**:141–7.

Tsafrir Z, Korianski J, Almog B, et al. Effects of fatigue on residents' performance in laparoscopy. *J Am Coll Surg* 2015;**221**:564–70.

Van Dongen HPA, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;**26**:117–26.

Venkatraman V, Chuah YML, Huettel SA, Chee MWL. Sleep deprivation elevates expectation of gains and attenuates response to losses following risky decisions. *Sleep* 2007;**30**:603–9.

Wesensten NJ, Belenky G, Thorne DR, et al. Modafinil vs. caffeine: effects on fatigue during sleep deprivation. *Aviat Space Environ Med* 2004;**75**:520–5.

Wilkinson RT. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *J Exp Psychol* 1961;**62**:263–71.

A complete reference list can be found online at ExpertConsult.com.

Sleep and Performance Prediction Modeling

Steven R. Hursh; Thomas J. Balkin; Hans P.A. Van Dongen

Chapter 72

Chapter Highlights

- Sleep loss and circadian misalignment cause neurobehavioral performance impairment and contribute to errors, incidents, and accidents. Biomathematical models may be used to help manage fatigue risk in operational settings.
- Most currently available biomathematical models predict sleepiness or neurobehavioral performance impairment based on three basic components: circadian variation, homeostatic sleep-wake regulation, and sleep inertia.
- Recent advances in biomathematical modeling include accounting for nonlinear interaction between circadian and sleep-wake homeostatic processes, prediction of the cumulative effects of chronic sleep restriction on neurobehavioral performance, and extension of model predictions from group averages to individuals.
- Biomathematical models are gaining acceptance in operational settings as components of fatigue risk management systems.

Alertness and cognitive performance vary as a function of time of day, time awake, and a variety of situational factors. Prescriptive rules for hours of service do not capture these intrinsic temporal influences and therefore do not fully protect against performance deficits. Accordingly, a number of "biomathematical models of fatigue" have been (and continue to be) developed for anticipating and avoiding performance impairment; targeting optimal timing and dosing of countermeasures; improving work schedules, productivity, and safety; and informing accident investigations.

Most biomathematical performance prediction models are based on quantification of two primary and one secondary neurobiologic processes underlying variation in alertness and performance over time: the circadian process and the homeostatic process (see Chapter 37) as well as the sleep inertia process.[1-7] A variety of other factors affecting alertness and performance, including stimulant use, light exposure, distractions, and motivation, are currently not accounted for in most models. Instead, these factors are generally considered to effect transient deviations from the general trend that temporarily mask but do not alter the circadian or homeostatic processes that drive those trends. As such, current biomathematical models are useful for predicting normative performance for specific times of day, durations of time awake, and recent sleep history, with the understanding that actual, moment-to-moment alertness and performance of individuals may vary as a function of a broad array of factors.

One of the primary processes on which physiology-based, biomathematical models are built is the circadian process, which influences both performance and sleep regulation. This process is presumed to reflect physiologic activation apparent in measurements of body core temperature and certain hormones (such as melatonin). Performance increases (improves) with circadian activation, whereas propensity to sleep decreases (and vice versa).

The other primary process tracks the brain's level of sleep debt (level of physiologic need for sleep), which in turn affects alertness and performance capacity such that it is depleted while awake and replenished while asleep. This sleep-wake homeostatic process is dependent on the number of hours of recent sleep obtained (prior day), the number of hours of wakefulness (time since awakening), and current overall sleep debt (which is dependent on the amount of sleep loss that has accumulated over days, weeks, or perhaps longer). Performance capacity decreases (degrades) with homeostatic depletion, and propensity to sleep increases. The circadian and homeostatic processes continuously interact to influence observed performance and propensity to sleep.

A third process, called sleep inertia, reflects the temporary degradation in performance that is seen immediately after awakening. The magnitude of this effect depends primarily on depth of sleep at the time of awakening (which is itself affected by both the circadian and homeostatic processes). Sleep inertia dissipates with time since awakening, and recovery of performance is typically essentially complete within about 20 minutes of awakening. [8]

Recent advances have refined the mathematical representations of these three components, have added interaction terms among the components, and have added factors to account for chronic sleep restriction and dynamic variations in circadian phase (including the influence of light exposure). Additionally, methods have been developed to estimate patterns of sleep under specific work schedules.[5,9-13]

COMPONENTS OF THREE-PROCESS BIOMATHEMATICAL MODELS

To illustrate the modeling approach shared in general form by several three-process models,[2-5] the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model will be used in this

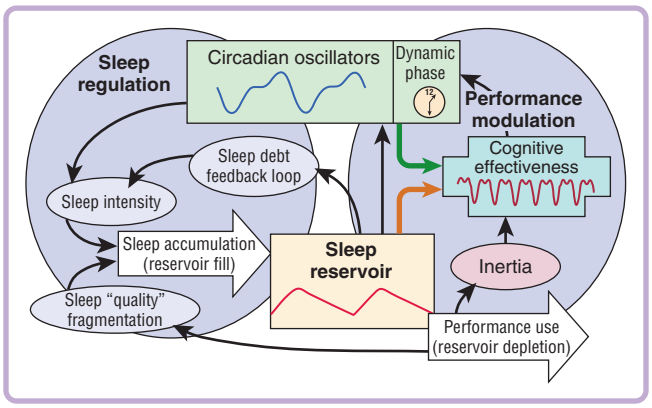


Figure 72-1 Schematic of the Sleep, Activity, Fatigue and Task Effectiveness (SAFTE) model.[5,9]

chapter. The components of this model are diagrammed in Figure 72-1.

Circadian Oscillator

Performance while awake and the drive to sleep are both controlled, in part, by a circadian process.[3,14,15] For a person entrained to a sleep period of 11 PM to 7 AM, performance reaches a peak in the early evening at approximately 7 PM and—if it could be measured during sleep—would fall to a minimum at approximately 4 AM. There is a secondary peak of performance in the morning at about 10 AM, and a secondary minimum (dip) in the early afternoon at about 2 PM. Negatively correlated with this alertness pattern is a tendency to fall asleep, which reaches a peak at about the same time performance and alertness reach a trough.

The existence of both a major and a minor peak in performance with corresponding troughs between them is often modeled as the result of two linked harmonic oscillations (cosine functions), one with a period of 24 hours and the other with a period of 12 hours. This results in a combined function of the form shown in Figure 72-2.[16] More dynamic models of the circadian process use a limit cycle or Van der Pol oscillator that yields a function of similar form.[17]

A simplifying assumption of most biomathematical models is that the same underlying arousal oscillator drives the variations in both cognitive performance and sleep propensity. The amplitude of the circadian process is dependent on the level of sleep debt modeled by the sleep-wake homeostatic process.[6,18,19] The phase (timing) of the circadian process is driven largely by environmental factors, most notably the timing of sunlight exposure,[12,20] but also has an individual trait component expressed as morningness and eveningness.[21]

Sleep-Wake Homeostatic Regulation

The control of sleep and its influence on cognitive capacity is usually modeled as a homeostatic process.[15,22-24] One way to conceptualize this process is through a sleep reservoir (see Figure 72-1). A fully rested person has a certain performance capacity (represented by the sleep reservoir capacity). During wakefulness, units are subtracted from the sleep reservoir according to a use function. During sleep, units are added to the sleep reservoir and the capacity to perform and be alert is restored.

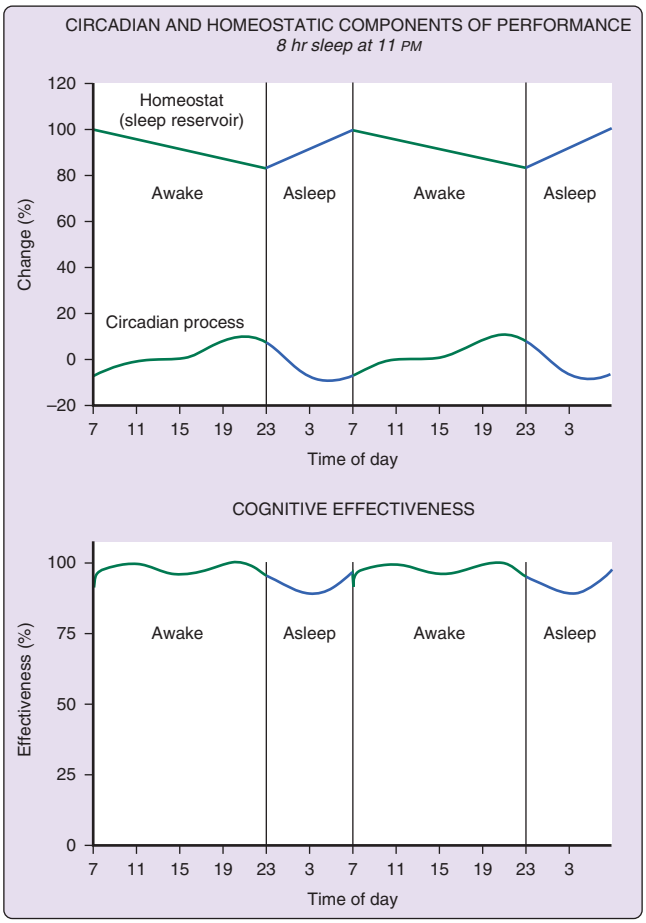


Figure 72-2 The sleep-wake homeostatic (*top*) and circadian (*middle*) processes affecting performance (cognitive effectiveness, *bottom*). Graphs are based on SAFTE model simulations of a 16-hour wake/8 hour sleep schedule, with sleep starting at 11 PM. (Performance predictions during sleep are suppositional because they cannot actually be observed during the sleep state.)

The rate of unit accumulation during sleep is driven by two factors: circadian variation and current overall sleep deficit, which is the shortage in the current level of the reservoir. Because the rate of unit accumulation is in part regulated by the current level of the reservoir, the process is homeostatic. The top panel of Figure 72-2 shows the decrease in performance units during wakefulness and the increase during sleep.

Results from laboratory studies have revealed that chronic sleep restriction (anything less than approximately 8 hours per 24 hours) leads to cumulative alertness and performance deficits.[25-27] For sleep durations of more than 4 hours per night, performance declines across days but eventually reaches a suboptimal plateau or equilibrium level.[13,26] The achievement of this suboptimal equilibrium state implies a feedback-modulated control system. Modeling this phenomenon has led to much new activity in model development.[5,11,13,28,29] In the SAFTE model, this phenomenon is implemented by reducing the reservoir capacity each time sleep restriction occurs (see later).

Accumulation of performance units does not start immediately on falling asleep. There is a 5-minute delay from sleep onset until performance units begin to accumulate. This delay

accounts for the approximate time required to return to restorative sleep following a brief arousal and results in a penalty during recuperation (see Figure 72-1) in an environment that leads to frequent interruptions (sleep fragmentation).

Sleep Inertia

A third factor diagrammed in Figure 72-1 is the transient performance impairment that often occurs immediately following awakening (i.e., sleep inertia).[8,22] It is typically modeled as an exponentially decreasing performance deficit.[5,30,31] Because of the relatively short duration of sleep inertia, it is relevant mainly in operational contexts in which individuals may be required to perform immediately on awakening (e.g., first responders), and therefore it is not included in some biomathematical models.[6,10,32]

COMBINED EFFECTS: PREDICTED PERFORMANCE

Alertness and performance are modeled as the combined effect (in most models calculated as the sum) of the mathematical functions for the circadian oscillator, the sleep-wake homeostatic reservoir, and sleep inertia immediately following awakening. The sum of the homeostatic process with the circadian process produces a performance function with two nadirs, one in the late night to early morning and a smaller one in the early afternoon. These two processes are shown in the top panel of Figure 72-2. Their combined effects on performance are shown in the bottom panel of Figure 72-2.

During the first part of a daytime waking period, the circadian process increases (growing activation) and the homeostatic process decreases (reservoir depletion). When summed, the changes in these two processes are approximately offsetting each other, leading to relatively constant daytime performance. After the 7 PM (approximately) evening peak in the circadian process, both the circadian process and homeostatic process decline, leading to a precipitous loss of alertness and performance at approximately 11 PM that facilitates sleep onset.

During nighttime sleep, the circadian process continues to decrease (declining activation, which facilitates sleep maintenance) and the homeostatic process increases (restoration of reservoir level). After the early morning nadir in the circadian process, both the circadian and homeostatic processes rise, culminating in spontaneous awakening at approximately 7 AM.[1] Combined (summated), the two processes produce the performance profile depicted in the bottom panel of Figure 72-2.

QUANTIFYING RECUPERATION DURING SLEEP

During sleep, the sleep reservoir is replenished. In the SAFTE model, the rate at which the reservoir is filled depends on the prior long-term sleep debt (as reflected in the current level of the reservoir). A large reservoir deficit leads to an increased rate of replenishment; a small deficit leads to a reduced rate of replenishment. The rate of replenishment also appears to be modulated by the circadian variation in sleep propensity[33] and may follow the circadian pattern of sleep propensity[34,35] (the inverse of the circadian pattern of activation in the top panel of Figure 72-2). The net effect is a near-linear homeostatic reservoir accumulation through a normal night (see top panel of Figure 72-2), which is in line with results that show an approximately linear relationship between sleep duration and performance recuperation.[25,36]

The homeostatic regulation of sleep[37] leads to equilibrium states under certain conditions of chronic sleep restriction.[8] If a subject is scheduled to take less than an optimal amount of sleep each night (defined here as approximately 8 hours) and instead obtains, for example, 5 hours per day, the reservoir initially loses more units during the waking period than are made up during the sleep period. This results in a sleep debt at the end of the sleep period that accumulates over days.[24,25] However, because the rate of accumulation during sleep increases with sleep debt, eventually the rate of accumulation increases so that 5 hours of sleep makes up for 19 hours of wakefulness. At this point, the reservoir reaches an equilibrium state over days, and no further debt is accumulated. However, the reservoir remains at less than 100% replenished as long as the person remains on this reduced sleep schedule.

The sleep-wake homeostatic process is not infinitely elastic, and results show that there is a limit to the rate of accumulation and thus to whether a given sleep-wake schedule permits equilibrium. In the SAFTE model, any schedule that provides less than 4 hours of sleep per day will not reach an equilibrium state, and performance capacity will gradually deplete entirely. Results of biomathematical model analyses have provided further evidence that there is such a bifurcation (i.e., qualitative change in behavior) for the homeostatic process when sleep is reduced to below approximately 4 hours.[8]

The long-lasting effects of chronic sleep restriction on performance impairment and reduced capacity to recuperate[24,25] suggest that some aspect of sleep-wake homeostasis undergoes a gradual change that is slow to recover.[2,27] This phenomenon has prompted modelers to revise the sleep-wake homeostatic process by adding a long-term process.[2,8,27,28,38,39] Within the context of the SAFTE model, this process is instantiated as a gradual downregulation of the sleep reservoir capacity during chronic sleep restriction. When sleep durations return to the nominal value of approximately 8 hours per day, the downregulation of the reservoir capacity is gradually reversed.

The implication is that after a period of chronic sleep restriction, it takes multiple days of recovery sleep to restore performance to baseline levels.[8,24] Laboratory evidence has shown that extending sleep to more than 8 hours per day for an extended period of time provides some resilience to the effects of subsequent sleep restriction.[33] This recent finding implies that it should also be possible to moderately increase the capacity of the sleep reservoir.[6] In the SAFTE model, more than 8 hours of sleep per day will hasten the recovery of the reservoir set point to the full value but will not increase capacity beyond the nominal limit for a person consistently sleeping 8 hours per day, so the findings that sleep extension confers performance benefits during subsequent sleep restriction may require further model refinement (as has already been pursued in another model[6]).

The time of day also affects the recuperative potential of a sleep period. For an individual given 8 hours of sleep per day from 11 AM until 7 PM, waking performance reaches a peak about 4 hours after awakening (9 PM). It then rapidly declines during the late night and early morning hours to a

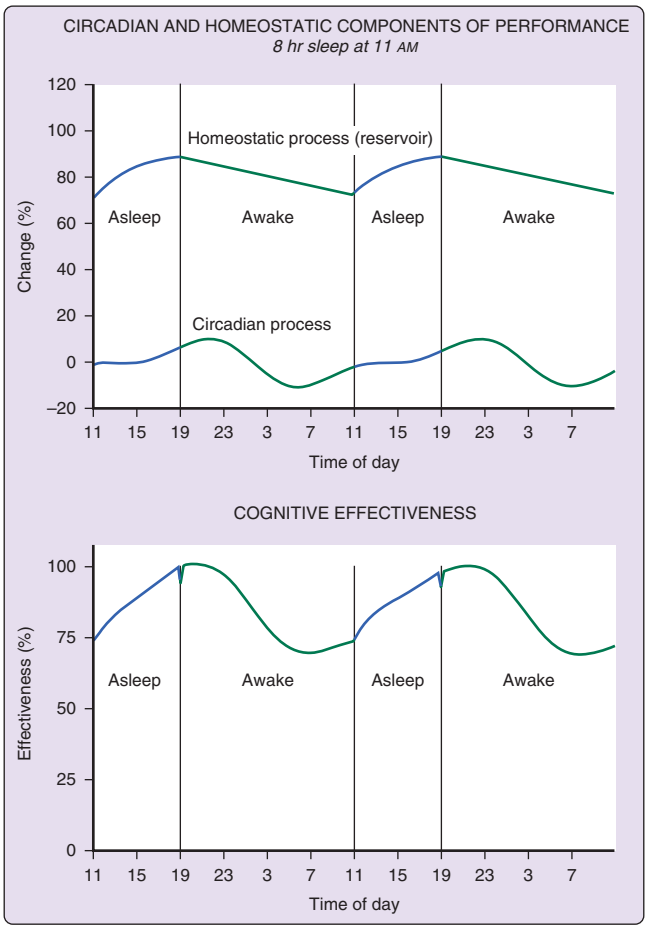


Figure 72-3 The sleep-wake homeostatic (*top*) and circadian (*middle*) processes affecting performance (cognitive effectiveness, *bottom*) when 8 hours of sleep occurs during the day starting at 11 AM. Graphs are based on SAFTE model simulations.

deep minimum at about 5 AM, as illustrated in the bottom panel of Figure 72-3. This pattern is substantially different from the performance pattern seen after nighttime sleep (see Figure 72-2) for two reasons. First, owing to the circadian rhythm in sleep propensity, daytime sleep periods exhibit less physiologic sleep and therefore less homeostatic reservoir replenishment, that is, the reservoir level does not return to 100% (see Figure 72-3, top panel). Second, circadian activation reaches its minimum in the early morning hours, at the same time that the sleep-wake homeostatic reservoir is increasingly depleted.

This pattern has negative implications for performance under shift schedules that require daytime sleep. It is well documented that most mistakes on the night shift occur during the early morning hours,[34,35] and biomathematical models predict this outcome. Note also that the pattern illustrated in Figure 72-3 is likely to be a "best case scenario" in that it was presumed that 8 hours of sleep was achieved during the 8-hour daytime sleep period. However, results indicate that shift workers seldom achieve 8 hours of sleep during the daytime hours even if they are sleep deprived. That is, sleep propensity (debt) tends to be insufficient to completely overcome countervailing circadian wake drive and environmental factors such as daylight and social activities.[36,37]

SLEEP ESTIMATION

The accuracy of biomathematical model predictions is dependent on accurate measurement or prediction of sleep times and durations. Sleep timing and duration can be estimated mathematically using the two-process model of sleep regulation[38] (see Chapter 36) and detailed models of sleep neurobiology.[40,41] However, these models do not account for social and other nonbiologic constraints on sleep.[42] For this reason, in situations in which biomathematical models are applied to work schedules, it may be necessary to make use of an alternative method for estimating potential sleep given the available sleep opportunities.

One approach to developing a sleep estimation algorithm involves modeling the observed likelihood of sleep from results of field studies in the operational environment in which sleep was recorded by means of diaries or actigraphy.[43] In the SAFTE model, it is assumed that the occurrence of major sleep episodes is largely the result of a decision to go to bed, conditioned by the person's typical need for sleep and opportunities to sleep, reflective of social, cultural, and professional factors. A sleep estimation algorithm simulates the decision process that governs when a person chooses to sleep. Sleep is assumed not to occur during work or commuting to work because these periods limit sleep opportunities.

Within sleep opportunities, further decisions about when to sleep are made based on sleep habits. These habits might include such factors as the preferred bedtime, the normal duration of sleep on work and rest days, the minimum duration of a nap, and any time during the day that a person normally uses for personal activities and not sleep. The sleep estimation algorithm combines these parameters with circadian factors (described earlier) to generate estimates of the timing and duration of sleep during opportunities afforded by the work schedule.[11,42]

In recent studies with airline pilots (36 domestic and 15 international) and shift workers (147 workers with fixed and rotating shift patterns with irregular shift extensions and overtime), the algorithm was trained on actigraphically recorded sleep-wake to achieve 85% to 87% accuracy for predicting sleep and wakefulness for individual subjects. The algorithm estimated average total sleep per 24 hours to within 1 minute of the actual group average.[44] Although there remained considerable individual differences in sleep patterns, from an aggregate risk assessment perspective, the SAFTE model predictions based on such data were as good as inputting actual wrist actigraphy data. The optimal settings of the algorithm were nearly identical for these two very different work populations, a function of the fact that the sleep-wake behaviors of all humans are governed by essentially the same physiologic and environmental factors.

CIRCADIAN PHASE SHIFTING

When people move to another time zone or alter work patterns so that sleep and wake consistently occur systematically at new times of day, the internal circadian oscillator that modulates alertness and performance shifts to this new schedule. During this period, individuals experience performance degradation, disrupted mood, and feelings of dysphoria, collectively termed jet lag (see Chapters 35 and 37).[45-48] Several models simulate this process and adjust the phase of the

circadian rhythm to coincide with the new activity pattern. This feature is critical for the accurate prediction of the effects of moving to a new time zone or changing to a new and regular work schedule, such as changing from the day shift to the night shift.[37]

The factors that mediate phase shifts vary across models. Research suggests that a major driver of circadian phase is exposure to sunlight or bright light[49]—especially the blue part of the spectrum[50] (see Chapter 35). In some models, direct measurements of light exposure are required as input and are used to predict changes in circadian phase.[3] Other models use a surrogate for sunlight measurement.[5,9,11,51]

In the SAFTE model, it is assumed that the probability of bright light exposure is proportional to the time spent awake during daytime hours. Hence, the model adjusts the rate of phase change based on the proportion of waking time that occurs during daytime hours. This simplification eliminates the need to take continuous measurements of light to drive the circadian process, but it also means that there is no mechanism to input light exposure as a deliberate countermeasure.[52-54] Given that the interaction between the circadian oscillator and the sleep-wake homeostatic process appears to be bidirectional,[18] the timing of the sleep period may in and of itself also contribute to circadian adjustment.[11,55]

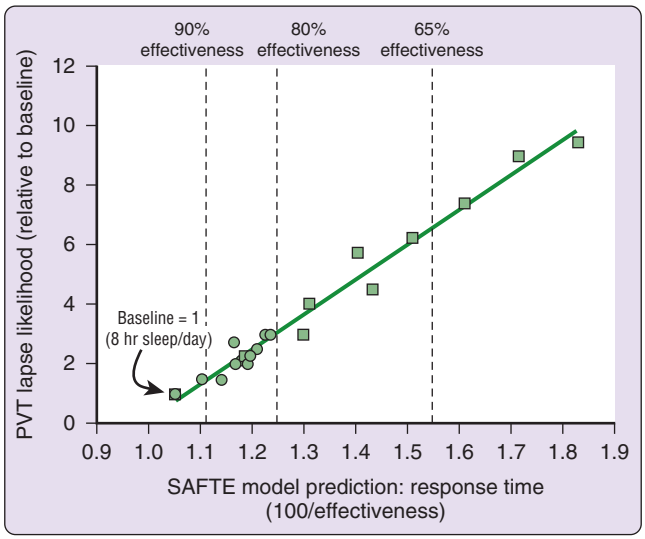


Figure 72-4 Linear relationship of lapse likelihood on a psychomotor vigilance test (PVT) to predicted response time (100/effectiveness) from the SAFTE model. PVT data were recorded during days of chronic sleep restriction to 3 hours per day (○) or 5 hours per day (■).[26]

PERFORMANCE PREDICTION

As illustrated in Figures 72-2 and 72-3, biomathematical performance prediction models simulate the dynamic interplay of circadian variation in alertness and sleep propensity, homeostatic sleep-wake regulation supporting performance capacity, and short-term sleep inertia following awakening. Specific models differ in the manner in which these three factors are represented and mathematically combined. The predictions of such models also depend on the metrics (units) of the models. For example, the units expressed by the SAFTE model are percent changes in cognitive speed from baseline performance when fully rested (i.e., 8 hours of sleep per 24 hours) and diurnally oriented (i.e., sleep from 11 PM to 7 AM). Cognitive speed corresponds to speed of response on a psychomotor vigilance test (PVT).[56] Studies that have included other cognitive tests along with the PVT have shown them to be highly correlated, with the PVT showing the greatest sensitivity to daily sleep amounts.[57]

When predicting group averages (but not individual performance[41]), translation functions can be used to calculate other performance metrics such as lapse likelihood, reaction time, and mean cognitive throughput (correct responses per unit of time) on other cognitive tests, such as serial addition-subtraction, choice reaction time, logical reasoning, and code substitution. Figure 72-4 shows output from a translation function that expresses performance as PVT lapse likelihood, which increases with time awake and mirrors response speed.

To illustrate the potential of these models to accurately predict performance under a variety of sleep-wake-work schedules, two examples are offered. The first example concerns performance changes on a variety of cognitive tasks during a period of total sleep deprivation in a laboratory.[58] As depicted in Figure 72-5, these measures were compared with model predictions and found to conform to the group means of the performance observations with 98% of the variance explained.

The second example deals with chronic sleep restriction as encountered in demanding schedules that allow for less than optimal nightly sleep durations over extended periods of time. Figure 72-6 shows performance observations (daytime averages) obtained during a laboratory dose-response study involving 7 days of sleep restricted to 3, 5, 7, or 9 hours per day, preceded by 3 baseline days and followed by 3 recovery days, each with 8 hours of time in bed per night.[26] As shown in the figure, the group means of the objective performance measures were compared with the predictions of the model and found to explain 94% of the variance.

Biomathematical models may predict performance on a selected, standardized performance task, such as the PVT, or on an abstract performance metric, such as a sleepiness scale from, say, 0 to 100. Mathematical modeling may be applied to make absolute predictions, which are compared against a threshold (e.g., that distinguishes acceptable from unacceptable performance[59]). In these applications, the specific performance metric used and the value of the threshold are crucially important.

More commonly, biomathematical modeling is used to make relative comparisons of predicted performance during two or more alternative sleep-wake or work scheduling options. In such cases, modeling enables selection of the scheduling alternative that is less fatiguing, or less fatiguing as well as more productive or cost-effective.[60] In this application, the specific performance metric used is not critical. However, there is debate about whether "better versus worse" should be quantified based on maximum or minimum level of the metric (i.e., highest risk) or on some other calculation such as a combination of both level and duration of performance impairment (i.e., risk exposure).[61]

Ongoing mathematical modeling efforts are aimed toward predicting, in greater detail, the effects of sleep loss on performance on a diverse range of tasks.[62,63]

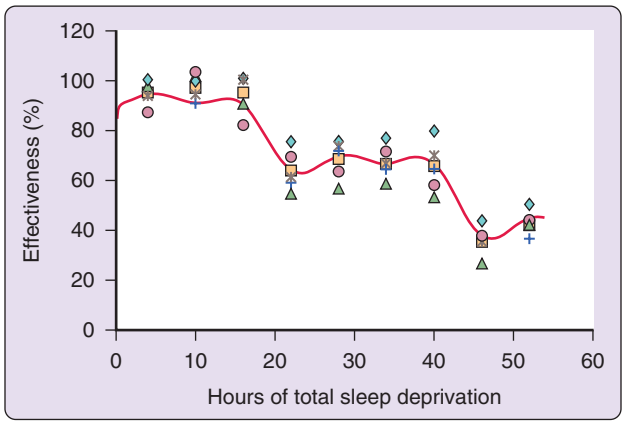


Figure 72-5 Fatigue model predictions for group-average (n = 12) cognitive performance during 54 hours of total sleep deprivation (beginning at 7 AM) compared with actual observations[58] for several cognitive measures (different symbols) and mean cognitive performance (*squares*) expressed relative to baseline. Model predictions were made with the SAFTE model (*solid line*) with a reservoir depletion rate of 1.1% per hour.[5] ◆, Serial reaction time; +, decoding problem performance; ж, encoding problem performance; ●, auditory vigilance; ▲, logical reasoning.

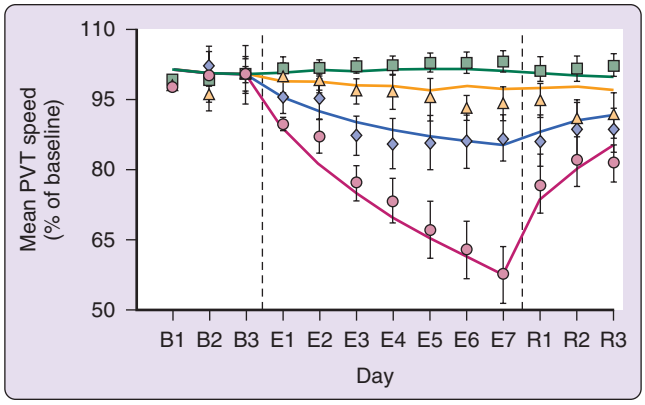


Figure 72-6 Fatigue model predictions for group-average psychomotor vigilance test (PVT) performance (averaged within each day) across 3 baseline days (B1–B3) with 8 hours time for sleep; 7 experimental days (E1–E7) with daily sleep restricted to 3 hours (●, n = 13), 5 hours (◆, n = 13), 7 hours (▲, n = 14), or 9 hours (■, n = 16); and 3 recovery days (R1–R3) with 8 hours time for sleep.[26] Model predictions were made with the SAFTE model (*solid line*).[5]

MODELING APPLIED TO OPERATIONAL SETTINGS

Most biomathematical performance prediction models are optimized to predict changes in cognitive performance as measured by standard laboratory tests performed under controlled laboratory conditions. It is assumed that these tests measure changes in the fundamental capacity to perform a variety of tasks that rely, more or less, on the cognitive skills of attention, detection, discrimination, reaction time, mental processing, reasoning, or decision making. However, the extent to which any specific operational task relies on these cognitive skills is generally not known. Thus deficits in cognitive capacity seen in the laboratory may not always predict deficits in the capacity to perform different operational tasks.

It is reasonable to assume, however, that changes in task performance will be correlated with changes in underlying

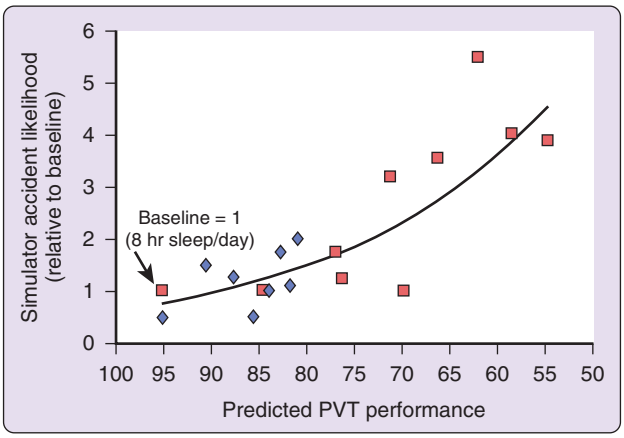


Figure 72-7 Systematic relationship (*solid line*) between model predictions optimized to predict psychomotor vigilance test (PVT) performance and group-average accidents observed in a driving simulator.[64] Model predictions were made with the SAFTE model.[5] ■, 3 hours sleep per day (n = 13); ◆, 5 hours sleep per day (n = 13).

cognitive capacity and therefore that changes in one task will be correlated with changes in another task if both tasks tap the same underlying capacity. The validity of this assumption has been demonstrated in job task simulator experiments. For example, Figure 72-7 illustrates that there is a systematic relationship between predicted performance for the PVT (generated from SAFTE) and driving simulator accident data from subjects given varying amounts of daily sleep for a week.[64]

Biomathematical modeling (prediction) of operational task performance could be used to design better work schedules, alter the timing of critical tasks to coincide with periods of predicted optimal performance, or time the application of countermeasures to prevent errors and accidents. Unlike prescriptive hours-of-service regulations, these models provide physiology-based, flexible, and quantifiable ways to optimize safety and performance in operational environments.

Current uses of biomathematical modeling alone or as a component of fatigue risk management systems in operational settings include the following:

- Prediction of risk for impairment associated with a given work-rest or wake-sleep schedule, allowing for selection of optimal duty scheduling
- Guidance for effective implementation of countermeasures
- To inform, supplement, or substitute hours-of-service policies and regulations
- As an aid in post hoc incident or accident investigations to determine the extent to which fatigue may have been a contributing factor
- As an educational tool for understanding the operational consequences of insufficient sleep and circadian misalignment

A relatively new application of biomathematical modeling involves the use of model predictions as a surrogate for performance measurements in settings where sleep can be measured (e.g., with actigraphy) but measuring performance would interfere with performance of critical tasks.

Biomathematical modeling plays a key role in the approval of exceptions to flight and duty time limitations under the fatigue risk management system approval process of the U.S. Federal Aviation Administration (FAA) as implemented in

2014.[65] Using a multistep approval process, the FAA accepts applications for exceptions to prescriptive flight and duty time limitations that combine longer duties with offsetting rest and time-of-day limitations. To obtain final approval, the aviation certificate holder must collect data to verify that sleep and performance under the alternative means of compliance (AMOC) are equivalent to what might be observed under the prescriptive rules.

To gain temporary approval to conduct operations to collect the necessary data, the aviation certificate holder is allowed to provide biomathematical modeling results or actual sleep and performance data (when available) to the FAA in order to establish a safety case for the AMOC. The FAA bases its final approval for the AMOC on statistical equivalence[66,67] of actual sleep and performance findings compared with a control condition (i.e., one that meets the prescriptive limitations). The integration of biomathematical modeling into a regulatory framework that allows for AMOC reflects the demonstrated utility of modeling for simultaneously enhancing both fatigue risk management or operational safety and efficiency of flight operations.

INDIVIDUAL DIFFERENCES

Individual differences in responses to sleep loss and circadian timing (e.g., in shift work)[68] represent a substantial source of error variance in most biomathematical models, which were developed first and foremost to predict group-average performance. Individual differences can be accounted for broadly by reflecting the expected distribution of performance outcomes in the model predictions, based on the individual variability of performance changes observed in laboratory studies.[11,69] This approach provides information on the range of effects to be expected under a specific sleep-wake-work schedule, but it does not provide information about where any specific individual would fall within that range.

Individual differences in vulnerability to sleep loss have been demonstrated to be traitlike[70] and thus predictable on a subject-specific basis.[71] Two complementary strategies could be implemented to enhance individual-level prediction.[41] The first is to incorporate predictors of performance vulnerability, such as morningness-eveningness,[72] basal sleep need,[73] or specific genetic polymorphisms,[74-76] into the model equations.[77] To date, none of these predictors has been demonstrated to account for a substantial portion of interindividual variability, but research efforts are ongoing[76,78] and novel candidate predictors continue to be identified.

The second strategy is to tailor the parameters of the model to the individual based on real-time or near-real-time performance observations obtained from the individual. Algorithms for this have been developed,[41,79,80] with confidence intervals indicating the statistical certainty level of the predictions.[79]

LIMITATIONS OF BIOMATHEMATICAL MODELS

Biomathematical modeling has become a valuable tool for fatigue risk management (see Chapter 71). However, it is also based on evolving science, and thus there are limitations that should be noted.[77,81,82] Some limitations to consider when evaluating a model include the following:

- Accuracy of model predictions relative to observations obtained from laboratory versus field studies
- Generalizability, that is, to what extent the model is applicable to sleep-wake-work scenarios not used in its development
- Balance between sensitivity and specificity, that is, the extent to which predictions are liberal or conservative with regard to identifying periods with meaningfully increased or decreased performance

These and other modeling issues were discussed at the international Fatigue and Performance Modeling Workshop (June 13 to 14, 2002, Seattle, Wash.) and were documented in the proceedings of that meeting.[83] Since then, biomathematical models have undergone further development, rendering some of the information in the proceedings outdated. Nevertheless, the volume continues to provide valuable documentation of biomathematical modeling as a developing science.

The accuracy of a model depends on the accuracy of the major inputs to that model. A key input is the pattern of sleep-wake under a given work schedule. In laboratory studies, precise polysomnographic quantification of sleep-wake is possible. Such data are seldom available in operational settings, but a variety of other techniques can be used to obtain sleep data, such as wrist actigraphy (see Chapter 73), sleep diaries, or an algorithmic sleep estimator (discussed previously). Although algorithmic sleep estimation can be useful for prediction of elevated accident risk,[84] the accuracy of such predictions depends on the accuracy of the sleep estimates that serve as input to the model. Such estimates may suffice when the goal is prediction of the average performance of large groups of workers, but they may be problematic when the goal is prediction of individual performance.

Similarly, as noted previously, measures of circadian rhythm are rarely available in the operational environment, and thus estimates of the phase and amplitude of the circadian process must be generated from the timing of sleep, light exposure, or other environmental drivers of biologic rhythms. Again, such estimates may suffice for large groups but are likely to be less useful for predicting the performance of individuals who may, for example, be extreme morning or evening types.[72] Additional information (e.g., recent sleep-wake history) would be required to reduce this source of error.

Another major input to performance prediction models is "initial state." This refers to the sleep debt (reservoir level), circadian phase, and (depending on the specific model) other relevant variables that provide the starting point from which predictions are made. If the initial states are unknown, assumptions regarding the likely sleep-wake history and circadian phase need to be made. The accuracy of these assumptions initially determines the accuracy of the model predictions. However, the influence of the initial state estimates diminishes over time,[85] and the accuracy of the model improves as actual daily sleep measures (e.g., actigraphic sleep measurements) or work schedule–based sleep estimates accrue and are used as input to the model.

CONCLUSION

Biomathematical models predict physiology-based cognitive capacity as a function of sleep-wake history, circadian rhythm, and sleep inertia. Cognitive capacity affects the ability to perform specific tasks and thereby affects the risk for making an error or the probability of rare events like accidents.[84]

It is important to understand assumptions underlying biomathematical models. In most three-process models, it is assumed that the rate and magnitude of cognitive performance impairment are independent of the nature of activities performed while awake (e.g., task load, work pace). However, it is known that performance decreases as a function of time on task,[86,87] the number of flight segments during an aircrew duty period,[88] successive duty periods in shift work,[89] and other factors. Conversely, performance is temporarily restored by rest breaks and sustained by days off.[39,90] Current performance prediction models do not account for these factors, nor do they account for the effects of pharmacologic countermeasures (e.g., caffeine, although model development to address caffeine is underway[91]).

As a rule of thumb, biomathematical model predictions should be interpreted as representing the fundamental cognitive capacity of a group or person unaided by pharmacologic countermeasures, that is, a cap on performance as determined by underlying brain physiology. However, actual performance will depend on additional factors, such as the nature of the tasks being performed and the circumstances at hand,[92] which are not accounted for in the model prediction.

CLINICAL PEARL

Sleep loss and circadian misalignment produce deficits in cognitive performance. Biomathematical models that predict performance based on these factors are valuable tools that are increasingly being applied in both operational and nonoperational settings. For personal use, these models can be applied to determine the effects of sleep-wake duration and timing on performance and alertness and thus serve to guide clinical, professional, and personal decision making regarding sleep habits and use of countermeasures. For employers, regulatory agencies, and practitioners in occupational medicine, these models are useful to guide the design of better work schedules, reduce performance errors and accidents, improve the health and well-being of employees, and advance public safety.

SUMMARY

Because insufficient sleep (which is prevalent in modern society) leads to lapses of attention, slowed reaction time, and impaired reasoning and decision making, it is a major proximate cause of errors and accidents in both industrial and military operational settings. Biomathematical models have been developed to predict sleep- and alertness-mediated performance in laboratory and field environments. Such models are already being incorporated into scheduling tools to anticipate and avoid performance impairment in operational settings. Most available models predict performance based on three basic components: circadian variation in alertness and sleep propensity, homeostatic sleep-wake regulation, and sleep inertia. These processes vary over time as a function of sleep-wake patterns and combine to produce changes in subjective alertness and cognitive performance capacity. Current models provide accurate predictions of average performance of groups under variations in sleep opportunities or work schedules. Efforts are underway to further refine these biomathematical models to enhance accuracy of prediction for individuals.

Selected Readings

Åkerstedt T, Garefelt J, Richter A, et al. Work and sleep—a prospective study of psychosocial work factors, physical work factors, and work scheduling. *Sleep* 2015;**38**(7):1129–36.

Borbély AA, Achermann P. Concepts and models of sleep regulation: an overview. *J Sleep Res* 1992;**1**:63–79.

Darwent D, Dawson D, Roach GD. Prediction of probabilistic sleep distributions following travel across multiple time zones. *Sleep* 2010;**33**:185–95.

Folkard S, Åkerstedt T. Trends in the risk of accidents and injuries and their implications for models of fatigue and performance. *Aviat Space Environ Med* 2004;**75**:A161–7.

Hursh SR, Redmond DP, Johnson ML, et al. Fatigue models for applied research in warfighting. *Aviat Space Environ Med* 2004;**75**:A44–53.

Ingre M, Van Leeuwen W, Klemets T, et al. Validating and extending the three process model of alertness in airline operations. *PLoS One* 2014;**9**(10):e108679.

McCauley P, Kalachev LV, Smith AD, et al. A new mathematical model for the homeostatic effect of sleep loss on neurobehavioral performance. *J Theor Biol* 2009;**256**:227–39.

Raslear TG, Hursh SR, Van Dongen HPA. Predicting cognitive impairment and accident risk. *Progress Brain Res* 2011;**190**:155–67.

Van Dongen HPA, Belenky G. Model-based fatigue risk management. In: Matthews G, Desmond PA, Neubauer C, Hancock PA, editors. *The handbook of operator fatigue*. Farnham, UK: Ashgate; 2012. p. 487–506.

Wesensten NJ, Belenky G, Thorne DR, et al. Modafinil vs. caffeine: effects on fatigue during sleep deprivation. *Aviat Space Environ Med* 2004;**75**:520–5.

Whitney P, Hinson JM, Jackson ML, Van Dongen HP. Feedback blunting: total sleep deprivation impairs decision making that requires updating based on feedback. *Sleep* 2015;**38**(5):745–54.

A complete reference list can be found online at ExpertConsult.com.

Chapter 73

Fatigue Risk Management Systems

Philippa H. Gander; Lora J. Wu; Margo van den Berg; Amanda Lamp; Laura Hoeg; Gregory Belenky

Chapter Highlights

- Fatigue-related performance impairment is widely recognized as a workplace hazard. The traditional approach to managing fatigue risk has been to limit work hours and set minimum rest breaks in government regulations or industrial agreements. This chapter introduces a flexible alternative approach—fatigue risk management systems (FRMSs), which are based on scientific understanding of the dynamic effects of sleep loss and recovery on performance, and of the modulating influence of the circadian system on performance capacity.
- The core of an FRMS is a closed-loop process consisting of (1) ongoing monitoring of fatigue levels, (2) identification of situations in which fatigue may constitute a hazard, (3) risk assessment, and (4) introduction of risk mitigations and strategies when needed. Multiple defensive strategies can be incorporated, tailored to the workplace, the level of operator fatigue, and the associated safety risk. This enables more robust safety management than the single defensive layer of one-size-fits-all prescriptive regulations on work hours and rest breaks.
- Implementation of FRMSs is expanding, particularly in the transport sector. Their effectiveness depends on dialogue between scientists and people with operational expertise in the workplaces(s) where these systems are implemented. This is an exciting new area of application of sleep and circadian science that highlights issues for which better scientific understanding is needed.

Limits on maximum working hours and minimum breaks (hours-of-service regulations) are the traditional approach to managing workplace fatigue, with a history dating back to the industrial revolution in Britain and the first Factory Acts in the early 19th century.[1] This chapter describes a new approach: fatigue risk management systems (FRMSs). FRMSs have evolved from advances in sleep and behavioral science and in safety science, leading to better understanding of the causes of human error and its role in the etiology of accidents, and to improvements in the practice of risk assessment and management. Central concepts in FRMS are as follows:

- Human functional capacity is variable (both between individuals and in the same person across time).
- Some of this variability is predictable from previous sleep history and the phase of the circadian pacemaker.
- To maintain safety, systems are needed to monitor and mitigate both the level of fatigue-related impairment and its potential consequences.

The FRMS approach is developing rapidly in the transportation sector in response to several converging trends. The complexity and scope of operations and vehicles continues to increase in all modes of transportation. As an example, pilot fatigue was historically managed through prescriptive regulatory limits on flight and duty times, including a 16-hour flight time limit for non-stop flights in commercial aviation.[2] However, advances in aircraft technology now allow flights much longer than 16 hours, which not only enables different route structures and duty schedules but also requires new approaches for managing pilot fatigue. With the accelerating evolution of transport operations, it is not surprising that the effectiveness of prescriptive hours-of-service regulations is being questioned. The United States introduced prescriptive limits for rail in 1907, for trucking in 1937, and for aviation in 1938. Nevertheless, in 2011, addressing human fatigue was ranked second on the U.S. National Transportation Safety Board's most wanted list of safety improvements across all modes of transportation. In parallel, increasing scientific understanding of the causes and consequences of fatigue-related impairment has provided an evidence base for more sophisticated fatigue risk management approaches.

In this chapter, the focus is on fatigue risk management in the aviation industry, where implementation of FRMSs is rapidly expanding. As a foundation, a comprehensive definition of fatigue is presented that identifies the key physiologic causes of fatigue-related impairment and its potential safety consequences.[3] The essential elements and processes of FRMSs are then described. This highlights the importance of being able to measure operator fatigue and its consequences

for safety in a given work context. Examples of what to measure and how to measure it are considered.

Measures have been proposed as safety performance indicators (SPIs) for FRMSs in commercial aviation, based on their demonstrated sensitivity to sleep history and circadian phase in laboratory protocols. An evaluation is presented of how reliably the proposed SPIs reflect these key determinants of performance capacity in the complex environment of very long commercial flights crossing multiple time zones.[4,5] In these operations, pilots have only one flight per duty period and have the opportunity for in-flight sleep. As more data become available, this evaluation can be expanded to other types of commercial aviation operations that involve different causes of pilot fatigue—for example, short-haul operations that involve multiple flights during long duty days with no opportunity for in-flight sleep. Short-haul operations may include night flying but typically cross only one or two time zones.

The chapter also introduces scientific and safety management principles for FRMSs, which can be adapted to other environments. In addition to these more generic principles, FRMSs require operational knowledge and expertise that are specific to each work situation, so they are not "one-size-fits-all." With increasing scientific knowledge and more widespread FRMS implementation, FRMSs will continue to evolve. The chapter concludes with a discussion of ways to improve the effectiveness of FRMSs.

DEFINING FATIGUE

The definition of fatigue has been the subject of academic debate,[6] although the term is commonly used in occupational safety. For example, in New Zealand, the Health and Safety in Employment Amendment Act (2002) specifically identifies workplace fatigue as a hazard, which means that it is a potential source of harm to employees that must be managed by all employers. The International Civil Aviation Organization (ICAO) is the United Nations agency responsible for regulating the global aviation industry. In 2011, the ICAO released a regulatory framework for FRMSs based on the following comprehensive definition.[3]

> Fatigue is a physiological state of reduced mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian phase, or workload (mental and/or physical activity) that can impair a crew member's alertness and ability to safely operate an aircraft or perform safety-related duties.

In this definition, fatigue is recognized as a physiologic state. The main factors affecting performance capability are identified, as well as the potential consequences of fatigue-related impairment for operational safety (for pilots, their ability to safety operate an aircraft; for cabin crew, their ability to perform safety-related duties).

A feature of fatigue that differentiates it from most other workplace hazards (such as noise, vibration, or toxic substances) is that exposure is not limited to the workplace. Fatigue also is affected by activities and choices made outside of work, sometimes described as a "whole of life issue." Thus effective fatigue management must be a shared responsibility between employers and employees. In analyzing fatigue-related safety incidents, consideration needs to be given to the actions (or inactions) of the individual(s) involved, work demands leading up to the event, and the safety systems present in the workplace. This approach is consistent with the principle that human error should be the starting point for analysis of safety incidents and accidents, rather than the end point.[7]

INTERACTING CAUSES OF FATIGUE-RELATED IMPAIRMENT

Evidence from laboratory studies shows that sleep loss and circadian phase have interactive effects on performance. For example, a study using 28-hour forced internal desynchrony protocols with either a 9.33-hour sleep opportunity or a 4.67-hour sleep opportunity found that the adverse effect of sleep restriction on psychomotor vigilance task (PVT) response speed was greatest during "biological night" (the part of the circadian cycle when sleep normally would be occurring).[8]

Some evidence from field studies supports interactions between circadian phase and time on duty and/or time awake. For example, an early German study recorded failures of locomotive engineers to interact with an on-board vigilance system which required them to extinguish a light within 2.5 seconds of its appearing on the console; otherwise, an acoustic warning sounded.[9] Data were gathered from 10 trains tracked for at least 30 days each (a total of 6304 work hours accumulated by approximately 1000 drivers). Figure 73-1 shows the hourly mean frequency of the acoustic warning sounding across the 24-hour day, comparing hours 1 to 3 of shifts with hours 4 to 6 of shifts. Clear peaks in signals missed were evident at the expected circadian times (early morning and early afternoon) only when engineers had been on duty for more than 3 hours. In a large field study of fatigue ratings self-reported by air traffic controllers at the end of each period of active controlling (interspersed with breaks) across duty periods, the following factors were independently associated with the level of fatigue: time of day, self-rated workload, time since coming on duty, duration of continuous wakefulness, and duration of the active operational duty period.[10] The effects of time on duty and time of day appeared to be additive, with highest fatigue ratings at the end of long duty periods that finished around 6 A.M.

The ICAO definition of fatigue includes workload as a potential cause of fatigue-related impairment; however, no clear operational definition of workload has emerged. In principle, it could include both factors relating to the nature of the work (e.g., time on task, task difficulty and complexity, and work intensity and whether this is paced by task demands or by the individual worker) and factors relating to the performance capacity of the individual worker (e.g., experience and skill level, sleep history, and circadian phase). Few studies have attempted to investigate the influence of workload on fatigue or potential interactive effects between workload and other causes of fatigue-related impairment. In the field study of fatigue ratings by air traffic controllers,[11] some evidence pointed to interactive effects of self-rated workload and time-on-task on fatigue. When self-rated workload was low, fatigue ratings remained relatively stable for continuous work periods up to 4 hours. However, when workload was high, rapid increase in fatigue was seen after 2 hours of continuous work. These effects of workload became more evident after controllers had been awake for at least 12 hours. The time-of-day

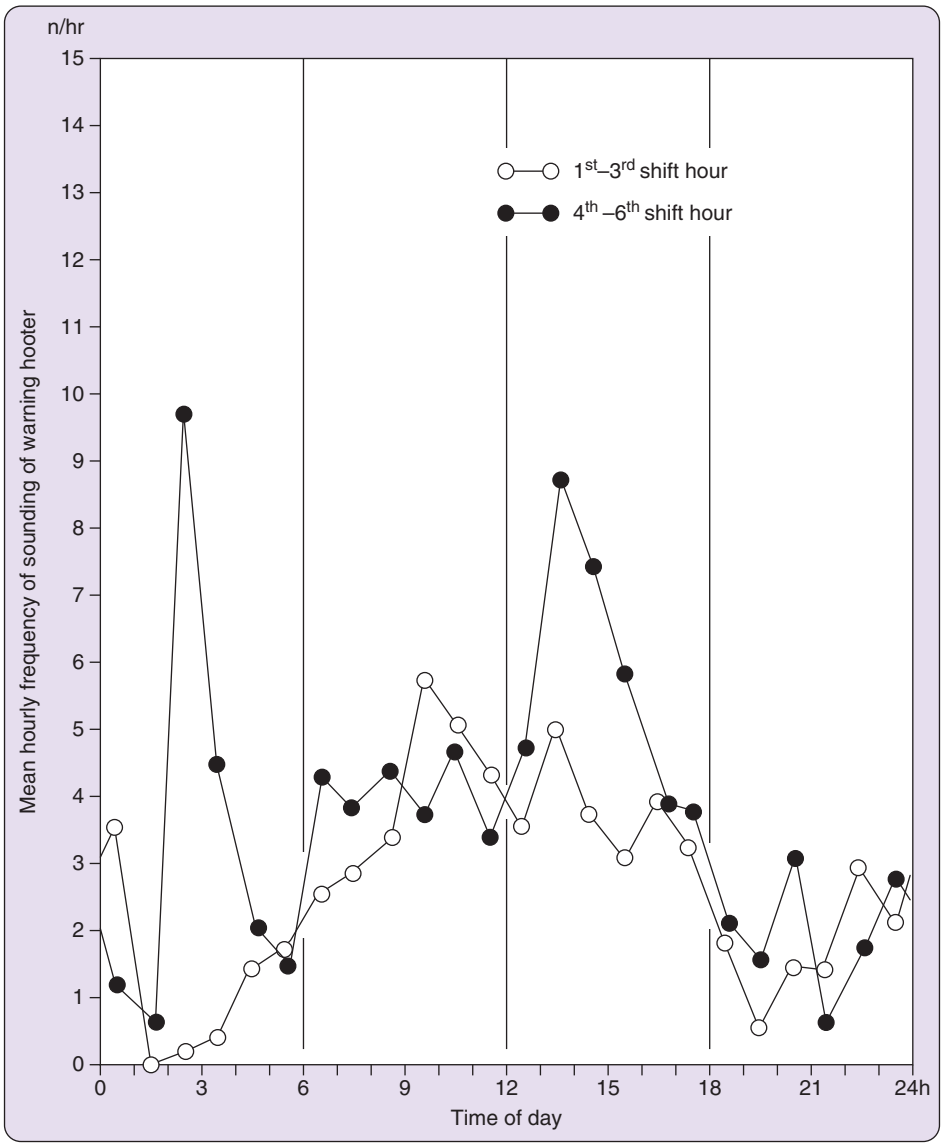


Figure 73-1 Locomotive engineers' failures to interact with a vigilance device across the 24-hour day, comparing hours 1 to 3 of shifts with hours 4 to 6 of shifts. (From Hildebrandt G, Rohmert W, Rutenfranz J. 12 and 24 h rhythms in error frequency of locomotive drivers and the influence of tiredness. *Int J Chronobiol* 1974;2:175–80, with permission.)

variation in fatigue ratings also was influenced by workload, being more marked at low and high levels of workload than at intermediate levels. To manage fatigue-related risk, it is difficult (if not impossible) to design prescriptive rules that address the interacting effects of duty history, sleep history, circadian phase, and workload.

The interactions of factors causing fatigue can have major effects at the population level. A robust, case-controlled study has estimated that motor vehicle crashes resulting in injury could be reduced by 19% on the Auckland metropolitan road network if drivers did not drive when they felt sleepy, or when they had slept 5 hours or less in the last 24 hours, or between 02:00 and 05:00.[12] This degree of improvement in road safety is not expected to be achieved by separate campaigns to reduce speed, enforce seat belt use, or eliminate drunk driving.

TYPES OF FATIGUE-RELATED IMPAIRMENT

Many aspects of performance are subject to fatigue-related impairment, and the consequences in a workplace can be complex.[13] For example, very little research has addressed the effects of individual team member fatigue on team performance. A flight simulation study with 67 Boeing B747-400 crews demonstrated that sleep loss increased the number of errors made by these two-pilot crews.[14] Paradoxically, however, when crew position was considered, greater sleep loss and higher subjective fatigue ratings among first officers were associated with improved rate of error detection, whereas greater sleep loss among captains led to a higher likelihood of failure to resolve detected errors. Greater sleep loss and higher subjective fatigue ratings also were associated with slower

decision making and a tendency toward choosing lower-risk options.[15]

These findings are consistent with evidence from laboratory studies indicating that more complex cognitive tasks, particularly those involving executive functions dependent on the prefrontal cortex, are more severely affected by sleep loss than are simpler tasks.[15-17] Other support derives from brain imaging studies suggesting that the regions involved in higher-order complex cognitive performance are most affected by sleep deprivation and have the greatest need for sleep-mediated recuperation.[18-21]

A compelling illustration of the possible safety consequences of fatigue-related impairment comes from debriefings after friendly fire incidents during Operation Desert Storm.[22] Sleep deprivation was identified as contributing to Bradley fighting vehicle and M1 tank crews becoming misoriented to the battlefield (a complex mental task involving situational awareness enabled by working memory), who then mistook friend for foe. They were, however, able to shoot straight (a simple task involving putting cross-hairs on the target and squeezing the trigger). This combination of degradation in complex task performance coupled with relative preservation of simple task performance led to their firing upon and destroying friendly vehicles.

PROCESS AND COMPONENTS IN A FATIGUE RISK MANAGEMENT SYSTEM

Rather than relying on a priori decisions about the factors most likely to be causing fatigue (the basis of prescriptive rules), FRMS is based on real-time monitoring and management of actual fatigue levels and their associated safety risk in an operation. The processes underlying FRMSs are summarized in Figure 73-2, which highlights that these processes are intended to be data-driven.[3]

Data for Fatigue Risk Management Systems

In ultrasafe systems such as commercial aviation operations, safety-critical errors and incidents are unpredictable, rare events. In FRMSs, they must be analyzed to identify whether operator fatigue was a contributing factor. Because they are rare, however, safety-critical errors and incidents cannot be used in FRMSs for routine tracking of fatigue levels and identification of potential hazards. A variety of additional sources of data are needed, from which a set of SPIs can be identified, together with acceptable values or ranges for each SPI. The use of multiple SPIs is recommended for robust safety management.[23] For commercial aviation, two different types of SPIs have been proposed: operational SPIs, which monitor the duty-related causes of fatigue, and SPIs based on monitoring actual levels of crew member fatigue.[3]

Operational SPIs can be derived from routinely collected data—for example, by comparing planned versus actual schedules. In commercial aviation, potential fatigue hazards may be identifiable from data such as how often flight duty periods end at least 30 minutes later than scheduled; how often flight duty periods start or end within the window of circadian low; or how often reserve crew members are called out on particular flights at a particular crew base. Acceptable limits might consist of the following: If report time is earlier than 05:00, then the actual flight duty period will not exceed the scheduled flight duty period on more than 10% of days in any 28-day period; or for crew members who do not have the opportunity for in-flight sleep, no duty period longer than 9 hours will be scheduled to end in the window of circadian low. Operational SPIs and their acceptable limits need to reflect the specific causes of fatigue risk in different operations, such as early starts and multiple flights during long duty days in domestic short-haul operations versus single flights during long duty days that cross multiple time zones with resulting circadian desynchrony in international long-haul operations.[23,24]

Monitoring operator fatigue as a source of data for FRMSs is relatively resource-intensive and time-consuming compared with using routinely collected operational data. In view of the fact that fatigue affects diverse aspects of waking function, it is considered inadvisable to make operational decisions based on any single measure of functional status (or based only on thresholds applied to biomathematical model predictions of functional status).[3,4] Measures of crew member fatigue are considered in more detail later in this chapter.

An essential source of fatigue monitoring data in FRMSs is voluntary fatigue reporting (hazard identification) by staff. In the ICAO FRMS approach, crew members must be able to voluntarily report potential fatigue hazards without fear of retribution, in addition to the mandatory reporting required for more serious safety events.[3]

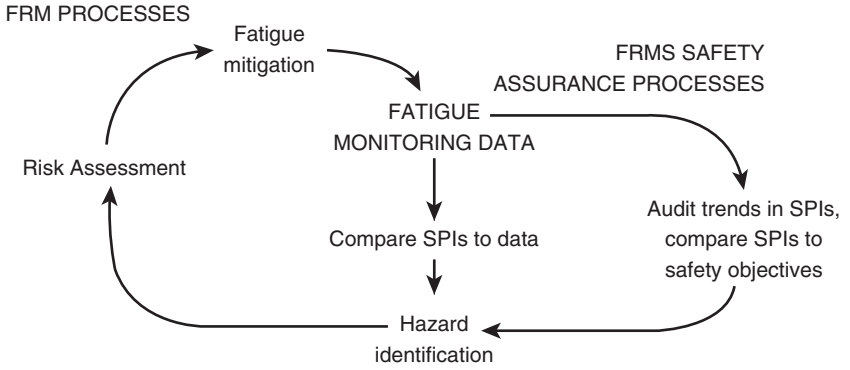


Figure 73-2 Summary of the core activities in fatigue risk management systems (FRMSs), as defined by the International Civil Aviation Organization (ICAO).[3] The left-hand loop (fatigue risk management [FRM] processes) deals with day-to-day management of fatigue in an organization. The right-hand loop (FRMS safety assurance processes) ensures that the FRM processes are delivering an acceptable level of fatigue risk and provides input for continuous improvement of the FRM processes. SPIs, Safety performance indicators.

Hazard Identification, Risk Assessment, and Mitigation

When data of any type suggest that fatigue is a potential hazard, the safety risk associated with that hazard needs to be evaluated (the fatigue risk management processes loop on the left in Figure 73-2). In airlines, risk assessment often is carried out by a specialized safety team that evaluates risk across all types of hazards, including crew member fatigue. As appropriate, additional fatigue mitigations may need to be implemented. Although a very common response is to change work patterns (rosters) to reduce fatigue hazards, other mitigations usually are possible and in some instances are preferable. For example, the main pilot fatigue mitigation on long flights is to have additional pilots and provide on-board crew rest facilities with horizontal bunks. This enables each pilot to take one or more rest breaks away from the flight deck, with the opportunity and environment for sleep. In the ICAO approach to fatigue risk management, the effectiveness of fatigue mitigations is continuously monitored by tracking the fatigue SPIs, so the fatigue risk management processes in Figure 73-2 constitute a closed loop.

Safety Assurance and Continuous Improvement

The FRMS safety assurance loop on the right hand side of Figure 73-2 represents another layer of organizational defenses against fatigue risk. This loop is designed to check whether the fatigue risk management processes are delivering an acceptable level of fatigue risk that meets internal and external standards set by company policy and/or by regulations. Data-derived SPIs inform decision making in the FRMS safety assurance loop. This loop also monitors external changes that could affect fatigue risk in the organization. In aviation, examples of such external changes might be the introduction of a new route or aircraft type, or changes made to flight schedules by the marketing department in response to a new competitor entering the market. The feedback provided by the FRMS safety assurance processes drives continuous improvement of the fatigue risk management processes.

Other Components of the Fatigue Risk Management System

The two loops in Figure 73-2 are considered the core operational activities of the FRMS.[3] In the ICAO regulatory framework, they are supported by additional required elements—namely, an FRMS policy (which includes an identified manager responsible for the resourcing and effectiveness of the FRMS), documentation of all FRMS activities, fatigue management education for crew members and other relevant personnel (schedulers, managers, and so on), and a communications plan.

MONITORING OPERATOR FATIGUE

Monitoring operator sleep history and waking function will sometimes be necessary in an FRMS, for example, when more in-depth information is needed about a particular fatigue hazard, or as an additional defensive layer when a new type of operation is being introduced. As a general rule, the type of monitoring undertaken should be commensurate with the anticipated level of fatigue and safety risk. Options range from retrospective surveys to full-scale studies monitoring sleep history before, during, and after a roster, with performance testing and subjective assessments during work periods. Resources for data collection (financial, logistical, and human) usually are limited. It also is important to consider the burden on staff asked to participate in data collection from time to time, particularly because sleep and performance monitoring often occur outside the workplace as well as at work. Ongoing commitment and cooperation of staff are critical not only for study participation but for all aspects of an effective FRMS.

The study design and measures used need to be adequate to provide scientifically defensible answers to the operational questions being asked about fatigue and safety. In the aviation context, recommended criteria for selecting measures include the following: They have been validated, to confirm that they measure what they purport to measure; they do not jeopardize a crew member's ability to perform operational duties; and they have been widely used in aviation, allowing data to be compared between different types of operations.[3]

Measuring Sleep

Sleep-wake history arguably provides the most valuable information about fatigue status, because insufficient sleep has a negative impact on diverse aspects of waking function, only some of which are captured by subjective ratings or objective performance tests.[5] Polysomnography, the laboratory gold standard, has been used to monitor sleep for short periods of time in field studies.[25-31] However, with current recording equipment and analytical procedures, this procedure is intrusive for participants as well as time-consuming and costly to analyze.

Actigraphy is a widely used alternative to polysomnography for objective sleep monitoring in field studies.[2,4,5,29,30,32–35] It can reliably measure total sleep time but cannot identify sleep stages.[34,36] An actigraph is an unobtrusive wristwatch-size device, worn on the nondominant wrist, that sums movements across a specified time period (usually a 1-minute epoch).[36] An algorithm (validated against polysomnography) is used to determine whether the wearer is asleep or awake in each epoch across the record. The actigraph can record sleep-wake history reliably[34,36] for weeks to months with minimal inconvenience to the wearer, and data analysis is much less complex and time-consuming than for polysomnography. Figure 73-3 shows an example of an actigraphy record from a commercial airline pilot.

Sleep diaries can be used to gather subjective data prospectively on sleep timing and quality and have the advantages that they are easy to implement (in paper form or electronically) and inexpensive. They typically are used to aid interpretation of actigraphy records and can be used to gather other relevant information such as duty-rest times. Subjective estimates of sleep duration in flight and on the ground were shown to be correlated with polysomnographic measurement in a group of 25 pilots.[34] However, there was wide variability among pilots in the accuracy of their subjective estimates, particularly for in-flight sleep duration.

Measuring Subjective Sleepiness and Fatigue

Two validated subjective rating scales have been recommended for use in airline FRMSs.[3] On the Karolinska Sleepiness Scale (KSS), participants rate their current level of sleepiness from 1 to 9, where 1 = "extremely alert;" 3 = "alert;" 5 = "neither

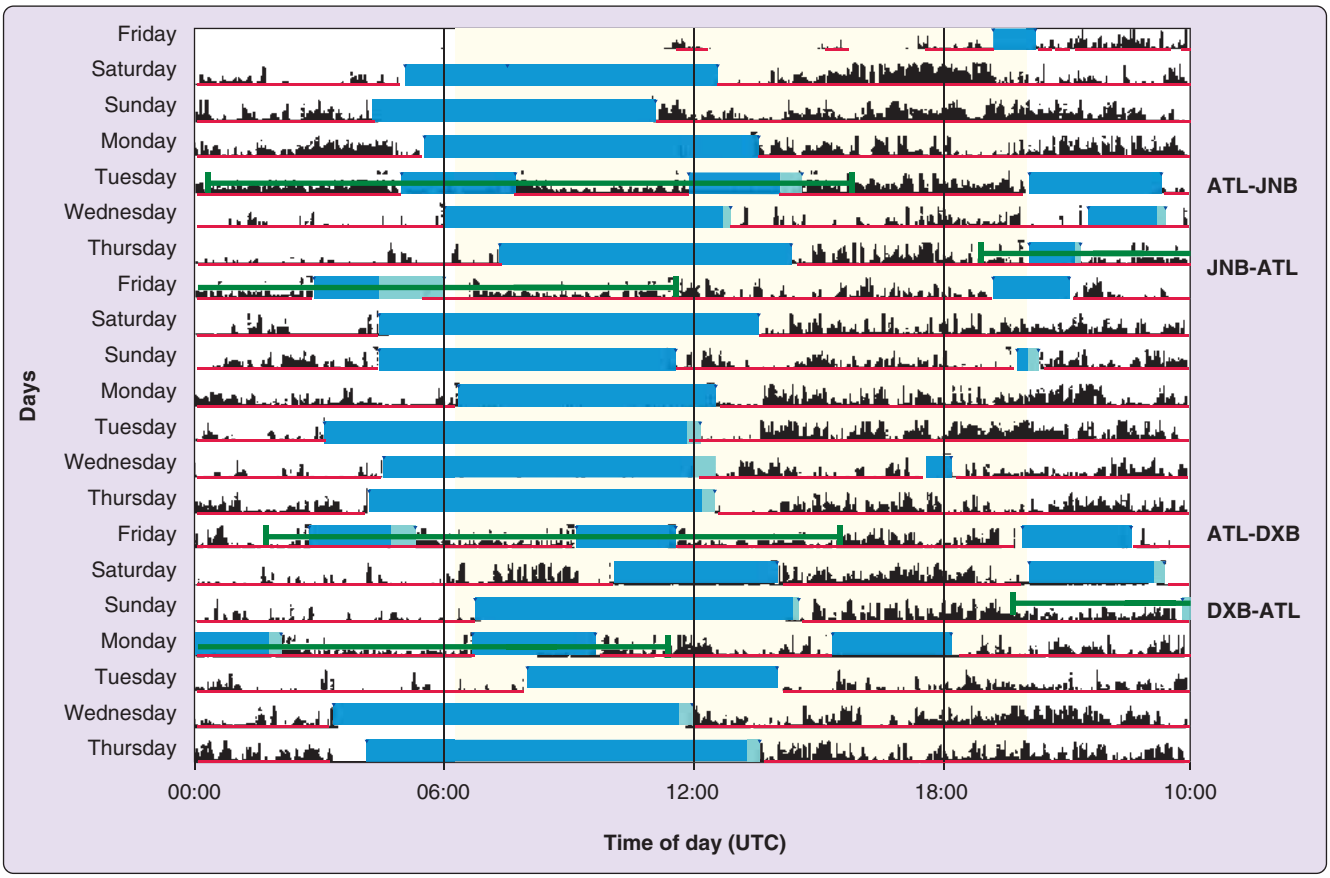


Figure 73-3 Actigraphy record from an airline pilot over a 3-week period during which he flew a return trip between Atlanta (ATL) and Johannesburg (JNB) (crossing six time zones) and a return trip between Atlanta (ATL) and Dubai (DXB) (crossing eight time zones). *Black vertical bars* are minute-by-minute activity counts. *Blue horizontal bars* indicate sleep as identified by actigraphy. *Turquoise horizontal bars* indicate times attempting to sleep but not asleep by actigraphy. *Red horizontal bars* indicate times awake as identified by actigraphy. *Green horizontal lines* are times on duty. Note *blue horizontal bars* for in-flight sleep periods during the flights. UTC, Coordinated universal time.

sleepy nor alert;" 7 = "sleepy, but no difficulty remaining awake;" and 9 = "extremely sleepy, fighting sleep."[37] The KSS was developed and validated in a laboratory context[38–40] and has been used in a variety of operational settings. Ratings of 7 to 9 have been associated with the onset of polysomnographically recorded microsleeps.[37] In a 28-hour forced desynchrony protocol with sleep restriction (4.7-hour sleep opportunity, 23.3 hours awake), KSS ratings were found to increase across wakefulness and to have significant circadian variation, being highest during biological night.[40]

On the Samn-Perelli Crew Status Check (also known as the Samn-Perelli Fatigue Scale), participants rate their current fatigue level from 1 to 7, where 1 = "fully alert, wide awake;" 2 = "very lively, responsive, but not at peak;" 3 = "okay, somewhat fresh;" 4 = "a little tired, less than fresh;" 5 = "moderately tired, let down;" 6 = "extremely tired, very difficult to concentrate;" and 7 = "completely exhausted, unable to function effectively."[41] The scale was developed for military airlift operations and has been widely used in aviation studies.[31,42] The original scale explicitly linked ratings to the likelihood of performance impairment.[41] At scores of 1 to 3, no fatigue-related performance impairment is expected. For a score of 4, "performance impairment is possible but not a significant factor." A score of 5 is equated with "moderate fatigue, performance impairment possible, flying duty permissible but not recommended unless urgent." A score of 6 is equated with "severe fatigue, performance impairment probable, flying duty not recommended." A score of 7 is equated with "severe fatigue, performance definitely impaired, flying duty not recommended, safety of flight in jeopardy." These proposed relationships between scores and performance impairment do not appear to have been systematically validated against objective measures of performance or quantified in other operational settings. The B747-400 simulator study mentioned earlier suggests the possibility of complex relationships between the Samn-Perelli fatigue scores of individual pilots and the flight deck performance of two-pilot crews. Higher Samn-Perelli fatigue ratings were associated with improved rates of error detection among first officers,[14] but with slower decision making by the crew and a tendency toward choosing lower-risk options.[15] A laboratory study using 28-hour forced desynchony protocols has demonstrated circadian variation in presleep and postsleep Samn-Perelli ratings, as well as cumulative effects of sleep restriction on postsleep Samn-Perelli ratings.[43]

Factors influencing KSS and Samn-Perelli ratings at top of descent (TOD) on very long flights have been analyzed using combined data from four airlines (237 pilots in four-pilot crews, monitored on 730 out-and-back flights between 13 city pairs, with 1- to 3-day layovers).[5] Mixed model

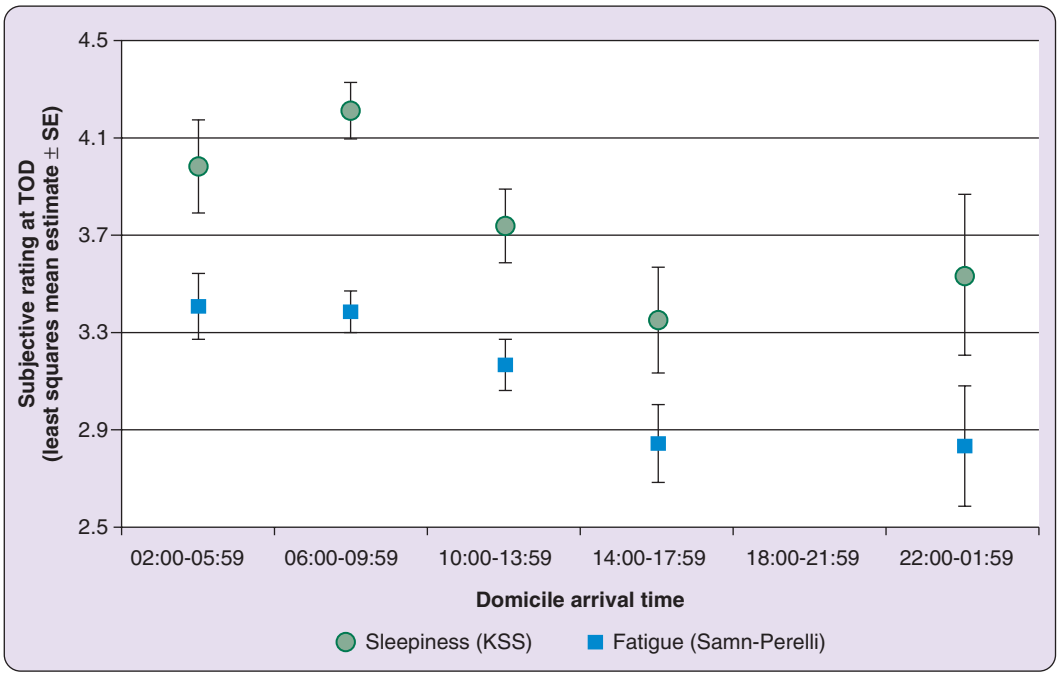


Figure 73-4 Subjective sleepiness (KSS) and fatigue (Samn-Perelli scale) ratings at TOD on flights arriving at different domicile times. Only eight flights arrived between 18:00 and 21:59, so this time bin was excluded from the analyses. KSS, Karolinska Sleepiness Scale; TOD, top of descent. (From Gander PH, Mulrine HM, van den Berg MJ, et al. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *J Sleep Res* 2015;24:110–9, with permission.)

analysis of variance was used to examine relationships between ratings at TOD, sleep-wake history, and circadian phase, after controlling for within- and between-subjects variability and operational factors (flight direction and duration). Circadian phase was approximated by local (domicile) time in the city where each trip began and ended.

For every 1-hour increase in continuous time awake by TOD, KSS ratings increased by an estimated 0.2 point and Samn-Perelli fatigue ratings increased by an estimated 0.1 point. For every 1-hour increase in total in-flight sleep, KSS ratings decreased by an estimated 0.3 point and Samn-Perelli fatigue ratings decreased by an estimated 0.2 point. Neither subjective rating was significantly associated with total sleep in the 24 hours before TOD. The independent effect on TOD ratings of the (domicile) time of arrival is illustrated in Figure 73-4. The pattern of variation is consistent with the circadian variation in these measures observed in 28-hour forced desynchrony protocols.[40,43]

Measuring Performance

A variety of approaches have been developed for measuring performance in field studies, each with strengths and weaknesses. Well-validated laboratory tasks such as the psychomotor vigilance task (PVT)[44,45] can easily be implemented on smartphones for field studies.[3-5,13,46,47] These added performance measures, however, interrupt the normal flow of work and could therefore be considered intrusive.[48] In addition, little is known about how an individual's performance on the PVT relates to performance on more complex tasks, or to that person's contribution to the performance of a workplace team. The flight simulator study with 67 B747-400 crews found no relationship between individual pilots' PVT performance before the 70-minute simulator session and the two-pilot crew's ability to detect threats or manage them effectively, or the number of errors made by the crew members and their ability to detect and manage those errors.[14]

Embedded measures, defined as measures taken unobtrusively as a part of the normal course of operations (e.g., aircraft-handling parameters), have the potential to monitor individual and crew performance during work. They have high ecologic validity and do not interrupt the normal flow of work. However, this approach can encounter participant resistance if it is seen as an assessment for clandestine purposes other than safety. In addition, well-established causal links between fatigue-related impairment of individuals and changes in embedded measures of workplace performance are needed if they are to be used as a source of data in FRMSs. Despite considerable effort, to date no embedded measures have been identified that are consistently related to available fatigue SPIs.

Simulator studies offer a middle ground, with greater ecologic validity than laboratory tests and greater experimental control than embedded measures in the workplace. They can be expensive, however, and even high-fidelity simulators are open to the criticism that participants' knowledge that they are in a simulator may alter their behavior. For example, deciding to divert to a different destination in a flight simulator has very different consequences from those associated with deciding to divert to a different destination with an airplane full of passengers.[15]

The PVT performance test has been recommended for use in FRMSs in aviation.[3] A 5-minute version of the task can be administered on a smartphone.[46,47] Participants are instructed to respond by button press when the stimulus appears (e.g., a bull's eye) on the screen. The time from presentation of the

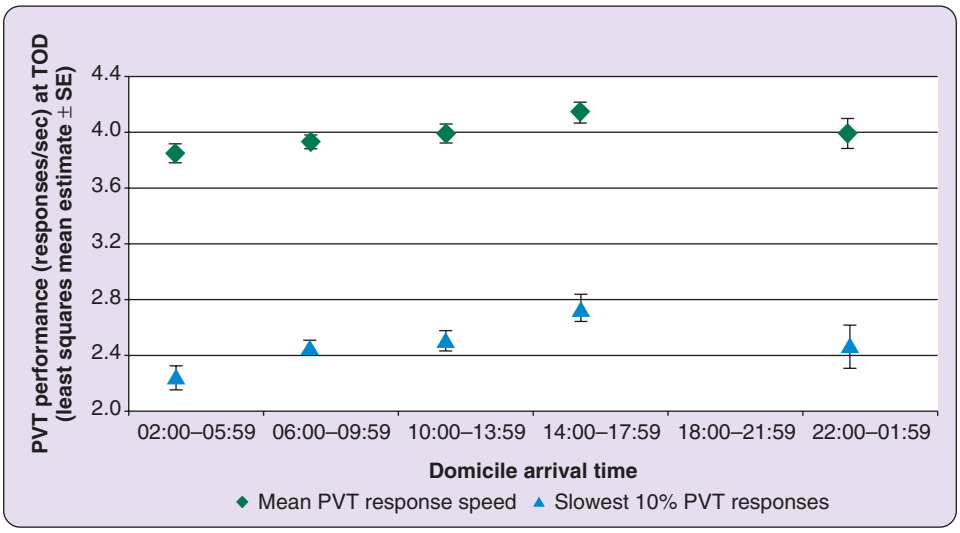


Figure 73-5 PVT performance at TOD on flights arriving at different domicile times. Only eight flights arrived in the period 18:00 to 21:59, so this time bin was excluded from the analyses. PVT, Psychomotor vigilance task; TOD, top of descent. (From Gander PH, Mulrine HM, van den Berg MJ, et al. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *J Sleep Res* 2015;24:110–9, with permission.)

stimulus to button response is recorded as reaction time. The stimulus reappears at random intervals across the duration of the task, so vigilance is required to maintain reaction times. Laboratory studies using forced internal desynchrony protocols have shown that PVT performance reaches its circadian nadir shortly after the circadian temperature minimum and degrades with increasing time awake.[8,49–53] Laboratory sleep restriction studies have shown cumulative sleep-dose–dependent slowing in PVT responses.[8,54,55]

Factors influencing PVT performance at TOD on very long flights were analyzed using the combined data from four airlines, as described previously.[5] Mixed model analysis of variance was used to examine relationships between performance at TOD, sleep-wake history, and circadian phase, after controlling for within- and between-subjects variability and operational factors (flight direction and duration). Circadian phase was approximated by local (domicile) time in the city where each trip began and ended. No significant relationships were found between PVT performance speed and time awake at TOD, total sleep in the 24 hours before TOD, or total in-flight sleep.

Compared with laboratory studies using a 10-minute version of the PVT,[54,56] 5-minute PVT response speeds of these pilots at TOD were relatively fast (median = 4.0 responses/second; in one of the four studies, pilots undertook a 10-minute PVT). This is noteworthy, considering that participants in the laboratory studies had 7 to 8 hours of time in bed for sleep at night, whereas most pilots had multiple sleep episodes in the 24 hours before TOD and median total sleep was 4.75 hours (range, 0.23 to 11.50 hours). Possible explanations for the lack of association between PVT response speed at TOD and previous sleep history might be that the levels of fatigue at TOD are relatively modest on these flights, owing to in-flight sleep (median time awake from last in-flight sleep period to TOD = 2.33 hours; range, 0.00 to 11.68 hours) and/or that these pilots represent a highly selected group with unusually fast PVT performance. The findings raise the question of whether PVT performance is a discriminating measure of fatigue-related impairment in pilots, supporting the need for additional objective performance measures to be developed and validated for use in airline FRMSs.

The independent effect of the domicile time of arrival on PVT performance at TOD is illustrated in Figure 73-5. The pattern of variation is consistent with the circadian variation observed in 28-hour forced desynchrony protocols.[8,49–51,53]

Routine Operator Monitoring and Biomathematical Modeling

In some contexts, particularly military operations or other high-risk environments, it may be appropriate to continuously record sleep-wake patterns in the field using actigraphy. These data could potentially be used to predict performance minute by minute, using biomathematical models that integrate the effects of sleep-wake history and circadian phase and, in the future, workload.[13] Ideally, models also would be tailored to reflect individual differences in chronotype ("morningness"/"eveningness") and resilience to sleep loss.[57–60] On the other hand, in scheduled commercial operations, routine operator monitoring may be unnecessary, and people may well be resistant to being monitored 24/7, including while off-duty, to improve their performance at work.

In the ICAO model of FRMS, biomathematical models are an optional tool for predicting potentially hazardous levels of fatigue in schedules or rosters. They also could be seen as a fatigue mitigation strategy if they aid the design of less fatiguing rosters. Current biomathematical models, however, aim to predict measures of operator fatigue (performance and/or subjective ratings), not the safety consequences of that fatigue in specific operational environments. This fact highlights why "bright lines"—model output values above which an individual worker is proposed to be "safe" and below which the person jeopardizes operational safety—are too simplistic for fatigue risk assessment in FRMSs. Biomathematical models are not a stand-alone substitute for the processes outlined in Figure 73-2. If models are used within the fatigue risk management process loop, then their effectiveness (along

with all other fatigue mitigations) is monitored by the SPIs. In this context, models are only one defensive strategy, and their limitations are less likely to inadvertently increase fatigue-related safety risk.

IMPROVING FATIGUE RISK MANAGEMENT

Scientific Challenges

Fatigue risk management aims to minimize fatigue-related impairment in the workplace *and* its operational consequences. Strategies to minimize fatigue-related impairment focus primarily on minimizing sleep loss (or its effects) and improving recovery sleep. Addressing important research questions in these areas can be expected to lead to improved FRMSs.

Sleep restriction is the most common sleep disturbance associated with 24/7 operations and is associated with both short-term consequences for productivity and safety[61-63] and long-term consequences for health.[64-70] Whether or not an individual is sleep-restricted, their performance capacity fluctuates across the circadian body clock cycle. There are stable trait-like differences between people in the extent to which their performance is impaired by sleep restriction and at suboptimal times in the circadian cycle.[58-60] Better understanding of the basis of this differential sensitivity may provide insights into the neurophysiologic mechanisms underlying the effects of fatigue on waking function. In very-high-risk occupations, it also may be appropriate to screen for people with high resilience to fatigue-related impairment. This approach does not have widespread applicability as a fatigue mitigation strategy, however, because it would probably exclude too many people from working in 24/7 operations.

Additional research is needed to be able to provide robust scientific advice on the amount of sleep needed to recover from chronic sleep restriction, in terms of both the duration of recovery sleep episodes and the number of consecutive nights of sleep required.[54,55,71] The recovery value of split sleep versus a single consolidated sleep also is an important question for fatigue risk management. Laboratory studies suggest that having a restricted sleep period at night plus a daytime nap has equivalent recovery value to the same total amount of sleep taken in one consolidated block at night.[72-75] However, these are short-term studies that take place in a quiet laboratory sleep environment with no distractions, and nighttime sleep is not always an option in 24/7 working environments.

A recent field study with junior doctors demonstrated that a split sleep schedule with napping during the night shift resulted in the same amount of total sleep per 24 hours (7 hours) as was obtained when working the day shift. Measures of vigilance, verbal and spatial learning, and memory were also comparable on the night shift and the day shift.[76]

For pilots on long commercial flights, in-flight sleep results in split sleep patterns. Split sleep also is typical on layovers between transmeridian flights. Figure 73-3 shows an example in which the participant was part of a four-pilot crew and had two opportunities per flight for horizontal sleep in an onboard crew rest facility. For two-pilot crews, planned 40-minute nap opportunities in the flight deck seat have been shown to provide an average of 23 minutes of sleep (confirmed by polysomnography) and improved alertness (by polysomnography) and PVT performance at TOD, with no apparent effect on subsequent layover sleep (as measured by actigraphy).[77,78] Controlled flight deck napping has been recommended by the ICAO as a fatigue mitigation strategy[3] and has been incorporated into aviation regulations in a number of countries. Although the Federal Aviation Administration (FAA) in the United States funded this research and the drafting of an advisory circular on how to implement controlled flight deck napping, it has not yet approved its use as a fatigue countermeasure for U.S. airlines.

The safety hazards associated with fatigued employees in different workplaces remain poorly understood. The relationship between individual fatigue and system safety is likely to be complex in ultrasafe systems[79] such as commercial aviation, where multiple layers of operational defenses (e.g., automation, teamwork, standardized procedures) reduce the probability of having an accident attributable uniquely to an error made by a fatigued individual. A comprehensive research approach is needed that integrates (1) experimental studies focusing on the effects of sleep restriction on different types of performance and on communication; (2) simulation studies to address the impact of fatigued individuals on the performance of workplace teams; (3) studies identifying causal links between fatigue-related impairment of individuals and teams and changes in embedded measures of workplace performance; and (4) careful investigation and analysis of fatigue-related incidents and accidents. Over time, this research should identify and validate better objective performance measures for use as safety performance indicators in FRMSs.

In field studies, the interacting causes of fatigue-related impairment—sleep loss, circadian modulation of performance capacity, and workload—are affected by a range of operational factors.[3] In long-haul aviation, for example, these include departure times, flight direction and duration, number of time zones crossed, crew complement, and layover timing and length.[80] Field studies usually are targeted to address a particular set of operational issues, so typically they include only limited combinations of operational factors. In addition, the number of participants is restricted because fatigue-monitoring studies in the field are expensive and time-consuming. These constraints can be overcome by combining data from multiple studies to evaluate the independent contributions of the different operational factors that influence fatigue.[4,5,80] This is possible only when studies use a standard set of core measures.[3] Meaningful data merging also requires detailed contextual knowledge about each data set, and issues of data ownership and confidentiality need to be addressed.

In addition to sleeping and working at inappropriate times in the daily cycle of the circadian pacemaker in the hypothalamus, people working in 24/7 operations are likely to be eating at inappropriate times in the cycle of the circadian master clock in the hypothalamus. Increasing evidence suggests that this mistiming can lead to desynchrony between rhythms in the circadian clock genes in the liver, fat, pancreas, and muscle, and those in the hypothalamic master clock. Better understanding is needed of the consequences of mistimed eating, both for short-term digestive complaints and for longer-term health consequences, including increased risk of obesity and type 2 diabetes[64,81-83] In the context of FRMSs, these may be exacerbating factors that affect an operator's functional capacity, and dietary advice may be an important part of fatigue education for some workforce groups. Advances in biomedical monitoring techniques hold promise for noninvasive field monitoring of circadian rhythms and sleep-wake patterns over long time periods.[13] For FRMSs in aviation, this will increase

understanding of the effects of rosters and combinations of transmeridian flights on the circadian system.[32,80]

Implementation Challenges

The legislative and regulatory context is a key factor in the implementation of FRMSs, and differs among countries and industries.[84–86] The FRMS approach is consistent with performance-based occupational health and safety legislation (required outcomes are specified, but not the means for achieving them) and with shared responsibility of employers and employees for occupational health and safety. In addition to the general health and safety legislation, in public transportation an additional expectation is that governments have a duty to regulate the activities of companies to limit the societal risk posed by fatigue-impaired workers.

Prescriptive hours-of-service regulations require regulators to clearly define the limits of acceptable performance, or the levels of fatigue that cause performance to become unacceptable, in every context across a broad range of operations.[11,85,86] These regulations can be difficult to defend from a scientific perspective.[87] From a regulatory perspective, however, hours-of-service regulations provide clear standards for judging violations and deciding culpability in legal contexts, whether or not the prescribed limits are scientifically defensible. FRMSs need to convince both governments and the public that the level of protection they offer is at least equivalent to that provided by hours-of-service regulations. Statistical testing for equivalence has significant potential as an adjunct to traditional testing for differences when comparing safety performance indicators in FRMSs versus prescriptive hours-of-service regulations.[88]

Regulatory frameworks permitting FRMSs as an alternative means of managing operator fatigue are spreading across the transport sector. Because they are relatively new, FRMSs remain relatively untested compared with prescriptive hours-of-service limits. However, FRMSs are designed with multiple defensive strategies to manage uncertainties, along with processes for continuous improvement. These characteristics enable them to respond to changes in rapidly evolving industries such as transportation. By contrast, traditional hours-of-service regulations offer only a single rigid, brittle defensive layer against fatigue-related risk, and regulatory change often is too slow to keep up with changes in industry practice.

In commercial aviation, there has traditionally been provision for airlines to be able to apply to the regulator for exemptions to exceed the prescriptive limits in specific situations. This provision can lead to ad hoc changes with potentially compounding effects that are difficult to evaluate. In 2014, the FAA introduced a new set of prescriptive flight and duty time limits and an FRMS regulatory framework for managing pilot fatigue.[86] To exceed the prescriptive limits, an airline regulated by the FAA now must have an approved FRMS and present a specific safety case for the exemption sought, to demonstrate that it can deliver an equivalent level of safety to flights compliant with the prescriptive limits.

FRMSs shift the locus of responsibility for safety away from the regulator and toward employers and employees. This approach is consistent with the philosophy that permitting industry to self-regulate fosters greater ownership of safety by industry and that those who create the risk are best placed to manage it.[89] An important subtext is that safety considerations should be separated from industrial bargaining and remuneration, because cooperation is necessary to manage fatigue as a "whole of life issue." Voluntary fatigue reporting (hazard identification) by staff is an essential source of fatigue monitoring data in FRMSs. An effective safety reporting culture requires a clear understanding by all parties of the delineation between intentional violation of policies or procedures (disciplinary matters dealt with outside the FRMS) and fatigue-related impairment and errors (a normal subset of human behavior).[3]

To meet their responsibilities, all parties involved in FRMS implementation need to have adequate understanding of the scientific principles on which it is based, of the specific causes and consequences of fatigue in their particular operational context, and of their role in the FRMS. Accredited, competency-based education and training programs are needed, and some regulators are providing detailed advice on training content.[90] Evaluation of the effectiveness of training also is a neglected area, with conflicting findings that may be linked to the type of training and the presence or absence of other ongoing strategies to maintain awareness of fatigue risk in an organization.[91,92]

The costs of implementing and running an FRMS may be disproportionately heavy for smaller operators, which could be seen as inequitable if an FRMS gives larger companies a competitive edge. In principle, the complexity and cost of an FRMS should be commensurate with the complexity of the operations that it covers and the level of fatigue-related risk. However, there are few detailed examples of how FRMSs can be scaled to the needs of smaller organizations.[93] Maritime New Zealand has introduced a fatigue management program for vessels as part of the safe ship management system. The program has three levels: for crew and the skipper, for vessel owners and managers, and for safe ship management companies and maritime safety inspectors who are involved in implementing the program. Tailored guidance material is available for different roles and types of operations.[94]

As awareness of the importance of sleep in the management of fatigue risk increases, transport regulators are increasingly turning their attention to regulatory requirements for managing people with clinical sleep disorders. Clinical sleep services clearly have a role in FRMSs, in the diagnosis, treatment, and return to work (when possible) of employees whose sleep disorders put them at elevated risk for fatigue-related impairment.[95] On the other hand, long waiting lists and high costs can be a significant disincentive for disclosure by an employee who is experiencing chronic sleep problems. Employees also may be facing loss of their livelihood if they fail to meet medical standards for fitness to work (e.g., in commercial driving and aviation).[96] The sleep medicine community needs to engage in active dialogue with regulators about these issues to ensure that any regulatory requirements imposed are appropriate and workable.

CONCLUSIONS AND FUTURE DIRECTIONS

FRMSs represent a new approach that seeks to integrate advances in sleep science, circadian science, and safety science to reduce the occupational hazard of workplace fatigue. As implementation and development of FRMSs expand, there will be increasing opportunities for scientists with expertise in sleep science and chronobiology to contribute. Such participation requires a willingness to enter into respectful

transdisciplinary collaborations with operational experts to find workable solutions for managing different fatigue hazards. Scientists also have a role in developing best practice standards for data collection and analysis in FRMSs, including when and how independent ethical approval should be obtained for data collection, and advocating for publication of findings in the open peer-reviewed literature whenever possible. Combining data from multiple studies enables evaluation of the independent contributions to fatigue-related impairment of the different operational and physiologic factors and has exciting potential for increasing scientific understanding and improving FRMSs.

CLINICAL PEARL

Fatigue risk management systems (FRMSs) are a new approach that integrates advances in sleep science and safety science to reduce the occupational hazard of fatigue-related performance impairment in the workplace. The core of an FRMS is a closed-process loop comprising (1) ongoing monitoring of fatigue levels, (2) identification of situations where fatigue may constitute a hazard, (3) risk assessment, and (4) when needed, introduction of risk mitigation measures whose effectiveness is monitored by step 1. This is an exciting new area of application of sleep and circadian science that highlights areas where better scientific understanding is needed.

SUMMARY

The traditional approach to managing workplace fatigue risk has been to impose regulatory limits on maximum work hours and minimum breaks within and between work periods (hours-of-service regulations). This approach, however, does not address all of the known causes of fatigue-related performance impairment and represents a single-layer defensive strategy. FRMSs are a flexible alternative approach based on scientific understanding of the dynamic effects of sleep loss and recovery on performance and of the modulating influence of the circadian system on performance capacity. The core of an FRMS is a closed-loop process that includes (1) ongoing monitoring of fatigue levels, (2) identification of situations where fatigue may constitute a hazard, (3) risk assessment, and (4) introduction of risk mitigations when needed. The effectiveness of mitigation strategies is in turn tracked by the ongoing monitoring of fatigue levels. This closes the FRMS process loop, which relies on measuring or estimating operator fatigue. Multiple defensive strategies can be incorporated, tailored to the operational environment, the level of operator fatigue, and the associated safety risk. Improving FRMSs requires dialogue between scientists and people with operational expertise in the environment(s) where these systems are implemented. FRMSs will improve as scientific knowledge grows and implementation expands.

ACKNOWLEDGMENTS

The ideas and constructs presented in this chapter have been developed through collaborative projects and discussions with many academic colleagues, industry partners, and regulatory agencies over a number of years. Responsibility for the content of the chapter rests with us alone.

Selected Readings

Banks S, Van Dongen HP, Maislin G, Dinges DF. Neurobehavioral dynamics following chronic sleep restriction: dose-response effects of one night for recovery. *Sleep* 2010;**33**:1013–26.

Darwent D, Dawson D, Paterson JL, et al. Managing fatigue: it really is about sleep. *Accid Anal Prev* 2015;**82**:20–6.

Ferguson SA, Paech GM, Sargent C, et al. The influence of circadian time and sleep dose on subjective fatigue ratings. *Accid Anal Prev* 2012;**45**:50–4.

Gander P, Hartley L, Powell D, et al. Fatigue risk management: organizational factors at the regulatory and industry/company level. *Accid Anal Prev* 2011;**43**:573–90.

Gander PH, Mangie J, van den Berg MJ, et al. Crew fatigue safety performance indicators for fatigue risk management systems. *Aviat Space Environ Med* 2014;**85**:139–47.

Gander PH, Mulrine HM, van den Berg MJ, et al. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *J Sleep Res* 2015;**24**(1):110–19.

International Civil Aviation Organization. *Fatigue risk management systems: manual for regulators*. Montreal: International Civil Aviation Organization; 2012 [Report No.: Doc 9966].

McDonald J, Potyk D, Fischer D, et al. Napping on the night shift: a study of sleep, performance, and learning in physicians-in-training. *J Grad Med Educ* 2013;**5**:634–8.

Mollicone DJ, Van Dongen H, Dinges DF. Optimizing sleep/wake schedules in space: sleep during chronic nocturnal sleep restriction with and without diurnal naps. *Acta Astronaut* 2007;**60**:354–61.

Randolph SA. Fatigue risk management. *Workplace Health Saf* 2015;**63**:236.

Zhou X, Ferguson SA, Matthews RW, et al. Sleep, wake and phase dependent changes in neurobehavioral function under forced desynchrony. *Sleep* 2011;**34**:931–41.

A complete reference list can be found online at ExpertConsult.com.

Chapter
74

Drowsiness in Transportation Workers

Pierre Philip; Patricia Sagaspe; Jacques Taillard

Chapter Highlights

- For years, fatigue has been associated with increased risk of accidents, but the causes of fatigue were unclear. More recently (over the past 20 years), drowsiness and falling asleep at the wheel have been identified as causes of some fatal crashes and traffic accidents.
- Sleep physicians tend not to pay sufficient attention to the problem of drowsy driving in their clinical evaluation of nonprofessional drivers, and better awareness in the health community will help in reducing the risk of accidents.
- Objective measures of sleepiness such as the Maintenance of Wakefulness Test (MWT) seem to be the most suitable tools to assess fitness to drive, compared with subjective self-assessments.
- Public campaigns integrating information on sleep hygiene and countermeasures as well as promotion of sleep awareness within the medical community to combat drowsy driving will significantly improve driving safety.

Traffic accidents are an increasing cause of death and injury around the world, and several countries have launched road safety campaigns in recent decades to decrease mortality and morbidity on their roads.[1] Alcohol and excessive speed have been highlighted in past years as major threats, but quite surprisingly, the sleep health status of drivers has received little attention.

For many years, fatigue has been associated with an increased risk of accidents, but the causes were unclear. Extended or nocturnal work or driving was correlated with accidents, but few reports differentiated fatigue (which usually is seen as due to driving time) from sleepiness (which is due to reduced sleep),[2] extended time awake, and/or being awake at the circadian trough,[3] alone or in combination with the influence of drugs.

Results from epidemiologic studies from the 1990s showed that sleep-related accidents represent up to 20% of all traffic accidents in industrial societies.[4-6] Although drowsiness[7-9] has been identified as the reason behind fatal road crashes and many industrial accidents,[10] many people drive when alertness is at its lowest level. Results from one case-control study[6] has shown that driving between 2 and 5 o'clock in the morning multiplies 5.6-fold the risk of traffic accidents and that being sleepy at the wheel multiplies eightfold the risk of accidents. These statistics provide clear measures of the risk associated with drowsy driving.

Results from an Internet-based survey[11] showed that inappropriate line crossings related to drowsy driving was a good predictor of future sleep-related accidents. Similarly, a Swedish epidemiologic study showed that more than 40% of the reported incidents involved crossing the far right line (on the passenger side of the automobile) before awaking and 16% involved crossing the center line.[12]

Although both the European Union and the United States have launched public campaigns to make their citizens aware of the risk of drowsy driving (e.g., the U.S. National Sleep Foundation "Drive Alert–Arrive Alive" campaign, the European Sleep Research Society's "Wake Up Bus" campaign), major problems remain in (1) the identification of patients who are at risk (or behaviors that put individuals at risk) for traffic accidents and (2) the best way to reduce these risks by appropriate countermeasures.

This chapter presents an update on (1) the relationships among extrinsic or intrinsic sleep disorders, drug intake, and traffic accidents; (2) the present state of knowledge; and (3) studies needed to improve safety.

PREVALENCE AND ASSOCIATED RISKS

Impact of Rest-Activity Patterns on Driving

Behavioral changes affecting the sleep-wake pattern can lead to sleepiness-related accidents. Studying large populations of drivers,[13,14] we demonstrated that long-distance driving frequently was associated with sleep curtailment (advanced wake-up time). Results from several experimental studies conducted by our group showed that sleep restriction[15] or nocturnal driving[16] impaired driving performance; the number of inappropriate highway line crossings was significantly increased by both.

Sleep deprivation affects not only the general population of automobile drivers but also many professional (commercial motor vehicle) drivers all over the world. Results from a study of professional truck drivers[17] revealed a mean duration of sleep of 4.78 hours/day in a 5-day period. Within that period, 56% of drivers displayed at least 6 noncontinuous minutes of sleep as recorded on the electroencephalogram while driving. The vast majority of these "micro" sleep episodes occurred during the late night and early morning.

Health care workers also are affected by sleep loss and frequently must drive home after a night on call. Results from

a prospective survey[18] in 2737 medical residents provided detailed information about work hours, work shifts of an extended duration, and documented motor vehicle accidents. The odds ratio (OR) for reporting a privately owned motor vehicle accident and for reporting a near-miss driving incident after an extended work shift was 2.3 (95% confidence interval [CI], 1.6 to 3.3), compared with an OR of 5.9 (95% CI, 5.4 to 6.3) for a shift that was not of extended duration.

Results of another study showed that extensive nocturnal driving worsens driving performance.[19] Eight hours of nocturnal driving from 9 P.M. to 5 A.M. increased six-fold the number of inappropriate highway line crossings over that for a 2-hour nocturnal driving session (3 to 5 A.M.). These results suggest that the fatigue caused by extended driving duration is amplified at night (implying that maximal driving duration should be shorter for nighttime driving than for daytime driving). The effects of driving duration on lane-tracking can be appreciated by comparing them with impairments in lane-tracking caused by (higher) blood alcohol concentration (BAC) levels. For example, 2 hours of continuous nocturnal driving were sufficient to produce driving impairment comparable to a BAC of 0.05%; after 3 hours of continuous nocturnal driving, impairment corresponded to a BAC of 0.08%.[20]

Impact of Sleep Disorders on Driving

Sleep Apnea

Of all sleep disorders, obstructive sleep apnea syndrome is possibly the most-studied pathologic process with regard to traffic accidents. Indeed, results from several studies performed in the past 20 years showed a clear positive relationship between sleep disorders and traffic accidents.[21-27,27a]

In a study to evaluate the additional accident risk related to sleep-disordered breathing,[28] patients with an apnea-hypopnea index (AHI) of 10 or higher had an OR of 6.3 (95% CI, 2.4 to 16.2) for having a traffic accident as compared with those without sleep apnea. Researchers in another study[29] found increased accident risk in 460 apneic patients, although only the most severely affected patients (AHI ≥30) presented an accident risk factor higher than that of the control subjects. In a third study,[30] the investigators performed an integrated analysis of recordings of sleep-related breathing disorders and self-reported automotive and company-recorded automotive accidents in 90 commercial long-haul truck drivers. In this study, truck drivers with sleep-disordered breathing had a twofold higher accident rate per mile than drivers without sleep-disordered breathing. In contrast with the study cited earlier,[29] accident frequency was not dependent on the severity of the sleep-related breathing disorder.[31] In another study on professional drivers,[8] more than 20% of long-haul drivers reported having dozed off at least twice while driving. Near-misses due to dozing off occurred in 17% of these drivers.

Debate is ongoing regarding the best predictive symptom for risk of sleep-related accidents. Surprisingly, in one study, excessive daytime sleepiness as measured by the Epworth Sleepiness Scale (ESS) was not associated with increased accident risk in apneic patients.[28] This finding could be explained by a low percentage of chronically sleepy drivers in the studied population (only 9% scored above 10 [i.e., borderline or higher sleepiness] on the ESS). In a case-control study[32] of a series of 189 consecutive patients and a control group of 40 hospital staff workers, the best predictors of traffic accidents were self-reported sleepiness while driving (OR, 5; 95% CI, 2.3 to 10.9), having quit driving because of sleepiness (OR, 3; 95% CI, 1.1 to 8.6), and being currently working (OR, 2.8; 95% CI, 1.1 to 7.7). Similarly, a sample of 4002 randomly selected drivers[33] were interviewed to define the prevalence of habitual sleepiness among drivers. The habitually sleepy drivers reported a significantly higher frequency of vehicular accidents than control subjects (adjusted OR, 13.3; CI, 4.1 to 43.0) and a significantly higher prevalence of respiratory sleep disorders than in control subjects. The authors concluded that habitually sleepy drivers are a large group (1 in 30 drivers) who are involved in several-fold more vehicular accidents than control subjects. More recently, in a meta-analysis[34] of sleep apnea and driving risk, it was found that results from 23 of 27 studies and 18 of 19 studies with control groups revealed a statistically significant increased risk of having a driving accident associated with sleep apnea, with results from many of the studies indicating a two-fold to three-fold increased risk.

Narcolepsy and Other Hypersomnias

Narcolepsy also is associated with excessive daytime sleepiness, and it also has been studied as a risk factor for traffic accidents. Narcoleptic patients are at higher risk for sleep-related accidents than those with sleep apnea.[22] The proportion of persons with sleep-related accidents is 1.5- to 4-fold greater in hypersomnolent patients than in control subjects. Apneic and narcoleptic persons account for 71% of all drivers involved in sleep-related accidents. Surprisingly, results of Multiple Sleep Latency Tests (MSLTs) did not correlate with the rate of accidents among sleepy patients. However, the number of patients in the study who underwent sleep latency testing was limited (46 apneics, 22 narcoleptics, 17 other patients diagnosed with various causes of excessive daytime sleepiness), suggesting a lack of statistical power. It is worth noting that regardless of diagnosis, sleep-disordered victims of accidents presented with sleep latencies shorter than those of control subjects.

Elevated risks for motor vehicle accidents due to sleepiness and cataplexy have been reported for persons with untreated narcolepsy.[35] In one study,[36] researchers compared performance on a simulated driving task (the Divided-Attention Driving Test) for 21 male patients with obstructive sleep apnea, 21 sex-matched control subjects, and 16 narcoleptic patients. Narcoleptic patients were younger and sleepier than the patients with obstructive sleep apnea. Tracking error (measured as distance in centimeters from the center of the driving lane) was greater in both patient groups than in control subjects (228 ± 145 cm for obstructive sleep apnea vs. 196 ± 146 for narcolepsy versus 71 ± 31 for control subjects; $P < .001$). Under actual driving conditions,[37] driving performance (measured as inappropriate line crossings and standard deviation of lateral position of the vehicle) of narcoleptics and patients with idiopathic hypersomnia was worse than performance of control subjects. Results from this study confirm that patients with untreated hypersomnia (regardless of etiology) are at higher risk for sleepiness-related driving accidents than persons without hypersomnia.

Insomnia and Attention Deficit–Hyperactivity Disorder

Although not always associated with excessive daytime sleepiness, insomnia also has been evaluated as a risk factor for

driving accidents. Results from a cross-sectional study conducted in 10 different countries showed high rates of car accidents related to sleep disturbances among patients with insomnnia; this association was independent of any adverse effects associated with hypnotic treatment.[38] As expected, reduced total sleep time was one factor explaining the high risk of car accidents for people who complain of insomnia.

Results from a recent study of patients with attention deficit–hyperactivity disorder[39] showed that a significant number of these patients exhibited altered levels of alertness in addition to attention problems. Patients in this study were drug-free. Results from driving simulator measurements combined with electrophysiologic measures showed that subjects with sleep latencies shorter than 19 minutes had significantly worse driving performance than the other patients and control subjects.

Impact of Hypnotics and Other Drugs on Driving Risk

Although results from numerous publications[40-44] have shown an association between central nervous system drugs and risk of accidents, very few results show a direct link between sleep-related accidents and drug intake. According to the French medication classification system, an increased risk of being responsible for a road traffic crash for users of prescribed medicines is defined as a level 2 ("be very careful") or level 3 ("danger: do not drive") risk of driving impairment. The percentage of road traffic crashes attributable to level 2 and level 3 medicine use was 3.3% (2.7% to 3.9%).[45] Results from a study conducted under actual driving conditions demonstrated that a single lorazepam dose (2 mg given by mouth) caused higher standard deviation of lateral position increases than a BAC above 0.05%.[46] Antiepileptics[47] and antidepressants themselves were not associated with risk of a road traffic crash, but this risk increased on *initiation* of antidepressant treatment (OR, 1.49; 95% CI, 1.24 to 1.79) and after a *change in* antidepressant treatment (OR, 1.32; 95% CI, 1.09 to 1.60).[48] Hypnotics with long half-lives (i.e., medium- and long-acting benzodiazepines and antihistamines) carry an increased risk of accidents when subjects drive in the morning.[49] The types of drugs associated with traffic accidents (i.e., hypnotics and benzodiazepines) suggest that insomnia-related sleepiness could be a major underlying cause of drug-related accidents, but further evidence is needed. Users of methadone and buprenorphine were more likely to be responsible for injurious road traffic crashes. This increased risk could be explained by the combined effect of drug addiction, risky behaviors, and its treatment.[50]

EVALUATION OF DRIVING RISK IN PATIENTS WITH SLEEPINESS

As noted earlier, some debate remains regarding the best predictor of risk of a sleep-related accident. Results of two studies[32,33] showed that simply asking a patient specifically about excessive sleepiness while driving may better predict which patients with sleep-disordered breathing are at risk for accidents (as opposed to asking about overall sleepiness). Thus a thorough clinical interview can serve to evaluate a patient's driving risk in most cases. Nevertheless, this strategy requires a truthful subjective self-assessment by the patient. Patients who are drivers dependent on their driving license for their job may be less willing to be truthful if they believe that their commercial driving license may be revoked.

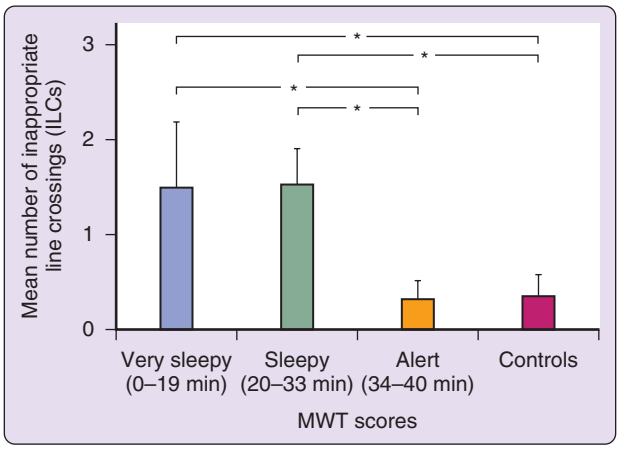


Figure 74-1 Mean number of inappropriate line crossings (ILCs) during real driving (mean ± SE) in the three sleep latency groups on the Maintenance of Wakefulness Test (MWT) and in healthy control subjects. *$P \leq .05$. (From Philip P, Sagaspe P, Taillard J, et al. Maintenance of Wakefulness test, obstructive sleep apnea syndrome, and driving risk. *Ann Neurol* 2008;64:410–6.)

Few studies have investigated the relationship between objective measurement of sleepiness (or Maintenance of Wakefulness Test [MWT] scores) and driving performance. In a study[51] of the Wisconsin Sleep Cohort, the investigators found a correlation between MSLT scores and driving accidents in male apneic drivers: Lower MSLT scores (indicating greater sleepiness) were associated with higher number of reported driving accidents. In another study,[52] comparing MWT scores with performance on a driving simulator in healthy sleep-deprived volunteers, the authors reported the first evidence of the predictive value of the MWT for driving performance. When additional studies were performed in a driving simulator and under actual driving conditions, the authors confirmed that lower MWT scores (indicating greater sleepiness) are significantly associated with impaired driving (i.e., number of inappropriate highway line crossings)[53,54] (Figure 74-1). This study showed that the MWT (administered four times per day using a 40-minute protocol) is a suitable clinical tool to assess fitness to drive in patients with hypersomnias of central origin (narcolepsy or idiopathic hypersomnia) as well as in patients with obstructive sleep apnea.[55] It seems reasonable to consider that patients with MWT sleep latency times under 19 minutes are "unsafe" to drive. Physicians should consider both objective measures (MWT scores) and clinical evaluation (self-reports of sleepiness at the wheel) when providing recommendations about fitness to drive in these clinical populations. Further data are needed to confirm the predictive value of the MWT for accident risk in larger cohorts of apneic and narcoleptic patients.

IMPACT OF TREATMENT AND COUNTERMEASURES ON ACCIDENT RISK

In view of the established sleepiness-related risk of a driving accident, a critical question is how accidents involving these patients can be reduced.

Uvulopalatopharyngoplasty has been evaluated as a therapeutic strategy to treat obstructive sleep apnea and thereby

reduce risk of a driving accident.[56] In one study, the investigators compared the rate of car accidents in 56 apneic patients for 5 years after surgery versus 5 years immediately before surgery. Risk in apneic patients was compared with that in a control group of subjects followed for nasal surgery. Eighty-seven percent ($P \leq .001$) of apneic patients who had reported habitual sleepiness while driving preoperatively reported no sleepiness while driving postoperatively. Accident risk reduction (corrected for mileage) in apnea patients was almost four times greater than the reduction in control subjects ($P \leq .001$) after surgery.

Results from several studies investigating the impact of continuous positive airway pressure (CPAP) on traffic accidents[57-59] confirmed that this therapy is associated with reduced risk of motor vehicle accidents due to obstructive sleep apnea.

For the first time, results from a study conducted under actual driving conditions demonstrated that modafinil (400 mg) improved on-road driving ability (mean number of inappropriate line crossings and standard deviation of lateral position of the vehicle) in narcoleptic and idiopathic hypersomnia patients although lane-tracking was not improved to the level seen in healthy control drivers.[37] This important finding confirms the usefulness of alerting drugs for improving driving safety. The results also indicate that systematic driving evaluations are useful for measuring the efficacy of sleep disorders treatments.

Although cold air or listening to the radio has not demonstrated any efficacy,[60] coffee and naps are effective for combating sleepiness at the wheel.[16,57-59] We have shown that the efficacy of specific countermeasures varies according to age and individual physiology.[61] Indeed, a cup of coffee containing 200 mg of caffeine significantly improved performance in both young (20 to 25 years) and middle-aged participants (40 to 50 years) on nighttime highway driving performance, whereas a 30-minute nap was more effective in younger than in middle-aged drivers. Recently, we showed that in-car continuous nocturnal blue light exposure could be used to mitigate nocturnal sleepiness at the wheel in blue light–tolerant drivers, regardless of age.[62] Critically, nighttime coffee intake or blue light exposure did not modify the quality, quantity, and timing of three subsequent nocturnal sleep episodes. During daytime driving, conversation with a passenger also was shown to contribute to safer lane-keeping in a driver in whom sleepiness was induced by benzodiazepine ingestion.[63] Whether cell phones and other electronic devices that enable social contact while driving may actually improve driving safety (as opposed to serving as a distractor) under some conditions (e.g., in solitary drivers) has yet to be confirmed.

DRIVER'S LICENSE REGULATIONS

Excessive daytime sleepiness and several sleep disorders have been targeted by experts as medical conditions impairing driving skills. As part of the European Cooperation in Science and Technology Action B-26 (the main objectives of which are to assess the role of obstructive sleep apnea syndrome as a possible cause of increased cardiovascular risk and coordinate studies on pathogenetic mechanisms of increased cardiovascular risk associated with several diseases—see http://www.cost.eu/domains_actions/bmbs/Actions/B26)—a review paper[64] was generated in which driving license regulations in 25 European countries was assessed. Excessive daytime sleepiness was mentioned as a medical handicap for driving in 9 countries (Belgium, Finland, France, Germany, Hungary, the Netherlands, Spain, Sweden, and the United Kingdom), whereas sleep apnea syndrome was mentioned in 10 countries (Belgium, Finland, France, Germany, Hungary, the Netherlands, Spain, Sweden, the United Kingdom, and Poland). In all these European countries, a patient with untreated obstructive sleep apnea syndrome is considered unfit to drive. Recently, the Commission Directive 2014/85/EU (amending Directive 2006/126/EC on driving licenses) was adopted. Introducing obstructive sleep apnea syndrome in the Annex for fitness to drive, this directive represents a major step forward in the official recognition of sleep disorders as a major factor in driving safety.

Implications for patients and physicians differ in each country. In the vast majority of European Union countries, once a diagnosis of sleep disorder is made it is the physician's responsibility to inform the administrative authorities issuing driving licenses of the driver's condition. This is not the case in four countries (Belgium, France, Germany, and The Netherlands), where the physician is expected to inform the patient, but not the authorities, that he or she is unfit to drive. Medicolegal authorizations for driving once a patient is diagnosed with a sleep disorder require a certificate delivered by a general practitioner or specialist (pulmonologist or neurologist) in eight countries (Belgium, Finland, France, Germany, Hungary, Spain, the United Kingdom, and Poland). This certificate is based on the patient's clinical improvement and therapeutic compliance, but in two countries the final decision for fitness to drive is determined by the patient's self-evaluation.

Despite available scientific evidence,[34] many countries in Europe do not yet include sleep apnea and excessive daytime sleepiness as risk factors for traffic accidents. A unified European directive seems desirable and should include several sleep disorders and excessive daytime sleepiness in the list of medical conditions adversely affecting driving skills.[65,66]

Several states within the United States require physicians to report to the Motor Vehicle Administration if a professional driver is diagnosed with a sleep disorder. Such a reporting requirement can have dramatic negative consequences for the driver's continued employment. A 2014 conference on Neurological Disorders and Commercial Drivers recommended that a person with a diagnosis of hypersomnia and/or restless legs syndrome should be disqualified from driving commercially.[67] In addition, narcoleptic patients and those with untreated apnea should be disqualified from commercial interstate driving.

This restrictive attitude regarding clinical conditions and driving fitness is not followed in Europe.

Owing to the potentially devastating consequences of disqualifying a professional driver, it is critical that a driver suspected of having sleep apnea be diagnosed by a physician specializing in sleep medicine based on polysomnography, preferably in an accredited sleep laboratory. A full-night study should be conducted unless a split-night study is indicated (severe obstructive sleep apnea identified after at least 2 hours of sleep). Treatment (e.g., CPAP) should be started as soon as possible (i.e., within 2 weeks of the sleep study). After a minimum of 2 weeks after initiating therapy (and preferably within 4 weeks), the driver should be reevaluated by the sleep specialist and compliance (as well as resolution of sleepiness)

assessed. If the driver is compliant, the driver can return to work but should be initially certified for no longer than 3 months, at which time continued compliance and resolution of sleepiness are assessed. Such an approach has been shown to reduce health care and disability costs.[59]

In Europe, driving regulations are similar to U.S. recommendations for apneic commercial drivers, with minor changes in terms of period of evaluation (e.g., 1 year in France).

Finally, whereas in some countries sleepiness while driving is considered the major problem regarding driver safety, only France requires an objective quantification of alertness (the MWT) to evaluate fitness to drive. A mean sleep latency score under 19 minutes associated with severe episodes of sleepiness at the wheel should disqualify patients from driving.[37] This specific requirement thus addresses the frequent combination of sleep disorders and poor sleep hygiene in truck drivers—that is, the requirement specifies the required outcome (objectively demonstrated alertness) rather than identifying the specific cause of sleepiness. To ensure legal protection for drivers and physicians, French experts determined that it was mandatory to demonstrate objectively that patients respond to treatment before allowing them to drive again. In case of a sleep-related accident in treated drivers, physicians could not be sued for insufficient efficacy of treatment and patients should not be prosecuted for misreporting the beneficial effects of treatments.

Based on recent data[37,54,55] collected for treated and untreated patients suffering from nocturnal breathing disorders and also hypersomnia, it seems reasonable to consider that patients presenting with mean sleep latencies under 19 minutes on the MWT are at major risk for sleep-related traffic accidents.

FUTURE CONSIDERATIONS

Although much has already been done in this field, many issues remain. First, at the diagnostic level, no simple objective measure to quantify sleepiness is available, as are available for other accident risk factors (e.g., using a breathalyzer for alcohol testing). Ideally, a "somnotest" is needed to quantify sleepiness-related driving risk, but driving simulators and electroencephalogram measures provide indirect and variable estimation of driving risk and have not been adapted for field use.

Second, treatments other than CPAP (or uvulopalatopharyngoplasty), such as mandibular advancement devices, could provide alternatives to treat sleepiness in lower-risk (less-impaired) patients. The efficacy of oral appliances has not yet been studied specifically regarding driving risk, but they have been shown to reduce sleepiness in patients with mild sleep apnea.[68]

Evaluating the impact of driving durations (and at different times of day) in treated and untreated patients is also important because of the high prevalence of sleep-disordered breathing in professional drivers. These studies are needed not only to define the phenotype of apneic persons involved in traffic accidents but also to determine when these drivers are at greatest risk for a sleepiness-related accident. Even though only 1 patient in 30 with sleep-disordered breathing is involved in a sleepiness-related accident, tracking these persons over time may better inform driving recommendations (e.g., no nocturnal driving) in this population.

Prescription stimulants may have some utility as adjunct treatments for sleepiness due to sleep-disordered breathing,[69] but their impact on driving risk is not known. Pharmacologic countermeasures should be tested and dosing recommendations developed. In addition, guidelines and recommendations regarding fitness to drive in patients with other disorders in which sleepiness is a symptom are needed.[70] Finally, public campaigns on the risks of drowsy driving and the effectiveness of naps and coffee as countermeasures need to be mounted in every country.

CLINICAL PEARLS

- Sleepiness at the wheel is now identified as a cause of many fatal traffic accidents.
- Sleep restriction and/or sleep disorders have been identified in up to 20% of traffic accidents. Use of sedative drugs affecting the central nervous system also has been associated with increased risk of driving accidents.
- Caffeine and naps have demonstrated efficacy for decreasing driving impairment in sleepy drivers.
- Medical treatments such as CPAP are efficacious in drivers with sleep apnea.
- Prescription stimulants may prove to be useful adjuncts for decreasing accident risk in sleepy patients.
- In many countries, it is a legal requirement to systematically evaluate medical disorders among professional drivers, but criteria are variable; an effort toward standardization would be beneficial. Objective measurements of sleepiness such as the MWT can be used to determine fitness to drive. Useful information also can be obtained directly from the patient by specific inquiries regarding subjective sleepiness while driving.
- Finally, public awareness of the danger of drowsy driving needs to be reinforced. Physicians can and should play a key role in educating their patients on risks associated with driving while sleepy.

SUMMARY

Extended work hours and nocturnal driving have been associated with driving accidents—falling asleep at the wheel associated with insufficient sleep and nocturnal driving have been identified as causal factors in up to 20% of traffic accidents. Drugs affecting the central nervous system (i.e., benzodiazepines, antidepressants, narcotic analgesics, and antihistamines), insomnia, nocturnal breathing disorders, and narcolepsy also have been associated with an increased risk of driving accidents. Self-reported drowsy driving is associated with increased risk of accidents. Objective measures of sleepiness such as the MWT can predict accident rates and deteriorated driving performance. Even if not easy to implement in daily clinical practice, the MWT remains the more suitable tool to assess fitness to drive. An MWT sleep latency under 19 minutes combined with self-reported sleepiness at the wheel characterizes status of a patient who is not capable of safe driving. Results from a recent study of actual driving showed that modafinil significantly reduces driving impairment in diagnosed narcoleptic and hypersomniac patients. Further studies of objective and subjective sleepiness before and after treatment are required in patients with periodic leg movement syndrome, insomnia, and circadian rhythm disorders (and other disorders associated with sleepiness) to determine cutoff criteria that can be used to determine safe driving. Treatments

improving daytime vigilance (e.g., CPAP in obstructive sleep apnea) or drugs that promote alertness significantly reduce the risk of traffic accidents for a reasonable cost. Many countries systematically evaluate medical disorders in professional drivers, but the criteria are variable. Efforts to achieve standardization would be beneficial, especially in Europe and in the United States, where sleepy drivers cross several countries or states, each of which may maintain different driving regulations. Regardless of cause, sleepiness at the wheel—a key factor associated with increased risk for motorvehicle accidents—remains underdiagnosed.

Selected Readings

Connor J, Norton R, Ameratunga S, et al. Driver sleepiness and risk of serious injury to car occupants: population based case control study. *BMJ* 2002;**324**:1125.

Duffy JF, Zitting KM, Czeisler CA. The case for addressing operator fatigue. *Rev Hum Factors Ergon* 2015;**10**(1):29–78.

Ellen RL, Marshall SC, Palayew M, et al. Systematic review of motor vehicle crash risk in persons with sleep apnea. *J Clin Sleep Med* 2006;**2**: 193–200.

Gander PH, Mulrine HM, van den Berg MJ, et al. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *J Sleep Res* 2015;**24**(1):110–19.

Hakkanen H, Summala H. Sleepiness at work among commercial truck drivers. *Sleep* 2000;**23**:49–57.

Lloberes P, Levy G, Descals C, et al. Self-reported sleepiness while driving as a risk factor for traffic accidents in patients with obstructive sleep apnoea syndrome and in non-apnoeic snorers. *Respir Med* 2000;**94**: 971–6.

Mitler MM, Miller JC, Lipsitz JJ, et al. The sleep of long-haul truck drivers. *N Engl J Med* 1997;**337**:755–61.

Orriols L, Delorme B, Gadegbeku B, et al. Prescription medicines and the risk of road traffic crashes: a French registry-based study. *PLoS Med* 2010;**7**:e1000366.

Philip P, Chaufton C, Taillard J, et al. Maintenance of Wakefulness Test scores and driving performance in sleep disorder patients and controls. *Int J Psychophysiol* 2013;**89**:195–202.

Philip P, Chaufton C, Taillard J, et al. Modafinil improves real driving performance in patients with hypersomnia: a randomized double-blind placebo-controlled crossover clinical trial. *Sleep* 2014;**37**:483–7.

Philip P, Taillard J, Moore N, et al. The effects of coffee and napping on nighttime highway driving: a randomized trial. *Ann Intern Med* 2006;**144**: 785–91.

Philip P, Vervialle F, Le Breton P, et al. Fatigue, alcohol, and serious road crashes in France: factorial study of national data. *BMJ* 2001;**322**: 829–30.

Sagaspe P, Taillard J, Bayon V, et al. Sleepiness, near-misses and driving accidents among a representative population of French drivers. *J Sleep Res* 2010;**19**:578–84.

Taillard J, Capelli A, Sagaspe P, et al. In-car nocturnal blue light exposure improves motorway driving: a randomized controlled trial. *PLoS ONE* 2012;**7**:e46750.

A complete reference list can be found online at ExpertConsult.com.

Chapter 75

Shift Work, Shift-Work Disorder, and Jet Lag

Christopher L. Drake; Kenneth P. Wright, Jr.

Chapter Highlights

- Circadian misalignment during shift work and jet lag negatively affects sleep, cognition, and work productivity and increases the risk of accidents on the job and during the commute to and from work.
- Shift work negatively affects mood and cardiovascular and gastrointestional health and is associated with an increased risk for developing cancer; jet lag is associated with gastrointestinal distress.
- Clear individual differences (genetic and behavioral) are recognized in the ability to adapt to shift work and time zone travel.
- This chapter emphasizes circadian science principles and pharmacologic strategies for the treatment of shift work disorder and jet lag.

The increased use of technologies such as artificial light and increased air travel have increased exposure to sleep-wake schedules that oppose internal circadian physiology. The ensuing circadian rhythm–mediated disruptions in sleep-wake and other physiologic processes are associated with significant morbidity and mortality. Other chapters in this book address the physiologic basis for conditions such as jet lag and shift work disorder (SWD) in terms of biologic regulation of the circadian system. The discussion in this chapter bridges basic science and laboratory studies and describes the impact of the underlying physiology on health. A summary is provided of results from available laboratory and field-based occupational health studies that can inform the clinician and the patient about clinical issues central to SWD and travel across time zones (jet lag).

SHIFT WORK

Prevalence

The current total working population of the United States is approximately 146 million.[1] Estimates regarding prevalence of shift work vary depending on the definition employed and the region studied, but U.S.-based estimates indicate that nearly 20% of employed adults are shift workers. The proportion is higher if workers engaged in early-morning shifts and infrequent or irregular shifts are included. In the United States today, 17.7% to 25.9% of the total workforce start their shifts between 2 PM and 6:30 AM.[2] These data suggest that between 25.8 and 37.8 million U.S. adults engage in shift work on a regular or rotating basis. Data from other countries also indicate that a high prevalence of the population is engaged in shift work: In Slovenia, the estimated prevalence is 32%; in the United Kingdom, 22%; in Australia, 16%; in Greece, 25%; and in Finland, 25%.[3,4]

Not all persons who engage in shift work develop SWD. Factors including scheduling differences, shift frequency, shift duration, family and social responsibilities, and differences in sleep and circadian physiology affect an individual's response to shift work and hence the development of SWD. These same factors also are important influences on development of jet lag and its impact on work performance and health.

Types

Although the literature is not always precise in defining shift work in terms of start times, the following classifications are based not only on statistics from the U.S. Department of Labor but also on differences in circadian physiology.

Night Shift Workers

Night shift workers with regular start times between 6 PM and 4 AM make up an estimated 4.3% of the total U.S. workforce.[5] This is a conservative estimate, however, because it does not include workers on variable shift schedules. Some have speculated that permanent night work may have benefits in terms of circadian adjustment (compared with variable shift schedules) and accordingly have advised shift workers to stay on a night shift schedule on days off. Little support for this contention can be mustered, however.[6] Indeed, results from both objective and subjective metrics show that night shifts result in greater loss of total sleep time than evening and slow-rotating shift schedules.[7-9] Sleep loss accumulates over successive night shifts, resulting in a buildup of homeostatic sleep debt. The latter, combined with the effects of circadian misalignment, negatively affects productivity and safety in shift workers. Sleep loss alone has been shown to impair alertness and performance (including driving ability), with impairment comparable to that associated with up to 0.19% breath ethanol concentration.[10] Not surprisingly, the night shift produces the greatest degree of sleepiness relative to daytime work, evening shifts, and even rotating shifts, with the sleepiness greatest during the early morning hours close to commute times.[11,12]

Early-Morning Shift Workers

The *International Classification of Sleep Disorders* (ICSD3) classifies early-morning shifts as those starting between 4 AM and 7 AM.[13] This is the most common alternate work shift,

with at least 18.1 million U.S. workers (12.4% of the workforce) falling into this category.[5] Given these start times, many early-morning shift workers awaken before 5 AM. As a consequence, these workers are on the road during the circadian trough and may be partially sleep-deprived because of their early time of rising.[14] These consequences are consistent with the high rate of excessive sleepiness reported in this population.[15] Objective measures of sleep show that early-morning shift workers accrue significantly less sleep than day workers; stage 2 and rapid eye movement (REM) sleep were particularly reduced.[15] Indeed, sleep disturbance is close to that in permanent night workers.[16] These factors, coupled with severe sleep inertia at that hour,[17] suggest that early-morning shift workers are at highest risk of all workers for motor vehicle accidents. Further examination of the prevalence of excessive sleepiness and accidents in this specific population is needed.

Evening/Afternoon Shift Workers

Evening shift workers with regular start times between 2 and 6 PM make up 4.3% of all U.S. workers.[5] This category of workers is at risk for social isolation and consequent reduced quality of life.[18] Unlike early-morning shift workers (who obtain less total sleep time), the average evening-shift worker sleeps 7.6 hours/night,[8] which is more than most day workers obtain (6.8 to 7.0 hours/night).[12] Because the human circadian pacemaker has an intrinsic period that is on average slightly longer than 24 hours,[19] the resulting tendency to delay internal rhythms combined with schedules that allow later morning wakeup times may account for the increased total sleep time observed in evening-shift workers. However, some evening shift workers may have shortened sleep times owing to family obligations that require earlier wakeup times on days off; such reduced sleep time could result in significant impairment over time.

Rotating Shift Workers

The U.S. population is estimated to include 4 million rotating shift workers (2.7% of the total workforce),[2] but nearly all shift workers could be considered to have rotating schedules, because most revert to daytime wakefulness and nocturnal sleep on days off. Nonetheless, even on their days off, rotating shift workers remain sleepier than daytime workers.[20] In a meta-analysis of sleep patterns, the amount of sleep reduction in rotating shift workers was almost as much as in permanent night workers, relative to day workers.[8]

Workers with regular rotating shift schedules face additional challenges related to the speed and direction of shift rotations. Rapid shift rotations (e.g., multiple rotations within a week) are associated with reduced total sleep duration compared with slower rotations (e.g., at least 3 weeks per shift schedule).[8] With respect to direction, both rapid clockwise and counterclockwise rotations negatively reduce total sleep duration and increase circadian misalignment.[21] These effects are thought to be less severe for clockwise rotation, because of the natural tendency of the circadian clock to delay to a later time,[19] and the increased time between shifts with such scheduling. Shift direction and duration may interact, however; findings from one controlled study showed that rapid clockwise and counterclockwise rotations were not significantly different in terms of total sleep duration or degree of excessive sleepiness.[22] The latter finding may have been due to napping; before a counterclockwise rotation to the midnight shift, 80% to 90% of workers nap. By contrast, before a clockwise rotation to the midnight shift, only 40% to 60% of workers nap. In short, napping before a counterclockwise shift to midnights may help to ameliorate some of the impairments in sleep and sleepiness that otherwise would be expected. This interpretation also is consistent with numerous studies demonstrating the beneficial effects of napping among shift workers.[23,24]

Circadian Misalignment and Effects of Light Exposure

The internal circadian clock promotes sleep and related functions during the biologic night and wakefulness and related functions during the biologic day.[25] The circadian clock in humans must be reset on a daily basis to remain entrained to the 24-hour day.[26] Light is the dominant environmental time cue that entrains the human circadian clock to the 24-hour day,[27] and the timing of light exposure will determine whether the internal clock is phase-delayed or phase-advanced[28] (Figure 75-1). Circadian misalignment can be caused by shift work and time zone travel—both of these rapidly alter the temporal relationship between the internal circadian clock and environmental time cues. That is, both shift work and time zone travel require wakefulness during the biologic night and sleep during the biologic day. The resulting circadian misalignment can produce significant morbidity in the form of disturbed sleep, impaired alertness, gastrointestinal disturbances, and so on.

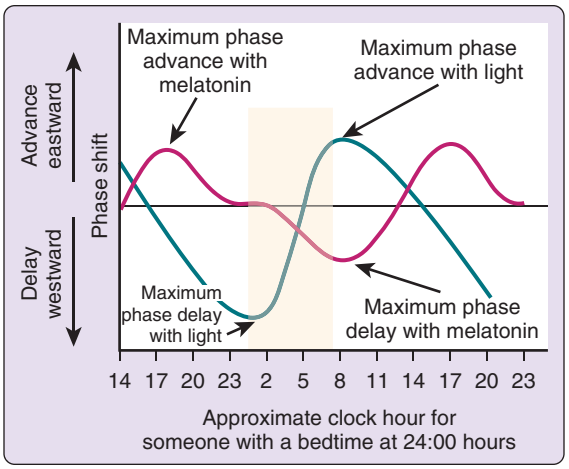


Figure 75-1 Schematic representation of the phase response curves to 1 day of light exposure (6.7 hours) (*blue line*) and 3 days of 3 to 5 mg of exogenous melatonin administration (*red line*) when the circadian system is entrained to local environmental time. The circadian phase resetting response to light and melatonin depends on the internal biologic time of exposure. Generally, bright light exposure before habitual bedtime and several hours thereafter will induce the largest westward phase delays, whereas bright light exposure just before the habitual time of awakening and several hours thereafter will induce the largest eastward phase advances. The time at which phase delays cross over to phase advances is, on average, approximately 2.5 hours before the habitual time of awakening in young adults and 2 hours in older adults.[164] Therefore bright light exposure close to the crossover point may shift the circadian phase in a direction opposite to what is desired. Opposite to the effects of light, ingestion of exogenous melatonin in the late afternoon will induce the largest eastward phase advances, whereas melatonin ingestion shortly after the habitual time of awakening and several hours thereafter will induce the largest westward phase delays. The time at which melatonin-induced phase delays change to phase advances is, on average, in the early afternoon.[106,107]

Morbidity Associated with Shift Work

Sleepiness and Insomnia

Among the most common problems experienced by shift workers are excessive sleepiness and insomnia.[29,30] For example, measures of sleepiness in anesthesia residents show impaired alertness after a single night shift; sleepiness levels were similar to levels observed in patients with diagnosed sleep disorders.[31] Shift workers may further exacerbate the problem by spending less time in bed during the day[32] to attend to everyday domestic activities. These latter issues are in addition to difficulties maintaining daytime sleep (i.e., when the circadian clock is promoting wakefulness).[29,30] Shift workers also may remain awake longer when changing from one shift schedule to another, producing a further accumulation of sleep debt. The known slow rate at which the circadian clock shifts (requiring days or even weeks) combined with erratic exposure to light in shift workers[33] results in impaired adaptation of biologic rhythms to shift work schedules. Thus, in some workers, both sleep disturbance and sleepiness continue even after months or years of shift work or even after discontinuation of long-term shift work.[34]

Reduced Alertness and Accidents

A high proportion of shift workers are involved in safety-sensitive operations, such as transportation, and one of the most well-documented and long-established effects of night shift work is an increased rate of motor vehicle accidents. Converging evidence from controlled laboratory settings, large epidemiologic studies, and clinical samples has established an incontrovertible link between shift work and accidents.[35,36] In a classic study, Smith and colleagues demonstrated that working a night shift increased the frequency of on-road accidents by 50%.[37] In a study of a large sample of nurses, 79.5% of those working the night shift reported at least one drowsy driving incident, equal to an increased odds ratio (OR) of 3.96 (95% confidence limit [CI], 3.24 to 4.84) relative to data for nurses working a day shift.[38] Residents who have frequent on-call schedules are involved in 6.7 times more motor vehicle accidents than those experienced by those working less-demanding call schedules.[39] Other studies have shown that risk increases nearly 10% for every extended shift worked in a month.[40] Driving home from the night shift is a time of particularly high risk, consistent with its proximity to the circadian nadir in alertness.[41,42] Increased rate of accidents after a night shift also is likely to be related to accumulated sleep loss. One study of night and rotating shift nurses showed that the odds for a hazardous driving event was eight times greater during the commute home from a night shift than during the pre-shift commute.[41]

Sleepiness-related impairment is not confined to motor vehicle accidents; for example, among medical personnel, an increased number of accidental injuries with sharp instruments (percutaneous injuries),[43] medication administration and diagnostic errors,[44] and increased patient death rates has been documented in those working extended and unconventional shift schedules.[45] These findings are not surprising, because driving impairment in residents on heavy call rotations in the hospital is similar to that in residents on light call rotations whose blood alcohol concentration is 0.05%.[46] In industrial settings, accidents and injuries increase on the night shift, and over successive night shifts, and with successive hours on a shift.[35]

Major catastrophes such as the event at Three Mile Island, the American Airlines flight 1420 crash, and the Chernobyl disaster occurred during the night shift, drawing increased attention to both the risks and costs associated with shift work schedules.[47,48] Cost of sleepiness-related accidents in the United States is estimated at up to $40 billion per year, representing 24% of the total cost of traffic accidents in the United States.[49] These data suggest that the economic savings associated with shift work for specific industries should be carefully weighed against the overall cost to society.

Work Productivity and Quality of Life

The negative impact of shift work is not limited to major adverse events and catastrophes. It also affects productivity. An association has been demonstrated between shift work and reduced dexterity and efficiency,[50] impaired threat detection,[51] and lower productivity.[52] Thus worker performance is significantly reduced at night in a broad range of occupational settings.[47] Additional evidence shows increased absenteeism in night workers compared with day workers, particularly among those experiencing insomnia and/or excessive sleepiness.[12]

Shift work negatively affects the worker's family as well as the worker's quality of life,[53] as evidenced by 57% higher divorce rates,[54] reduced job satisfaction,[55] and limitations on family and social interaction.[12] After controlling for a number of demographic variables, findings from a 5-year longitudinal study demonstrated a relationship between parental shift work and poor school performance and behavioral problems in children 5 to 12 years of age.[56]

Health Effects of Shift Work

A wealth of information documents the negative health effects associated with shift work. Notable are results from large prospective studies showing a 36% to 60% increased risk of breast cancer.[57,58] The latter effect is particularly evident with increasing years of night shift exposure. Additional findings include a fourfold increased risk of duodenal ulcers (verified by endoscopy)[59] and increased cardiovascular morbidity and mortality,[60-62] with specific disorders including atherosclerosis and myocardial infarction.[60] Poor eating habits[63] and other adverse health behaviors among shift workers may account for some of the increased morbidity.

Genetic Contributions to Shift Work Tolerance

People on the same shift schedule differ dramatically with regard to two of the most common consequences, excessive sleepiness and insomnia.[12,64] This large interindividual variability may be related in part to findings that most night workers' internal circadian rhythms (as indexed by endogenous melatonin levels) do not adapt to their adjusted sleep-wake schedule.[65,66,67] The subgroup of nonadapted workers show reduced daytime sleep relative to that in persons whose melatonin profiles showed rapid adaptation in response to night work.[68] Such individual differences in the circadian system itself (i.e., period, amplitude, response to light) may account for the adverse response (insomnia, excessive sleepiness) to shift work.

Several lines of research suggest that the aforementioned individual differences in adaptation to shift work are at least

partially genetically determined. Results from one study suggest approximately 50% genetic heritability for vulnerability to insomnia[69] (which can be elicited by a circadian misalignment[70]). Likewise, vulnerability to sleep loss during nocturnal hours is associated with a variable number tandem repeat (VNTR) polymorphism in the circadian clock gene *PER3*.[71] People possessing the *PER3*$^{5/5}$ genotype were shown to be more vulnerable than those possessing the *PER3*$^{4/4}$ genotype. It recently was shown that night shift workers who are carriers of the 5-repeat allele (*PER3*$^{5/5}$) experience a high degree of excessive sleepiness during the night shift, whereas *PER3*$^{4/4}$ homozygotes do not.[72] One predictor of shift work tolerance is circadian preference (i.e., "morningness" versus "eveningness"),[73] a heritable trait also linked to the VNTR polymorphism in the *PER3* gene[74] and the intrinsic period of the circadian clock.[75] Morning-type people tend to have a reduced tolerance for night shift work compared with evening-type people.[76] Morningness also has been associated with the *PER3*$^{5/5}$ genotype.

SHIFT WORK DISORDER

As described earlier, large individual differences in circadian and sleep-wake system responses to shift work have been observed.[77] In some cases, inability to adapt to shift work is extreme and constitutes shift work disorder (SWD). Diagnostic criteria for SWD are listed in Table 75-1.[13] People who are unable to tolerate the effects of a shift work schedule present with symptoms of excessive sleepiness or insomnia despite adequate time in bed (i.e., 7–8 hrs) and an absence of other sleep disorders. Although few studies have polysomnographically documented the insomnia associated with SWD, the sleep of patients with SWD is likely to be characterized by sleep fragmentation and early final awakenings. Such persons are unable to remain awake during the early-morning work/commute hours.[78] These deficits can adversely affect job performance, driving safety, quality of life, work satisfaction, and health.[12,64] In addition to better documentation of the sleep and alertness of persons with SWD, a major challenge will be to determine which specific forms of morbidity are associated with insomnia, excessive sleepiness, or circadian misalignment, or some combination of these three factors.

Prevalence

Clinical evaluation of insomnia and excessive sleepiness in shift workers is necessary to determine the actual prevalence of SWD, but few population-based studies are available. In one representative sample, the prevalence of the disorder has been conservatively estimated at 14% to 32% among night shift workers and 8% to 26% among rotating shift workers.[12] Findings from other studies provide similar estimates, with approximately 23% of a sample of oil rig workers in the North Sea,[79] approximately 38% of a sample of 5400 nurses working more than half-time,[80] and approximately 32% of night workers from a random population sample affected.[81]

Table 75-1 Diagnostic Criteria for Shift Work Disorder

***International Classification of Sleep Disorders* (ICSD3) Criteria: General Criteria for Any Circadian Rhythm Sleep-Wake Disorder**

A. A chronic or recurrent pattern of sleep-wake rhythm disruption due primarily to alteration of the endogenous circadian timing system or misalignment between the endogenous circadian rhythm and the sleep-wake schedule desired or required by the person's physical environment or social/work schedules.
B. The circadian rhythm disruption leads to insomnia symptoms, excessive sleepiness, or both.
C. The sleep and wake disturbances cause clinically significant distress or impairment in mental, physical, social, occupational, educational, or other important areas of functioning.

Specific Criteria for Circadian Rhythm Sleep-Wake Disorder, Shift Work Disorder (ICD-9-CM code: 327.36)*

A. There is a report of insomnia and/or excessive sleepiness, accompanied by a reduction of total sleep time, which is associated with a recurring work schedule that overlaps the usual time for sleep.
B. The symptoms have been present and associated with the shift work schedule for at least three months.
C. Sleep log and actigraphy monitoring (whenever possible and preferably with concurrent light exposure measurement) for at least 14 days (work and free days) demonstrate a disturbed sleep and wake pattern.
D. The sleep and/or wake disturbance are not better explained by another current sleep disorder, medical or neurologic disorder, mental disorder, medication use, poor sleep hygiene, or substance use disorder.

Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: DSM-5*: Diagnostic Criteria for Circadian Rhythm Sleep Disorder (307.45)

A. A persistent or recurrent pattern of sleep disruption that is due primarily to an alteration of the circadian system or to a misalignment between the endogenous circadian rhythm and the sleep-wake schedule required by the person's physical environment or social or professional schedule.
B. The sleep disruption leads to excessive sleepiness or insomnia, or both.
C. The sleep disturbance causes clinically significant distress or impairment in social, occupational, and other important areas of functioning.

Specify Type:

Shift work type: insomnia during the major sleep period and/or excessive sleepiness (including inadvertent sleep) during the major awake period associated with a shift work schedule (i.e., recurring unconventional work hours).

*ICD-10-CM code: G47.26 circadian rhythm sleep-wake disorder, shift work type.

Morbidity

In addition to sleep-wake disturbances, persons with SWD show significant impairments in neurophysiologic measures of attention and memory compared with night shift workers without SWD.[82] In comparisons of shift workers with and without SWD, those with SWD also show a higher prevalence of ulcers, absenteeism, and difficulties with family and social activities and higher rates of major depression and mood problems.[12,83] A relationship also has been found between reduced satisfaction with the work schedule and excessive sleepiness on the night shift.[64] Not surprisingly, persons with SWD report higher rates of sleepiness-related accidents[84] than shift workers without SWD or day workers. Although increased rates of heart disease have been found in shift workers, it is not clear whether this morbidity is directly related to the excessive sleepiness or sleep disturbance characteristic of SWD.[12] Additional studies with patients meeting SWD diagnostic criteria are needed before definitive conclusions regarding unique morbidity in SWD (as opposed to shift work per se) can be determined.

Clinical Evaluation

SWD can be diagnosed by a thorough sleep history without polysomnography (PSG), although PSG may be indicated if there is a high level of suspicion for presence of another sleep disorder (e.g., obstructive sleep apnea).[85] The clinical evaluation for SWD is similar to that for other sleep disorders, particularly the other circadian rhythm sleep-wake disorders covered in Chapter 40. Evaluation of the shift worker, however, also requires careful attention to effects of the work-sleep schedule on cognitive, social, and health-related functioning. Ability to maintain wakefulness, particularly during sedentary activities (e.g., commute home from work), should be carefully assessed, because patients with SWD still have a twofold increase in sleepiness-related accidents compared with shift workers who do not meet diagnostic criteria.[12] The use of a sleep diary and actigraphy monitoring (when feasible and preferably with concurrent light exposure measurement) with a focus on regularity, duration, and timing of sleep periods for at least 2 weeks are helpful for determining the degree and pattern of sleep disturbance and circadian disruption.[86] This information should be collected for both days off and days on the work schedule that produces the symptom complex.

Fatigue often is confused with sleepiness and is a common presenting complaint.[87] If fatigue is present, an important consideration in the differential diagnosis in the shift worker is major depressive disorder.[12] Patients complaining of mental or muscle fatigue (as opposed to sleepiness) may find that sedentary activity or rest without sleeping leads to symptomatic improvement. On the other hand, patients with excessive sleepiness often state that sedentary activities or "rest" periods exacerbate their sleepiness. This is an important distinction to make because patients frequently do not recognize that dozing off in such situations is abnormal. Brief, well-validated tools can be used for assessing sleepiness.[86] The Epworth Sleepiness Scale (ESS) is particularly useful in this regard and can be easily administered in most clinical settings. An ESS score of greater than 10 is considered clinically important. With excessive sleepiness defined by the ESS cutoff score of greater than 10, prevalence rates of this symptom of up to 44% have been reported in shift workers,[12] compared with 24% to 33% in representative samples of daytime workers.[12,88,89]

Assessment tools such as the Insomnia Severity Index and the Pittsburgh Sleep Quality Index are useful for determining the extent and relative impact of sleep disturbance.[90] Practical issues regarding their implementation are discussed in detail in Chapter 83. Extensive clinical guidelines for the general evaluation of insomnia, including its relation to functioning, have been published and should be considered in evaluating a shift worker.[91] Whether the patient reports difficulty initiating sleep, maintaining sleep, or both is important to ascertain before considering appropriate treatment strategies. A thorough history of the patient's recent work schedule is essential to diagnosis, because many types of shift schedules (particularly rapidly backward-rotating schedules) cause significant sleep disturbance and excessive sleepiness.[92] If an evening shift worker presents with excessive sleepiness or insomnia, causes other than the effects of circadian misalignment due to shift work should be considered, because evening shifts are unlikely to lead to SWD (however, SWD might occur in evening workers with a strong morning circadian phase preference or high social or domestic demands).

The clinician also must be aware of potential mental health (e.g., depression), gastrointestinal, cardiovascular, and other health risks associated with shift work, and workers should be encouraged to undergo regular physical examinations to rule out these conditions.[93] In addition, health risks related to use/abuse of substances to relieve insomnia (e.g., illicit drugs and alcohol), poor diet, and nicotine use, which are common in shift workers, also should be addressed during the clinical evaluation. Finally, educating the shift worker on appropriate sleep behaviors (sleep hygiene), with an emphasis on adequate sleep opportunity, is beneficial.

Treatment

A cardinal feature of SWD is that symptoms are directly linked to the shift work schedule and thus are likely to remit after returning to daytime work.[13] Observation of such improvement indicates that circadian misalignment is the underlying pathophysiologic mechanism. Short of going off the night shift, occupational adjustments such as slower rotations (with weeks spent on a particular shift), moving to an evening shift, changing from backward to forward shift rotation, and incorporating increased worker control by allowing "self-scheduling" of shifts may be of some benefit.[94] In most cases, however, clinical intervention is needed, because occupational constraints generally prevent scheduling adjustments or a return to a diurnal schedule that would correct the circadian misalignment. Accordingly, treatment is necessarily targeted at the two symptoms of SWD: reducing excessive sleepiness and/or improving sleep.

Even before the diagnosis of SWD is made, the clinician must first address all potential safety concerns—and the threshold for immediate treatment intervention should be lower in a shift worker presenting with excessive sleepiness while driving and for occupations in which performance is critical for individual or public safety.[13] Shift workers typically have a chaotic sleep-wake schedule and may drastically curtail their time in bed in an attempt to meet their social, occupational, and daily obligations. Potential contributing factors include a short shift transition, long overtime hours to keep up with work demands, a second job, or staying awake to

engage in normal social activities during the day. All of these factors can make it difficult for the shift worker to get enough sleep. The clinician needs to address these issues with each patient.

Circadian Interventions

Appropriately timed bright light exposure to shift endogenous rhythms has been extensively studied in shift workers. Exposure to bright light in the evening near habitual bedtime and for several hours thereafter will induce a phase delay of internal biologic time, whereas exposure to bright light in the morning from approximately 2 hours before habitual wake time and thereafter will induce a phase advance (see Figure 75-1). In general, for every hour of properly timed exposure to bright light, there is a 0.5-hour shift in internal biologic time.

Using these principles, interventions aimed at producing phase delays of the circadian pacemaker (i.e., bright light during the first half of the night shift followed by daytime darkness) have been shown to improve daytime sleep and nocturnal functioning.[95,96] In a landmark study, Czeisler and colleagues treated circadian misalignment due to shift work with exposure to 7.5 hours of bright light (7,000 to 12,000 lux) on 4 consecutive nights in the laboratory and scheduled sleep in darkness between 9 AM and 5 PM at home.[95] In comparison with a control group exposed to room light (150 lux) and unscheduled sleep, those treated exhibited a 9.6-hour delay (i.e., to a later hour) of their circadian rhythms (i.e., temperature, cortisol, alertness). They also averaged 2 hours more sleep during the day and showed improved nocturnal alertness and cognitive performance. Additional details of circadian interventions are available in an American Academy of Sleep Medicine review.[97]

Despite the ability to achieve large phase shifts in controlled laboratory environments, investigators have been less successful in producing complete circadian adaptation in shift workers under field conditions.[98-102] Phase shifting requires near-complete control over the timing of exposure to light and darkness (which requires bright light and light-blocking glasses, respectively), which may not be practical in most real-world settings. Furthermore, complete circadian phase reversals could be maladaptive in that patients could not participate in family and social obligations (i.e., revert to a diurnal schedule) on their days off.

Recognition of practical limitations regarding continuous bright light exposure (e.g., 6 to 8 hours at 10,000 lux) and the limitations of complete phase shifts has led some investigators to study the sleep and performance effects of using brief intermittent light exposure to induce a "compromise" circadian phase (i.e., moderate but stable delays) compatible with permanent night work and common day schedules on days off. In a study by Smith and Eastman,[103] subjects engaged in three simulated night shifts (11 PM to 7 AM) followed by two "weekend" days off, with a final four additional night shifts, closely simulating the variable sleep-wake schedule of shift workers that often occurs in the real world. Subjects were exposed to brief light pulses during the night (15 minutes each hour), with total light exposure duration of 1.25 hours per shift, and wore sunglasses while outside, with the goal of achieving a moderately delayed or "compromise" circadian phase position in which the lowest point of alertness would occur just a few hours after their shift work ended (10 AM). The intervention produced a larger delay in onset of melatonin secretion (from 9 PM to 4:30 AM—a 7.5-hour delay) than that in the control group (3.5-hour delay). Sleep duration (determined by actigraphy and sleep logs) improved in the delayed group participants relative to the control subjects. Treatment also resulted in better and more consistent psychomotor performance. Importantly, delayed circadian phase was positively correlated with improved sleep even in the control group ($r = 0.65$), suggesting that some persons benefit from even small delays in circadian phase. These findings emphasize potentially important individual differences in the response to large and abrupt circadian shifts of the sleep-wake cycle. The effectiveness of this intervention for SWD needs to be studied. Not surprisingly, the benefits of light-induced phase shifts to improve sleep and wakefulness in shift workers appear to be particularly strong for people who commonly experience sleep-wake difficulties when exposed to night shift work schedules.[104]

Compounds with phase-shifting properties (e.g., chronobiotics) also may be beneficial for shift workers. Exogenous melatonin is perhaps the strongest nonphotic time cue agent in humans. Normally, endogenous melatonin levels rise approximately 2 hours before habitual bedtime,[105] remain high across the night, and decline again near the habitual time of awakening. Use of exogenous melatonin to shift circadian rhythms generally follows the reciprocal of the phase response curve for light (see Figure 75-1). Thus melatonin ingested during the biologic day can be used to phase advance or phase delay the circadian clock.[106,107] Properly timed exogenous melatonin administration for 3 days can lead to an approximate 1.5-hour shift in internal biologic time.[106] However, for melatonin to be an effective phase-resetting agent, control over light exposure appears to be necessary, because bright light can counter the phase shift produced by melatonin. Properly timed combinations of melatonin and bright light also can be used to induce larger phase shifts than either alone.[108]

Four days of treatment with 1, 2, or 4 mg of the prescription melatonin agonist ramelteon[109] or 3 days of treatment with 100 mg of the prescription melatonin agonist tasimelteon[110] had significant phase-advancing effects (80 to 120 minutes) when normal sleep-wake schedules were abruptly shifted earlier by 5 hours. Although these effects may have more practical applications for circadian rhythm disorders such as delayed sleep phase syndrome and jet lag,[111] in which modest shifts in the circadian pacemaker may improve symptoms, melatonin or melatonin agonists may still provide benefits in patients with SWD through the sleep-promoting and phase-shifting properties of these compounds.[112,113,110] In studies of shift work, melatonin (0.5 to 3 mg) improved circadian adaptation through phase delays or advances, depending on the time of administration.[96,112] Of note, however, although chronobiotic compounds may benefit shift workers, these benefits will be counteracted by inappropriate exposure to the more powerful zeitgeber—that is, daylight during the early morning hours. Melatonin has been shown to cause performance impairment at doses as low as 5 mg, suggesting caution if wakefulness is intended to be maintained for more than 30 minutes after administration.[112] Finally, no large-scale clinical trials have been conducted to evaluate the effectiveness or safety of melatonin use. Therefore melatonin use should be discussed with the patient's physician. In addition to prescription-only melatonin agonists, in the United States,

melatonin also is available over the counter as a dietary supplement. Some melatonin preparations are certified for purity and dosage level, and such preparations are recommended. Use of melatonin to induce and maintain sleep as opposed to phase-shifting is reviewed later.

Findings from studies have also demonstrated that exercise interventions delay circadian rhythms.[114] Without adequate control over daytime light exposure, however, the usefulness of such interventions alone may be limited owing to the extensive time required for significant phase-shifting effects.

Improving Diurnal (and Nocturnal) Sleep

Shift workers should be encouraged to attempt sleep immediately after the night shift and to maintain a sleep-conducive environment during their sleep time by using light-blocking shades, ear plugs, and a comfortable eye mask. If near-complete circadian alignment to the night shift is achieved, improvements in daytime sleep occur.[95,115] However, the practical limitations of circadian interventions often prevent complete alignment (e.g., diurnal light exposure), thus necessitating interventions directly targeting sleep improvement. This lack of circadian adjustment in shift workers manifests as sleep disturbance during the latter half of daytime sleep.[115] Use of two sleep periods—(1) an "anchor" sleep period of approximately 4 to 5 hours that represents a time of day (e.g., 8 AM to noon when the shift worker is instructed to "always sleep" regardless of its being a workday or day off and (2) another 3- to 4-hour period of sleep taken at irregular times, depending on the work schedule—may help to stabilize circadian rhythms and increase sleep duration for a given 24-hour period.[116] Results from one study showed that splitting sleep into two phases provided more daytime sleep than that obtained with an equal daytime duration of consecutive time in bed.[117]

Although melatonin is less likely to produce circadian phase shifts when light exposure is not controlled or when endogenous melatonin is present,[112] exogenous melatonin (and melatonin agonists), administered during the biologic day when endogenous melatonin levels are low, increases total sleep time even at doses as low as 0.3 mg.[113,110,118,119] In shift workers, some improvements in sleep have been shown with melatonin doses between 5 and 10 mg.[120,121] By contrast, with use of subjective measures of sleep[122,123] or lower doses (2 mg), soporific effects have not been found.[124] Studies in patients diagnosed with SWD are needed, because some subjects in previous studies may not have had significant sleep disturbance.

Benzodiazepine receptor agonists have been shown to improve daytime sleep in simulated shift work environments,[125,126] yet these medications do not normalize nocturnal alertness.[125,127,128] Use of the benzodiazepine triazolam was shown to improve sleep during the daytime hours (by 30 to 60 minutes/day) under conditions of simulated night shift work; however, these effects did not translate into substantially improved alertness during the night.[125,126] Use of newer benzodiazepine receptor agonists have produced similar results in subjective reports of daytime sleep[129] and slight improvements in nocturnal performance; however, evidence for worsened mood also has been reported.[127] Patients diagnosed with SWD may show greater benefits from benzodiazepine receptor agonists, but this has yet to be demonstrated. In contrast with medium-acting compounds (half-life of 5 to 12 hours), short-acting hypnotics (i.e., half-life of 1.5 hours) may be of little benefit in shift workers with isolated sleep maintenance problems (i.e., difficulty staying asleep with no problems falling asleep). However, the possibility of residual sedation should be considered with use of longer-acting hypnotic medications, because shift workers frequently have shorter sleep periods than those for day workers. Behavioral treatment of insomnia in shift workers may be helpful, but specific approaches (e.g., relaxation therapy, stimulus control techniques) have not been systematically evaluated.

Despite its sedative properties, the use of alcohol as a hypnotic in SWD should be strongly discouraged owing to fragmentation of sleep in the second half of the night after alcohol consumption (a time when sleep is particularly vulnerable to disruption in this population).[130]

Napping during the day before the night shift and for brief episodes during the night has been effective for improving alertness and performance.[23,24,131] In two recent studies, the combination of an evening nap and caffeine (250 to 350 mg) 30 minutes before the night shift was particularly beneficial for improving alertness and performance for up to 3 nights.[132] Findings from one study in shift workers who were professional drivers demonstrated that a clinically feasible combination of brief naps (two at 20 minutes each) and a short light exposure period (10 minutes at 5000 lux) reduced polysomnographically measured number of episodes of falling asleep while driving.[24]

Pharmacologically Enhancing Alertness

Many people who experience sleepiness use caffeine to combat the problem. Caffeine can be used to improve wakefulness and performance during the biologic night.[132,133] Although prenap caffeine may be beneficial in reducing the performance-impairing effects of sleep inertia,[134] this approach has yet to be tested in patients with SWD. In a study by Wyatt and colleagues, low-dose caffeine (0.3 mg/kg/hour) administered over periods of extended wakefulness (29 hours) and circadian misalignment helped subjects remain awake and improved memory and psychomotor performance.[135] Overall, evidence supports a role for the use of caffeine to enhance alertness in shift workers.[136] Although other alerting agents have been used, Schedule II stimulant medications such as amphetamines and methylphenidate have several disadvantages, including high abuse potential, that offset their ability to enhance alertness in the context of shift work.[137,138]

Medications indicated for enhancing alertness in SWD have been used to promote nocturnal wakefulness in patients.[85] A study on the use of modafinil for treatment of SWD provides evidence for its therapeutic benefit in occupationally related outcomes, including sleepiness while driving. In that study, 204 patients who met the criteria for SWD were given either 200 mg of modafinil or a placebo for 3 months during the clinical field trial.[139] The group taking modafinil showed significant improvement over those taking the placebo in psychomotor vigilance and drowsy driving on the commute home, among other end points. Patients also exhibited significant reductions in objectively defined sleepiness on the Multiple Sleep Latency Test during the night shift, although alertness levels did not normalize to those seen during typical daytime assessments. Importantly, modafinil did not have detrimental effects on subsequent daytime sleep when taken at the beginning of the night.[139] Armodafinil, the longer-lasting isomer of modafinil, also has been shown to be effective for the treatment of excessive sleepiness in SWD and has reduced

the level of sleepiness during the commute home.[140] In a placebo-controlled driving simulator study in patients with SWD, armodafinil improved driving performance including reduced standard deviation of lateral position and reduced off-road deviations during the night shift up to 9 AM.[141] It has been FDA-approved for use in SWD and has been shown to significantly reduce excessive sleepiness and to improve overall clinical condition and performance in this patient population.[140] Additional studies regarding pharmacologic treatment of circadian desynchrony are reviewed in Chapter 40.

Enhancing Alertness with Combined Treatments

Several investigations have examined the efficacy of combined treatments for promoting wakefulness during the biologic night in normal, healthy volunteers. Combined treatments studied include caffeine and bright light,[133] caffeine and naps,[132] and naps and modafinil.[142] Such wakefulness-promoting treatment combinations have been reported to be of greater benefit than the use of either treatment alone (Video 75-1). With respect to efficacy, these nonprescription alternatives have not been compared with prescription agents (such as modafinil) for treatment of SWD.

Management Guidelines for Shift Work Sleep Disorder

In light of the paucity of clinical tools for assessing and treating SWD, we have proposed a brief set of clinical guidelines, presented in Table 75-2. These guidelines are based on established circadian principles, the literature on morbidity associated with shift work, and data from the few currently available studies on the assessment and treatment of SWD and tolerance to shift work. Several useful clinical case reviews from initial evaluation through treatment follow-up are presented in Chapter 76, with specific reference to high-risk occupations.

Jet Lag

Many millions of people travel by jet plane each year. Jet lag results from a mismatch between internal biologic time and environmental time caused by rapid eastward or westward travel across multiple time zones. Disturbed and/or shortened sleep duration before and during travel also may contribute to jet lag symptoms.

Severity and duration of symptoms (i.e., tolerance to travel) are dependent on (1) the direction and number of time zones traveled, (2) the ability to obtain sufficient sleep while traveling, and (3) exposure to environmental circadian time cues in the new time zone. Symptoms include daytime fatigue and sleepiness and insomnia in the new time zone. Gastrointestinal disturbance is common and may be related to intake of food at a biologic time when the body is not prepared to undertake this function.[143,144] Although the circadian principles used in the management of jet lag are similar to those of shift work, environmental cues (e.g., daytime light exposure) often support adaptation after travel to a different time zone. Thus jet lag symptoms typically subside in a few days (although in some cases symptoms can last for weeks).

Symptoms of jet lag occur because immediately after eastward jet travel, the traveler attempts sleep in the new time zone before his or her internal biologic night. Sleep in the new time zone is thus disturbed, and such sleep disruption contributes to subsequent daytime sleepiness and reports of fatigue. Daytime sleepiness and fatigue also occur because the traveler must stay awake during his or her internal biologic night. Individual differences in the ability to sleep during the biologic day (vulnerability to insomnia) and to maintain alert wakefulness during the biologic night also may contribute to jet lag symptoms. The cognitive impairments associated with jet lag can have serious consequences, resulting in drowsy driving or flying accidents, impaired decision making for the business traveler, and impaired athletic performance. In other cases, jet lag can be an inconvenience leading to difficulty staying awake during sightseeing, attending the theater, or having a meal. Sleepiness and fatigue also increase the performance-impairing effects of alcohol.[145-147] Jet lag often is reported to be worse after eastward than westward travel. Westward travel may be easier because the average period of the circadian clock in humans is longer than 24 hours, so the biologic tendency is for later bedtime and wake time. From 20% to 25% of people, however, have a shorter than 24-hour clock,[148] and such persons may find it easier to adapt to eastward travel.

Treatment

Successful treatment requires a detailed history and knowledge of circadian physiology and of countermeasures to improve sleep, wakefulness, and circadian adaptation. Most evidence for jet lag treatments comes from laboratory studies. Currently, no medications are approved by the U.S. Food and Drug Administration (FDA) for the treatment of jet lag disorder.

Promoting Sleep during Flight and in the New Time Zone

The traveler should be educated on environmental factors that can be controlled to promote sleep[149] (see Chapter 85). Most flights eastward from the United States to Europe are scheduled at night. Eyeshades and earplugs or noise-canceling headphones may help promote sleep during the flight. Alcohol consumption should be avoided during the flight. Although alcohol shortens sleep latency, it disrupts sleep continuity.[130] Direct overnight flights provide more opportunity to sleep during travel than do itineraries with multiple stops. On arrival, it often is recommended to immediately adapt to the new bedtime and awakening times of the new time zone. Exogenous melatonin can shorten sleep latency and increase sleep duration in a dose-dependent manner when taken during the biologic daytime.[112,118,119] Thus, if sleep is attempted during the biologic daytime while in flight and/or in the new time zone, melatonin may be used to improve sleep quality and duration. Melatonin should be tried in the home time zone before use during travel to determine the response to the chosen dose. Prescription melatonin receptor agonists, such as ramelteon[113] and tasimelteon,[110] are reported to improve sleep during the biologic daytime, so they may be useful in the treatment of jet lag. In a clinical trial, ramelteon (1 mg) was reported to shorten the latency to sleep compared with placebo after eastward air travel across five time zones (from Hawaii to the East Coast of the United States). In addition, 4 mg decreased some daytime symptoms of jet lag.[150] Neither ramelteon nor tasimelteon has been FDA-approved for treatment of jet lag.

Over-the-counter sleep aids have not been tested to determine their effectiveness in treating jet lag. Several prescription benzodiazepine receptor agonists approved for insomnia have been tested in simulated and actual jet lag trials. Findings from these studies indicate that sleep is improved during jet

Table 75-2 Clinical Guidelines for Assessment and Management of Shift Work Disorder

Assessment

I. Determine circadian misalignment (sleep diaries and actigraphy with concurrent light exposure).
II. Assess sleep disturbance.
  A. Determine difficulty falling asleep, staying asleep, or having nonrestorative sleep (both during daytime and nighttime sleeps).
  B. Measure degree of alertness.
  C. Assess falling asleep during inappropriate circumstances or times (using Epworth Sleepiness Scale [ESS]), with special attention to drowsy driving.
  D. Determine important job-related factors: duration of commute after shift, number of consecutive shifts, type of shift, time between shifts.
III. Determine impact on social and domestic responsibilities.[165]

Management[93]

I. Shift workers should have regular physical examinations with attention to psychological (e.g., depression), gastrointestinal, cardiovascular, and potential cancer risks associated with shift work.[93]
  A. Sleep-related comorbidity: Determine risk of sleep-disordered breathing, restless legs syndrome, or other potential sleep disorder.
  B. Other comorbidity: Identify medical or psychiatric disorders that may contribute to the symptoms of insomnia or excessive sleepiness.
II. Determine if removal from shift work is appropriate or practically feasible. If patient meets criteria for a diagnosis of shift work disorder, cessation of the shift work schedule should be the first option discussed with the patient.
III. Determine patient-specific therapeutic approach.
  A. Circadian adaptation
    1. Consider individual difference factors (e.g., age, phase preference).
    2. Consider compromise phase position (e.g., partial phase delay using bright light during first half of night and increased darkness during daytime).
    3. *Night workers*: On days off, adopt a late sleep schedule (i.e., bedtime of 3 to 4 AM).
  B. Symptom management
    1. Insomnia
      a. Good sleep behaviors
        i. Target inappropriate sleep behaviors and encourage use of eye mask, ear plugs, and light-blocking shades during daytime sleep.
      b. Sleep maintenance a primary concern
        i. Consider intermediate-acting hypnotic (half-life of 5 to 8 hours).
        ii. Consider melatonin treatment for daytime sleep (~3 mg).
      c. Sleep initiation problems
        i. Consider short-acting hypnotic.
      d. Sleep problems on days off
        i. Consider fixed sleep-wake schedule and consider anchor sleep.
    2. Excessive sleepiness (i.e., ESS score <10)
      a. Address sleep disturbance if present.
      b. Consider wake-enhancing medication before shift (e.g., modafinil, armodafinil) or off-label stimulants (e.g., amphetamine, methylphenidate).
      c. Prophylactic nap before work shift is recommended.
      d. Judicious use of brief to moderate-length naps (30 to 60 minutes), with recognition of risk of sleep inertia (consider prenap caffeine to reduce sleep inertia).
      e. Consider combined treatment strategies during the work shift (alerting medications, bright light, anchor sleep and naps).
IV. Address additional work, social, and domestic factors.
  A. Social/family/psychological: Improve balance between family/social, work, and sleep time and treat psychosocial stress, depression, or marital discord if present,[165] educate patient's family regarding shift workers need of protected time for sleep.
  B. Health and safety: Promote improved healthy eating habits with respect to regularity and timing relative to the major sleep period (not within 2 to 4 hours of bedtime), reduce inappropriate substance use, increase exercise at appropriate times (not within 2 to 4 hours of bedtime), educate on risks of drowsy driving and critical times of performance vulnerability.
  C. Work-related: Reduce number of consecutive shifts (<4),[165] reduce shift duration (<12 hours), use clockwise rotation,[166] ensure adequate time between shifts (>11 hours),[165] move heavy workload outside circadian nadir (4 to 7 AM), address commute time (longer = greater accident risk), move to day or evening shift,[8] consider incorporation of a shift work awareness program.[167]

travel,[151,152] but there is little evidence that subsequent wakefulness is improved by these or other sleep medications. In addition, the side effects of both over-the-counter and prescription sleep-inducing agents must be considered (e.g., cognitive and balance impairments).

Promoting Wakefulness during Flight and in the New Time Zone

Most flights westward from Europe to the United States are scheduled during the daytime. Staying awake until bedtime in the new time zone should promote sleep. Naps when in flight during westward travel and in the new time zone after eastward or westward travel are likely to be effective in promoting subsequent wakefulness.[153,154] Caffeine is perhaps the most commonly used self-selected countermeasure to promote wakefulness during jet travel. As indicated by sleep deprivation and simulated jet travel studies, caffeine can help to promote wakefulness during the biologic night.[133,135,155]

Modafinil has been shown to improve several, but not all, aspects of cognitive function during sleep deprivation combined with circadian misalignment[156]—which are factors common to jet travel. In a clinical trial of patients with jet lag disorder, armodafinil (150 mg) was reported to increase sleep latency on the MSLT and reduce patient ratings of jet lag severity compared with placebo after eastward air travel across six time zones (from the East Coast of the United States to France).[157] However, wakefulness after sleep onset also was higher during the sleep episode in the armodafinil versus placebo condition, and several subjects reported insomnia as a side effect. Thus patients should be monitored for sleep disturbance if armodafinil is used to combat jet lag disorder; alternatively, the shorter-acting compound modafinil may be preferable. Other side effects of armodafinil included headaches, nausea, and heart palpitations.

Circadian Adaptation

Adjustment of the circadian system to a new time zone often takes days. It has been estimated that complete circadian adjustment may require a day or more for each time zone crossed. Proper timing of bright light exposure and dim light or darkness can quicken adaptation of the internal circadian clock to the new time zone (see Figure 75-1). Exogenous melatonin also may help to shift the circadian clock during jet travel. The timing of melatonin administration to induce phase shifts is opposite to that of light (see Figure 75-1). For example, if a westward phase delay is desired, exposure to dim light-darkness and melatonin in the morning is likely to facilitate a delay induced by exposure to bright light at night (Figures 75-2 and 75-3). Inappropriate timing of light and darkness during and immediately after jet travel can shift the circadian clock in the wrong direction, thereby increasing the duration of jet lag symptoms from days to more than a week.

Partial Preflight Circadian Adaptation

Partial preadaptation of the jet traveler's circadian clock before travel[158,159] may shorten the duration of jet lag symptoms. Partial preadaptation involves going to bed and waking up earlier for 1 or 2 days before eastward flight or going to bed and waking up later before westward flight, combined with properly timed exposure to light (see Figure 75-3). Earlier awakening before eastward flight should be combined with exposure to bright light (e.g., a sunrise walk or turning up the lights in the house) and exposure to dim light at night. Later bedtimes and wake times before westward flight should be combined with exposure to bright light in the evening and exposure to dim light in the morning. For a full case presentation of successful assessment and treatment using this approach, the reader is referred to the description by Wright[160]; clinical management recommendations have been summarized by Sack.[161]

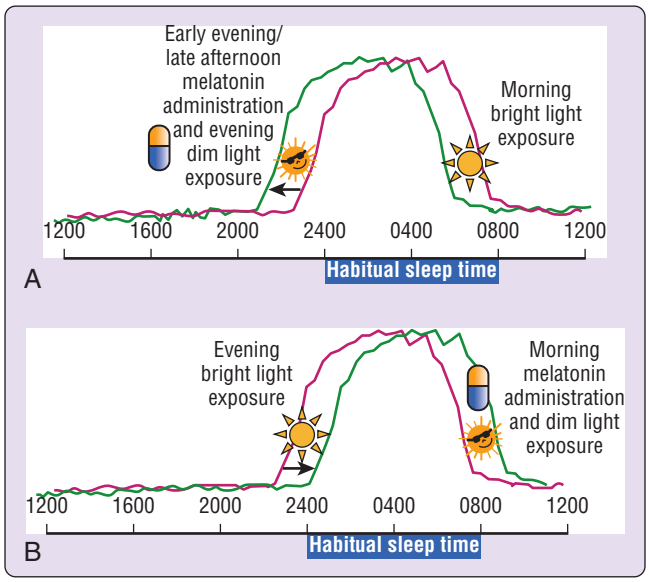


Figure 75-2 Using light and melatonin to phase shift the internal biologic time. **A,** When the traveler is entrained to local time, light exposure during the morning and melatonin ingestion in the afternoon will induce an eastward phase advance shift of the internal circadian clock. *Red line* shows initial melatonin rhythm and *green line* shows advanced melatonin rhythm after treatment. **B,** Light exposure during the evening and melatonin administration in the morning will induce a westward phase delay shift of the internal circadian clock. *Red line* shows initial melatonin rhythm and *green line* shows delayed melatonin rhythm after treatment. Adjustment to eastward travel should include exposure to dim light or darkness in the evening. This can be achieved by wearing sunglasses if the sun has not set and by turning down the lights in the house. Adjustment to westward travel should include exposure to dim light or darkness in the morning. Administration of melatonin when the jet traveler is required to be awake (e.g., for work or driving) should be avoided, because melatonin has been reported to impair performance.

CONCLUSIONS

Shift work can have debilitating consequences from chronic sleep disruption, social isolation, and circadian misalignment.[162] The negative effects of shift work may involve the gastrointestinal, cardiovascular, and other physiologic systems.[163] Although the direct relation between disruptions of the sleep and circadian systems and corresponding morbidity needs further study, circadian and sleep adaptation to shift work can be improved using interventions that target these systems. Impaired alertness may require management strategies beyond circadian and sleep-related interventions. In people with SWD, treatment should be directed at circadian adjustment to the shift work schedule and, if necessary, treatment of excessive sleepiness and/or sleep disruption symptoms. Although data on patients diagnosed with SWD are limited, successful treatment approaches have included wake-promoting medications, prophylactic naps, appropriately

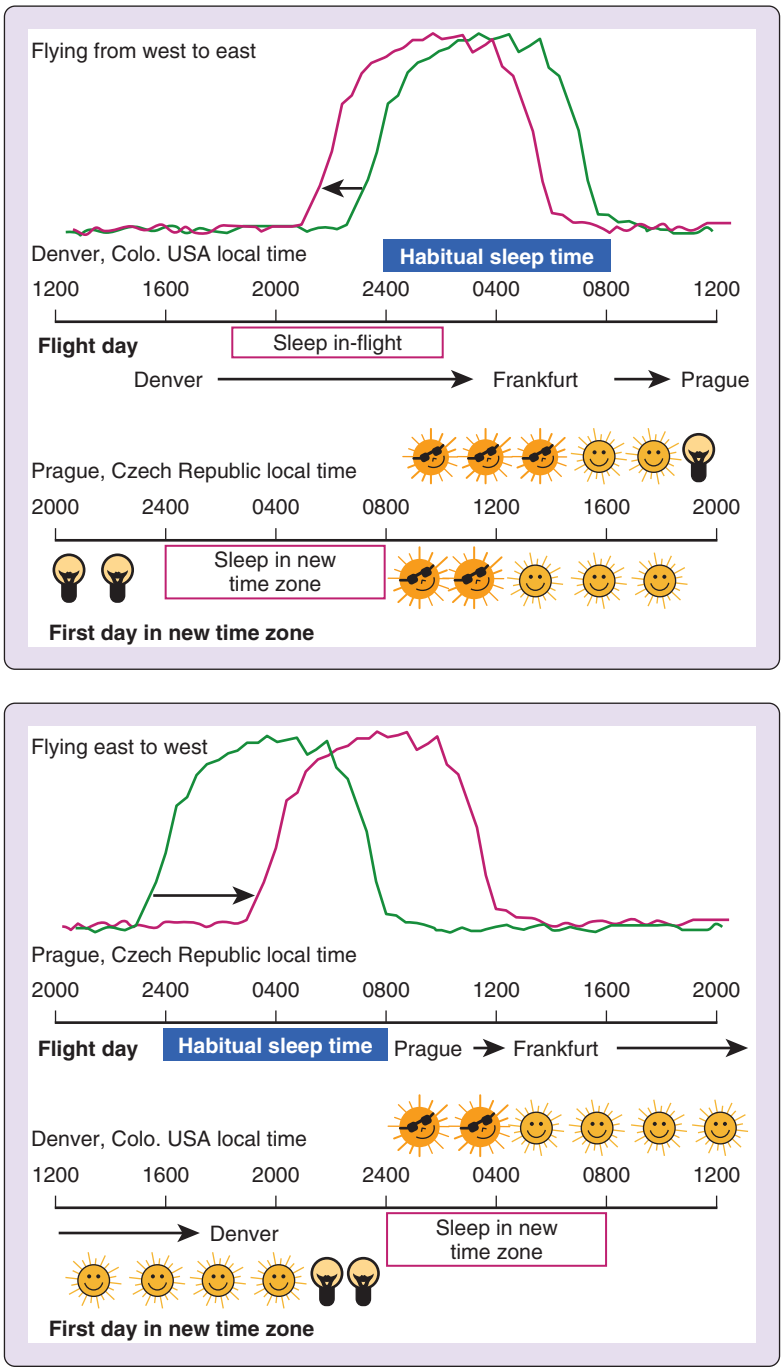


Figure 75-3 Circadian adaptation during jet travel. *Top:* When traveling eastward, the jet traveler needs to advance the timing of his or her sleep schedule as well as the phase of the internal circadian clock so that both occur earlier. In the current example of a trip from Denver, Colorado, to Prague, Czech Republic, the primary plane flight to Frankfurt (*long black arrows*) occurs during the beginning of the biologic night, when endogenous melatonin levels are high (*green line; top panel*). The traveler should sleep as much as possible on the plane flight to reduce the negative impact of sleep deprivation during jet travel. Exogenous melatonin taken shortly after boarding the plane may help the traveler fall asleep earlier than normal when endogenous melatonin levels are low. In this example, the traveler spends several hours in Frankfurt before the flight connection to Prague. Light exposure during most of the biologic night will induce a westward phase delay, opposite to what is needed for an eastward shift; accordingly, the traveler should avoid exposure to bright light (*sun with sunglasses*) and wear an eye mask or sunglasses until after the habitual time of awakening in the home time zone. Subsequent exposure to bright light (*sun without sunglasses and light bulb*) will facilitate the eastward phase advance of the traveler's circadian clock (*red line; top panel*). Sleep timing should be consistent with the local time zone in Prague, and melatonin may again help phase shift the clock and promote sleep onset if taken when endogenous melatonin levels are low. The next day, the traveler should at first continue to avoid exposure to bright light. Each day, the time of exposure to bright light can be moved earlier by 1 to 2 hours per day. *Bottom:* When traveling westward, the jet traveler needs to delay the timing of his or her sleep schedule as well as the phase of the internal circadian clock (*green line; bottom panel*) so that both occur later. In the example of a trip from Prague to Denver, plane flights occur across the biologic day when endogenous melatonin levels are low, and the traveler should remain awake for much of the trip, assuming that sleep was adequate before travel. Caffeine and/or a nap can help to support subsequent wakefulness. Exposure to bright light should occur across the entire plane flight and until just before bedtime in the new time zone, to facilitate a westward phase delay (*red line; bottom panel*).

timed exposure to bright light exposure and darkness, and combined countermeasures (when appropriate). Sleep disturbance, short sleep duration, and circadian misalignment can have a negative impact on productivity, performance, and safety during and immediately after jet travel. Sleep and circadian science principles indicate that appropriately timed exposure to light and darkness before and during the flight and on arrival can hasten adjustment to the new time zone, whereas inappropriately timed exposure to light and darkness during and immediately after jet travel can shift the circadian clock in the wrong direction, thereby increasing the duration of jet lag symptoms.

CLINICAL PEARLS

Shift Work

Shift work disrupts sleep and circadian rhythms and is associated with increased risk for cardiovascular disease, gastrointestinal disorders, and cancer. Careful application of circadian principles (appropriate timed bright light-darkness exposure and melatonin) can improve adjustment to shift work. Melatonin and prescription hypnotic medications may provide additional benefit in some people. Treatment with FDA-approved alerting medication often is useful in patients with SWD.

Jet Lag

Rapid eastward or westward travel across multiple time zones induces circadian misalignment, which produces insomnia and daytime sleepiness with impaired driving performance and cognition. Circadian adaptation to the new time zone requires appropriately timed exposure to light and darkness. Inappropriately timed exposure to light can shift the circadian clock in the wrong direction, prolonging jet lag symptoms. Wakefulness- and sleep-promoting countermeasures can be used to address symptoms of daytime sleepiness and sleep disruption, respectively, during travel and in the new time zone.

SUMMARY

Shift work and travel across time zones are commonplace. These factors pose circadian challenges because they require abrupt and often large shifts in the timing of sleep-wake schedules. Individual differences exist in the ability to adapt to this mismatch between circadian physiology and sleep-wake behavior that in turn impact sleep-wake, cardiovascular, and gastrointestinal system functioning. Insomnia and excessive sleepiness, which contribute to other morbidity (e.g., accidents), are the defining symptoms of SWD. Effective treatments for symptoms of SWD include use of nocturnal bright light and daytime darkness, maintenance of anchor sleep with naps to increase 24-hour sleep time, and sleep- and wake-enhancing medications. Jet lag symptoms include gastrointestinal disturbance, daytime fatigue, sleepiness, and insomnia. Cognitive impairments can have serious consequences, including impaired driving and impaired decision making. Interventions such as appropriately timed bright light and darkness can improve circadian adaptation to time zone changes; sleep-promoting agents and melatonin and its agonists (during the biologic daytime) may promote sleep but may not improve wakefulness in the new time zone. Preadaptation of the circadian clock and use of caffeine and brief naps in the new time zone are useful countermeasures to promote wakefulness.

ACKNOWLEDGMENTS

We thank Drs. Gregory Belenky, Karen Roy, and Thomas Roth for their helpful comments on the manuscript.

Work on which this chapter is based was supported by National Institute of Mental Health (NIMH) grant 068372 and National Institute of Nursing Research (NINR) grant 013959 and Henry Ford Hospital (for C.L.D.) and by National Institutes of Health (NIH) grants HL109706, HL111478, and DK092624 and the University of Colorado (for K.P.W.).

Selected Readings

Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med* 2005;**352**:125–34.

Burgess HJ, Sharkey KM, Eastman CI. Bright light, dark and melatonin can promote circadian adaptation in night shift workers. *Sleep Med Rev* 2002;**6**:407–20.

Crowley SJ, Eastman CI. Phase advancing human circadian rhythms with morning bright light, afternoon melatonin, and gradually shifted sleep: can we reduce morning bright-light duration? *Sleep Med* 2015;**16**(2): 288–97.

Czeisler CA, Johnson MP, Duffy JF, et al. Exposure to bright light and darkness to treat physiologic maladaptation to night work. *N Engl J Med* 1990;**322**:1253–9.

Drake C, Belcher L, Howard R. Length polymorphism in the Period 3 gene is associated with sleepiness and maladaptive circadian phase in night-shift workers. *J Sleep Res* 2015;**24**:254–61.

Drake CL, Roehrs T, Richardson G, et al. Shift work sleep disorder: prevalence and consequences beyond that of symptomatic day workers. *Sleep* 2004;**27**:1453–62.

Eastman CI, Burgess HJ. How to travel the world without jet lag. *Sleep Med Clin* 2009;**4**:241–55.

Gumenyuk V, Roth T, Drake CL. Circadian phase, sleepiness, and light exposure assessment in night workers with and without shift work disorder. *Chronobiol Int* 2012;**29**:928–36.

Juda M, Vetter C, Roenneberg T. Chronotype modulates sleep duration, sleep quality, and social jet lag in shift-workers. *J Biol Rhythms* 2013;**28**: 141–51.

Monk TH, Buysse DJ, Billy BD, et al. Polysomnographic sleep and circadian temperature rhythms as a function of prior shift work exposure in retired seniors. *Healthy Agin Clin Care Elder* 2013;**5**:9.

Sack RL, Auckley D, Auger RR, et al. Circadian rhythm sleep disorders: part I, basic principles, shift work and jet lag disorders. An American Academy of Sleep Medicine review. *Sleep* 2007;**30**:1460–83.

Vetter C, Fischer D, Matera JL, Roenneberg T. Aligning work and circadian time in shift workers improves sleep and reduces circadian disruption. *Curr Biol* 2015;**25**:907–11.

Wright KP Jr, Bogan RK, Wyatt JK. Shift work and the assessment and management of shift work disorder (SWD). *Sleep Med Rev* 2013;**17**: 41–54.

A complete reference list can be found online at ExpertConsult.com.

Chapter
76

Sleep Problems in First Responders and in Deployed Military Personnel

Bryan Vila; Charles Samuels; Nancy J. Wesensten

Chapter Highlights

- Fatigue associated with inadequate sleep and prolonged work shifts is one of the most common health and safety hazards faced by police officers and other first responders and deployed military personnel. Whether caused by extended-duty hours, night work, or circadian disruption, fatigue contributes to morbidity (and possibly mortality) in these occupational groups.
- The impact of insufficient sleep on first responders can extend to bystanders and noncombatants because this deficit tends to degrade cognitive performance, including the ability to make sound judgments, decide on appropriate courses of action, and exercise restraint in the face of threat and provocation. This impairment is particularly problematic in civilian police work, used in this chapter as an exemplar for other classes of first responders.
- The social, organizational, and individual causes of insufficient sleep within these occupational groups are inextricably linked, and the interactions of these various factors must be considered to address the problem of insufficient sleep (and sleep disorders) in these populations.
- By working with operational agencies, physicians and other health care providers can improve individual patient treatment and public health outcomes.

Police officers in the United States, Canada, and other industrialized nations often are exposed to long and erratic work hours, shift work, and insufficient sleep. These factors are likely to contribute to the elevated morbidity observed among police officers. Fatigue-related impairments in officer performance and decision making generate unique social and economic costs, because an error on the part of a police officer (e.g., in distinguishing between an innocent bystander and an aggressor) can result in injury or death.[1]

Managing fatigue in a police force requires balancing the biologic and social needs of officers against the requirements of the organizations that employ them and the safety of the communities they serve. Communities must have sufficient officers on duty at any moment to respond to emergencies, prevent crime, and arrest offenders, but not so many that public resources are wasted. To complicate matters, the need for police services fluctuates across the day, week, and season.

Officers impaired by insufficient sleep are less alert, their cognitive and physical abilities decline, and their mood worsens—all of which translate into degraded ability to effectively cope with stressors. This detriment in turn reduces both public and officer safety because risks of job-related accidents, injuries, errors, and misconduct increase. Over the long term, chronic insufficient sleep may make officers more vulnerable to physical and psychiatric illness and may corrode the quality of family and social interactions that buffer the impact of a stressful work environment. Preventive measures and treatment require consideration of the processes that lead to insufficient sleep and interfere with recovery from exposure to stressors.

PREVALENCE OF SLEEP LOSS, MORTALITY, AND MORBIDITY AMONG POLICE OFFICERS

Compared with workers in the general population, officers whose sleep is chronically disrupted, as a consequence of either personal factors such as medical disorders, intolerance to shift work and schedule changes, or involvement in off-duty activities or because of employer scheduling and work-hour practices, suffer from disproportionately high levels of cardiovascular, gastrointestinal, and metabolic diseases; chronic insomnia, sleep apnea, and other sleep disorders; and psychological disorders including depression, suicide, and family dysfunction.[2-16] Research linking long and erratic work hours to these sorts of disorders is substantial and the findings are compelling.[1,12] Shorter-term links between sleep loss and the sorts of on-the-job accidents and injuries that most frequently kill or seriously harm police officers also are well documented.[1,12,13]

Figures 76-1 and 76-2 show the sources of on-the-job injury and death for police officers. As is evident in Figure 76-1, felonious killings of police officers declined steadily from 1980 to 2013. This decline probably was a consequence of improvements in training, tactics, and soft body armor. Over the same period, accidents declined to a lesser extent despite major improvements in vehicle safety such as crash-resistant designs, air bags, shoulder harnesses, radial tires, antilock braking systems, and disk brakes. Since 1997, accidents have accounted for an average of 57% of all officer deaths, versus 43% for felonies. Reported suicide rates from 1984 to 1998 (the only period for which these official data were available nationally[17]) also declined—and, up to 1998, suicides

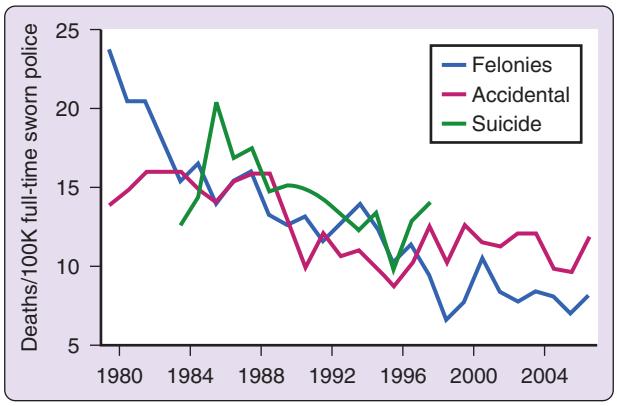


Figure 76-1 Police officers killed per capita in the United States, by cause. (*Felonies and accidents:* Data from Federal Bureau of Investigation. *Law enforcement officers killed and assaulted* [LEOKA]. Uniform Crime Reports. <https://www.fbi.gov/about-us/cjis/ucr/ucr-publications>; 1980–2013. *Suicides*: Data from Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. National occupational mortality surveillance [NOMS] data 1984 to 1998; and Web surveillance data for 2008 to 2012 [see text].)

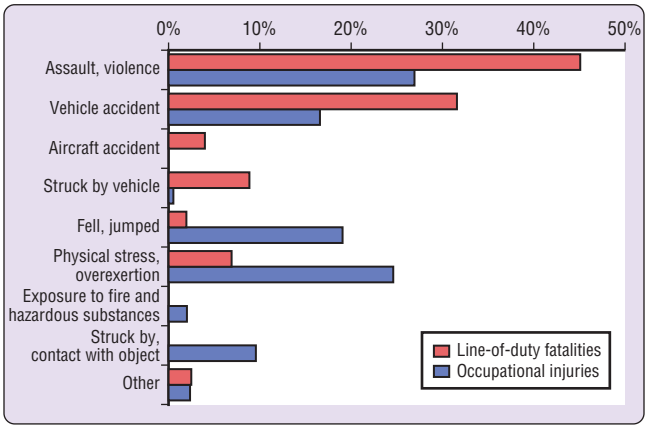


Figure 76-2 Causes of police injuries and fatalities in which work time was lost. (Data from National Law Enforcement Officer Memorial Fund website, 1995 to 2002.)

accounted for slightly more officer deaths annually than either felonious killings or accidents. More recent surveillance data using extensive Web-based reviews of news reports showed average officer suicide rates for 2008 to 2012 that were 28% higher.[18,19] Another recent analysis, using occupational and mortality data from 23 states during 1999, 2003 to 2004, and 2007 to calculate proportionate mortality ratios, found a 69% greater risk of suicide for police officers than for other adult workers in the United States.[15]

Figure 76-2 illustrates common causes of police occupational injuries that were sufficiently serious to result in lost work time and police on-the-job fatalities. It is worth noting that of the various causes, (1) nonfelonious vehicular accidents are the second most common cause of both fatalities and serious injuries (second only to assault and violence); and (2) operational deaths within both police and military forces tend to be split roughly evenly between deadly encounters and operational accidents, especially vehicle collisions.[20] Because the probability of vehicular accidents is significantly increased by sleep loss and disruption[21-23] as well as circadian factors,[14,24] it is highly likely that these two factors increase the likelihood of nonfelonious vehicular accidents in police officers.[24a,24b] These problems are compounded by the substantial amounts of overtime worked by police officers in the United States and Canada.[25-29] Officers assigned to patrol and detective assignments may work 16 or more consecutive hours—and sometimes more than 24 hours straight. As with most occupations, a small proportion of the officers in most departments work a large proportion of the overtime. Extreme examples of officers working more than 3000 hours of overtime per year have been reported regularly during the past decade.[30] On top of these practices, many officers also work second jobs, about which almost no national data are available.

Shift work, overtime, and long work hours interfere with the social and domestic lives of officers, increasing logistical and scheduling problems that affect nearly every aspect of their lives. The burden borne by officers and their families as a consequence of work hour practices is similar to that for other occupational groups.[1] However, unlike occupations in which shift work and long, erratic work hours tend to be limited to the first few years after initial training (e.g., physician, public defender), many police and first responders (e.g., firefighters, emergency medical technicians) are exposed to these types of work schedules throughout their careers. Moreover, they do so in unstructured and unpredictable situations that require expert judgment, decision making, and self-restraint. Dr. William Dement summarized this issue as follows[31]:

> Police work is the one profession in which we would want all practitioners to have adequate and healthful sleep to perform their duties at peak alertness levels. Not only is fatigue associated with individual misery, but it can also lead to counterproductive behavior. It is well known that impulsiveness, aggression, irritability and angry outbursts are associated with sleep deprivation.[31]

SLEEP PROBLEMS IN MILITARY PERSONNEL

Military personnel deployed to theaters of operation (to include both combat missions such as Operation Desert Storm and humanitarian missions such as Operation Restore Hope) are exposed to stressors that preclude sufficient sleep (e.g., night operations) or are thought to cause sleep disturbances (e.g., witnessing the injury or death of a "battle buddy"). In-theater surveys of military personnel serve as the main source of data regarding sleep problems associated with military operations. Since 2003, the U.S. Army's Mental Health Advisory Teams (MHATs) have conducted approximately yearly in-theater surveys of deployed United States military personnel (in active duty, reserve, or National Guard units) to assess personnel behavioral health. Although early MHAT reports did not include survey items pertaining to sleep, such items have been included in more recent years. For the MHAT-5 to -9 surveys (conducted in 2007, 2009, 2010, 2012, and 2013, respectively), personnel were asked how many hours of sleep they obtained per day. In a separate item, they were asked how many hours of sleep per day they needed in order to feel well rested. Possible responses for both items were 4 or fewer, 5, 6, 7, and 8 or more hours. For the MHAT-V (as originally designated) survey, soldiers reported obtaining an average of 5.6 hours of sleep per day and needing 6.4 hours

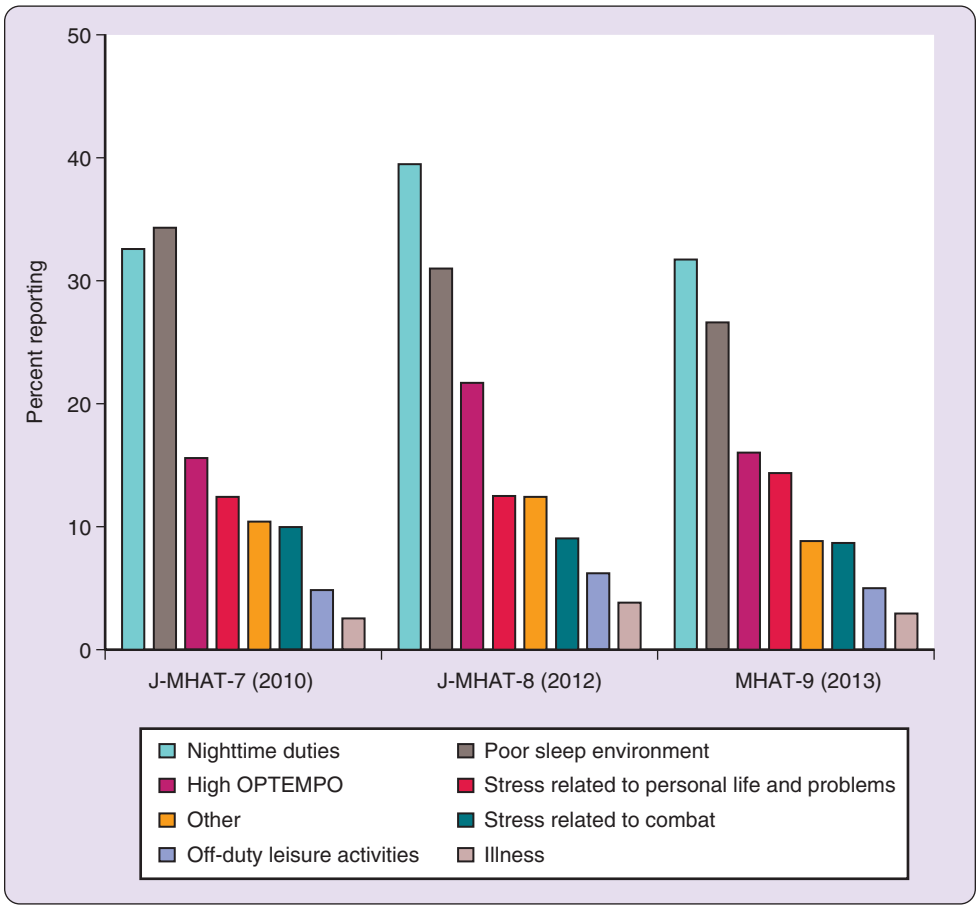


Figure 76-3 Percentage of respondents reporting that a particular factor interfered with their sleep on "more than half the nights" or "nearly every night over the past 30 nights." (Modified from U.S. Department of the Army, Mental Health Advisory Team 9 [MHAT 9]. *Operation Enduring Freedom [OEF] 2013 Afghanistan: report,* Table 5.6.1. Office of The Surgeon General United States Army Medical Command, and Office of the Command Surgeon Headquarters, US Army Central Command [USCENTCOM], and Office of the Command Surgeon US Forces Afghanistan [USFOR-A]. <http://armymedicine.mil/Documents/MHAT_9_OEF_Report.pdf>; 2013.)

sleep to feel well rested (results for these two items were not provided in subsequent MHAT reports).[32] Another item that was included in MHAT surveys conducted in 2010, 2012, and 2013 pertained to the frequency with which various factors interfered with sleep over the past month. Possible responses were "not at all," "few or several nights," "more than half the nights," and "nearly every night." The percentages of personnel responding "more than half the nights" or "nearly every night" are shown in Figure 76-3.[33] Consistently across MHAT reports, the two most frequently reported factors interfering with sleep were nighttime duties and poor sleep environment.

Several MHAT surveys also included an item in which personnel were asked to report how often (during the current deployment) that they had fallen asleep on a convoy, on guard duty, or during a briefing. Possible responses were "never," "seldom," "sometimes," and "often." Figure 76-4 illustrates results for those responding "sometimes" or "often."[33] Response rates were highest for falling asleep on a convoy, with almost 50% reporting that they had fallen asleep on a convoy in 2013. Rates for reporting falling asleep on guard duty and during a briefing were 20% or less. MHAT surveys conducted in 2009, 2010, 2012, and 2013 also contained an item in which personnel were asked to think about their experiences on the current deployment and rank how much trouble or concern had been caused by several factors, one of which was not getting enough sleep. Results for personnel responding that they had medium, high, or very high concern about not getting enough sleep are shown in Figure 76-5.[33] For all years except 2013, greater than 50% of personnel indicated medium, high, or very high concern about not getting enough sleep.

The U.S. Navy also has conducted behavioral health needs assessments. One such assessment consisted of Navy active duty and reservists (n = 3175) serving in a ground combat zone (Operation Enduring Freedom, Afghanistan).[34] The Navy survey was modeled after the Army's MHAT survey and included similar sleep items. Results from the Navy survey indicated that personnel reported obtaining 5.9 ± 1.1 hours of sleep per day, on average, as well as requiring 6.8 ± 1.0 hours of sleep per day in order to feel well rested. Personnel also were asked to rate their difficulty falling asleep and difficulty staying asleep in the past 2 weeks, using a 5-point scale ranging from 1 = "none" to 5 = "very severe." Mean difficulty falling asleep was rated as "mild" (2.1 ± 1.0), as was mean difficulty staying asleep (2.1 ± 1.1).

Results of other analyses from both the Army and Navy assessments revealed relationships between sleep and behavioral health. For example, in the Navy survey, sleep duration

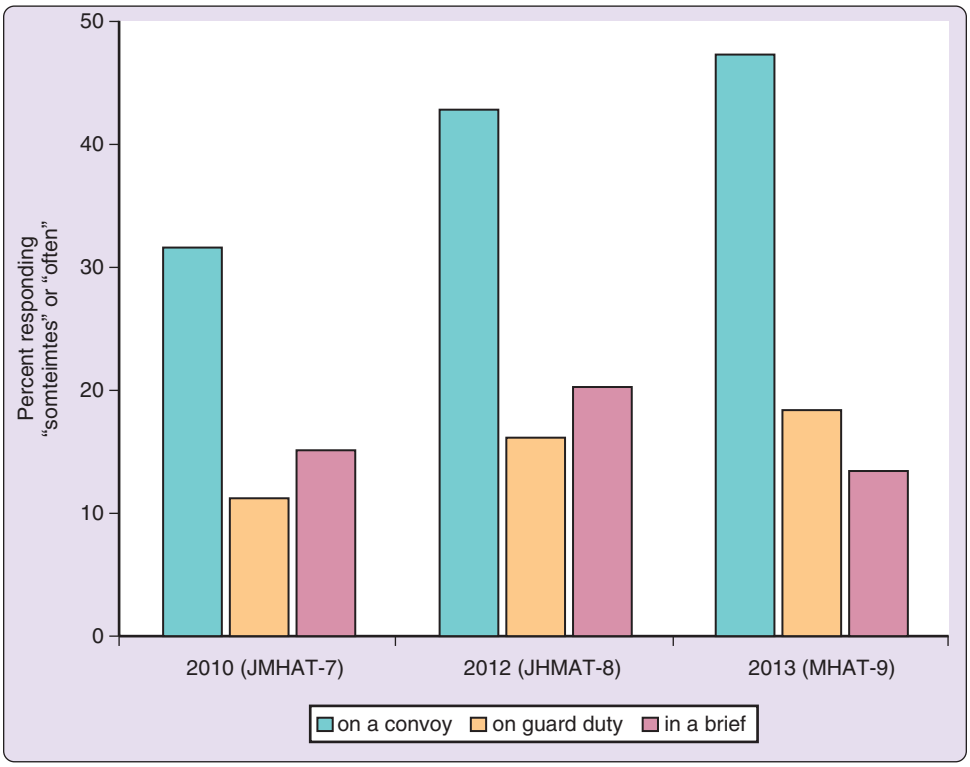


Figure 76-4 Percentage of respondents reporting that they had "sometimes" or "often fallen asleep (even briefly)" during a particular activity throughout the current deployment. (Modified from U.S. Department of the Army, Mental Health Advisory Team 9 [MHAT 9]. *Operation Enduring Freedom [OEF] 2013 Afghanistan: report,* Figure 5.6b. Office of The Surgeon General United States Army Medical Command, and Office of the Command Surgeon Headquarters, US Army Central Command [USCENTCOM], and Office of the Command Surgeon US Forces Afghanistan [USFOR-A]. <http://armymedicine.mil/Documents/MHAT_9_OEF_Report.pdf>; 2013.)

of 6 hours or less, difficulty falling asleep, and difficulty staying asleep were associated with significantly higher likelihood of experiencing symptoms associated with generalized anxiety disorder, major depressive disorder, and posttraumatic stress disorder. For MHAT-V, a sleep debt index was calculated by subtracting reported hours of sleep obtained (e.g., 4) from reported hours of sleep needed to feel well rested (e.g., 6 hours). This sleep debt index was then compared against responses for items pertaining to behavioral health problems. A positive (and approximately linear) relationship was found between amount of sleep debt and percentage of soldiers screening positive for depression, anxiety, or acute stress: Whereas less than 15% of soldiers with zero hours of nightly sleep debt screened positive for a mental health problem, just over 40% of soldiers with 4 hours of nightly sleep debt scored positive for a mental health problem. In addition, a higher percentage of soldiers whose nightly sleep debt score was 1 hour reported that stress or emotional problems over the past month had limited their ability to do their job, caused them to perform their work less carefully, or caused their supervisor to be concerned about their performance, compared with soldiers whose nightly sleep debt score was zero hours.

The MHAT-6 to -9 surveys also included an item regarding use of medications for a sleep problem—specifically, "Have you taken any medication for a sleep problem during this deployment?" Across the MHAT-6 to -9 reports, 9.6%, 11.3%, 6.4%, and 11.4% of respondents, respectively, indicated that they had taken medication for a sleep problem during that particular deployment. Whether using a medication for a sleep problem was associated with hours of sleep (or other sleep-related survey items) was not reported. For example, it might be expected that people working night operations would be more likely to use a medication for daytime sleep problems. Some MHAT assessments also included surveys of health care providers. For the MHAT-4 and -5 surveys, 30% and 52% of primary care providers, respectively, reported prescribing medications for sleep problems on a weekly basis.

In sum, results from surveys of personnel deployed to military theaters of operation indicate that chronic, insufficient sleep is common—and is associated with symptoms of degraded behavioral health. Some evidence suggests that deployed personnel use medications to help with sleep problems. Finally, the two most commonly reported factors leading to disrupted sleep are night operations and poor sleep environment.

OPERATIONAL PERFORMANCE CHALLENGES RESULTING FROM INSUFFICIENT SLEEP

Police and personnel deployed to combat zones may be seen as model occupational groups for (1) identifying factors that may be driving insufficient sleep and (2) recognizing those capabilities that could be negatively affected by insufficient sleep in other first responders such as firefighters and field emergency medical personnel.[35,36] As in police work and small-unit military operations, the tactical and strategic

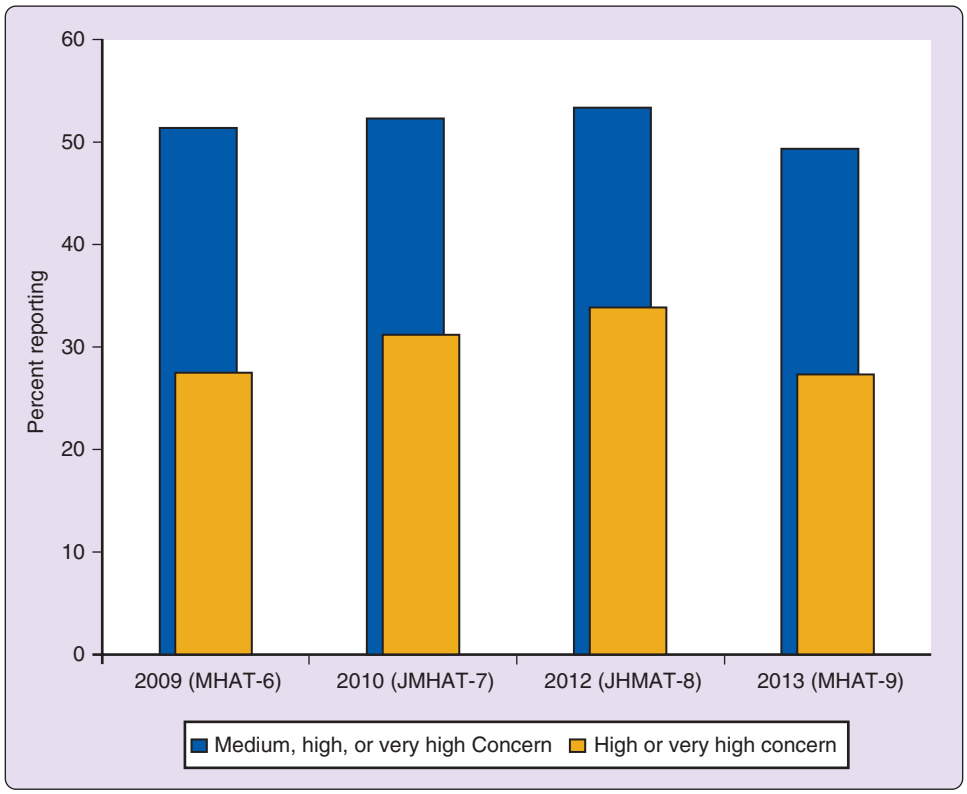


Figure 76-5 Percentage of soldiers reporting "medium" or "high"/"very high" concerns about not getting enough sleep during current deployment. (Modified from U.S. Department of the Army, Mental Health Advisory Team 9 [MHAT 9]. *Operation Enduring Freedom [OEF] 2013 Afghanistan: report,* Figure 5.6a. Office of The Surgeon General United States Army Medical Command, and Office of the Command Surgeon Headquarters, US Army Central Command [USCENTCOM], and Office of the Command Surgeon US Forces Afghanistan [USFOR-A]. <http://armymedicine.mil/Documents/MHAT_9_OEF_Report.pdf>; 2013.)

success of these endeavors requires a balance between the use of aggression and restraint.[30,31,37,38] Success often hinges on the performance of low-ranking infantry and security personnel operating in small, relatively autonomous units. Failure to strike an effective balance between aggression and restraint can lead to serious tactical, operational, and strategic setbacks.

Most of what is known about the effects of insufficient sleep on mental abilities is based on controlled laboratory-based studies (see further on). The most basic (and universal) effect is slowed response time—which on the job translates into impaired decision making, particularly under conditions of time pressure. Extrapolating from such results, it can be anticipated that the most proximal impact of insufficient sleep–related performance decrements in police, deployed military personnel, and other first responders is increased risk of injury (or death) in operators, their peers, and bystanders.

Sleep loss reduces the ability to think clearly, handle complex cognitive tasks, and solve problems.[13,39] Basic skills (e.g., marksmanship, emergency driving, routine enforcement or public safety activities) that are acquired through rote training or repetition or require choosing from only a small number of learned alternative courses of action also are degraded by sleep loss, but to a lesser extent.[40-42] Thus the impact of sleep loss in police and first responders is most likely to manifest in stressful, rapidly evolving, and complex situations characterized by ambiguity, high risk, and personal threat.

SYSTEMATICALLY MANAGING, PREVENTING, AND TREATING SLEEP DISORDERS AND SLEEP DEPRIVATION IN POLICE OFFICERS

Because the frequency, types, and locations of events that drive police judgments, decisions, and actions tend to differ with the daily, weekly, and seasonal rhythms of society, demands for service must be matched with efficient staffing and scheduling patterns that, in turn, limit officer opportunities for sleep and determine how much circadian disruption they must endure. This system of relationships must be taken into account to effectively manage, prevent, and treat police officers' sleep deprivation and sleep disorders.

Shift Work: Interaction Between the Organization and the Individual

The quality of fit between police scheduling and staffing and the moving target of demand for services directly impacts sleep: Poorer fit increases disruption of work schedules, generates more overtime work, and increases instances in which officers work longer shifts or multiple shifts or lose days off. All of these effects increase sleep loss, promote circadian disruption, generate additional stress in daily life, and reduce opportunities for sleep and recuperation. They also contribute to unexpected staff vacancies associated with illness, accidents, injuries, burnout, and early retirement. (Some police agencies in the United States report average daily absenteeism rates that range from 30% to 45%, attributed to sick

call-ins, occupational injuries, and similar causes.) Filling these vacancies generates more overtime and circadian disruption, creating a vicious circle in which long and erratic work hours reduce the number of staff available for work, which in turn increases the prevalence of long and erratic work hours.

Until relatively recently, this sort of human resource management problem could be solved by hiring more officers whenever the costs associated with overtime premium pay exceeded the overhead associated with recruiting and training new officers.[43] Now, however, demographic shifts and increasing demands for police involvement in solving a wider array of community problems have left managers "fishing" for more qualified recruits than are available within the human resource pool.[30,44,45]

Researchers are engaged with police agencies in an attempt to elucidate the health and human performance consequences of long and erratic work shifts in police work. A major goal of this research is to provide evidence-based policy guidance on deployment strategies and standards of practice.[31,46] Until now, demand for services and staffing levels generally were the only variables considered in assessing how efficiently an organization was deploying its personnel. Increasingly, however, managers and policy makers are recognizing that officer performance, health, and safety also are critical for community protection and safety. The essential insight here is that tired cops tend to be less able to serve their community than those who are adequately rested. This correlation puts officer sleep loss and disruption front and center as a research and policy issue. At the heart of this investigative endeavor is the need to explore the complex interaction among the community, police agencies, and police officers.

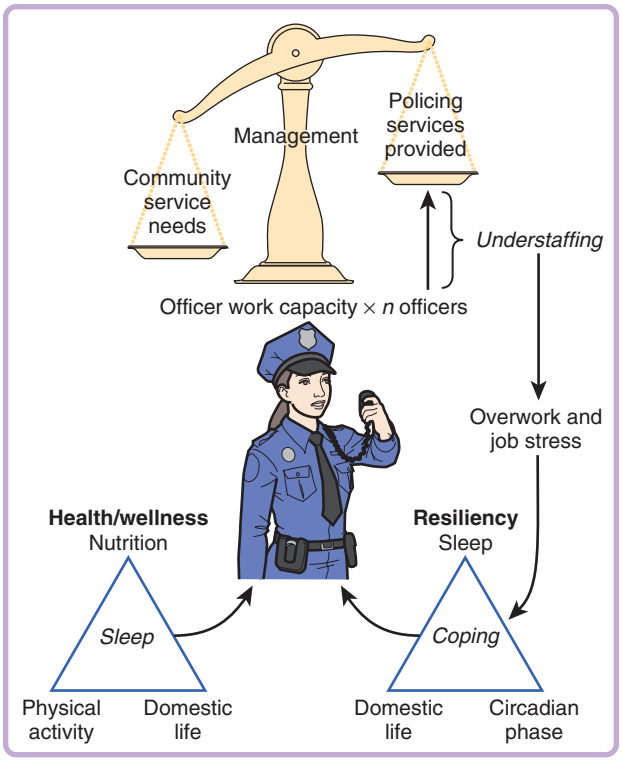


Figure 76-6 Conceptual model of interactions among needs of officers, police organizations, and communities.

Figure 76-6 provides a heuristic for understanding the relationship between individual police officer needs and management needs. As can be seen, the ability of management to provide sufficient police services to the community is a function of both the number of officers assigned to work at a particular time and the work capacity of those officers. Understaffing results in more work load and longer work hours—that is, greater stress to the individual officer. In turn, an officer's ability to cope with the demands of shift work and long work hours is affected by sleep, circadian, and domestic factors.[47] Shift work, long work hours, overwork, and job stress also tend to undermine overall health and well-being because they interfere with good nutrition, physical activity, and a connected social life:[16,48] Late-night shift workers often have difficulty obtaining nutritious food; shift work and long or erratic work hours make it difficult to find sufficient time to sleep, recover, exercise, and interact with others. The combined effects of these various stressors reduce work capacity by diminishing overall health and wellness (Box 76-1). Repeated exposure to the sleep- or circadian rhythm–related stress of shift work and long, erratic work hours throughout an officer's career may affect his or her longevity in the profession, in addition to health, safety, and survival.[11] When treating sleep

Box 76-1 KEY POINTS

- Research evidence to date identifies three key factors associated with negative health and human performance outcomes in shift work: (1) extended hours of service, (2) limited recovery time, and (3) time on task.
- The social, organizational, and individual causes of sleep loss among police and similar occupational groups are systematically linked, so they should be dealt with by striking a careful balance between the needs of society and the impact of shift work and long work hours on officers' health and performance.
- Shift length and rotation policies should be developed to fit the unique needs of each community and its police officers. There is no one ideal system for shift work.
- The development of shift scheduling strategies and staff deployment models in police agencies requires both operational expertise and human factors expertise. Sleep medicine physicians should participate in the scheduling/deployment process both as advocates for police officers and as experts on health and human performance issues. Their input can be crucial to the prevention of health problems and the optimization of wellness, with consequent improvement in officer performance.
- For diagnosing and treating sleep disorders in police officers, an understanding of how occupational factors affect resilience, health, and wellness is essential. Excessive overtime, frequent shift changes, and secondary employment are especially problematic issues in conjunction with sleep disorders.
- Determination of "fitness for duty" requires the expertise of occupational medicine physicians and is their responsibility. Sleep specialists should provide occupational medicine physicians with appropriate fitness-to-work recommendations regarding shift work–related sleep disorders.
- Appropriate management of sleep factors, circadian factors, and sleep disorders can mitigate the need for occupational accommodation and leaves of absence.

problems in police officers, sleep medicine physicians typically focus on sleep and circadian factors. To better address sleep problems in this group and similar first responders, it may be necessary for the clinician to broaden the scope of management to include a discussion with the police officer regarding his or her domestic demands, work scheduling, and similar critical factors.

Increased awareness on the part of sleep medicine physicians is only part of the overall solution. Similarly, incomplete approaches that ignore prevention and treatment issues associated with sleep disorders and related health problems ignore critical opportunities for managing risks and human resources more effectively. Comprehensive models (such as that shown in Figure 76-6) that consider demand for services, staffing, and human factors can improve individual health care, public health, and the management of scarce public resources. As Figure 76-6 emphasizes, adequately rested officers multiply the effective strength of a police force—just as adequate sleep can drastically improve the quality of officers' lives.[49,50]

The comprehensive approach that we advocate suggests measurable, objective outcomes that clinicians, practitioners, and policy makers can use to manage shift work and scheduling in a way that maximizes health, safety, and performance. For example, interventions in a comprehensive fatigue management program could be assessed by measuring independent variables such as timing and duration of sleep (using wrist actigraphy), hours worked (using timekeeping data), and regularity of work hours (using overtime and absenteeism data). Dependent variables such as overwork and job stress, resilience, nutrition, and physical activity all can be measured using standard assessment instruments. This comprehensive approach also supports a systematic attempt to optimize community well-being, organizational efficiency, and worker health and wellness. From a practical standpoint, this approach makes it easier to manage the competing interests of community, management, labor organizations, and individual officers because it provides all involved with a common ground for negotiating changes that help mitigate or minimize the unhealthy effects of shift work. For example, police agencies that take a comprehensive approach to staffing that manages sleep and work-hour issues effectively can conserve human capital by improving retention and reducing lost work days, disability claims, and requests for accommodation. They also may improve recruitment—often from agencies with less desirable work-hour practices.

Shift Work: The Individual Officer

In accordance with the proposed model, an officer's resilience to repeated exposure to sleep restriction, circadian dysrhythmia, and domestic disturbance resulting from shift work and erratic work hours can be enhanced by addressing (1) intrinsic sleep disturbance in the form of primary sleep disorders and (2) extrinsic factors related to the maintenance of general health. The relationships among sleep and shift work, cardiovascular disease, diabetes mellitus, and obesity have been described[51,52,53] in the sleep medicine, occupational medicine, and general medical literature.[54,55] Figure 76-7 conceptualizes how these factors may be causally related.[7]

The high prevalence of cardiovascular and metabolic disease among police officers is currently the focus of three major studies: the Buffalo Cardio-Metabolic Occupational Police Stress (BCOPS) study, the Calgary Police Service Health and Human Performance Research Initiative (CPS/HHPRI), and the Harvard Work Hours and Safety Group Police Study (HWHSGPS). Results from these studies highlight the prevalence of cardiovascular and metabolic diseases and sleep-disordered breathing in this population.[6,7,9,47] HWHSGPS researchers have established a high prevalence of primary sleep disorders (38.4%) and, in particular, sleep-disordered breathing (35.1%) on the basis of a survey conducted in 4471 police officers in the United States and Canada.[9] This result has been corroborated in a smaller pilot study in the Calgary Police Service[56] and in the BCOPS study.[6,7] Table 76-1 demonstrates the strong relationship between standard measures of obesity and sleep disordered breathing in police officers, while Table 76-2 demonstrates that night shift workers tend to snore more and controlling for BMI did not alter this relationship appreciably.

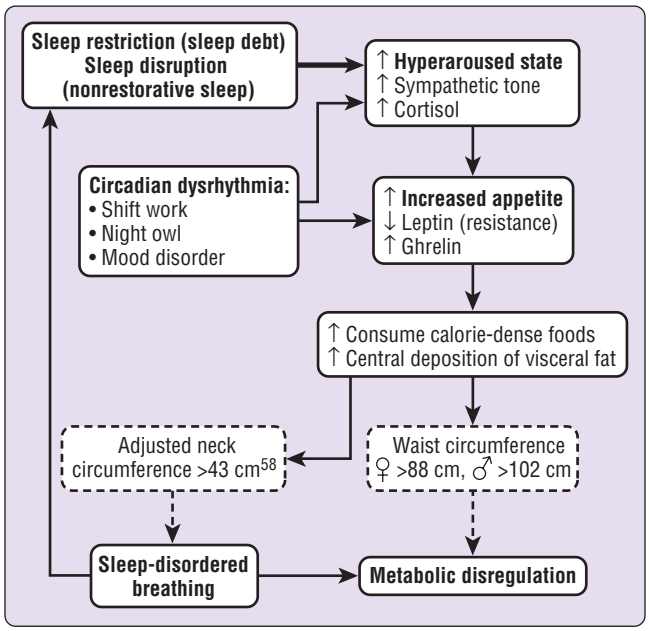


Figure 76-7 Conceptual model of relationship of sleep restriction and disruption to metabolic dysregulation, obesity, and sleep-disordered breathing. (From Samuels CH. Sleep and weight control: a wake-up call. *Can J Diagn* 2005;June:75–9.)

Officers working night shifts also have been shown to obtain less sleep per day (44% higher prevalence of less than 7 hours per day than among day shift officers) and to have a higher prevalence of snoring, a risk factor for sleep-disordered breathing (16% higher prevalence than among day shift officers), both of which are associated with increased risks of vascular disease.[6,7] Primary care, occupational medicine, and sleep medicine physicians can help minimize and mitigate the negative effects of shift work and erratic work hours in police by implementing the following measures:

- Screening clinically for primary sleep disorders
- Screening specifically for sleep-disordered breathing
- Educating officers about the importance of managing chronic cumulative sleep debt
- Educating officers about the relationships among weight control, obesity, metabolic syndrome, cardiovascular disease, and sleep or circadian factors

Table 76-1 Adjusted Mean Values of Anthropometric Obesity Measurements among Police Officers (Models 1 and 2) Stratified According to Symptoms of Sleep-Disordered Breathing

Anthropometric Obesity Measure*	Frequency of Sleep Problems	Snoring			Stop Breathing			Gasping for Breath		
		No.	Model 1 Mean (SE)	Model 2 Mean (SE)	No.	Model 1 Mean (SE)	Model 2 Mean (SE)	No.	Model 1 Mean (SE)	Model 2 Mean (SE)
Body mass index	Never	15	25.9 (1.0)	25.4 (1.2)	71	28.1 (0.5)	27.7 (0.7)	78	27.8 (0.5)	27.1 (0.6)
	<1–2 times/wk	50	27.5 (0.6)	26.8 (0.7)	4	28.3 (2.2)	28.6 (2.0)	13	29.0 (1.2)	28.9 (1.4)
	3–7 times/wk	24	30.9 (0.8)	29.6 (1.0)	4	30.8 (2.2)	29.9 (2.2)	3	32.0 (2.5)	32.3 (2.5)
	P_{trend}		<.001	.002		.235	.312		.108	.036
Hip circumference	Never	15	102.8 (2.2)	100.5 (2.6)	71	107.8 (1.1)	107.0 (1.5)	78	106.8 (1.1)	104.9 (1.4)
	<1–2 times/wk	50	106.2 (1.2)	105.6 (1.6)	4	105.4 (4.6)	106.6 (4.5)	13	107.8 (2.6)	106.5 (3.1)
	3–7 times/wk	24	112.3 (1.8)	109.9 (2.2)	4	112.1 (4.6)	111.0 (4.7)	3	115.9 (5.4)	116.7 (5.5)
	P_{trend}		.001	.003		.366	.391		.105	.036
Waist circumference	Never	15	84.9 (3.0)	85.2 (2.8)	71	91.6 (1.5)	91.6 (1.8)	78	90.6 (1.4)	89.8 (1.6)
	<1–2 times/wk	50	90.2 (1.7)	89.2 (1.7)	4	93.2 (6.5)	93.4 (5.3)	13	91.1 (3.6)	92.8 (3.4)
	3–7 times/wk	24	99.0 (2.4)	95.6 (2.3)	4	92.7 (6.5)	92.7 (5.6)	3	99.2 (7.5)	101.0 (6.1)
	P_{trend}		<.001	.002		.868	.845		.265	.068
Abdominal height	Never	15	19.0 (0.7)	18.7 (0.8)	70	20.9 (0.4)	20.5 (0.5)	77	20.6 (0.4)	20.2 (0.4)
	<1–2 times/wk	50	20.6 (0.4)	20.1 (0.5)	4	21.1 (1.6)	21.0 (1.4)	13	20.6 (0.9)	20.5 (0.9)
	3–7 times/wk	23	22.8 (0.6)	21.7 (0.7)	4	22.6 (1.6)	22.1 (1.5)	3	23.8 (1.8)	24.3 (1.6)
	P_{trend}		<.001	.002		.283	.254		.091	.014
Waist-to-hip ratio	Never	15	0.82 (0.02)	0.85 (0.02)	71	0.85 (0.01)	0.85 (0.01)	78	0.85 (0.01)	0.85 (0.01)
	<1–2 times/wk	50	0.85 (0.01)	0.85 (0.01)	4	0.88 (0.04)	0.87 (0.03)	13	0.84 (0.02)	0.87 (0.02)
	3–7 times/wk	24	0.88 (0.02)	0.87 (0.01)	4	0.83 (0.04)	0.84 (0.03)	3	0.86 (0.05)	0.87 (0.03)
	P_{trend}		.038	.225		.665	.622		.764	.606
Waist-to-height ratio	Never	15	0.49 (0.01)	0.50 (0.02)	71	0.52 (0.01)	0.53 (0.01	78	0.52 (0.01)	0.52 (0.01)
	<1–2 times/wk	50	0.52 (0.01)	0.52 (0.01)	4	0.52 (0.03)	0.52 (0.03)	13	0.52 (0.02)	0.53 (0.02)
	3–7 times/wk	24	0.56 (0.01)	0.55 (0.01)	4	0.53 (0.03)	0.54 (0.03)	3	0.57 (0.04)	0.58 (0.04)
	P_{trend}		<.001	.006		.847	.844		.200	.089
Neck circumference	Never	13	38.0 (1.1)	36.8 (1.0)	67	39.4 (0.5)	38.4 (0.6)	74	39.2 (0.5)	38.5 (0.5)
	<1–2 times/wk	48	38.9 (0.6)	37.8 (0.6)	4	40.3 (2.0)	39.7 (1.7)	13	40.3 (1.2)	41.0 (1.1)
	3–7 times/wk	24	42.0 (0.8)	40.3 (0.8)	4	42.0 (2.0)	40.3 (1.8)	3	43.3 (2.4)	43.2 (2.0)
	P_{trend}		.004	.002		.206	.275		.096	.023
Neck-to-height ratio	Never	13	0.22 (0.006)	0.21 (0.010)	67	0.22 (0.003)	0.22 (0.004)	74	0.22 (0.003)	0.22 (0.003)
	<1–2 times/wk	48	0.22 (0.003)	0.22 (0.003)	4	0.23 (0.011)	0.23 (0.010)	13	0.23 (0.006)	0.24 (0.007)
	3–7 times/wk	24	0.24 (0.004)	0.23 (0.004)	4	0.24 (0.011)	0.23 (0.011)	3	0.25 (0.013)	0.25 (0.013)
	P_{trend}		.004	.010		.156	.316		.057	.042

*Adjusted mean values of anthropometric obesity measurements by symptoms of sleep disordered breathing.
Data from Charles LE, Violanti J, Burchfiel C, et al. Obesity and sleep: the Buffalo Police Health Study. *Policing* 2007;30:203–14.

- Motivating officers to optimize behaviors over which they have at least partial control (i.e., physical activity, nutrition, rest/sleep)

Police agencies also can contribute to this effort by conducting educational programs that reinforce the importance of getting adequate sleep. This combination of physician and agency interventions can help officers focus on critical factors that will improve their ability to withstand the rigors of shift work in a challenging work environment.

Shift Work: The Organization

Sleep physicians should be aware that both management and labor share responsibility for mitigating the impact of shift work and erratic work hours. Policy makers, police executives, and officers and their families need to understand the implications of empiric evidence to make good decisions about work-hour issues. Efforts to translate knowledge into practice should emphasize the impact of shift work and erratic work hours on human performance and health. Moreover, they

Table 76-2 Association of Sleep Quality and Quantity* in Police Officers Who Work Night Shifts Compared with Those Who Work Day/Afternoon Shifts

	Model 1		Model 2		Model 3	
Reported Description of Sleep Quality/Quantity	PR	95% CI	PR	95% CI	PR	95% CI
"At night, my sleep disturbs my bed partner's sleep."	0.93	0.65–1.33	0.99	0.68–1.45	0.94	0.62–1.43
"I am told I snore in my sleep."	1.26	1.12–1.41	1.21	1.06–1.38	1.16	1.00–1.33
"I am told I stop breathing in my sleep."	0.53	0.07–3.95	0.67	0.08–5.68	0.59	0.02–14.3
"I suddenly wake up gasping for breath during the night."	0.27	0.04–1.89	0.25	0.03–1.89	0.27	0.03–2.67
"I have or have been told that I have restless legs."	1.26	0.64–2.45	0.91	0.46–1.80	1.02	0.53–1.95
"I feel tired upon awakening and want to go back to sleep."	0.99	0.83–1.18	0.97	0.82–1.16	0.97	0.81–1.16
"I am very sleepy during the daytime and struggle to stay awake."	0.91	0.64-1.30	0.83	0.58–1.19	0.90	0.62–1.32
Hours of sleep per 24-hour period during the previous week (<7 vs. ≥7)	1.46	1.09–1.94	1.35	0.99–1.85	1.44	1.00–2.07

*Results obtained from Poisson regression models: Model 1, unadjusted; Model 2, adjusted for years of service; Model 3, adjusted for years of service, depression, body mass index, physical activity, and gender.
CI, Confidence interval; PR, prevalence ratio.
Data from Charles LE, Burchfiel CM, Fekedulgen D, et al. Shift work and sleep: the Buffalo Police Health Study. *Policing* 2007;30:215-27.

should provide practical solutions that help strike a balance between the sometimes competing needs of officers and the communities they serve.

Shift-work management strategies must go beyond the current focus on identifying an "ideal shift." Wellness and health promotion programs need to broaden their focus to include sleep and circadian factors as well as nutrition, activity, and stress reduction. As Figure 76-7 illustrates, the rigors of shift work in police work require a broad-based and comprehensive approach to be effective.

Exemplary Unified Approach

An example of the unified approach we advocate is the Calgary Police Service (CPS) Health and Human Performance Research Initiative. This effort, which grew out of the 2001 CPS Patrol Officers Shift Schedule Review (POSSR),[49,50] is a long-term research and knowledge transfer project. One of the key POSSR recommendations was for enlistment of a sleep specialist and a knowledge transfer expert to help the service develop and evaluate new shift schedules that were sensitive to officer health and human performance issues. Perhaps the most unique aspect of this initiative is the knowledge transfer component, which uses new scientific evidence as it emerges to develop educational programs. These educational programs guide CPS efforts to change how police officers, labor leaders, and managers view shift-work management and related health and human performance issues.

The CPS Health and Human Performance Research Initiative is a comprehensive, collaborative effort by all stakeholders that recognizes that each has a stake in developing strategies for managing shift work and that cooperation is vital for long-term success. Another important aspect of this initiative is that it has support from high-ranking labor and management leaders, who have been instrumental in maintaining its momentum. The success of this approach has drawn the attention of other police agencies, several of which are developing fatigue management strategies that fit their own resources, needs, and organizational cultures. In each of these efforts, sleep and occupational medicine specialists, along with knowledge transfer professionals, can provide critical expertise and advice.

CLINICAL PEARLS

A comprehensive program of fatigue risk management for police officers and other first responders should include:

- Screening all officers for primary sleep disorders, and specifically for sleep-disordered breathing, owing to its high prevalence in this population
- Emphasizing the importance of managing day and night sleep environments, cumulative sleep debt, early signs of chronic sleep deprivation, and symptoms of shift-work sleep disorder
- Providing educational material for officers and their families for the purpose of preventing and mitigating the health and human performance consequences of shift work.

SUMMARY

Fatigue associated with sleep loss, circadian disruption, and night operations is one of the most common health and safety hazards faced by police officers and other first responders and by deployed military personnel. Sleep-related fatigue contributes to morbidity (and possibly mortality) in these occupational groups. It also degrades cognitive performance, affecting judgment and decision making as well as emotional responses to situational challenges. Because the social, organizational, and individual causes of insufficient sleep within these occupational groups are inextricably linked, the various individual, social, organizational, and environmental processes, as reviewed in this chapter, should be considered in addressing sleep-related problems in these populations.

Physicians and other health care providers can improve individual and public health outcomes by using comprehensive approaches that systematically consider all of these issues. From a practical standpoint, this approach makes it easier to

manage the competing interests of individual police officers/first responders/military personnel and their families, their organizations, and the larger communities they serve. Expert advice from appropriate health care and other professionals will help these competing stakeholders find a common ground for negotiating changes that help mitigate or minimize the unhealthy effects of shift work.

Selected Readings

Chang JH, Huang PT, Lin YK, et al. Association between sleep duration and sleep quality, and metabolic syndrome in Taiwanese police officers. *Int J Occup Med Environ Health* 2015;**28**(6):1011–23.

Charles LE, Burchfiel CM, Fekedulgen D, et al. Shift work and sleep: the Buffalo Police Health Study. *Policing* 2007;**30**:215–27.

Folkard S, Tucker P. Shift work, safety and productivity. *Occup Med (Lond)* 2003;**53**:95–110.

James SM. Distracted driving impairs police patrol officer driving performance. *Policing* 2015;**38**(3):505–16.

James SM, Vila B. Police drowsy driving: predicting fatigue-related performance decay. *Policing* 2015;**38**(3):517–38.

Rajaratnam SM, Barger LK, Lockley SW, et al. Sleep disorders, health, and safety in police officers. *JAMA* 2011;**306**(23):2567–78.

Taylor MK, Hilton SM, Campbell JS, et al. Prevalence and mental health correlates of sleep disruption among military members serving in a combat zone. *Mil Med* 2014;**179**(7):744.

U.S. Department of the Army, Mental Health Advisory Team 9 (MHAT 9). *Operation Enduring Freedom (OEF) 2013, Afghanistan: report.* Office of the Surgeon General, United States Army Medical Command; Office of the Command Surgeon, Headquarters, US Army Central Command (USCENTCOM); and Office of the Command Surgeon, US Forces Afghanistan (USFOR-A). <http://armymedicine.mil/Documents/MHAT_9_OEF_Report.pdf>; 2013.

Vila B. *Tired cops: the importance of managing police fatigue*. Washington (D.C.): Police Executive Research Forum and National Sleep Foundation; 2000.

Vila B. Impact of long work hours on police officers and the communities they serve. *Am J Ind Med* 2006;**49**:972–80.

Violanti JM, Burchfiel CM, Hartley TA, et al. Atypical work hours and metabolic syndrome among police officers. *Arch Environ Occup Health* 2009;**64**(3):194–201.

A complete reference list can be found online at ExpertConsult.com.

Chapter
77

Sleep, Occupational Stress, and Burnout

Torbjörn Åkerstedt; Aleksander Perski; Göran Kecklund

Chapter Highlights

- Results from prospective epidemiologic studies show that work demands are associated with sleep disturbances.
- Results from polysomnography studies show reduced sleep continuity or reduced slow wave sleep in relation to bedtime stress.
- Long-term exposure to occupational stress is related to substantial reductions in sleep continuity and slow wave sleep.
- Rumination, to include worry regarding next-day events, seems to be a key factor related to sleep disturbances.

Disturbed sleep usually is attributed to daily life stress.[1] Among people with insomnia, stress (and depression) is a correlate of disturbed sleep[2] and increased levels of arousal.[3] Before examining work stress and sleep, it is necessary to define stress and work life.

STRESS

Stress generally refers to the observation that repeated exposure to some event without sufficient recovery between exposures leads to physiologic activation that causes wear and tear.[4] Although both the duration and type of rest between bouts of stress may be important factors, sleep is one of the major physiologic means of recovering from stressors. The amount of available data on stress and sleep is relatively modest; however, some of the findings are summarized here.

From a general perspective, *psychosocial* stress refers to the rate of wear and tear in the organism, and *biologic* stress refers to the nonspecific response to any demand[5] that increases the chances of successful handling of a threatening situation. Contemporary physiologic stress models derive from the pioneering work of Cannon[6] and of Selye.[5] Cannon[6] developed the concept of the "fight-or-flight" response, which linked the emotional, central nervous system perception of a threat to physiologic changes in the peripheral nervous system. Selye proposed the *general adaptation syndrome* model of stress, which was composed of three stages—alarm, resistance, and exhaustion—that corresponded to the physiologic nonspecific response to a challenge. The resistance stage of the general adaptation syndrome has extreme physiologic energy requirements, which if persistent over time deplete available resources and lead to exhaustion.

Markers of the fight-or-flight response are the catecholamines epinephrine and norepinephrine and other physiologic indicators associated with the autonomic nervous system.[4] The hypothalamic-pituitary-adrenocortical (HPA) axis is fundamental to the stress response. When the hypothalamus is activated, neuropeptides such as corticotropin-releasing hormone (CRH) and vasopressin are released. They in turn stimulate the pituitary to release adrenocorticotropic hormone (ACTH) into the general blood circulation. ACTH further stimulates adrenal cortex to release hormones including the glucocorticoid cortisol. Cortisol in turn activates mobilization of energy resources.

The ability of the body to increase or decrease the activation level of vital functions to new steady states, dependent on the characteristics of the challenge and the person's emotions and appraisal of the event, is referred to as *allostasis*.[4] *Allostatic load* refers to the cumulative cost to the body of responding to stressful events. It is suggested that serious pathophysiology can occur if allostatic load is not relieved in some way.[4] One of the outcomes may be insomnia.

With respect to *working life,* nonphysiologically based, systematic theories have emerged, and one leading theory in work stress research is the so-called *demand-control model* (Figure 77-1, *top*).[7] This model is based on empiric evidence, such as the finding that high work demands and low decision latitude are predictive of cardiovascular and other types of stress-related diseases.[8] Another work stress model is the *effort–reward model* (Figure 77-1, *bottom*), which focuses on the contrast between demands and resources.[9] Resources include salary, career opportunity, and other rewards. In this model, *immersion* plays an important role and represents major commitment of effort. Immersion (or commitment) involves experiences such as not being able to stop thinking of work in the evening or starting to think of work immediately on awakening. Another important work-related factor may be the amount of social support received at work. Results from several studies indicate that lack of social support is associated with development of cardiovascular disease, depression, and other negative health outcomes.[4]

THE CONNECTION BETWEEN STRESS AND SLEEP: EVIDENCE FROM CROSS-SECTIONAL STUDIES

Considering the physiologic activation involved in the stress response, it seems logical to expect a connection with sleep disturbances. In fact, stress is considered the primary cause of persistent psychophysiologic (primary) insomnia.[3] The evidence supporting a relationship between stress and insomnia is, however, surprisingly modest. No longitudinal systematic studies of causal relationships between stress and sleep have

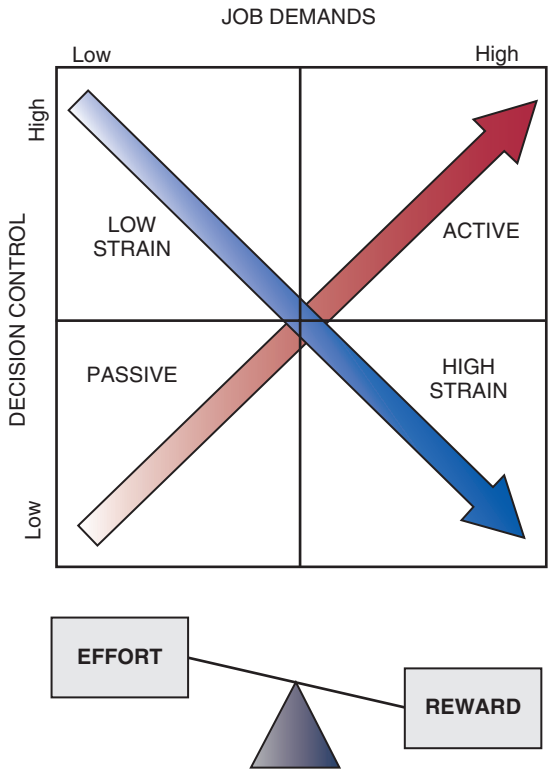


Figure 77-1 Schematics of the demand-control-support model (*top*)[7] and the effort-reward model (*bottom*).[9] In the four-component box at top, degree of decision control and perceived job demands interact to determine the worker's job situation. The ideal situation is "active," where high job demands coupled with high decision control result in active engagement—which in turn is associated with positive health outcomes. By contrast, high job demands with low decision control are associated with poor health outcomes. In the simple balance scale depicted at bottom, the direct relationship between effort (e.g., responsibilities, time pressure) and reward (e.g., salary, career opportunity) affects health and well-being. High commitment of effort is referred to as "immersion," which is associated with rumination (a feature of insomnia).

been conducted. Cross-sectional epidemiologic studies are readily available, however, and survey results indicate a strong link between self-ratings of stress and sleep (as reported by Kompier and colleagues,[10] for example).

Studies of stress and sleep in which sleep was objectively quantified using polysomnography (PSG) are far more limited. Shaver and associates[11] studied women with insomnia and found that stress ratings in the insomniac women did not differ from those in good sleepers. Women with insomnia did show evidence of neuroendocrine activation (and similar measures) suggestive of increased stress. By contrast, Akerstedt and associates found a positive relationship between ratings of daily stress and increased levels of polysomnographically recorded stage 1 sleep.[12] Using actigraphy, Mezick and associates showed that high reported stress level was associated with more variability in sleep duration and increased sleep fragmentation.[13]

WORK DEMANDS

The demand component of the demand-control model of stressful work conditions also has been shown to be associated with reported sleep impairment: Early findings by Urponen and coworkers[14] showed that men rated work-related pressure as a critical factor that impaired the ability to fall asleep and quality of sleep. Åkerstedt and associates[15] found that the strongest item of the demand index was "having to exert a lot of effort at work"; the latter was distinct from simply "having too much to do." The distinction suggests that amount of work per se is not the key factor; rather, the response to the work situation is more important. Also, when the item "not being able to stop thinking about work in the evening" was added to the regression, this variable became the most important predictor of disturbed sleep. Again, it is the response to the work situation that predicts sleep impairment. Imbalance between effort and reward also has been related to disturbed sleep.[16]

Socioeconomic Status

With regard to socioeconomic group, sleep complaints are more frequent in blue-collar workers: Partinen and coworkers[17] investigated several occupational groups and found disturbed sleep to be most common among manual workers and much less so among physicians or managing directors. In a retrospective study in older people (older than 75 years), Geroldi and associates[18] found that former white-collar workers reported better sleep than former blue-collar workers. On the whole, blue-collar workers report higher levels of stress than do white-collar workers, but the source of the stress appears to be related more to living conditions in general (e.g., economy, social situation, neighborhood) than to work stress in particular.

Social Support

In demand-control models, social support is viewed as a countervailing force pushing back against the effects of high demands.[7] Lack of social support may thus function as a risk factor for disturbed sleep. Alfredsson and associates[19] demonstrated an interactive effect on sleep between social support, high demands at work, and lack of control. Poor social support also has been associated with sleep complaints in Vietnam War veterans.[20]

THE CONNECTION BETWEEN STRESS AND SLEEP: EVIDENCE FROM LONGITUDINAL STUDIES

The question of whether stress actually *causes* impaired sleep can be resolved only by implementing longitudinal studies. In one of the first studies of this type, Ribet and colleagues[21] interviewed more than 21,000 subjects in France on two occasions separated by 5 years. They used a five-item sleep disturbance index and provided a variety of occupational questions. Results from the logistic regression analysis indicated that reporting "having to hurry" during the first interview predicted sleep disturbances during the second interview. In another field study approach, Dahlgren and colleagues[22] assessed white-collar workers before and during a period of intense work stress as well as a period without intense work stress. Reported sleep duration, sleep quality (assessed by sleep diary), and evening alertness were reduced during the period of intense work stress, and reported restlessness at bedtime was increased. In addition, the daily cortisol pattern was flattened during intense work stress. In another study of college students, Sadeh and associates[23] found reduced amounts of sleep

(assessed by actigraphy and diary ratings) during a period of examination stress. In both the Dahlgren et al. and Sadeh et al. studies, it is not clear whether reduced sleep was partly due to simply trading sleep time for more work[22] or study[23] time, rather than stress directly impairing the ability to sleep.

Linton[24] studied employees with no reported initial sleeping problems and found that 14.3% developed a sleeping problem during the ensuing year. Even with institution of controls for possible confounding factors, stress in the form of a poor psychosocial work environment doubled the risk of developing a sleep problem. Similarly, Jansson and Linton[25] showed that among people with no insomnia at baseline, high work demands increased the risk of developing insomnia 1 year later. Among participants with insomnia at baseline, high leader support decreased the risk of still reporting insomnia at follow-up. Finally, low (weak) influence over decisions and high work demands were related to the maintenance of insomnia. Results from several large prospective epidemiologic studies also have linked high work demands with subsequent poor sleep.[26-29] Results from a recent systematic review confirmed that both high job demand and low job control are associated with poor sleep quality.[30]

Using a 5-year longitudinal study that included participants with and without sleep problems at baseline, Vahtera and colleagues showed that a self-rated tendency to respond strongly to stress (which they referred to as "liability to anxiety") predicted later sleep disturbances.[31] These results point to a consistent pattern of disturbed sleep in response to stress. Similarly, Drake and coworkers utilized the known "first-night effect" on sleep as a naturalistic stressor—that is, sleep generally is disturbed on the first night that subjects spend in the sleep laboratory. These workers showed that subjectse who reported a higher habitual sleep vulnerability to stress also displayed longer sleep latency and lower sleep efficiency on the first night in the sleep laboratory.[32] These findings suggest that sleep vulnerability to stress is a definitive trait. In the earlier-mentioned study by Sadeh and associates,[23] a stronger sleep duration response to examination stress was seen in those subjects who showed a high emotion-focused coping strategy.

Åkerstedt and colleagues[33] assessed self-rated daytime stress and polysomnographically recorded sleep at home in 50 participants (thereby precluding the first-night effect) on four occasions spaced across several weeks. These investigators showed that nighttime ratings of stress and worries at bedtime were associated with reduced sleep efficiency, increased wake time after sleep onset, and increased latency to slow wave sleep. Self-ratings of "stress" (which the participants recorded every 3 hours) were increased both on the day before and the day after impaired nighttime sleep. Also, interindividual vulnerability was evident: Those participants who showed an increase in stress ratings scored a significantly higher level of depression on the Hospital Depression and Anxiety Scale. In a polysomnographic study of teachers, Petersen and colleagues[34] showed a decrease in sleep continuity and sleep efficiency during high stress periods in persons who reported that their sleep was vulnerable to stress (as determined by the Ford Insomnia Response to Stress questionnaire). In another study, 50 volunteers completed a daily sleep-wake diary and responded to daily questions on sleep and stress for 42 days.[35] Results showed that day-to-day variability in sleep quality was related to the level of stress and worries in the evening.

In other studies, sleep was polysomnographically recorded the night before an important examination[36] and also before a day of skydiving.[37] The results indicated a slightly negative effect on sleep efficiency and the amount of deep sleep. In addition, a number of laboratory studies of stress and sleep have been performed, but the stressors were artificial (e.g., viewing an unpleasant movie). Perhaps not surprisingly, the results from such studies are mixed, suggesting that the stressor must be of some direct significance to the subject to have any effect on objective (PSG-based) measures of sleep. However, results from other laboratory-based studies have demonstrated that anticipatory stress (see further on) can have rather pronounced effects on sleep.[38,39]

Rumination and Anticipation

Results from of the studies just described suggest that it is not the stress experience itself that impairs sleep but rather that the *anticipation* of untoward events and worries about the immediate future was the critical factor. Examinations, high work demands, and external threats are likely to cause this type of negative anticipation. For example, Cropley and colleagues[40] showed that the relationship between strain and poor sleep in teachers was moderated by rumination (i.e., compulsively thinking about a situation, usually a negative situation). Results from the aforementioned cross-sectional study by Åkerstedt and associates[41] showed that not being able to turn off thoughts of work in the evening was significantly associated with subjectively disturbed sleep. Inability to turn off work-related thoughts is a form of rumination, and rumination is seen as a major cause of disturbed sleep.[42] In a cross-sectional study, Hall and coworkers[2] showed that intrusive thoughts at bedtime are related to increased alpha and beta power in the subsequent sleep electroencephalogram (EEG). Similarly, increased cognitive arousal at bedtime is related to increased sleep latency.[43]

Another form of rumination is worrying before sleep about having to awaken early before an early-morning shift. In one study, Akerstedt and colleagues showed that such rumination was associated with less slow wave sleep during subsequent sleep.[44] Their findings support the notion that anticipation of future difficulties also evokes a stress reaction. A similar study was carried out with machine officers on container ships.[45] Sleep EEGs recorded during a night on call (but without any call occurring) showed reduced slow wave sleep and increased stage 2 sleep. Also, heart rate was increased. These results were interpreted as stress activation associated with the possibility that the alarm would sound, which happened on average every second night on call. On the other hand, increased mental activity has been shown to increase sleep intensity.[46] Thus *acute* stress (with associated increased brain activity) could *improve* subsequent sleep so long as there is no lingering negative anticipation or rumination. This hypothesis has yet to be explored.

SLEEP AND LONG-TERM EXPOSURE TO OCCUPATIONAL STRESS

The foregoing studies dealt with relatively acute and modest levels of stress. Assessing the impact of severe stress on sleep in a longitudinal or repeated-measures study (pre- and then poststressor) is difficult because intentionally exposing subjects to severe stressors is considered unethical. Nonetheless,

a critical question is how long-term exposure to stress affects sleep. Little work on this topic has been published. An increasing number of reports of stress-related sickness among otherwise healthy and high-performing persons have emerged, however.[47] Symptoms include profound fatigue, problems with memory and concentration, aches, irritability, anxiety, and a feeling of being emotionally drained—all attributed to long-term exposure to stress. The overall effect, which is relatively long-lasting, is associated with difficulties living up to job expectations. People who are prone to developing symptoms seem to be those who accept high workloads and increasing overtime work to handle increased work demands, and they view work as a positive challenge. The phenomenon also is sometimes referred to as "occupational burnout."[48] Many such patients are diagnosed with depression. A more appropriate diagnosis, however, may be "reaction to severe stress and adjustment disorder" according to the *International Classification of Diseases* (ICD10) (F43), with "work-related" as a further criterion.

The physiologic mechanism underlying the fatigue associated with occupational burnout has not been identified, but the HPA axis is likely to be involved. Results of several studies revealed a blunted cortisol response in the dexamethasone suppression test, suggesting impairments in the feedback loop.[47] Reduced hippocampal size also was observed. The latter finding is interesting in view of the relationship between circulating cortisol levels and impaired cognitive function (a characteristic symptom of occupational burnout). The other consistent finding is an exaggerated cortisol awakening response, suggesting increased anticipation of stress. This exaggerated cortisol awakening response has been mainly observed among female patients with exhaustion syndrome.[49] Results from other studies have shown (1) a functional disconnect between the amygdala and the anterior cingulate cortex/medial prefrontal cortex, (2) reduced 5-hydroxytryptamine HT_{1A} receptor binding potential in the anterior cingulate, the insula, and the hippocampus,[50] and (3) gray matter shrinkage in the anterior cingulate cortex and the dorsolateral prefrontal cortex, caudate, and putamen. Volumes of the caudate and putamen also correlated inversely with the degree of perceived stress.[51]

Although some physiologic (endocrine and brain structural) links between chronic stress exposure and occupational burnout have been established, the results do not necessarily "explain" or indicate direct causation of the symptoms.[52] A mediating factor could be disturbed sleep. To date, only a few PSG studies of sleep have been performed in persons with long-term exposure to stress. In one investigation of young people in the information technology industry with high burnout scores,[53] an increased level of sleep fragmentation was found; subjective sleep problems also were markedly increased. In addition, sleepiness ratings (taken every 3 hours during waking) were increased during the working week (compared with those in subjects who did not exhibit high burnout scores), and sleepiness during days off fell only marginally. In subjects without high burnout scores, dramatic decreases in sleepiness were observed on days off. In the same company, referral for treatment for stress-related disorders was predicted by fatigue and burnout scores at 1 year before referral but even more so by reported "insufficient sleep."[54]

In a study of workers on long-term sick leave because of long-term work stress, a similar increase in sleep fragmentation was seen, as well as lower sleep efficiency, more wake time after sleep onset, a doubled sleep latency, and reduced amounts of stage 3 and stage 4 sleep.[55] The degree of fragmentation of sleep was positively correlated with morning levels of cortisol, lipids, and other stress-related variables.[56] In interviews with a subsample of these patients, they described long periods of stress before becoming sick-listed; they identified the use of self-imposed sleep reduction to obtain more time to carry out work tasks; and they reported suddenly stopping sleeping.[57] The poor sleep was followed by extreme fatigue (the symptom that brought them to medical attention), resulting in a diagnosis of occupational burnout. Follow-up evaluation was performed at an average of 1.5 years after the first investigation; by that time, most PSG measures (as subjective ratings) had improved, as had stress, sleep, and fatigue ratings. Of interest, the PSG variable that best predicted fatigue 1.5 years later was reduction in sleep fragmentation. The group of patients who showed greatest reduction in sleep fragmentation also was the group that had returned to work.[55]

If a causal link exists between occupational burnout and sleep, the latter should predict the former in longitudinal studies. In one study, such a relationship was demonstrated.[58] Those persons who benefited poorly from sleep at baseline exhibited higher exhaustion levels at follow-up evaluation compared with those who benefited from sleep. Trouble falling asleep and less refreshing sleep at baseline were associated with inability to resume work. Armon and colleagues[59] further showed that occupational burnout and insomnia recursively predict each other's development and intensification over time, suggesting that either might be a risk factor for the other. Of note, however, the latter study was conducted with previously healthy workers and persons with occupational burnout.

SLEEP PHYSIOLOGY THAT SEEMS TO LINK SLEEP WITH STRESS

Insights into the connection between sleep and stress also may be obtained by comparing the physiologic changes during sleep versus after sleep loss with those during stress. Reductions in rectal temperature, rate of breathing, heart rate, blood pressure, and so on that occur during non-REM (NREM) sleep are normal aspects of sleep physiology.[60] Also, in adults, the first part of sleep is characterized by increased growth hormone (GH) release (together with increased slow wave sleep and low levels of REM sleep) and suppressed secretion of hormones of the HPA system[61] (i.e., CRH, ACTH, and cortisol are suppressed). Suppressed HPA axis hormone secretion during early portions of sleep is the opposite of that seen after stress. When sleep is prevented, cortisol secretion increases,[62] as it does subsequent to stressor exposure. Thus sleep reduction has effects similar to those of stress (i.e., increased cortisol secretion). Also, secretion of GH is strongly reduced if no sleep occurs.[62] GH promotes protein synthesis, which means that it is essential for growing and for repairing tissue. In other words, insufficient sleep may itself constitute a stressor.

Sleep not only regulates stress hormone secretion but also is affected by it. For example, the rate of sleep fragmentation (manifested as microarousals) increases with increased levels of cortisol.[56] Thus the quality of sleep partly depends on an interaction of GH-releasing hormone and CRH.

Reduced glucose clearance after sleep loss[62] also possibly reflects changes that constitute a stress response. The immune system is strongly suppressed by stress, mediated by, for

example, increased activation of the HPA axis. Similarly, reduced or fragmented sleep is associated with increased levels of proinflammatory cytokines such as interleukin-1β, interleukin-6, and tumor necrosis factor alpha.[63] Essentially, lack of sleep is linked with the metabolic syndrome, as are other stressors—further evidence that lack of sleep itself is a stressor and, moreover, exacerbates the physiologic response to other stressors.[64]

SIMILARITY OF MORBIDITY DUE TO STRESS AND TO SLEEP LOSS

Poor sleep is associated with an increased prospective risk for development of diseases that are considered to be stress-related, including myocardial infarction[65]and hypertension.[66] In women undergoing rehabilitation from a myocardial infarction, the risk of recurrent myocardial events is increased in self-reported poor sleepers.[67]

In addition, type 2 diabetes (also considered to be stress-related) has been shown to correlate with poor sleep.[68] For example, patients with type 2 diabetes report more sleep problems than nondiabetic subjects (although this association could be confounded by obesity or by obstructive sleep apnea). However, in a prospective follow-up study of healthy middle-aged men from Malmö, Sweden, it was shown that the 12-year risk of developing type 2 diabetes was independently predicted by self-reported difficulties with falling asleep and by elevated resting heart rate, after adjusting for obesity, lifestyle factors, and other risk factors.[69] Obstructive sleep apnea was not measured in the Malmö study and presumably could mediate the relationship between type 2 diabetes and sleep. Another possibility is chronic low-grade inflammation, which is linked to both insomnia and risk of type 2 diabetes. Finally, as noted earlier, short sleep duration also is involved in the metabolic syndrome.[64]

MEASURING STRESS

From the discussion so far, it is obvious that "stress" has many meanings. Earlier, stress was defined as an activating response to demands. In many of the aforementioned studies, stress was operationally defined as high levels on the demand-control questionnaire and/or on the basis of commonsense notions that the anticipation of an examination or of sky diving should be stressful. In short, a generally accepted operational definition of stress is lacking. Several attempts to provide one have been made, but these attempts reflected broad approaches that frequently included numerous subjective self-assessments of the consequences of stress. For example, the Perceived Stress Questionnaire[70] contains 20 items such as "tension," "worries," and "joy" in addition to items on pressure and demands. The Perceived Stress Scale[71] contains 14 items such as "upset," "on top of things," and "nervous" in addition to items on demands and stress. The Stress Response Inventory contains 39 items, including scales on depression, frustration, anger, and others.[72] The Calgary Symptoms of Stress Inventory[73] is a similar scale. Because these types of scales combine the cause (e.g., stress, demands, pressure) with the effects (e.g., anger, anxiety), they are not well suited for use in cause-and-effect studies of stress and sleep. Understanding the role of stress in sleep disorders requires the development of unambiguous instruments for quantifying stress.

CONCLUSIONS AND FUTURE DIRECTIONS

It is apparent from the foregoing that more detailed knowledge regarding the development of stress and sleep problems is needed—that is, longitudinal studies with frequent measures of stress and objectively measured sleep. These are labor-intensive and thus difficult studies for both researchers and participants; it is unlikely that many studies of this type will be forthcoming. In addition, longitudinal studies of the amount and duration of stress needed to affect sleep are needed as are longitudinal studies of recovery from stress-related sleep disturbances. Furthermore, information is needed regarding the contribution of both work and private life stressors to sleep and the impact of sleep on stress and work performance.

Although available evidence indicates that stress causes subjectively impaired sleep, little is known about the effects of stress on sleep architecture, at what level or by what type of stress sleep is impaired, and what duration of stress causes insomnia. Also lacking is information on trait factors contributing to vulnerability. Long-duration longitudinal studies are needed.

When meeting a patient who suffers from insomnia that may be stress-related, the clinician may find it useful to conduct a structured interview regarding the patient's work and personal life situations that may be stressful. In our experience, much information is gained from questions that focus on preoccupation with work, ruminations, and worries about the immediate future. Such a question may be phrased more specifically: "Are you often unable to stop thinking about work in the evening?"

Another area of exploration with patients is work pressure. Inquiries may be phrased variously, as follows: "Do you have to exert a lot of effort at work?" "Do you get enough time for sleep?" "Do you reduce sleep duration to make room for work?"

Specific treatment of sleep problems due to stress is covered in Chapter 79, on psychological treatment for sleep disturbances. It might also be useful, however, to refer the patient to a short stress management course (based on cognitive therapy), which often is effective for reducing sleep problems and may be combined with cognitive behavior therapy for sleep disorders. The stress management courses usually involve stress reduction techniques, teaching basic life hygiene, and relaxation techniques to improve bedtime practices. Patients with work-related occupational burnout may benefit from sleep treatment based on cognitive techniques (see Chapter 79).

SUMMARY

Long-term exposure to stress is viewed as a major cause of insomnia. Results from cross-sectional studies consistently support this view and demonstrate close relationships between reported occupational stress and impaired sleep. Prospective studies are rare, however, but results from the few that are available indicate that new cases of insomnia follow increases in previously already high work demands. Polysomnographic findings reveal minor effects, such as slightly increased sleep latency or slightly decreased sleep efficiency, after periods of increased day-to-day stress. The level of stress in these studies usually is rather modest, however. One key factor seems to be

worries (rumination) at bedtime, often related to upcoming difficulties during the subsequent day. In the few PSG-based studies of long-term exposure to work stress, results indicate that sleep is characterized by pronounced sleep fragmentation, increased sleep latency, reduced sleep efficiency, and reduced slow wave sleep. Also, subjective sleep self-assessments indicate major sleep problems. An interesting link between sleep loss and stress is that they exert similar endocrine and metabolic effects. Both involve increased levels of cortisol, lipids, and insulin resistance, for example. Furthermore, typical stress-related diseases such as cardiovascular disease or type 2 diabetes have been linked to previous sleep disturbance. Although correlational evidence that stress affects sleep negatively is substantial, little is known about the direction of the effects (which are likely to be bidirectional), or regarding details such as the critical levels and durations of stress necessary to cause insomnia and the process of the development of stress-related insomnia across time.

CLINICAL PEARL

Insomnia often is a result of long-term exposure to stress. Even modest amounts of day-to-day variation in stress may significantly affect sleep architecture. Patients with complaints of incapacitating fatigue (e.g., occupational burnout) should undergo appropriate investigation for disturbed sleep and long-term exposure to stress. From the standpoint of prevention, as well as treatment, it is important to recognize that impaired sleep results from *anticipatory* stress. Effective stress management techniques, applied to address and correct difficult work and general life issues, may even improve sleep.

Selected Readings

Åkerstedt T, Garefelt J, Richter A, et al. Work and sleep—a prospective study of psychosocial work factors, physical work factors, and work scheduling. *Sleep* 2015;**38**:1129–36.

de Lange AH, Kompier MA, Taris TW, et al. A hard day's night: a longitudinal study on the relationships among job demands and job control, sleep quality and fatigue. *J Sleep Res* 2009;**18**:374–83.

Ekstedt M, Soderstrom M, Akerstedt T. Sleep physiology in recovery from burnout. *Biol Psychol* 2009;**82**:267–73.

Querstret D, Cropley M. Exploring the relationship between work-related rumination, sleep quality, and work-related fatigue. *J Occup Health Psychol* 2012;**17**:341–53.

Van Laethem M, Beckers DG, Kompier MA, et al. Psychosocial work characteristics and sleep quality: a systematic review of longitudinal and intervention research. *Scand J Work Environ Health* 2013;**39**:535–49.

Vandekerckhove M, Weiss R, Schotte C, et al. The role of presleep negative emotion in sleep physiology. *Psychophysiology* 2011;**48**:1738–44.

A complete reference list can be found online at ExpertConsult.com.

Chapter
78

Optimizing Shift Scheduling

Göran Kecklund; Mikael Sallinen; John Axelsson

Chapter Highlights

- Sleep loss, transient insomnia, and excessive sleepiness are common consequences of shift work, in particular if the shift system involves night work. Among the organizational remedies, shift scheduling plays a key role. According to sleep-wake regulation theory, a schedule that (1) minimizes circadian disruptions and an accumulation of sleep loss work periods and (2) permits sufficient time for recovery during days off will maintain alertness and sleep quality, which in turn will be beneficial for occupational safety and long-term health.
- Operational requirements of companies and industries differ. The design of a shift schedule includes a number of components that can be adapted to meet different requirements in an optimal manner. The organizational elements of a shift system of greatest importance relative to sleep and sleepiness are (1) the duration and timing of shifts, (2) the number of consecutive shifts, (3) the distribution of days off in the schedule, and (4) the length of off-duty time between two consecutive shifts and shift cycles. An additional important aspect of system design is the choice between a permanent (fixed) and a rotating shift schedule.
- This chapter summarizes the current knowledge on the role of organizational elements—most notably, shift schedule characteristics—in determining the various factors affecting sleep and sleepiness and (when data are available) the work performance of shift workers. On the basis of this knowledge, general recommendations for organizing shifts to optimize sleep and on-the-job alertness are presented.

As discussed in Chapter 75, insomnia and excessive sleepiness are common consequences of shift work, in particular if the shift system involves night work.[1] The primary causes of these symptoms are (1) the mismatch between the internal circadian clock and the usual timing of sleep and wakefulness that is imposed by a shift schedule and (2) an altered ratio between sleep and wakefulness within a 24-hour or longer period. In night work, changes in the usual relationships between circadian and homeostatic processes of sleep and sleepiness are often most dramatic, as the worker is required to stay awake and alert during the biologic night and try to sleep during the biologic day.

Because disturbed sleep and consequent increased sleepiness pose substantial health and safety hazards, numerous attempts have been made to develop effective and feasible remedies for these problems in the context of shift work (Table 78-1). Generally speaking, three levels of remedies are in use: regulatory (e.g., hours of service regulations), organizational (e.g., a fatigue risk management system), and individual (e.g., napping strategies). These remedies constitute important elements of health and safety management in shift work.[2] *Regulatory* remedies are inflexible and not under the control of a given operational entity and are not considered further here. Some aspects of hours of service regulations are discussed in Chapter 73, on fatigue risk management. *Individual* remedies are highly flexible but also not under the direct control of a given operational entity; these are discussed in Chapter 75. *Organizational* remedies can be adapted to meet the unique requirements of a given operational entity and constitute the basic content for this chapter.

Among the organizational remedies, shift scheduling plays a key role. According to Borbely's two-process model of sleep-wake regulation (see Chapter 72), a schedule that (1) minimizes circadian misalignment, (2) limits, to the exent possible, buildup of sleep loss over a single shift and across a shift cycle, and (3) permits sufficient recovery during days off will best maintain alertness and allow for sufficient sleep. In turn, alertness and sleep are necessary for occupational safety and long-term health.

Because the operational requirements of companies and industries often significantly differ from each other, a wide range of shift systems have been devised. The design of a shift schedule includes a number of components that can be adapted to meet specific company/industry requirements in an optimal manner. Adaptable elements of a shift system that are of importance in addressing sleep and sleepiness include (1) the duration and timing of shifts, (2) the number of consecutive shifts, (3) the distribution of days off in the schedule, and (4) the length of off-duty time between two consecutive shifts and shift cycles. In addition, either a permanent (fixed) or a rotating shift schedule can be used. If work shifts rotate between different shift types (e.g., morning, evening, and night shifts) within a certain period of time, a number of additional issues, such as the regularity, speed, and direction

Table 78-1 Remedies Related to Health and Safety Management in Shift Work

Regulatory Remedies	Organizational Remedies	Individual Remedies (Countermeasures)
Hours of service regulations (e.g., limiting maximum work hours, imposing minimum number of off-duty breaks)	Fatigue (health) risk management systems	Napping
Occupational health and safety regulations	Shift scheduling	Bright light therapy
	Education/training programs	Hypnotics (to improve sleep)
		Melatonin

of the shift rotation, also need to be addressed. These issues are discussed next.

PERMANENT VERSUS ROTATING SHIFT SYSTEMS

One of the cornerstones in the design of a shift system is whether the schedule should be based on rotation or on permanent fixed shifts. In a *rotating* shift system, the worker alternates between various work shifts (e.g., morning, evening, and night shifts). An alternative to rotation is *permanent* shifts, in which the workers are divided into groups that work only one type of shift (e.g., permanent morning, permanent evening, and permanent night shifts).

From a theoretical standpoint, a shift system based on permanent shifts should facilitate circadian adaptation (e.g., to the night shift), which manifests as reduced sleep problems, reduced on-the-job sleepiness, and reduced likelihood of poor job performance.[3] The discussion of pros and cons for permanent versus rotating shift systems has focused on night work. Some experts claim that rapidly rotating shift systems (alternating among morning, evening, and night shifts) should be abolished because of the problems they create for sleep, sleepiness, safety, and health stemming from lack of circadian adjustment.[3] A compromise option is to minimize rotation and to alternate only between evening and night work, or between morning and evening work, because these compromises should produce less circadian misalignment between the internal circadian clock and the forced sleep-activity cycle.[3]

To date, few controlled intervention studies have been conducted to investigate how the change from a rotating shift system to a permanent shift system affects sleep and sleepiness.[4] However, results from two intervention studies on police officers showed that removal of rotation reduced self-reported sleep disturbances.[5,6] Turning to studies using an observational (cross-sectional) design, more research is available. Several investigators have compared rotating shift workers with permanent night workers. In a meta-analysis, Pilcher and colleagues estimated that the average (daytime) sleep length of permanent night workers was 6.6 hours, which was slightly longer than that of workers in rapidly rotating shift systems (approximately 6.5 hours).[7] Permanent day and evening workers obtain the most sleep (more than 7 hours). In several studies sleep was measured using objective methods. In one study, actigraphs were used to compare sleep amounts in permanent night workers with their day and evening counterparts.[8] Results showed that permanent night workers obtained the least sleep (4.7 hours, compared with 5.7 and 6.1 hours for the day and evening shifts, respectively) and reported poorer sleep quality than the other shift groups. Studies using polysomnography are rare, but Dahlgren showed that permanent night workers obtained slightly better sleep (less wake time after the onset of sleep) than rotating night shift workers. However, sleep duration in permanent night workers was still quite short (5.5 hours) and only marginally longer (10 to 15 minutes) than that of rotating shift workers.[9] Thus permanent night workers are not likely to obtain more sleep than regular three-shift workers when the entire shift cycle is taken into account (i.e., also morning and evening shifts and days off), as suggested by Folkard.[10]

Table 78-2 Prevalence of Insomnia Symptoms on Each Type of Shift (Day, Evening, Night, and Off-Duty) for Two Different Shift Schedules*

	Symptom Prevalence	
Insomnia by Shift Type	Rotating Three-Shift Group	Permanent Night Group
Day shift insomnia	35.6% (× 7 days = 249.2)	—
Evening shift insomnia	19.8% (× 7 days = 138.6)	—
Night shift insomnia	67.7% (× 7 days = 473.9)	41.7% (× 21 days = 875.7)
Rest-day (off-duty) insomnia	3.6% (× 7 days = 25.2)	11% (× 7 days = 77)
(Sum score/28 days) = Hypothetical mean over shift cycle	(886.9/28) = 31.7%	(952.7/28) = 34.0%

*The hypothetical mean over a 4-week (28-day) shift cycle was calculated. The latter calculation was generated by multiplying means for each shift by 7.
Data from Flo E, Pallesen S, Åkerstedt T, et al. Shift-related sleep problems vary according to work schedule. *Occup Environ Med* 2013;70:238–45.

In several studies, investigators have compared sleep quality between permanent night workers and rotating night workers. Table 78-2 summarizes the prevalence of insomnia-related sleep complaints associated with different shifts among permanent night shift workers versus three-shift workers.[11] Higher rates of insomnia were reported in the rotating three-shift group during night shifts (68%) than in the permanent night shift group (42%).[11] However, a greater percentage of the permanent night shift group reported rest-day (day off) insomnia (11%) compared with the rotating three-shift group (4%). The percentage of three-shift workers reporting insomnia during day shifts was 36%, compared with 20% for evening shifts. We have calculated a hypothetical average insomnia score over the shift cycle for the rotating three-shift group and the permanent night group (see Table 78-2). The "mean over shift cycle" insomnia score for the three-shift group (31.7%) was slightly lower than that calculated for the permanent night shift group (34.0%).

The observation that permanent night shift workers reported more sleep problems during days off (rest-day insomnia; see Table 78-2) has been noted in several studies.[12,13] Most permanent night shift workers revert to day schedules and nocturnal sleep on their days off. It has been shown that readaptation to sleeping at night is difficult and creates sleep problems if the circadian rhythm has adjusted (phase-delayed) to working at night.[14] On the other hand, results from other studies show that permanent night shift workers might be no worse off or even slightly less sleepy during the night than rotating shift night workers.[11,13] However, the levels of on-shift sleepiness reported by permanent night workers were higher than those reported by day and evening workers, suggesting that the circadian rhythm did not adapt to working at night. In the few studies comparing permanent night shft workers with rotating shift night workers with respect to cognitive and psychomotor performance, the results did not show significant group differences.[15,16]

Folkard summarized results from six studies related to circadian adaptation to permanent night work and calculated that only a small percentage of workers (approximately 3%) seem to show complete circadian adaptation (as indexed by the endogenous melatonin rhythm).[17] Partial circadian adaptation in permanent night shift workers is more common and might explain the small increase in daytime sleep (approximately 30 minutes) obtained by that group compared with rotating shift workers. However, longer daytime sleep durations also might be explained by a later circadian phase in permanent night workers. Permanent night shift workers are in many cases a highly self-selected group—that is, persons who have chosen such a shift system because they prefer to work at night and can sleep during the day. It is not likely that self-selection is going to solve the problem with night work, because few people prefer to work only at night because it fits their sleep-wake rhythm. Also, the choice of permanent night work may be based on social factors (e.g., the goal of obtaining more time with children).[18]

In summary, the differences in (1) the duration of daytime sleep, (2) insomnia over the entire shift schedule, and (3) nighttime sleepiness between permanent and rotating shift night workers seems to be small. These results indicate that permanent night shift workers suffer from circadian disruption, as do rotating shift workers during their night shifts. Thus no clear evidence exists showing that permanent night work is more advantageous than the rotating shift system. However, permanent night work may be a good solution for persons who have a late circadian phase and who can maintain a night-oriented rhythm during days off.

SPEED AND DIRECTION OF SHIFT ROTATION—NUMBER OF CONSECUTIVE SHIFTS

The speed and direction of shift rotation vary considerably between shift schedules. The speed of rotation refers to the number of consecutive shifts that are at the same time of day (e.g., morning) before rotating to the next shift. A shift schedule including only one to three consecutive shifts of the same type (e.g., two consecutive morning shifts before the ensuing two evening and night shifts) is considered to be a fast-rotating shift schedule, whereas a schedule including at least five consecutive shifts of the same type is considered to be a slowly rotating schedule.

The direction of shift rotation can be either counterclockwise or clockwise. *Counterclockwise rotation* (also referred to as backward or advancing rotation) means that shift start time advances as the worker changes from one shift type to the next one within a shift cycle (i.e., night to evening to day shift). *Clockwise rotation* (also referred to as forward or delaying rotation) means that shift start time delays as the worker changes to the next shift type within a shift cycle (i.e., day to evening to night shift). An example of counterclockwise rotation is seen in a shift cycle consisting of three night shifts, two days off, three evening shifts, two days off, and three morning shifts, in that order. An example of clockwise rotation is that of a shift cycle starting with morning shifts followed by evening shifts and then night shifts, in that order. Regarding the human circadian system, clockwise rotation can be considered preferable to counterclockwise, because for most adults, it is easier to extend subjective days (i.e., delay bedtime) than to shorten them (i.e., advance bedtime) (see chronobiology chapters, in Section 5).

In principle, both fast- and slow-rotating shift cycles have their advantages and disadvantages. Fast rotation means that early-morning shifts and night shifts (both of which are associated with sleep disruption and increased sleepiness[11,19]) do not occur more than three times in a row, so a significant buildup of sleep loss and sleepiness can be mitigated (which is not the case with slow shift rotation). On the other hand, slowly rotating shift cycles could be considered advantageous, particularly if the circadian rhythm of the worker showed significant adaptation to the work-sleep patterns required by the working hours. In relation to early-morning work, adaptation would mean a circadian phase advance; in relation to night shift, adaptation would mean a circadian phase delay.

Previous findings regarding the effects of the speed of shift rotation on sleep and sleepiness are mixed. Results from a meta-analysis of data from 36 studies showed that slowly rotating shift cycles were preferable to fast-rotating ones with regard to sleep length.[7] The difference between slow- and fast-rotating shift cycles was most pronounced during night shifts: Persons who worked on slowly rotating shift cycles obtained, on average, 1 hour and 18 minutes more self-reported sleep than those who worked on fast-rotating shift cycles. For morning and evening shifts, the differences between fast and slow shift rotations were marginal. Of note, the meta-analysis did not take direction of shift rotation into account.

In a later systematic review including 26 studies on organizational interventions, the findings regarding rotation speed were in disagreement with those in the aforementioned meta-analysis.[4] A change from slowly rotating shift cycles to fast-rotating ones was found to most consistently improve self-assessed sleep quality. Self-assessed sleepiness also was reduced to some extent.

A specific issue associated with the speed of shift rotation is the number of consecutive work nights, because sleep-wake disturbances are known to be most pronounced during night work.[11] Working several consecutive night shifts in a row has been associated with increased risk for occupational accidents and injuries.[20,21] In comparison with the first night shift, risk for occupational injuries and accidents was found to increase by 36% on the fourth night shift in a row.[20] In nurses, having worked one- to two-night shifts in the past 7 days was associated with a 1.26-fold increased risk for occupational injuries, whereas having worked three- to six-night shifts was

associated with a 2.90-fold increased risk.[21] Taken together, these results suggest that the maximum number of consecutive night shifts should be limited to three.

On the other hand, results from other studies do not support the view of limiting the number of consecutive night shifts to fewer than four. Results from a study of rotating shift workers found that those who worked four consecutive night shifts showed levels of cognitive functioning, as measured by tests tapping executive function and automatic and controlled attention processing (e.g., Digit Symbol Substitution Test, Symbol Searching Test, Wisconsin Card Sorting Test), either the same as or better than those in nurses who had worked only two consecutive night shifts.[22] In line with this field study of actual shift workers, results from a laboratory-based study of healthy, non–shift-working males and females showed that subjective sleepiness reached its peak on the first simulated night shift and then gradually diminished during the ensuing four night shifts.[23] A similar trend was observed for some of the performance measures.

Finally, the consequences of working several night shifts in a row also differ according to specific circumstances such as location of operations. The specific situation of working at remote locations (such as off-shore installations and mines) is discussed separately later in the chapter (see Shift Scheduling in Extended, Offshore, and Mining Operations).

With regard to the direction of shift rotation, the current findings are more consistent than those for the speed of rotation. In two published systematic reviews, the investigators concluded that clockwise-rotating shift cycles were preferable over counterclockwise-rotating ones.[4,24] These conclusions were based on studies comparing the results of self- and actigraphy-assessed sleep, subjective sleepiness, and vigilance performance between these two shift cycles.

In practice, improvements due solely to changing from a counterclockwise to clockwise shift rotation are difficult to estimate, because the change in the direction of rotation often is accompanied by changes in other shift characteristics (e.g., speed of rotation, shift length, time-off period).[24,25] In a narrative review of the results of observational and intervention studies regarding the speed and rotation of shift cycles, the investigators reported that especially fast- and clockwise-rotating shift cycles were associated with better sleep and on-the-job alertness than in slow- and counterclockwise-rotating cycles.[25] The reasons for this difference between "fast plus clockwise" versus "slow plus counterclockwise" may lie not only in the speed and direction of rotation themselves but also in time-off periods between two consecutive shifts. "Slow plus counterclockwise" rotating shift cycles often are characterized by short, 7- to 9-hour time-off periods (e.g., in connection with a combination of an evening shift followed by a morning shift), whereas "fast plus clockwise" rotating cycles are always characterized by at least 11 to 12 hours of free time.

To conclude, fast- and clockwise-rotating shift cycles seem to be an optimal choice in terms of sleep between shifts and alertness at work, but supporting evidence is limited. In addition, individual difference factors (such as age, gender, diurnal type, and personality characteristics) contribute to adaptation of the sleep-wake pattern to rotating shift schedules.[26,27] In practice, this means that fast and clockwise rotation would not be optimal for everyone because considerable interindividual differences in response to changes in the speed and direction of shift rotation are likely.

OFF-DUTY TIMES BETWEEN SHIFTS

The off-duty time between shifts varies considerably, ranging from the usual 15 to 16 hours off-duty between consecutive 8-hour shifts to 12 hours off-duty between consecutive 12-hour shifts, to only 8 to 9 hours off-duty between work shifts requiring "quick returns" (e.g., when an evening shift is followed by a morning shift) or in irregular/unpredictable work environment systems. More extreme split shift systems/watch-keeping systems may be characterized by 4, 6, or 8 hours off-duty between shifts.[28] The reasons for reducing off-duty time between shifts are many, ranging from compressing the working period (e.g., finishing a series of night shifts with an afternoon shift) to allowing for continuity between staff and patients (i.e., so that patients have the same staff caring for them in the evening and on the following morning).

A number of aspects affect the sleep duration during off-duty time between work shifts. Besides sleep, the workers must commute to and from work, eat meals, unwind, maintain hygiene, and carry out other household and social responsibilities. As discussed previously regarding night shift work, timing of the off-duty period also is critical because a worker is less likely to obtain sufficient sleep if the off-duty time occurs during the day as opposed to nighttime.

Although results from some studies indicate that 12 hours of off-duty time may be sufficient for many workers,[29,30] other results indicate that 12 hours may be too little and allow for obtaining less than 6 hours of sleep.[31] Accordingly, most authors recommend avoiding off-duty times of less than 12 hours in general, and off-duty times less than 10 hours (quick returns) in particular.[32] The main reasons for the conflicting findings on sleep duration are likely related to different commute/travel times, extent of social obligations, and poor versus optimal sleep conditions. Hence, an employer using schedules with off-duty times of less than 12 hours should arrange for workers to sleep on site or close by in a sleep-conducive environment. Good sleeping conditions, short travel time, and few social obligations may explain why 12-hour shifts allow for sufficient sleep duration among oil rig workers.[30]

Quick returns (less than 11 hours of off-duty time between shifts) are used as a transition between shift types (e.g., as a transition from a series of night shifts to a period of morning shifts). This quick return compresses working hours and therefore results in extra off-duty time at some point. The general finding is that quick returns reduce average sleep durations to between 4 and 6 hours.[25,33] A quick return from a morning shift to a subsequent night shift is most disruptive, resulting in an average sleep duration of 3 hours or less.[34]

Results from several questionnaire studies show that severe sleepiness is more common in schedules incorporating quick returns. Most of these studies were carried out among nurses, for whom a quick return between the afternoon shift and a subsequent morning shift is particularly common.[35,36] By contrast, results from studies in which the investigators have attempted to predict fatigue often failed to find quick returns to be a strong predictor of fatigue.[37] Some evidence indicates that quick returns increase the risk for accidents in steel manufacturing workers.[38]

To conclude, short off-duty times have a negative impact on sleep duration. Off-duty times of fewer than 12 hours typically reduce sleep duration to 4 to 6 hours, with more reductions in sleep duration if the off-duty period occurs

during the day. Many individual and work site–related factors, such as travel times and accommodations for sleeping on site, affect sleep duration. Inasmuch as short sleep duration increases fatigue and impairs cognitive functions, off-duty periods between shifts that are shorter than 12 hours in duration should be avoided.

SHIFT CHANGEOVER TIMES AND START TIME OF THE MORNING SHIFT

One challenge in shift design is to find the optimal changeover time—that is, the time when night shift workers go off-duty and morning shift workers come on-duty. On the one hand, an early changeover time (e.g., at 5:00 AM) might result in less circadian disruption and increased daytime sleep duration for the night shift workers. On the other hand, an early start for morning shift workers is often associated with decreased sleep length, impaired sleep quality, and increased on-shift sleepiness.[39] For example, results from a large epidemiologic study showed that sleep duration was less than 5 hours when the morning shift started between 3:00 AM and 4:30 AM; early-morning shifts were also associated with increased risk for (self-reported and nonspecific) accidents.[40]

The effect of changeover time on sleep and sleepiness in three-shift work was examined in a controlled intervention study carried out by Rosa and coworkers.[41] The intervention involved a 1-hour delay of the shift changeover time: The shift changes occurred at 7:00 AM, 3:00 PM, and 11:00 PM on the new schedule. After the change in shift changeover time, participants extended their (self-reported) sleep by 0.5 hour on morning shifts (6.2 hours before the change). No significant changes in sleep duration occurred on evening shift (7.2 hours) or night shift (5.8 hours) days. Subjective sleep quality improved on morning shift days and worsened to some extent on night and evening shift days. Self-rated sleepiness decreased during the morning shift and increased on the night shift. In a control work site where shift changeover times remained unchanged, no significant changes occurred in sleep or alertness. In all, the results suggest that minor changes in shift start and end times are important factors affecting sleep and sleepiness when such changes occur within a normal night sleep period. Similar results were obtained in another study in which changes in shift changeover times were compared.[42]

The negative effects of early start times for morning shifts have been demonstrated in several studies. Results from studies in which polysomnography was used to objectively document sleep-wake have shown that with early-morning work (defined as start times at 6:30 AM or earlier), sleep duration is only slightly longer than 5 hours.[43,44] Ingre and colleagues carried out an experimental study in train drivers using a within-subject design in which they compared three different start times of the morning shift: 5:49 AM, 7:49 AM, and 9:49 AM).[39] The results demonstrated strong effects of start time on early-morning work in irregular shift systems—on average, drivers obtained only 5.6 hours of sleep before the earliest shift start time (5:49 AM). By contrast, drivers obtained 7.7 hours of sleep before the latest shift start time (9:49 AM). Results also showed that the earliest shift resulted in higher degree of subjective sleepiness, and 82% of the drivers reported at least one period of severe sleepiness (defined as a Karolinska Sleepiness Scale score of 7 or greater) during the work shift.[39] The corresponding value for the latest shift was 53%. The observation that an early start time for the morning shift is associated with increased sleepiness also has been noted in several other studies.[45]

To summarize, the shift changeover time between night and morning shifts affects sleep duration and sleepiness. A later changeover time is beneficial for the morning shift but detrimental for the night shift. This situation illustrates the difficulty in designing a shift system to optimize sleep and alertness for all shifts because there does not seem to be an optimal changeover time between the night and the morning shift that does not negatively affect one or the other.

LONG WORK SHIFTS AND ON-CALL WORK

Long work shifts (defined as shifts longer than 10 hours) are used in a wide range of settings and are common in the medical field (e.g., among nurses and medical interns). Although results from several studies show that 12-hour work shifts reduce opportunities to obtain sufficient sleep between shifts,[31] one report indicated that oil rig workers can manage up to 14 consecutive 12-hour shifts with relatively few problems.[30] In this report, the oil rig workers displayed the worst sleep problems during the sleep period before the first night shift but thereafter were able to achieve an average of 6.5 to 7.0 hours of sleep per off-duty period, with a gradual reduction of sleepiness during the subsequent night shifts. These findings indicate that it is possible to manage 12-hour shifts and that many workers can adapt when conditions are optimal: The oil rig workers slept on site, with few social obligations, and their light exposure allowed for circadian adaptation.

In contrast to findings in oil rig workers, results from studies of 12-hour shifts in other occupations are mixed: Both poorer sleep and better sleep have been observed when 12-hour shift systems were compared with shift systems based on 8-hour shifts.[46] One factor that may explain the inconsistent findings is different starting times.[46] Also, workers may have more free days to recover between work periods in a 12-hour shift system and thus may be more motivated to obtain sufficient sleep before the shifts. A negative aspect of schedules with long work shifts is the difficulty in taking prophylactic daytime naps before the night shift.

According to some evidence, long shifts increase severe sleepiness and fatigue,[47,48] although results of several studies failed to show any differences in sleepiness or fatigue between 8-hour shifts and 12-hour work shifts.[29,46] Results from several studies support the notion that 12-hour shift systems may reduce severe sleepiness, suggesting that a good 12-hour shift system is better than a poor 8-hour shift system.[49] The evidence also indicates that fatigue accumulates over successive long work shifts, particularly when the shifts are as long as 15.5 hours.[48] The study authors argued that a main reason for the accumulation of fatigue was the short off-duty time (8.5 hours) between shifts (allowing workers to obtain only 5.5 hours of sleep) in a schedule with several concomitant long shifts.[48] In a study of 104 railway traffic controllers, the risk for severe sleepiness increased in long shifts, the increase being 15% for every hour on the shift.[37]

On-call work is a scheduling approach used to provide 24-hour coverage of personnel to deal with emergencies in a prompt manner when normal staff resources are not present (e.g., at night). The main reason for this working arrangement is that having employees on call is a more effective (i.e., less

costly) use of staff than maintaining full coverage during off-peak hours. This working arrangement is common in a wide range of occupations including the medical field (e.g., physicians, midwives, medical technologists), utilities (e.g., electrical technicians), and other fields in which full coverage over a 24-hour period would be cost-prohibitive (e.g., ship engineers, tugboat pilots). The characteristics of on-call shifts in these fields vary considerably and particularly with respect to whether the employee must be on site (as in the case for medical residents) or may remain off site (e.g., on-call utility workers).

The limited literature suggests that on-call work is associated with disturbed sleep.[50,51] On-call workers report experiencing difficulties falling asleep and reduced sleep amounts (as reported by 198 locomotive engineers[51]). Few researchers have investigated how on-call affects polysomnographic sleep, but in a small-scale study of ship engineers ($n = 5$), the sleep EEG showed less slow wave sleep and less REM sleep, associated with a higher heart rate, all of which might have been caused by the anticipation of having to wake up and work.[50] Sleep periods during on-call shifts typically are short but even within the same occupation may differ considerably between work sites: Actigraphically recorded sleep durations reported among emergency medical specialties in France averaged 4.75 hours on on-call shifts,[52] whereas medical residents in Chicago averaged 2.8 hours[53] and medical residents in Houston averaged 3.8 hours.[54] A strong inverse relationship has been found between the amount of working hours across a work week with extended shifts and the amount of sleep the residents obtain.[55]

Important evidence shows that long working hours also increase the risk of on-the-job accidents. The authors of a systematic review of work shift duration and accidents concluded that the increased risk is approximately doubled after approximately 12 hours of work, compared with the risk after 8 hours of work.[56] Landrigan and coworkers[57] carried out a randomized intervention study in medical interns working in two intensive care units. One unit kept the traditional schedule (which included several shifts lasting 24 hours or longer), and the other unit was assigned to an intervention schedule that eliminated the extremely long shifts, so that the longest shifts were approximately 16 hours. The objective was to determine whether the intervention schedule increased the interns' opportunities to sleep. Diagnostic error rates were significantly higher for the traditional schedule than for the intervention schedule (24 ± 18.6 versus 3 ± 3.3). Moreover, the rate of serious medical errors was higher for the traditional schedule as well (176 ± 136.0 versus 91 ± 100.1). The intervention schedule still contained several 16-hour shifts and 60 to 63 hours of scheduled work hours per week (plus overtime). Additional evidence suggests that the commute home from work is characterized by an increased risk of falling asleep and for being involved in a vehicle accident on schedules with extended shifts.[58]

In sum, long work shifts and on-call work are associated with increased sleepiness and an increased risk for accidents and errors on shift and during the commute home. Moderately long work shifts (i.e., 12-hour shifts) may not necessarily reduce sleep duration—and the ability to obtain approximately 6.5 to 7 hours of sleep in between successive shifts appears to be the key aspect allowing oil rig workers to manage long work periods with few fatigue problems. Shift timing is crucial for long working shifts, and early start times are to be avoided because the combination of short sleep and long working periods may lead to an accumulation of sleepiness. On-call work among residents disrupts sleep and may cause severe sleep loss if the weekly working hours are extended. Finally, in safety-critical operations, long work shifts are associated with fatigue, serious medical errors, and increased accident risk after approximately 8 to 9 hours on duty.

COUNTERMEASURES AND SHIFT SCHEDULING

On-Shift Napping

A countermeasure for reducing sleepiness in shift work (and especially during night shifts) is to implement rest breaks with an option for sleeping. Results from studies conducted in medical staff,[59] air traffic controllers,[60] and industrial workers[61] show that a nap (even for less than an hour) can reduce sleepiness on night and extended shifts. Results from a systematic review of data from 13 laboratory and field-based studies showed that napping during night shifts consistently reduced sleepiness and improved performance but was not likely to impair post-shift daytime sleep.[62] In these studies, nap start time varied, ranging from midnight to 4:00 AM, with nap break duration of 20 to 120 minutes and nap sleep duration of 17 to 117 minutes. In all, these findings propose that preplanned nap breaks constitute a useful sleepiness countermeasure for addressing sleepiness in shift work. The only putative negative consequence of taking a nap is sleep inertia (which lasts only 15 to 20 minutes). In general, the short-lived impairment of sleep inertia probably is outweighed by the larger, long-term benefit associated with nap sleep; especially in safety-critical occupations (e.g., airline pilot), it is important to schedule enough time for the process of awakening after a nap break to avoid sleepiness-related performance errors.[63]

Prophylactic (Pre-Shift) Napping

Another countermeasure for sleepiness during the night shift is to nap before the shift (and particularly before the first night shift), because otherwise the duration of sustained wakefulness will be approximately 24 hours by the end of the first night shift. Results from a study of policemen working on fast-rotating shift cycles showed that taking a nap before the night shift (which spanned 1:00 AM to 7:00 AM) was associated with a 48% reduction in on-duty car accidents.[64] The length of the nap sleep was approximately 1.5 hours, and the nap occurred either in the afternoon or closer to the beginning of the night shift.

Shift Scheduling in Extended, Offshore, and Mining Operations

The relationship of the shift schedules used at North Sea oil and gas installations to workers' sleep, sleepiness, and performance has been studied extensively. Two widely used shift schedules for a 2-week offshore work period (usually followed by 2 to 4 weeks of free time) are the fixed schedule and the swing shift rotation. With the *fixed schedule*, workers alternate between offshore tours of 14 consecutive night shifts and 14 consecutive day shifts. The *swing shift rotation* is composed of seven 12-hour night shifts followed immediately by seven 12-hour day shifts.

These extended schedules have different effects on circadian disruption, sleep, and sleepiness, during both the offshore period and during the break when workers return home.

Regarding sleep and sleepiness, the fixed schedule has proved to be less disruptive than the swing schedule.[65] This benefit is explained mainly by the fact that workers show significant or even complete adaptation to night shifts during the first work week—a phenomenon that does not occur in on-shore personnel doing comparable work.[66] Significant adaptation to consecutive night shifts in off-shore workers is mostly explained by only low-level exposure to morning light and a good opportunity for daytime sleep during a shift cycle. Adaptation to night shifts is evidenced by changes in cortisol and melatonin production rhythms and improvements in sleep, sleepiness, and performance.[67,68] The amount of sleep obtained per 24-hour period has been reported to be 6 to 7 hours, independent of whether the schedule consists of 14 consecutive day or night shifts.[69] Of note, a relatively large number of oil rig workers still report problems coping with long stretches of consecutive 12-hour shifts—23% have been classified as suffering from shift work sleep disorder[70] (see Chapter 75). The most important disadvantage of the fixed schedule system is that workers must readapt to a diurnal sleep-wake rhythm after 14 consecutive night shifts. The readaptation process takes at least a week.[67] When working the swing shift schedule, workers must adapt to night shifts and then readapt to a diurnal sleep-wake rhythm during each 2-week work period. This readaptation process has been shown to lead to more circadian disruption and sleep loss than with the fixed shift schedule.[65]

As with offshore oil and gas installations, extended periods of consecutive 12-hour shifts are typical of the mining industry. Investigators of one field-based study examined sleep during four types of shift schedules, differing in terms of (1) the number of consecutive night (6:00 PM to 6:00 AM) and day (6:00 AM to 6:00 PM) shifts (4 to 7), (2) the number of consecutive shifts (4 to 14), and (3) the number of days within a shift cycle (15 to 22).[71] The results showed that the shift schedules or shift types (day versus night) did not differ from each other in terms of self-reported sleep duration or quality on workdays. On average, workers obtained 6 hours of sleep per 24-hour period. The workers slept most (7.5 to 9.5 hours) before the first night shift on all schedules, which suggests that they prepared themselves for the upcoming night shifts and also, for some schedules, tried to recover as completely as possible from the preceding sequence of consecutive 12-hour shifts.

Results from another study in miners showed that in contrast with the data for offshore workers, physiologic adaptation to seven consecutive night shifts was limited.[72] This lack of adaptation also was reflected in progressive impairment in psychomotor performance over consecutive shifts. Taken together, the findings indicate that the effects of similar shift schedules on sleep are largely dependent on the prevailing working environment.

WHAT IS THE OPTIMAL SHIFT SYSTEM DESIGN FOR REDUCING SLEEP-WAKE PROBLEMS?

The recommendations for shift schedule designs that are beneficial for sleep and on-the job alertness are listed in Table 78-3. A shift schedule that (1) is fast and clockwise-rotating, (2) limits consecutive working days in a row, (3) eliminates extremely long (16 hours or longer) shifts, and (4) spreads out days off across the schedule decreases the risk of accumulated sleep loss and fatigue and minimizes circadian disruption.[73] If work is safety-critical, long shifts (lasting more than 10 hours should be avoided. Nevertheless, occurrence of acute sleepiness (e.g., toward the end of the night shift) is highly likely, because no circadian adaptation occurs on a fast rotation. Use of sleepiness countermeasures (e.g., on-shift or pre-shift napping) is therefore important to avoid severe work-related sleepiness. Of note, permanent night work rarely results in complete circadian adaptation, and the partial adjustment observed in some field studies seems to be due to sleep problems during days off.

Also of note, the scientific evidence supporting several of the shift design recommendations is insufficient, so definitive recommendations on how a shift system should be designed universally are few.[24] The lack of definitive shift scheduling recommendations is partly because of the difficulties of introducing a randomized controlled study designs in shift work settings. In addition, the shift workers themselves often participate in the development of the new shift system, so the results from some of the intervention studies might be related more to the shift workers' expectations and attitudes about the new schedule, rather than to the specific components of the shift system. Moreover, sleep and sleepiness are influenced by factors other than working hours; the internal validity of field studies, therefore, usually is lower in comparison with laboratory-based experimental research (on the other hand, laboratory studies often possess low ecologic validity and are not optimal for drawing conclusions regarding shift scheduling). Most field studies use subjective self-reported metrics, and relatively few studies have applied objective methods for measuring sleep and alertness. Furthermore, many of the intervention studies are based on small samples, and follow-up times are short (less than 12 months). These limitations require that the recommendations for shift scheduling be interpreted with some caution. Of note, the recommendations are based on the "one-size-fits-all" assumption. Individual differences in sleep need, circadian phase, and sleep flexibility probably are substantial among shift workers. As a consequence, some workers will encounter problems tolerating fast-rotating shift systems and might prefer a more regular (and fixed) shift system. Also of note, various domains and occupational settings differ significantly in terms of operational requirements and working conditions, which makes it difficult to apply exactly the same shift scheduling recommendations to all cases.

CLINICAL PEARLS

- Sleepiness and impaired sleep associated with shift work are due to (1) failures in circadian adaptation and (2) an accumulated sleep debt. Tolerance to shift work is therefore determined by the design of the shift schedule: A shift schedule that minimizes both circadian disruption and the buildup of sleep loss is easier to tolerate and may improve long-term health and safety.
- An optimal shift schedule should feature fast and clockwise rotation, few consecutive night shifts, at least 11 hours of off-duty time between shifts, no extremely long-duration (16 hours or longer) shifts, and avoidance of early (before 6:00 AM) starting times for the morning shift. Even with an optimal shift system, however, workers are likely to have a high level of sleepiness at the end of the night shift. It is therefore important to incorporate alertness-enhancing countermeasures (e.g., pre-shift or on-shift naps), to avoid sleepiness-related incidents at the job site as well as accidents during the commute home from work.

Table 78-3 Recommendations for Shift Scheduling

Shift Schedule Characteristic	Recommendation	Expected Result
Permanent or rotating shifts	Rotating shift system	Limits repeated circadian disruption and reduces the risk of accumulated sleep loss and build-up of severe sleepiness Less sleep and wake problems during days off
Speed and direction of rotation	Fast (maximum three similar shifts in a row) and clockwise (e.g., morning-evening-night) rotation	Limits circadian disruption and reduces the risk of accumulated sleep loss and buildup of severe sleepiness
Number of consecutive work days	Maximum 6 days in a row, but should be reduced if the length of work shift is >8 hours, the off-duty time between shifts is <11 hours, and if the work shifts include early morning work and night work	Reduces the risk of accumulated sleep loss and buildup of severe sleepiness
Duration of off-duty time between shifts	Minimum off-duty time between shifts is 11 hours	Reduces the risk of accumulated sleep loss and build-up of severe sleepiness
Duration of work shift	Very long (>16 hours) shifts should be eliminated. Shifts longer than 10 hours should be avoided if the work task requires sustained attention, is safety-critical, and does not contain breaks during work	Reduces the risk of accumulated sleepiness within the work shift May also reduce accident risk
Starting time of the shift	The start time of the morning shift should be later than 6:00 AM and preferably around 7:00 AM	Reduces severe sleep loss, insomnia, and sleepiness during the morning shift May cause more disturbed sleep related to the night shift
Distribution of off-duty days within the shift schedule	Days off should be spread out in the shift schedule to avoid compressed work hours	Decreases the risk of developing accumulated sleep loss and buildup of severe sleepiness

Modified from Knauth P, Hornberger S. Preventive and compensatory measures for shift workers. *Occup Med (Lond)* 2003;53:109–16.

SUMMARY

Sleep loss, disturbed sleep, and excessive sleepiness are common consequences of shift work, in particular if the shift system involves night work. Although misalignment between the internal circadian clock and the sleep-work pattern determined by the shift system is a key factor, certain shift scheduling factors also play an important role. According to sleep-wake regulation theory, a schedule that minimizes circadian disruptions and buildup of sleep loss over a single shift and a shift cycle, and permits sufficient recovery during days off, will optimize on-shift alertness and off-duty sleep quality, which in turn benefits occupational safety and long-term health. The recommendations for shift schedule design are to use fast and clockwise-rotating shift system with few consecutive working days in a row, to eliminate very long (16 hours and longer) shifts, and to spread out days off across the schedule. Such a schedule decreases the risk of developing accumulated sleep loss and fatigue and minimizes circadian disruption. One of the most well-studied and effective interventions is to eliminate extremely long (more than 16 hours) work shifts—and if work is safety-critical, long (more than 10 hours) shifts should be avoided. Nevertheless, the occurrence of acute sleepiness (e.g., toward the end of the night shift) is highly likely, because no circadian adaptation occurs on a fast rotation. It is therefore important to use sleepiness countermeasures (e.g., pre-shift and on-the-job napping) to avoid severe work-related sleepiness and consequent errors and accidents.

Selected Readings

Bambra CL, Whitehead MM, Sowden AJ, et al. Shifting schedules: the health effects of reorganizing shift work. *Am J Prev Med* 2008;**34**:427–34.

Bjorvatn B, Stangenes K, Øyane N, et al. Subjective and objective measures of adaptation to night work at an oil rig in the North Sea. *Sleep* 2006;**29**:821–9.

Flo E, Pallesen S, Åkerstedt T, et al. Shift-related sleep problems vary according to work schedule. *Occup Environ Med* 2013;**70**:238–45.

Folkard S. Do permanent night workers show circadian adjustment? A review based on the endogenous melatonin rhythm. *Chronobiol Int* 2008;**25**:215–24.

Ingre M, Kecklund G, Åkerstedt T, et al. Variation in sleepiness during early morning shifts: a mixed model approach to an experimental field study of train drivers. *Chronobiol Int* 2004;**21**:973–90.

Landrigan CP, Rothschild JM, Cronin JW, et al. Effects of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med* 2004;**351**:1838–48.

Pilcher JJ, Lambert BJ, Huffcutt AI. Differential effects of permanent and rotating shifts on self-report sleep length: a meta-analytic review. *Sleep* 2000;**23**:155–63.

Short MA, Agostini A, Lushington K, et al. A systematic review of the sleep, sleepiness, and performance implications of limited wake shift work schedules. *Scand J Work Environ Health* 2015;**41**(5):425–40.

Smith MR, Eastman CI. Shift work: health, performance and safety problems, traditional countermeasures and innovative management strategies to reduce circadian misalignment. *Nat Sci Sleep* 2012;**4**:111–32.

Vedaa Ø, Harris A, Bjorvatn B, et al. Systematic review of the relationship between quick returns in rotating shift work and health-related outcomes. *Ergonomics* 2015;1–14. [Epub ahead of print].

Viitasalo K, Puttonen S, Kuosma E, et al. Shift rotation and age—interactions with sleep-wakefulness and inflammation. *Ergonomics* 2015;**58**:65–74.

A complete reference list can be found online at ExpertConsult.com.

Chapter 79

Obstructive Sleep Apnea in the Workplace

Chunbai Zhang; Mark B. Berger; Albert Rielly; Atul Malhotra; Stefanos N. Kales

Chapter Highlights

- Obstructive sleep apnea (OSA) is highly prevalent among workers and a common cause of excessive daytime sleepiness at work.
- Untreated OSA negatively affects occupational health, safety, and productivity.
- Untreated OSA increases the risk of transportation accidents.
- Clinicians are likely to encounter employees in several situations for evaluation for possible OSA.
- In the occupational setting, subjective symptom reports are unreliable for screening and diagnosis of OSA.

OVERVIEW AND BACKGROUND

Obstructive sleep apnea (OSA) is characterized by repetitive cessations of or decrements in airflow through the upper airway during sleep, resulting in a variety of physiologic and metabolic disturbances, including frequent arousals from sleep.[1] Untreated OSA has been linked to excessive health, safety, and lost productivity costs in the range of $65 billion to $165 billion per year in the United States alone.[2] Motor vehicle accidents and other workplace injuries in safety-sensitive occupations are increased owing to the resulting excessive daytime sleepiness and decreased vigilance/attention associated with OSA.[3-6] In addition to lost productivity and increased absenteeism, individual and public health costs attributable to the cardiovascular and metabolic comorbidity associated with OSA have been well documented.[7,8]

OSA remains underdiagnosed and often goes untreated. Accordingly, sleep specialists and occupational physicians can expect to encounter several typical workplace referral scenarios in which patients or employees with potential OSA will require evaluation. One of the most common scenarios is that in which an employee is observed to be sleeping at work (as with an employee in an unsupervised position that requires close attention to the assigned task) or falling asleep repeatedly during group meetings. Supervisors may mistake such behavior for laziness or may recognize the possibility of a medical problem. In such cases, the employee may agree to be medically assessed or may be in denial regarding his or her impairment. A more urgent referral scenario is that in which the employee is required to present to the sleep clinic after falling asleep and causing an accident during performance of a safety-sensitive job (e.g., professional driver, airline pilot, health care worker, nuclear plant employee, public safety officer). In keeping with the increased public awareness of transportation accident risk associated with OSA, sleep professionals can expect to see occupational medicine professionals and employers refer transportation operator-employees who deny all symptoms of OSA yet are found to have objective risk factors (e.g., obesity, increased neck circumference and/or hypertension). These employees may be referred for evaluation to rule out OSA as a condition of employment or to medically qualify for an operating license. In all of these scenarios, the determination of a diagnosis of OSA and reports of treatment compliance data will have medicolegal relevance for the clinician and potential job security implications for the referred employee.

CONSEQUENCES OF OBSTRUCTIVE SLEEP APNEA IN TRANSPORTATION WORKERS

Experts estimate that between 7%[9] and 20%[10] of all large truck crashes are due to drowsy or fatigued driving.[11] As reported over the past decade (2003-2012), between 3454 and 9528 deaths and between 84,000 and 224,000 serious injuries (mostly among the traveling public) are likely attributable to sleep-related impairment in commercial motor vehicle (CMV) drivers in the United States alone.[9,10,12] OSA is the most common medical cause of excessive daytime sleepiness.[13] Not surprisingly, therefore, several high-profile transportation accidents related to insufficient sleep and OSA have been reported.[14] In one such accident, a pilot of an oil tankship collided with a general cargo vessel, which led to a spill of crude oil. The volume of the spill was estimated at 1000 to 11,000 barrels. The official investigation concluded that contributing to the accident was the pilot's fatigue caused by his untreated obstructive sleep apnea and his work schedule.[15] In 2008, a tour bus carrying passengers returning from a weekend ski trip crashed and killed nine people and injured 43 others. The driver was found to be suffering from OSA that was inadequately treated in the days before the accident. In 2009, two airline pilots on a flight in Hawaii dozed for at least 18 minutes during a midmorning flight, initially overshooting the specified destination. The captain subsequently was diagnosed with OSA. In December 2013, a southbound New York Metro-North Railroad train derailed as a consequence of excessive speed, killing four passengers and injuring 59 others. A diagnostic evaluation of the train's engineer as part of the accident investigation revealed that he was suffering from untreated severe OSA.[16]

PRINCIPLES OF OBSTRUCTIVE SLEEP APNEA MANAGEMENT IN THE WORKPLACE

Occupational medical programs that address OSA screening, diagnosis, and management should be accompanied by administrative controls related to hours of service, shift work, and other factors that affect worker fatigue. Taken together, these factors are referred to as fatigue risk management systems (FRMSs). In 2012, the American College of Occupational and Environmental Medicine (ACOEM) Presidential Task Force on fatigue risk management published a guidance statement to assist in the design and implementation of FRMSs.[17] A successful FRMS should be science-based, data-driven (with decisions based on collection and objective analysis of data), cooperative (designed together by all stakeholders), fully implemented (systemwide use of tools, systems, policies, and procedures), integrated (built into existing corporate safety and health management systems), continuously improved (for progressive reduction of risk using feedback, evaluation, and modification), adequately budgeted (justified by an accurate return-on-investment business model), and "owned" (responsibility accepted by senior corporate leadership).[17] For an FRMS to function efficiently, a senior manager must be assigned accountability for the program. Active engagement from everyone employed will make the program more successful, as will a culture of mutual trust between management and employees. Gander and colleagues provide an in-depth discussion of FRMSs in Chapter 73.

Sleep Disorder Management

The principles of fatigue management apply to OSA management and require a proactive approach rather than a reactive approach. In safety-sensitive positions, minimizing fatigue by actively screening, diagnosing, and managing OSA in advance, with implementation of corrective actions when necessary to address noncompliance with prescribed treatment, is expected to be more cost-effective than responding to a fatigue-related incident after it occurs. For most workplaces, the sleep disorder management program need not be extensive. A screening questionnaire that encourages follow-up evaluation with personal physicians for response profiles suggestive of sleep disorders may be adequate. In safety-sensitive occupations, however, objective physical assessments may be required because they are much more sensitive and reliable screening tools than most self-report questionnaires (as discussed further on).

Screening and Risk Factors

Nonmodifiable risk factors for OSA include older age, male gender, postmenopausal status in females, and ethnicity (Asian American or African American descent). Obesity/adiposity is the most significant risk factor and is modifiable.[18-21] Results from studies have shown that OSA is closely associated with higher body mass index (BMI) and larger neck and waist circumference.[21-24] These findings make the foregoing parameters key objective elements for OSA screening in an occupational setting. Visceral fat is significantly correlated with increasing OSA severity.[23,25-27] Higher waist-to-hip ratio also has been shown in some studies to be more predictive of OSA than is obesity in general. Among morbidly obese patients with a BMI of 40 kg/m^2 or higher, OSA is nearly universal. Among men with BMI of 32 kg/m^2 or higher, the prevalence of OSA is approximately 75%.[28] Persons with large neck circumferences (in men, greater than 17 inches; in women, greater than 16 inches) should raise clinical suspicion for presence of OSA.[29]

Women with OSA tend to report fewer "classic" daytime symptoms of OSA—for example, instead of reporting daytime sleepiness, they may report fatigue and lack of energy. In addition, women have different anatomic and functional upper airway properties and differences in control of breathing compared with men.[30,31] These diagnostic and biologic differences between men and women contribute to lower rates of sleep apnea diagnosis in women.

Among different races, obesity plays a variable degree of importance. For example, adult African Americans younger than 25 years or older than 65 years have higher prevalence of OSA than others.[20,32] In the East Asian population, although the prevalence of obesity is lower, the prevalence of OSA is similar to that for Western populations.[33,34] Ethnic differences in adipose tissue distribution (i.e., peripheral versus visceral) and predisposing craniofacial profiles such as crowded posterior oropharynx, shorter cranial base, and more acute cranial base flexure may be important in explaining the pathogenesis of OSA among certain nonobese populations.[18,30,35]

Finally, smoking and use of alcohol and other sedatives are important modifiable risk factors.

OBSTRUCTIVE SLEEP APNEA SCREENING METHODS IN THE WORKPLACE

Screening is defined in this chapter as risk assessment or stratification before referral for a diagnostic test. Employer screening refers to the use of questionnaires (such as the STOP-Bang questionnaire; for further information, see www.stopbang.ca), anthropometric measures, and other subjective and objective criteria applied to all employees to identify those who should be referred to a sleep disorders specialist, who will then confirm or exclude the diagnosis of OSA. "Diagnosis" and "diagnostic procedures" for OSA refer to sleep studies that measure or estimate the presence of sleep apnea (see later on). OSA screening modalities include (1) subjective/self-identified reports of perceived sleep disorders, daytime sleepiness, and sleep-related symptoms; (2) objective measures such as BMI and neck circumference, with cutoff values, and blood pressure; (3) guidelines that combine subjective and objective criteria; and (4) functional performance screens designed to detect impairment related to fatigue or sleepiness. Table 79-1 summarizes a number of these screening methods regarding characteristics of workplace performance (in this case, transportation operators) and effectiveness in detecting OSA.

Subjective Measures

Subjective screening modalities depend on the individual employee's self-report of previous OSA or daytime sleepiness. However, this approach to OSA screening in an occupational setting creates multiple challenges.[36-39] Unlike in a sleep clinic, where patients with undiagnosed OSA typically are actively seeking diagnosis and treatment for inadequate sleep, snoring, and/or excessive daytime sleepiness, employees in safety-sensitive positions (e.g., truck drivers, pilots, mariners) wish to avoid incurring an OSA diagnosis because of its potential

Table 79-1 Comparison of Various Obstructive Sleep Apnea (OSA) Screening Strategies in a Typical Population of Transportation Operators

	Estimated Performance during Occupational Medical Examinations of Transportation Operators				
Screening Criterion	Prevalence of Positive Screens (%)	OSA Case Yield* (%)	Sensitivity (%)	Positive Predictive Value (%)	Mean AHI in Cases Detected
U.S. Federal Commercial Drivers' License Exam Driver Sleep Question	0-3	0-2	0-7	—	—
ESS score >10	3.4	1.4	4	—	37± 28
SomniSage Questionnaire	30	21	75	68	40 ± 28
BMI ≥30 kg/m^2	50	19	68-70	38	41 ± 29
Joint Task Force Guidelines	12-13	10-12	36-46	79-≥95	42-49
BMI ≥ 40 kg/m^2	6-7	6-7	23	>95	51 ± 32

*Percent of commercial drivers who will screen positive and then be diagnosed with OSA (defined as AHI >10) by polysomnography.
AHI, Apnea-hypopnea index; BMI, body mass index; ESS, Epworth Sleepiness Scale.
Modified from Kales SN, Straubel MG. Obstructive sleep apnea in North American commercial drivers. *Ind Health* 2014;52:13-24.

negative economic and occupational consequences.[40] In fact, data from diverse sources confirm that commercial drivers generally do not report their symptoms and diagnoses because of the negative economic and occupational consequences (perceived or real) of an OSA diagnosis. These concerns range from the inconvenience of having to submit to a diagnostic workup after a positive screening result, to the negative impact on employee pay resulting from being pulled out of service during a medical evaluation, to potential loss of a job and/or medical certification for employment.[9,14,41] In view of this reluctance of employees to self-report symptoms of OSA, it is especially important that occupational medicine physicians take the time to implement objective measures (described further on) as well as to ask supplementary questions in evaluating high-risk drivers.

In the United States, commercial vehicle drivers are required to complete a federal medical form that contains the single yes-or-no question "Do you have sleep disorders, pauses in breathing while asleep, daytime sleepiness, loud snoring?" Among those drivers identified as being at high risk for OSA, as many as 85% answered "No" to this question.[38]

The Epworth Sleepiness Scale (ESS) is a widely used questionnaire designed to identify persons with excessive daytime sleepiness due to either lifestyle circumstances (e.g., chronic sleep deprivation stemming from demanding work or social schedules) or a sleep disorder.[42] Unfortunately, among commercial drivers, the ESS questionnaire may produce a high rate of false-negative results. Most commercial vehicle operators report very low ESS scores at driver certification examinations (ESS scores in the range of 2 to 4 indicate a low likelihood of abnormal sleepiness), lower than ESS scores from the general community.[40] Similarly, results from an anonymous survey of transportation operators conducted by the National Sleep Foundation revealed a mean ESS score of 5.2 among professional truck drivers assessed.[43] By contrast, in another anonymous survey of U.S. truck drivers, more than 20% reported falling asleep at traffic lights.[39]

Somni-Sage is a questionnaire that incorporates weighted values for BMI, neck circumference, hypertension, and other medical comorbid conditions, as well as heavy snoring, witnessed apneic episodes, and other manifestations of excessive daytime sleepiness, and calculates categories of relative OSA risk.[44] Results from a retrospective assessment of Somni-Sage Questionnaire validity in more than 19,000 drivers showed that almost 6000 of the respondents (30%) were at higher risk for OSA. Of more than 2000 higher-risk drivers who underwent PSG, 68% were diagnosed with definite OSA (AHI greater than 10) and 80% had at least probable OSA (AHI of 5 or greater). A conservative prevalence estimate for definite OSA (AHI greater than 10) was 21% among commercial drivers in the population studied.[44]

Although evidence-based data are lacking, our own clinical experience supports the practice of a more probing interview of higher-risk commercial drivers by an experienced physician to obtain critical information. Drivers who deny symptoms and diagnoses on questionnaires and self-report forms often divulge more information when skilled physicians ask repetitive and additional questions regarding sleep hygiene, symptoms of daytime sleepiness, comorbid conditions, and findings on previous evaluations.[40]

Objective Measures

Objective screening tools have the advantage over questionnaires that they are less subject to manipulation, deception, and underreporting by drivers. Therefore application of such tools to obtain objective physical findings (e.g., BMI cutoff values) should be the screening modality of choice for OSA evaluation of employees who are in safety-sensitive positions.

At least four groups have generated recommendations for objectively screening CMV operators: (1) Dagan and colleagues,[45] (2) the U.S. Department of Transportation's Federal Motor Carrier Safety Administration (FMCSA) Medical Review Board (MRB),[46] (3) the FMCSA Medical Expert Panel (MEP),[47] and (4) the Joint Task Force (JTF) of the American College of Chest Physicians, the ACOEM, and the National Sleep Foundation.[48,49] Dagan's group showed that 78% of commercial drivers with a BMI of

32 kg/m^2 or greater had polysomnography (PSG)-confirmed OSA, and almost 50% also had objectively confirmed excessive daytime sleepiness as measured by a Multiple Sleep Latency Test. Confirming our review finding (see earlier) that subjective measures may yield a high number of false negatives, Dagan and colleagues also found that 100% of these affected drivers denied symptoms of OSA or excessive daytime sleepiness.[45]

The FMCSA MEP consisted of experienced clinicians and researchers knowledgeable in evidence-based medicine. In its 2008 report, the panel recommended referral of all CMV operators with BMI of 33 kg/m^2 or greater for a sleep study.[47] The MEP took feasibility of implementation into consideration by aiming to identify those CMV operators who were most likely to have severe OSA. After reviewing all of the negative long-term medical outcomes from untreated OSA, the FMCSA MRB recommended that a lower BMI criterion (BMI of 30 kg/m^2 or higher) be used for referral.[46] Neither set of recommendations from the MEP and the MRB have been implemented by the FMCSA as requirements. Barriers to a federal mandate include political, financial, liability, and legal concerns.[36,38]

The 2006 JTF issued recommendations for OSA screening of commercial drivers at certification examinations performed by commercial driver medical examiners (CDMEs).[48,49] Their screening recommendations include self-reported historical findings and the ESS, but they emphasized objective physical examination findings such as BMI, neck circumference, and hypertension criteria. The JTF guidelines recommended a higher BMI threshold (BMI of 35 km/m^2 or higher) for sleep study referral compared with that issued by either the MEP or the MRB of the FMCSA. This threshold was designed to have a high positive predictive value and to identify more severe OSA cases while not removing too many drivers from service. In the absence of a clear federal mandate for OSA screening of commercial drivers, the JTF guidelines are viewed by many occupational medicine professionals as representing a minimum standard for OSA screening to be implemented by the occupational medicine community.

Functional Screening

Functional or performance-based screening is an emerging approach for OSA screening at work sites using techniques such as psychomotor vigilance testing (PVT) and driving simulation. PVT is attractive as an adjunct to screen for OSA[50]: This testing modality has been shown to detect decrements in performance due to sleep deprivation, and it requires only a few minutes to administer. However, PVT performance cutoff criteria and correlation with accident or safety risks associated with OSA have not yet been established. Likewise, results from studies of driving simulation consistently demonstrate decrements in performance among subjects with OSA and other sleep disorders, as well as in subjects who have been sleep-deprived. At present, however, robust evidence associating simulator performance with on-road driving performance is lacking.[48] Additionally, performance criteria for identifying operators with OSA have not been established, and testing generally takes at least 30 minutes to complete. The latter characteristic makes use of simulators less attractive as an addition to occupational medical examinations.[40]

DIAGNOSIS OF OBSTRUCTIVE SLEEP APNEA IN OCCUPATIONAL SETTINGS

A "diagnosis" of OSA incorporates the results obtained from different types of sleep studies. Sleep studies may be performed using portable monitors (PMs) or full-channel PSG to measure (or estimate) the subject's sleep-disordered breathing (if present) via the AHI, respiratory disturbance index (RDI), or other means.[1] Although the current gold standard for diagnosing OSA is a laboratory-based PSG study, in-laboratory PSG is more expensive, time-consuming, and frequently of limited availability. The high prevalence of obesity means that more employees will have a positive result on screening and require follow-up sleep studies; consequently, the up-front costs of performing sleep studies has become a challenge to more widespread OSA screening and detection.[51-56] Using laboratory-based PSG as the diagnostic standard is less feasible because U.S. health insurers have placed increasing restrictions on reimbursement for the use of such studies. For all of these reasons, there has been an increasing interest in and push for use of PMs for diagnosing OSA in occupational settings, and this topic recently has been reviewed in depth.[51] To date, comparative effectiveness studies in safety-sensitive occupations (e.g., truck drivers, pilots) are lacking.

In the occupational setting, the major concern regarding the use of PMs is that transportation operators may actively avoid incurring an OSA diagnosis because of its economic and occupational implications (as discussed previously).[51] It is therefore imperative that clinicians be aware of scenarios whereby employees could alter PM results. They may stay awake to avoid sleep-disordered breathing events, which would go undetected if the PM does not record or estimate sleep stages, or the device may be placed on a healthier family member to avoid diagnosis. Accordingly, occupational medicine experts advocate using PM devices with a so-called chain-of-custody feature (such as a bracelet or other identifier placed on the driver by a professional technician), which deactivates the PM if it is removed. The best diagnostic use of PMs in an occupational context is to confirm the presence of OSA in an employee who already has been deemed to be at high risk for the disorder on the basis of symptoms or other screening techniques. When OSA is confirmed using the PM, treatment can be recommended and implemented without further testing. On the other hand, a negative or indeterminate result from a PM should not be considered sufficient evidence to exclude OSA, particularly in a high-risk employee working in a safety-sensitive position (e.g., transportation).[51]

Other unique considerations have emerged as important in interpreting diagnostic sleep tests in the occupational setting. Unlike in the nonoccupational setting (where the presence of symptoms can be used to make the diagnosis or decide on the necessity of treatment in cases in which the AHI or RDI is relatively low), the absence of symptoms should not be used to rule out OSA or accident risk.[38,40,45] Second, no clear thresholds of OSA severity (i.e., based on AHI/RDI, oxygen desaturation, or other measures) below which OSA-affected employees are *not* at an increased risk for accidents have been established.[4] Accordingly, in cases of mild OSA, the bias should be toward treating the OSA in affected employees who work in safety-sensitive positions (e.g., transportation operators).

TREATMENT OPTIONS AND COMPLIANCE MONITORING

As in the nonoccupational setting, continuous positive airway pressure (CPAP) is the first-line and best treatment for OSA in an occupational context. For employees in non–safety-sensitive positions, other OSA treatments may be considered. For transportation operators, however, CPAP is the only nonsurgical therapy for which the effectiveness of treatment and compliance can be objectively monitored. All employees in safety-sensitive positions (including transportation operators who receive surgical treatment) should undergo follow-up PSG to document improvement.

For safety-sensitive workers, once CPAP is instituted, compliance should be documented for at least 2 to 4 weeks and then reassessed at least yearly. Compliance is based on objective measures using data downloads or printouts from the CPAP device. Most practitioners regard minimum adequate compliance to consist of at least 4 hours of CPAP use per sleep period (or nightly) on at least 70% of nights.[48]

Adjunctive treatment with prescription stimulants for residual excessive daytime sleepiness in safety-sensitive workers should be undertaken only after consultation with the company medical director and after review of applicable federal rules. Failure to successfully respond to CPAP therapy should be a red flag regarding overall fitness for duty (particularly among transportation operators). Furthermore, some federal agencies may prohibit stimulant use or advise caution (see further on).

FEDERAL REGULATIONS AND RECOMMENDATIONS

Transportation operators with known diagnoses of OSA generally should be disqualified when their condition is untreated. However, despite various calls from the National Transportation Safety Board (NTSB), expert panels, and other bodies, regulating agencies under the Department of Transportation do not contain explicit objective requirements for OSA screening. Selected relevant regulations and recommendations are summarized in Table 79-2.

In April 2012, the FMCSA published a request for public comments on its proposed recommendations on regulatory guidance for diagnosis and management of obstructive sleep apnea.[57] The proposed guidance was based on MRB recommendations, as reviewed previously. A final ruling is still pending—and subsequent legislation forbids the FMCSA from using guidance alone to mandate sleep apnea screening for drivers before formal federal rulemaking.[58] On the other hand, since May 2014, medical examiners (i.e., CDMEs) providing screening examinations must belong to a Federal Registry of Certified Medical Examiners, which requires meeting certain licensing, continuing education, and testing requirements. Previously, such screening could be conducted by almost any health care practitioner, and these assessments often were performed by providers with little knowledge of occupational medicine, sleep disorders, and fitness for driving.[59] It is now expected that more drivers will be subject to examinations conducted in a stringent manner by persons with training in objective OSA screening as part of their practice standards.

The Federal Aviation Administration (FAA) regulates pilots and air traffic controllers. Pilots who are identified by an aviation medical examiner as being at risk for OSA are granted a medical certificate but are then required to undergo OSA evaluation within 90 days.[60] Examiners may reissue an airman medical certificate under the provisions of an Authorization, if the applicant provides the following: a current report (performed within the last 90 days) from the treating physician that references the present treatment, whether the treatment has eliminated any symptoms—with specific comments regarding daytime sleepiness. If any question arises regarding response to or compliance with treatment or if the applicant has developed some associated illness (e.g., right-sided heart failure), then examiners must defer to the Aerospace Medical Certification Division or the Regional Flight Surgeon.[61]

For the U.S. Coast Guard, the Mariner Medical Standards apply for OSA screening. No objective screening criteria are required or recommended, but the mariner is asked to disclose voluntarily an OSA diagnosis, if present.[62] To be considered for a waiver for sleep disorders, the mariner must demonstrate compliance with and efficacy of OSA treatment. In the case of OSA treatment with dental devices or positional therapy, the mariner must undergo PSG while using a dental device or positional therapy to demonstrate efficacy. If the condition was treated with surgery, then a postoperative sleep study is requested to document resolution of the condition. Once the initial information has been reviewed and a determination made that a mariner qualifies for a sleep disorder waiver, then the mariner is required to submit to periodic evaluations that include compliance information.[62]

The Federal Railroad Administration does not have any specific regulations regarding OSA; however, it has issued a safety advisory, which is summarized in Table 79-2.[63]

The FAA disqualifies pilots using prescription stimulants.[64] The FMCSA does not disqualify drivers using modafinil but advises monitoring and other precautions.[65] The U.S. Coast Guard Mariner Medical Standards do not routinely allow use of stimulants for OSA treatment, and waivers are granted on a case-by case-basis.[62]

RISK FACTOR REDUCTION

Beyond screening, diagnosis, and treatment of OSA in the workplace, other risk factor reduction measures can further mitigate OSA-related sleepiness and its consequences (see also the earlier discussion of FRMSs).

Because obesity is a primary risk factor for OSA, effective employee fitness and wellness programs can produce a return on investment for the transportation industry. Such programs fall within the realm of productivity and health management, where employee health and employer costs are viewed from a holistic perspective—with the ultimate goal of reducing costs by improving health.

Shift work is another risk factor for accidents (see Chapter 75), and shift work is likely to produce synergistic impairment with OSA. Therefore we do not recommend rotating shifts or night shift driving or night operations for OSA-affected employees even when they are compliant with CPAP treatment. This is particularly true for OSA-affected transportation operators.[40]

When possible, the work environment can be modified to allow for short but frequent breaks and strategically timed

Table 79-2 Summary of Selected Federal Regulations* or Recommendations for Medical Examiners Regarding Obstructive Sleep Apnea (OSA)

Federal Agency	Regulation/Recommendation	Required/Recommended Reporting Specifically Related to OSA	BMI Threshold for PSG Referral
FMCSA	*Regulation(s)*: No established medical history or clinical diagnosis of respiratory or neurologic dysfunction likely to interfere with the ability to control and drive a commercial motor vehicle safely[a]	*Required*: "Do you have sleep disorders, pauses in breathing while asleep, daytime sleepiness, loud snoring?"[b]	*Required*: None
FRA	*Regulation(s)*: With the exception of examinations and minimum standards for vision and hearing, U.S. commercial railroad companies have discretion as to the content, frequency and extent of their medical screening programs.[c] *Recommended*: That railroads and representatives of employees working together, develop and implement policies such that, "when a railroad becomes aware that an employee in a safety sensitive position has an incapacitating or performance-impairing medical condition related to sleep, the railroad prohibits that employee from performing any safety-sensitive duties until that medical condition appropriately responds to treatment."[d]	*Required*: None *Recommended*: That "employees' medical examinations include assessment and screening for possible sleep disorders and other associated medical conditions"[d]	*Required*: None
FAA	*Regulation(s)*: Untreated OSA is a disqualifying medical condition. "If a pilot is diagnosed with OSA, an AME must submit all pertinent medical information and a current status report, a sleep study with a polysomnogram, and use of medications and titration study results to the FAA. The FAA will then decide if a special issuance medial certificate is appropriate."[f]	*Required*: None	*Recommended*: BMI >40 (high BMI not disqualifying by itself)[e]
U.S. Coast Guard	*Regulation(s)*: "Are of sound health; have no physical limitations that would hinder or prevent performance of duties; and are free from any medical conditions that pose a risk of sudden incapacitation, which would affect operating, or working on vessels."[g]	*Required*: None— relies on self-disclosure of a sleep apnea diagnosis[h]	*Required*: None
NTSB	*Recommended*: Elicit preexisting diagnoses of OSA; screen for OSA risk factors; and ensure operators with sleep apnea are effectively treated before granting unrestricted medical certification.[i]	*Recommended*: Develop standard medical examination forms to elicit diagnoses of sleep disorders and to screen for sleep disorders; require use of these forms[j]	*Recommended*: Not specified

*As of 2014. Data from cited sources.

[a]Qualifications of drivers and longer combination vehicle (LCV) driver instructors. *CFR* 2012;49, §391.41.

[b]Federal Motor Carrier Safety Administration. *Medical examination report for commercial driver fitness determination.* Form 649-F (6045). Washington (D.C.): Federal Motor Carrier Safety Administration; March 19, 2014. <http://www.fmcsa.dot.gov/regulations/medical/medical-examination-report-commercial-driver-fitness-determination>; 2014.

[c]U.S. Department of Transportation, Federal Railroad Administration. *Medical standards for railroad workers.* Final report.Washington (D.C.): Office of Safety; January 2005.

[d]Notice of Safety Advisory 2004-04; Effect of sleep disorders on safety of railroad operations. *Fed Reg* 2004;69(190):58995-6.

[e]Federal Aviation Administration. *Fact sheet—sleep apnea in aviation* [press release]. <http://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=15994>; 2014.

[f]Federal Aviation Administration. *Fact sheet—sleep apnea in aviation* [press release]. <http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=15474>; 2013.

[g]U.S. Department of Homeland Security, U.S. Coast Guard. *Merchant Mariner credential medical evaluation report.* Form CG-719K Rev. (01-09). <http://www.uscg.mil/forms/cg/cg_719k.pdf>; 2009.

[h]U.S. Department of Homeland Security, U.S. Coast Guard. *Medical and physical evaluation guidelines for Merchant Mariner credentials.* NVIC 04-08, COMDTPUB 16700.4. Washington (D.C.): June 7, 2013.

[i]National Transportation Safety Board. *Safety recommendation (M-09-14 through -16).* Washington (D.C.): October 20, 2009. <http://www.ntsb.gov/doclib/recletters/2009/M09_14_16.pdf>; 2009.

[j]National Transportation Safety Board. *Safety recommendation (A-09-61 through -66).* Washington (D.C.): August 7, 2009. <http://www.ntsb.gov/doclib/recletters/2009/a09_61_66.pdf>; 2009.

BMI, Body mass index; FAA, Federal Aviation Administration; FMCSA, Federal Motor Carrier Safety Administration; FRA, Federal Railroad Administration; NTSB, National Transportation Safety Board; PSG, polysomnography.

naps. As reviewed in Chapter 75, naps decrease subjective fatigue and improve objective alertness and performance.

CLINICAL PEARLS

- Employees in safety-sensitive positions (e.g., truck drivers, pilots, mariners) may specifically deny or underreport symptoms to avoid incurring an OSA diagnosis because of its potential negative economic and occupational consequences.
- Screening for OSA should include BMI, neck circumference, and other easily obtained objective criteria. Negative subjective symptom reports should be considered unreliable. A more probing interview performed by an experienced physician to obtain critical information often is required.
- Monitoring with portable devices can be used to confirm the presence of OSA so that appropriate treatment can be initiated. Negative or inconclusive PM-based findings warrant a follow-up evaluation with full in-laboratory PSG, particularly for high-risk employees working in safety-sensitive positions.
- CPAP remains the first-line treatment for OSA, and minimum adequate compliance consists of at least 4 hours of CPAP use per sleep period on at least 70% of nights.

SUMMARY

OSA is common in the workplace and often goes undiagnosed and thus untreated, with negative consequences for health, safety, and productivity. Transportation accidents constitute a particular concern. Sleep and occupational medicine professionals are likely to encounter a number of patients in several workplace referral scenarios that trigger evaluation for possible OSA, including falling asleep at work, on-the-job accidents, and screening or diagnostic assessment for safety-sensitive positions (especially in transportation). Because employees may be reluctant to disclose OSA-related symptoms, clinicians cannot depend on symptom reports for screening or diagnosis and must rely primarily on objective measures, tests, and demonstration of treatment compliance. Screening can be accomplished objectively through use of BMI and neck circumference, for which thresholds can be set for referring employees to a sleep medicine specialist.

Selected Readings

Berger M, Varvarigou V, Rielly A, et al. Employer-mandated sleep apnea screening and diagnosis in drivers. *J Occup Environ Med* 2012;**54**(8):1017–25.

Colvin LJ, Collop N. Commercial motor vehicle driver obstructive sleep apnea screening and treatment in the United States: an update and recommendation overview. *J Clin Sleep Med* 2015 Jun 11. pii: jc-00092-15.

Kales SN, Straubel M. Obstructive sleep apnea in North American commercial drivers. *Ind Health* 2014;**52**(1):13–24.

Perlman S. Sleep apnea and workplace safety. *BCMJ* 2014;**56**(2):94–6.

Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. *J Clin Sleep Med* 2009;**5**(6):573–81.

Zhang C, Berger M, Malhotra A, Kales SN. Portable diagnostic devices for identifying obstructive sleep apnea among commercial motor vehicle drivers: considerations and unanswered questions. *Sleep* 2012;**35**:1481–9.

A complete reference list can be found online at ExpertConsult.com.

Section 11

Insomnia

Chapter 80

Insomnia: Recent Developments and Future Directions

Daniel J. Buysse; Allison G. Harvey

Chapter Highlights

- Recent revisions of widely used classification systems now include one broad insomnia diagnosis category to cover most adult and child insomnias. In the *International Classification of Sleep Disorder,* third edition, this is termed chronic insomnia disorder, and in the *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition, it is insomnia disorder.
- Both classification systems include the specification of 3 months as the duration criterion for insomnia disorder and the specification of 3 times per week as a minimum frequency.
- There is now strong evidence supporting the relationship between insomnia and subsequent development of depression.
- The "insomnia with short sleep" phenotype is associated with hypertension, diabetes, neurobehavioral performance impairments, and mortality risk.
- Cognitive behavior therapy for insomnia and related techniques may be useful for a broader range of patients, including those with comorbid psychiatric conditions such as depression, bipolar disorder, posttraumatic stress disorder, and schizophrenia.

Insomnia is the most common sleep disorder, but not the most commonly addressed in research studies or in the development of clinical practice models. Although the manifestations and immediate consequences of insomnia are perhaps less dramatic than those of several other sleep disorders, accumulating evidence demonstrates the adverse effects of insomnia on long-term health, functioning, and productivity, as well as their associated health care costs. Important developments have occurred in the diagnosis and assessment of insomnia, understanding its long-term clinical consequences, identification of important physiologic features, and strategies for both psychological-behavioral and pharmacologic management. These new developments are addressed in each of the chapters in this section. Here, we highlight some of the most important new developments in insomnia research and clinical practice and outline several areas for future investigation.

DIAGNOSIS AND ASSESSMENT

Two new classifications for sleep disorders have been published, the *International Classification of Sleep Disorders*, third edition (ICSD3),[1] and the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5).[2] As discussed in Chapter 83, both classifications have changed not only the specific diagnostic criteria for insomnia but also its very conceptualization. Fortunately for clinicians, the two systems are consistent with each other, which will enhance the clinical utility and clarity of each system. The most important change embraced in each classification is the specification of one broad insomnia diagnosis category to cover most adult and child insomnias. In ICSD3 this is termed chronic insomnia disorder, and in DSM-5, insomnia disorder. ICSD3 also includes a specific category of short-term insomnia disorder, which is included as a subtype of insomnia disorder in DSM-5. These categories replace the previous distinction between "primary" insomnia and "secondary" insomnia diagnoses, recognizing the fact that it is often impossible to reliably determine directionality or causality between insomnia and comorbid conditions.[3,4] Previous editions of the ICSD included subtypes of "primary" insomnia," such as psychophysiological, idiopathic, and paradoxical insomnia; ICSD3 collapses these subtypes into a single diagnosis, given poor reliability and discriminant validity of these subtypes.[5,6] Importantly, the insomnia diagnoses in ICSD3 and DSM-5 apply to both children and adults, which has advantages for billing and coding of pediatric insomnia in clinical practice. Other important changes in ICSD3 and DSM-5 include the specification of 3 months as the duration criterion for insomnia disorder, the specification of 3 times per week as a minimum frequency, and the elimination of nonrestorative sleep as a sufficient symptom for the diagnosis of insomnia. Nonrestorative sleep, although a common symptom in patients with insomnia, infrequently occurs without sleep initiation or sleep maintenance difficulties, and it also occurs in many other sleep disorders.[7] Taken together, these changes may enhance the reliability of insomnia diagnoses and make it easier to diagnose insomnia across a wide range of medical and health care settings. The shift to a single broad diagnosis need not stifle research into potential phenotypes or subtypes, but it does prevent the promulgation of subtypes with inadequate reliability and validity.

CONSEQUENCES OF INSOMNIA

New epidemiologic studies have provided additional evidence for adverse health and functional consequences of insomnia. Numerous individual studies and an influential meta-analysis have demonstrated the relationship between insomnia and subsequent development of depression.[8] In addition, longitudinal studies have demonstrated a relationship between insomnia and incident cardiovascular disease and even mortality.[9] As discussed by Hall, Vgontzas, and colleagues, several studies have also demonstrated that the adverse health correlates of insomnia may interact with those of short objective sleep duration. This "insomnia with short sleep" phenotype is associated with hypertension,[10] diabetes,[11] neurobehavioral performance impairments,[12] and mortality risk.[13] These studies need to be replicated in independent samples and with other measures of sleep duration than overnight polysomnography to maximize their utility. However, the identification of a novel insomnia phenotype will surely open the door to further studies of insomnia mechanisms and treatment impact. In a larger sense, studies documenting the health correlates and consequences of insomnia provide support for assessment, diagnosis, and treatment of insomnia across the full range of health care settings.

ETIOLOGY AND PATHOPHYSIOLOGY

Perlis and colleagues provide a scholarly overview of various models of the etiology and pathophysiology of insomnia. Several psychological, behavioral, and cognitive models of insomnia have reached maturity, and their influence on research and clinical practice are evident. However, additional work on the etiology and pathophysiology of insomnia has also addressed other levels of analysis, ranging from genes[14] and specific neurotransmitter molecules[15] to brain morphology,[16,17] functional connectivity,[18] and functional responses to cognitive and affect tasks.[19,20] These studies have been the subject of several recent reviews.[21,21-23] Although published studies have provided tantalizing leads into the neurobiology of insomnia, they often suffer from low power, varying diagnostic criteria and imaging techniques, and lack of replication. Genetic studies have failed to identify specific genes associated with human insomnia, although findings from experimental studies in *Drosophila* species suggest a relationship among insomnia phenotypes, metabolism, and mortality.[24] Epigenetic mechanisms have also been discussed as a mechanism through which genes and environment may interact to produce sleep problems.[23] Current neuroimaging evidence implicates dysfunction in the hippocampus, anterior cingulate cortex, default mode network, and primary sensory networks.[21] These findings may help to explain the relationship between insomnia and affective and cognitive difficulties, as well as subjective-objective sleep discrepancies in insomnia. However, much more work is needed to identify consistent patterns of dysfunction in neural structures and circuits associated with insomnia or vulnerability to insomnia, as well as how such patterns may change with treatment. Such studies could help to develop and refine future treatments and potentially usher in the era of personalized insomnia medicine.

PSYCHOLOGICAL-BEHAVIORAL TREATMENTS

Psychological-behavioral interventions for insomnia have seen the development of new approaches and more aggressive efforts at dissemination of evidence-based treatments, as reviewed by Edinger and colleagues. Morin and colleagues draw our attention to the newer approaches that include the use of mindfulness-based techniques, either singly or in combination with cognitive-behavioral approaches,[25] and the investigation of cognitive therapy as a stand-alone intervention for insomnia.[26] Initial evidence suggests that each of these approaches may be efficacious, although further work is needed to identify which patients respond best to which approach.

Major efforts have been applied to the dissemination of evidence-based treatments for insomnia, mainly cognitive behavior therapy for insomnia (CBT-I). This is a great need and a major challenge given the substantial gap between

treatment development and upscaling for use in routine practice.[27] A national training effort was carried out in the Veterans Administration health system to teach CBT-I to behavioral health providers.[28-30] The impact of this education effort, as well as studies that it might engender, will be investigated for many years. Dissemination may also involve alternate modes of administration, including self-guided Internet-based approaches.[31,32] Several research groups have independently developed online CBT-I programs, which have been evaluated in empiric research studies and made available to the public. Although initial studies demonstrate the efficacy of these approaches, implementation of online CBT-I in actual practice settings may be more challenging and requires further study. For instance, one study identified several groups (e.g., older people, unemployed people) who responded better to a higher intensity therapist-delivered intervention than they did to online CBT-I.[33] Hence research breakthroughs in the anxiety disorders, involving combining computerized delivery of a CBT treatment but assisted by a mental health worker,[34] may be a combination also worth trying for insomnia.

As Edinger and colleagues note, stepped care approaches to the management of insomnia were investigated in early studies in single-payer health systems such as the United Kingdom but require further development and testing in private-public health systems such as those in the United States. Finally, dissemination efforts also include the use of CBT-I and related techniques for a broader range of patients, including those with comorbid psychiatric conditions such as depression, bipolar disorder, posttraumatic stress disorder, and schizophrenia.[35] The results of small initial trials have been promising, and the results of larger studies are now becoming available.

PHARMACOLOGIC TREATMENTS

After a period of intense development efforts by the late 1980s and early 1990s, the development of new pharmacologic agents for the treatment of insomnia has been relatively quiescent. One notable exception, reviewed by Krystal and colleagues, is U.S. Food and Drug Administration (FDA) approval of a first-in-class orexin receptor antagonist, suvorexant.[36] This medication has a novel mechanism of action, antagonizing orexin receptors and inhibiting wake-promoting centers of the hypothalamus and brainstem. Although clinical trials in support of the FDA filing demonstrate the short and long-term efficacy of suvorexant, understanding its role in clinical practice will require further experience. Additional histamine receptor antagonists are also under development, which may provide additional therapeutic options.

The mainstay of pharmacologic treatment remains benzodiazepines and similar drugs that antagonize benzodiazepine receptors. Use of these agents steadily climbed during the 1990s and 2000s. Additional concerns have been raised regarding potential adverse consequences of these agents, ranging from development of dementia[37] to increased mortality risk.[38] As Walsh and colleagues note, epidemiologic studies cannot demonstrate causation, and additional work is needed to understand the mechanism by which benzodiazepines may increase risk, or whether comorbidities and confounders may explain the observed associations.

CLINICAL MANAGEMENT

Additional developments have affected insomnia at the organizational and practice-based level. The American Academy of Sleep Medicine published quality measures for tracking patient care for several sleep disorders, including insomnia, in 2015.[39] The insomnia metrics focus on outcomes of improving sleep quality and satisfaction and daytime functioning and on process measures of assessing sleep quality, daytime function, and treatment-related side effects, as well as the provision of evidence-based treatment. Quality measures serve as a means of promoting appropriate care for patients with insomnia across a wide variety of practice settings. As further evidence accumulates on important outcomes and process measures, the quality metrics can be refined.

Although not directly related to insomnia, the National Sleep Foundation[40] published recommendations for sleep duration across the life span, and the American Academy of Sleep Medicine and the Sleep Research Society are preparing to publish recommendations for the adequate amount of sleep to promote health in adults. These guidelines are an important step in addressing sleep as a public health need. They set the stage for future recommendations that could focus on other aspects of sleep, including sleep timing, regularity, and quality—attributes that are critical to consideration of insomnia.

CHALLENGES AND FUTURE DIRECTIONS

The brief summaries provided previously note some of the challenges confronting future efforts in the research and clinical management of insomnia. Although genetic and imaging studies hold great promise for understanding the etiology and pathogenesis of insomnia, more investigators and research studies are needed, using consistent and convergent approaches. Techniques such as epigenetic and proteomic studies, magnetic resonance spectroscopy, and both resting state and event related functional magnetic resonance imaging will undoubtedly play a role. In addition, studies using high-density electroencephalography and magnetoencephalography may be useful for understanding both regional and temporal sources of dysregulation. In addition to defining broad categories of alterations in insomnia patients, these techniques may also be useful for developing and validating specific phenotypes, such as insomnia with short sleep duration.

Psychological and behavioral treatments are clearly efficacious for insomnia. However, further clarity is needed to define which patients are most likely to benefit from which type of treatment. Treatments of younger age groups and more "real life" patient populations, such as those with multiple medical problems and serious mental illness, are also needed. Cognitive behavioral interventions for insomnia are ready to be incorporated into dissemination, implementation, and patient-centered outcomes research studies. However, few insomnia studies of this sort have been supported by federal research agencies to date. Further effort is needed to understand the role of online treatment modalities and to develop methods to mount and train a workforce, including a variety of practitioners, to act as first-line insomnia therapists. As one successful example of the latter, the Improving Access to Psychological Therapies (IAPT) program trained 3600 new therapists to deliver CBT treatments for anxiety and

depression across the United Kingdom between 2008 and 2011.[41] By 2013 IAPT was treating approximately 400,000 patients each year, nearly half of whom had recovered by the end of treatment and many more of whom benefited.[42] In other words, rapidly mounting and training the required workforce can be achieved with the appropriate funding resources. The funding for IAPT was provided by the United Kingdom government on the basis of an economic analysis that concluded that the costs of providing these treatments would be fully covered by the savings from fewer people with anxiety and depression being unemployed.[43]

Finally, insomnia may benefit from being placed in a broader health context. Healthy People 2020[44] for the first time has sleep-related public health goals, and sleep organizations are actively partnering with the Centers for Disease Control and Prevention on these efforts. The expertise and techniques of insomnia researchers and clinicians may aid such efforts. For instance, addressing issues such as insufficient sleep at the population level would benefit from the accumulated expertise of insomnia researchers and clinicians who have been addressing insomnia from a behavioral health perspective. Using the techniques of behavioral sleep medicine to promote population health provides an exciting and meaningful challenge to the field.

CLINICAL PEARL

Insomnia is comorbid with several important psychiatric disorders. CBT-I and related techniques may be useful for a broader range of patients, including those with comorbid psychiatric conditions such as depression, bipolar disorder, posttraumatic stress disorder, and schizophrenia.

Selected Readings

American Academy of Sleep Medicine. *International classification of sleep disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014. p. 19–48.

Beaulac J, Vincent N, Walsh K. Dissemination of an internet-based treatment for chronic insomnia into primary care. *Behav Sleep Med* 2015;**13**: 124–39.

Conley S, Redeker NS. Cognitive behavioral therapy for insomnia in the context of cardiovascular conditions. *Curr Sleep Med Rep* 2015;**1**(3): 157–65.

Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health* 2015;**1**:40–3.

Riemann D, Nissen C, Palagini L, et al. The neurobiology, investigation, and treatment of chronic insomnia. *Lancet Neurol* 2015;**14**:547–58.

A complete reference list can be found online at ExpertConsult.com.

Insomnia: Epidemiology and Risk Factors

Chapter 81

Kenneth L. Lichstein; Daniel J. Taylor; Christina S. McCrae; Megan E. Petrov

Chapter Highlights

- The prevalence of chronic, clinically significant insomnia is about10%, but diagnostic and methodologic inconsistences render a definitive estimate difficult. Moderately frequent insomnia symptoms appear in a much larger percentage of persons. Advancing age, female gender, and low socioeconomic status are strong insomnia correlates.
- Some variant of hyperarousal and psychiatric or medical comorbidity are often associated with insomnia and are therefore deemed risk factors. Precise mechanisms and causal paths with respect to insomnia correlates and risk factors are poorly understood. Transient insomnia and genetic factors may also emerge as important risk factors.
- Insomnia is itself a risk factor for a wide variety of psychological, psychiatric, and medical disorders, including depression, substance abuse, and hypertension.
- The highest priorities for future research are establishing a widely accepted, consistent definition of insomnia, performing longitudinal studies, and cultivating a more comprehensive understanding of insomnia by expanded monitoring of psychological, psychiatric, and medical comorbidity, race, and genetic variables.

Insomnia is the most common sleep disorder and is among the most prevalent of all mental health disorders. However, at the core of the challenge in evaluating the epidemiology of insomnia is demarcating the border between insomnia complaints and clinical insomnia.

Twenty-six percent of people complain of difficulty falling asleep, and 42% complain of difficulty staying asleep at least a few nights a week.[1] These survey results illuminate the potential magnitude of insomnia, but they also serve to caution us that a sleep complaint alone does not necessarily meet formal diagnostic criteria for insomnia.

Numerous factors predict insomnia, and insomnia creates heightened risk for declining health. Although the identification of causal mechanisms remains elusive at this point, there is a robust insomnia risk literature. It should be immediately apparent that complexities impede efforts at clarifying the epidemiology of insomnia and associated risks. This chapter enumerates such obstacles and presents a summary of current knowledge on the epidemiology and risk factors of insomnia.

EPIDEMIOLOGY

Definition of Insomnia

Defining insomnia is a complex task for several reasons. First, insomnia can occur as a symptom, a disorder, or both. Second, insomnia that begins as a symptom of another disorder often evolves over time into an independent disorder. Third, as a disorder, insomnia is heterogeneous, with varying durations, types, and etiologies. Fourth, insomnia severity fluctuates over time. Epidemiology studies of insomnia have too often been secondary analyses of epidemiologic studies in which sleep or insomnia was assessed only with as few as one question, which could be current, past month, or even lifetime history of a sleep complaint. However, the most recent revisions of the *International Classifications of Sleep Disorders*, third edition (ICSD3)[2] and the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5)[3] both specify a sleep complaint that occurs at least three times per week for at least 3 months and is associated with daytime impairment to diagnose insomnia.

Given this complexity, it is not surprising that the literature contains numerous definitions of insomnia, ranging from the broad and inclusive to the narrow and exclusive. For example, some epidemiologic studies have relied on respondents to provide their own definitions by simply asking whether or not they have trouble sleeping,[4] whereas other studies have defined insomnia more narrowly using quantitative severity threshold criteria, such as greater than 30 minutes of sleep-onset latency or wake time after sleep onset.[5] Unfortunately, the study of insomnia has been hampered by such varied definitions of insomnia as evidenced by a review of epidemiologic studies in which prevalence rates in the general population ranged from 4% to 48% depending on the definition used.[6] Efforts to standardize insomnia criteria[7] might prove helpful in the future.

Evaluation

Epidemiologic surveys have traditionally relied on self-report data because large-scale data collection using objective methods (polysomnography [PSG] or actigraphy) is expensive and cumbersome. It may be relevant to note that in most clinical practice settings, routine insomnia assessment does not include PSG or actigraphy. Further, self-report

instruments better capture sleep perception (but are vulnerable to missing other confounding sleep disorders such as sleep apnea) than objective methods, and the subjective view may have greater import for insomnia than objective assessment.[7] A consensus sleep diary is now available,[8] and this should make an important contribution to elevating the quality of self-report sleep data and promoting consistency of such data across studies.

Self-reported accounts of one's sleep experiences are necessarily retrospective in nature. However, the time period over which published studies have asked subjects to provide sleep estimates has varied widely—from 1 year to the previous night.[5] Both recency and saliency effects are likely to contribute to inaccuracies in estimating one's average total sleep time, sleep-onset latency, and other sleep variables over a 1-week period, let alone up to a year. The term *prospective* refers to the future, but when used to compare with retrospective sleep data, prospective sleep data refer to sleep diaries collecting information on sleep from the night just completed and for multiple nights thereafter. This distinction is important because sleep diaries allow researchers to capitalize on the immediacy of retrospective recall.

The heavy reliance on retrospective assessment that requires a patient to average across multiple nights is problematic because of the increased risk for recall errors and unreliability when respondents are asked to provide sleep estimates, and averages, over long periods of time. The recall bias introduced by retrospective estimates may produce overestimates of insomnia symptoms. Indeed, retrospective estimates of total sleep time and number of awakenings are poorly correlated with parallel sleep diary estimates,[9] and PSG sleep variables correlate better with daily sleep diaries than with retrospective reports.[10] Additionally, retrospective estimates can introduce unreliability because single-point estimates may be susceptible to situational influences that are temporary. Momentary emotional states (e.g., work-related stress, depressed mood) tend to influence recall more than do routine experiences. Recency bias can also introduce unreliability because participants can exhibit a tendency to respond based on their previous night's sleep. Another problem plaguing some retrospective studies is the absence of a specified time period of assessment. For example, some studies asked respondents to report how they "usually" or "generally" sleep.

Prospective assessment using at least 1 week of sleep diaries, preferably 2 weeks, avoids or reduces many of the limitations inherent in retrospective assessment.[11] It also allows a richer characterization of sleep because sleep diaries provide detailed quantitative descriptions (e.g., sleep-onset latency, wake time after sleep onset, sleep efficiency) and information on type of insomnia (i.e., onset, maintenance, terminal). At the same time, it is also important to note that the time- and resource-intensive nature of obtaining prospective sleep diary data affects the feasibility of routinely collecting such data. Although prospective assessment is ideal for obtaining the most accurate prevalence estimates, examining insomnia-related items in publically available datasets helps to highlight the need for greater clinical and research focus on insomnia. However, such retrospective data should be interpreted with the knowledge of their limitations and the understanding that the estimates provided, although informative, may also overestimate the level of clinically relevant insomnia.

Table 81-1 Insomnia Prevalence Rates by Four Main Definitional Categories and Three Symptom Categories

Category	Prevalence (%)
Definitional Category	
DSM-IV insomnia diagnosis ($n = 5$)	4–6
Dissatisfaction with sleep quantity or quality ($n = 11$)	8–18
Insomnia symptoms plus daytime consequences ($n = 8$)	9–15
Insomnia symptoms* ($n = 21$)	10–48
Symptom Category	
Insomnia symptoms only	30–48
Insomnia symptoms plus frequency criteria (≥3 nights/week or often/always)	16–21
Insomnia symptoms plus severity criteria (moderately to extremely)	10–28

*Insomnia symptoms included difficulty initiating or maintaining sleep. Some studies also included nonrestorative sleep as a symptom.
DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition, text revision.
Modified from Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Med Rev* 2002;6:97–111.

Prevalence

The methodologic inconsistencies in defining and assessing insomnia have resulted in a wide range of insomnia prevalence estimates for the general population. As shown in Table 81-1, Ohayon reviewed more than 50 studies and found large decreases in prevalence as criteria stringency increased, resulting in rates as high as 48% (using a single symptom) to as low as 4% (using *International Classification of Diseases,* tenth revision [ICD-10] criteria for insomnia).[6] The America Insomnia Survey of more than 10,000 managed health care plan subscribers further highlights the effect that differences in several established classification systems criteria can make on these rates.[12] In that survey, insomnia prevalence varied considerably: 22% (based on DSM-IV-TR[13]), 15% (based on Research Diagnostic Criteria[14]), and 4% (based on ICD-10[15]).

European and Scandinavian findings indicate insomnia prevalence is on the rise. A Norwegian study[16] examined the 10-year trend in insomnia in two separate adult samples (1999 to 2000, $n = 2001$; and 2009 to 2010, $n = 2000$). Controlling for gender, age, socioeconomic status, and region, that study found increases in the prevalence of sleep-onset difficulties (13% to 15%), dissatisfaction with sleep (8% to 14%), and DSM-IV diagnoses (12% to 16%). Similarly, a study of middle-aged Swedish women[17] between 1968-1969 and 2004-2005 found that sleep problems almost doubled over that time period. Interestingly, an English study[18] of more than 20,000 adults revealed only modest increases of already high insomnia prevalence rates over a 15-year period from 1993 (35%) to 2007 (38.6%).

Consistent with most of the epidemiologic literature, these studies were all single-point retrospective estimates. Of the more limited number of studies using prospective assessment,

Lichstein and colleagues[5] collected 2 weeks of sleep diaries and multiple daytime impairment measures from at least 50 men and 50 women in each age decade from 20 to 80+ years. Based on standard diagnostic criteria (ICSD) supplemented by empirically derived quantitative criteria,[19] the overall prevalence of insomnia, weighted by gender and age, was 15.9%. This rate is higher than expected based on the rate decreases seen with increasing methodologic rigor in Ohayon's review but is consistent with the higher rates found by Roth and colleagues.[12]

Comorbidities

Comorbid insomnia is an important topic that is increasingly receiving attention in the epidemiology literature. This is attributable to changes in the conceptualization of insomnia that occurred after a 2005 National Institutes of Health State-of-the Science Consensus Panel[20] recommended that the term *comorbid insomnia* replace the term *secondary insomnia* when insomnia occurs in the context of another disorder (medical or psychiatric), indicating that the insomnia is at least a partially independent and separate disorder. This recommendation was based on a review of the evidence and expert consensus that insomnia initially occurring secondary to another condition often becomes an independent problem or a partially independent problem that shares a reciprocal relationship with the original primary disorder.[21] With comorbid insomnia accounting for 70% to 90% of insomnia in the general population,[22-24] increasing focus on comorbid insomnia in the epidemiology literature is a welcome change.

The most common comorbidities associated with insomnia are psychiatric disorders, including anxiety, depression, panic disorder, adjustment disorder, somatoform disorders, and personality disorders. Epidemiologic studies consistently demonstrate high comorbidity rates for anxiety and mood disorders. In the National Comorbidity Survey—Replication,[25] respondents with mood (45.5%), anxiety (45.6%), or comorbid mood and anxiety (62.8%) disorders were more likely to report insomnia symptoms than those with no mood or anxiety disorder (23.3%). In another study, Taylor and colleagues[26] found that 20% percent of people with insomnia displayed clinically significant depression and 19.3% demonstrated clinically significant anxiety. People with insomnia were 10 times more likely to have depression and 17 times more likely to have anxiety than normal sleepers.

Medical conditions that commonly co-occur with insomnia include arthritis, cancer, hypertension, chronic pain, coronary heart disease, and diabetes. In a cross-sectional study of 9000 persons 55 to 84 years of age, Foley and colleagues[27] found an association between history of insomnia and difficulties with activities of daily living, respiratory symptoms, and other health problems, including hypertension, heart disease, cancer, stroke, diabetes, and hip and other fractures. A 3-year follow-up (n = 6800) revealed that heart disease, stroke, hip fracture, and respiratory symptoms were associated with incident insomnia, meaning new cases of insomnia. Heart disease, diabetes, respiratory symptoms, and stroke were also associated with the maintenance of insomnia.[28] Such findings are not limited to older adults, as studies that have examined adults across the lifespan demonstrate. In a survey of 3161 adults[24] 18 to 79 years of age, those who reported insomnia symptoms were more likely to report having two or more health problems than were normal sleepers.

Table 81-2 Prevalence of Medical Problems in People with or Without Insomnia

Medical Problem	Prevalence of Medical Problem (%)* PWI	PNI	Adjusted Odds Ratio† (95% CI)
Heart disease	21.9	9.5	2.27 (1.13–4.56)‡
Cancer	8.8	4.2	2.58 (0.98–6.82)
Hypertension	43.1	18.7	3.18 (1.90–5.32)§
Neurologic disease	7.3	1.2	4.64 (1.37–15.67) ‡
Breathing problems	24.8	5.7	3.78 (1.73–8.27)‖
Urinary problems	19.7	9.5	3.28 (1.67–6.43)‖
Diabetes	13.1	5.0	1.80 (0.78–4.16)
Chronic pain	50.4	18.2	3.19 (1.92–5.29)§
Gastrointestinal problems	33.6	9.2	3.33 (1.83–6.05)§
Any medical problem	86.1	48.4	5.17 (2.93–9.12)§

*Percentage of people with or without insomnia who report that particular disease.
†Adjusted for depression, anxiety, and sleep disorder symptoms.
‡$P < .05$
§$P < .001$
‖$P < .01$
CI, confidence interval; PNI, people not having insomnia; PWI, people with insomnia.
Modified from Taylor DJ, Mallory LJ, Lichstein KL, et al. Comorbidity of chronic insomnia with medical problems. *Sleep* 2007;30:213–8.

Taylor and colleagues[29] focused on prevalence of medical problems in persons with insomnia in the general population. Controlling for anxiety, depression, and symptoms of other sleep disorders, they found that persons with insomnia had a higher prevalence of comorbid medical problems than did persons without insomnia (Table 81-2). Additionally, the adjusted odds of having any chronic medical problem were 5.17 times higher in persons with insomnia than in those without insomnia. From a different perspective, the same study calculated insomnia rates in the same medical disorders (Table 81-3). Insomnia prevalence varied between 44% in people with hypertension and 66.7% in people with neurologic disease, with an overall adjusted odds ratio of 5.26. Another study[30] (n = 3282) of the prevalence of insomnia in persons with self-reported medical diseases revealed similar findings. Controlling for age and gender, the investigators found higher prevalence of insomnia in persons reporting 13 different diseases than in those who did not.

Demographics

Gender

Insomnia is more common in women than in men. Lichstein and colleagues[5] reported prevalence for insomnia by gender derived from their review of 42 epidemiology studies (33 of which reported gender comparisons). Overall, insomnia prevalence was 50% greater in women: 18.2% in women and

Table 81-3 Prevalence of Insomnia in People with or Without Medical Disorders

Medical Problem	Insomnia Prevalence (%)* PHM	PNM	Adjusted Odds Ratio† (95% CI)
Heart disease	44.1	22.8	2.11 (1.07–4.15)‡
Cancer	41.4	24.6	2.50 (1.01–6.21)‡
Hypertension	44.0	19.3	3.19 (1.87–5.43)§
Neurologic disease	66.7	24.3	5.21 (1.22–22.21)‡
Breathing problems	59.6	21.4	2.79 (1.27–6.14)‡
Urinary problems	41.5	23.3	3.51 (1.82–6.79)§
Diabetes	47.4	23.8	2.03 (0.86–4.79)
Chronic pain	48.6	17.2	3.16 (1.90–5.27)§
Gastrointestinal problems	55.4	20.0	3.00 (1.66–5.43)§
Any medical problem	37.8	8.4	5.26 (2.82–9.80)§

*Percentage of people with or without that particular disease who report insomnia.
†Adjusted for depression, anxiety, and sleep disorder symptoms.
‡$P < .05$
§$P < .001$
CI, confidence interval; PHM, people who reported having the medical problem; PNM, people who did not report having the medical problem.
Modified from Taylor DJ, Mallory LJ, Lichstein KL, et al. Comorbidity of chronic insomnia with medical problems. *Sleep* 2007;30:213–8.

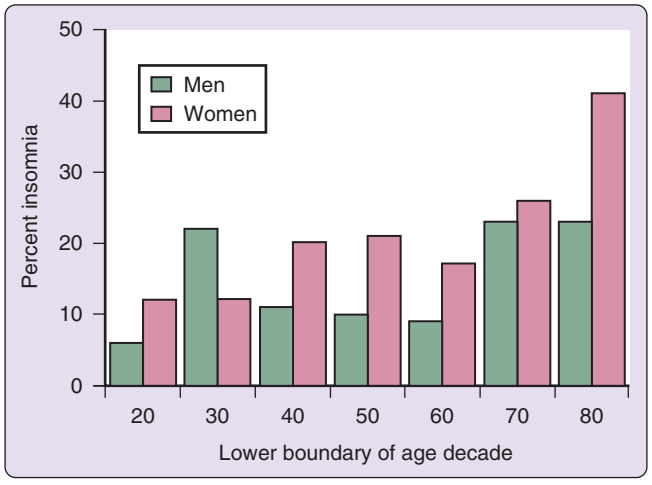


Figure 81-1 Insomnia prevalence by sex and age. Each value on the abscissa represents the beginning of an age decade as in 20 (to 29) and 30 (to 39). (Data from Lichstein KL, Durrence HH, Riedel BW, et al. *Epidemiology of sleep: age, gender, and ethnicity.* Mahwah, NJ: Erlbaum; 2004.)

12.4% in men. Results from the associated prospective epidemiology study further support the overall greater prevalence of insomnia in women (Figure 81-1). There was no reliable difference in the sleep (e.g., sleep-onset latency, total sleep time) of men and women with diagnosed insomnia. This study also provided a decade-by-decade breakdown of insomnia prevalence for women and men. With the exception of the decade 30 to 39 years, women exhibited greater prevalence of insomnia than men across the lifespan. Specifically, insomnia prevalence rates for women ranged from 12% (20 to 29 and 30 to 39 years) to 41% (80+ years). For men, rates ranged from 6% (20 to 29 years) to 23% (70 to 79 and 80+ years). Although prevalence rates are generally higher for women, insomnia may be particularly problematic for men because mortality risk has been associated with insomnia and short sleep duration (<6 hours) in men, but not women.[31]

A meta-analysis[32] examined 29 published epidemiologic studies (n = 1,265,015; female subjects = 718,828, male subjects = 546,187) to estimate sex differences in the risk for insomnia. Women exhibited a higher risk ratio (RR = 1.4) for insomnia than men. Women's risk ratio varied depending on the sample size and strength of methodology of the studies included in the analyses. Specifically, women exhibited a higher risk ratio in larger studies with more rigorous methodology (RR = 1.64) than in smaller, less rigorous studies (RR = 1.32). As in prior results,[5] this female preponderance in the risk for insomnia was evident across elderly (65 years or older), middle-aged (31 to 64 years), and young adult (15 to 30 years) age groups. This risk may begin as young as prepubescence; a study[33] of children (n = 700; 5 to 12 years of age) found that the prevalence of insomnia symptoms (controlling for anxiety and depressive symptoms) was 10% to 14% higher in girls 11 to 12 years compared with boys and younger girls.

Education and Socioeconomic Status

Epidemiologic studies have consistently shown that education and socioeconomic status inversely predict sleep quality.[6] In a study[34] of 159,856 U.S. adults (age 18 years or older), Grandner and colleagues found that the odds of sleep complaints were higher for those who were unemployed compared with those who were employed. Amongst the unemployed, men had a higher likelihood of sleep complaints than women. Gender also affected the education and sleep complaint relationship, such that the association between greater sleep complaints and lower education was stronger in men who did not finish high school.

Another study[35] investigated insomnia and related impairment. Persons with lower individual and household education were significantly more likely to experience insomnia and also reported greater insomnia-related impairment. Importantly, education level was related to insomnia even after controlling for race, gender, and age.

Research on the potential role of childhood economic status on insomnia in adulthood has produced mixed results. Lallukka and colleagues[36] found that economic difficulties in childhood were associated with adult insomnia complaints, but these results failed to replicate in another study[37] by the same research group.

Aging

The prevalence and severity of insomnia increases with advancing age.[5,6,38] Reporting at least one insomnia complaint of varying frequency is common in the United States[27,39] and in other countries,[40,41] with a similar prevalence, about 40%. In studies reporting on separate younger and older samples, average yearly incidence rates of chronic insomnia range from

3.6% to 12.3% in older adults,[41-45] whereas young to middle-aged adults report incidence rates ranging from 2.8% to 4.2%.[46,47] One of the studies was longitudinal[43] and determined that incident insomnia increases with age and nearly doubles in the 75 years and older group. Insomnia also appears to be more persistent with age. The odds ratio of having persistent insomnia increases 1.1-fold each decade of life.[44] Annual persistence rates of insomnia symptoms among older adults range from 15.4% to 22.7%.[41,45]

Researchers and clinicians have debated considerably about whether aging is an independent risk factor for insomnia. Evidence suggests that age, per se, does not inevitably herald the incidence and persistence of insomnia, but rather physical and mental health complications account for the correlation.[48,49] Indeed, consistent predictors of insomnia include mood disorders, poor perceived health, bodily pain, cardiovascular disease, respiratory symptoms, stressful life events, somatic complaints, and memory problems.[41,42,45]

Race

In the last couple of decades, the role of race/ethnicity as an important explanatory factor is a growing topic of interest given that data indicate there are sleep health inequities by ethnic group. For example, one study found that ethnicity may account for 20% of the variance in insomnia among adult women.[50] Most epidemiologic data on race/ethnicity and insomnia have focused on comparing non-Hispanic whites and non-Hispanic blacks. A meta-analysis of 13 studies comparing insomnia symptom prevalence found that whites were more likely to report difficulty maintaining sleep (Hedges g = –0.19; 95% confidence interval [CI] = –0.21, –0.17) and early morning awakenings (Hedges g = –0.07; 95% CI = –0.09, –0.04) than blacks.[51] There was no ethnic difference in difficulty initiating sleep. In the meta-analysis, age and gender were noteworthy moderators. The ethnic differences in insomnia symptoms were attenuated but still significant among samples of older adults. In samples of women, the ethnic difference was reduced for difficulty maintaining sleep, yet larger for early morning awakenings. Studies not included in this meta-analysis suggest that ethnic differences in insomnia symptoms are mixed.[52,53]

These discrepancies in findings may be partly accounted for by study design and insomnia definition differences. It is also possible that the interpretation of questions about sleep varies with ethnicity. Moreover, unexplored moderation by age, gender, and socioeconomic status may also be a factor.[32-35] For example, middle-aged blacks tend to have a higher rate of insomnia symptoms than whites,[17,54] whereas no ethnic differences were found among multiethnic adolescents and young adults.[55,56] Also, black women tend to sleep worse than black men and whites.[27,57]

Available data on ethnic differences in the prevalence of insomnia among other racial/ethnic groups remain minimal. A large, national survey found that blacks, Mexican Americans, and other Hispanic/Latino groups were less likely to report insomnia symptoms than whites.[58] Yet, blacks were more likely to report sleep-onset latencies of longer than 30 minutes. This result indicates that blacks may be at greater risk for insomnia but are likely to deny experiencing symptoms. Asian groups reported comparable likelihood in experiencing insomnia symptoms as whites. Overall, ethnic minority groups appear less likely to report symptoms of insomnia than whites. However, this does not necessarily mean that insomnia is not occurring in these groups, nor does it indicate that these groups are not at increased risk for insomnia, particularly among blacks.

RISK FACTORS

A risk factor is a disorder or characteristic that predicts future poor health status.[59] The behavioral model of insomnia[60] proposes that certain predispositions (i.e., risks) make people more susceptible to developing insomnia. Several studies have also begun to investigate insomnia as a risk factor for the development of other difficulties. Although it is costly and time consuming to collect the data, prospective or longitudinal research designs are the best way to determine risk, assessing presence of a condition or trait of interest (e.g., hyperarousal) before the onset of diseases or events of interest (e.g., insomnia). Cross-sectional and retrospective research designs sometimes purport to assess risk, but these methods are less reliable than the longitudinal studies because of the inability to determine temporal relationships and retrospective bias.[61] One caveat to this statement is where the correlate is a relatively static condition or status (i.e., gender, race, ethnicity, age, genetics).

Next we review primarily prospective studies investigating risk factors for the development of insomnia, as well as the converse, where insomnia serves as a risk factor for the development of other problems and disorders. More static risk factors for insomnia, such as gender and race/ethnicity, were reviewed previously.

Personality

Certain personality characteristics may predispose individuals to insomnia. For instance, cross-sectional studies show that neuroticism, internalization, and perfectionistic traits are associated with insomnia. Unfortunately, very little longitudinal data exist,[62] other than recent studies finding social introversion and low ego strength were predictive of incident insomnia.[63,64]

Hyperarousal

All of the models of insomnia (i.e., physiologic, behavioral, cognitive, and neurocognitive) assume some degree of hyperarousal as a predisposition or risk factor for the onset of insomnia. To date, virtually all of the evidence for this relationship has been cross-sectional.[65] The absence of longitudinal evidence to explicate these associations makes it impossible to delineate causal paths. It is possible that individuals with hyperarousal are predisposed to developing insomnia, but it is also possible that hyperarousal is a downstream effect of having insomnia.

Brief Insomnia Episodes

Although long neglected by researchers, three recent reviews of what is termed *acute* or *transient insomnia* are helping to focus attention on its potential clinical significance as a signal of insomnia predisposition.[66-68] Transient insomnia may be an important risk factor for chronic insomnia as it increasingly becomes independent of precipitating stress triggers and instigates perpetuating factors such as worry about sleep and maladaptive safety behaviors. Methodologic limitations that contaminate the findings on insomnia cited earlier, such as

inconsistent definitions of insomnia and unreliability of retrospective sleep reports, also taint the literature on transient insomnia.

Genetics

Familial Risk

Familial aggregation is a type of statistical analysis in which one determines whether a disorder (i.e., insomnia) occurs more commonly in members of a family than among nonrelated persons. To date, several studies of familial aggregation of insomnia have been reported, using a variety of different methodologies (e.g., adults with insomnia, children with insomnia, no-insomnia controls), definitions of insomnia and insomnia symptoms (e.g., severity scales, DSM-III, DSM-IV), and degree of control for potential confounding factors (e.g., age, gender, work schedules, comorbidities).[69,70] In one of the earliest studies, Hauri and Olmsted[71] found that 55% of adults whose insomnia started in childhood had a family member with insomnia compared with 39% of those whose insomnia started in adulthood, suggesting a stronger genetic predisposition for childhood-onset insomnia. Later studies, using similar methods but with less rigorous definitions of insomnia, have found rates of 20% to 73% of family history,[72-76] with this rate sometimes being higher in mothers than fathers.[72,73,76] In studies in which normal sleeping control groups were used, continued inconsistent results have been found. Rates have varied from 38% to 73% of people with insomnia having a relative with insomnia compared with 23% to 29% of their normal controls.[72,75,77]

Twin Studies

Several twin studies have now reported heritability of a composite insomnia-type syndrome of 28% to 57%.[78-83] A few of these studies also reported results on the frequency of specific insomnia symptoms. The adult studies reported heritability rates of 28% to 32% for sleep-onset difficulties and 33% to 45% for sleep maintenance difficulties.[78,79] The child study was more variable depending on parent (79%) or child (17%) report, both of which were somewhat questionable given the age of the children (8 years old) and the potential biases of parental reporting.[84]

Several research groups have now attempted to take the next step and identify specific genetic mechanisms or polymorphisms that may underlie the risk for developing insomnia or be a perpetuating mechanism. The most commonly studied genes are the period circadian clock gene (i.e., *PER*, *CLOCK*) and serotonin transporter genes (i.e., *5-HTTLPR*).[70] As with other areas of research in the epidemiology and risk domain, results are likely to be contaminated by heterogeneity of diagnostic criteria and insomnia type. Future research would benefit from greater effort to secure subject type purity (e.g., no history of insomnia vs. chronic insomnia).

Psychological and Psychiatric Comorbidities

Examining the risk relationship between insomnia and psychological disorders is complicated. Insomnia is a symptom of many psychological disorders, raising the possibility that these disorders cause insomnia, but there appears to be even more research showing that insomnia is a risk factor for the development of these disorders. This has led some to hypothesize a reciprocal relationship between insomnia and psychological disorders.[85]

Depression

Sleep problems are some of the most prevalent complaints in patients with major depressive disorder, with as many as 84% reporting insomnia versus on average 10% to 30% of the general population.[22,86,87] Several studies have now shown that depression or depressive symptoms increase the risk (odds ratio [OR] = 1.1 to 8.6) of developing insomnia across multiple age groups and populations.[28,44] Interestingly, only 23% to 29% of patients report their insomnia symptoms start after the occurrence of a mood disorder, whereas 41% to 69% report insomnia symptoms started before the depression, and 8% to 29% report they started at the same time.[88,89] A large body of research has established that insomnia is a significant risk factor for depression, as seen in a recent meta-analysis. Insomnia resulted in nearly two times greater risk (OR = 1.10 to 3.51) for developing depression.[90]

Suicide

There is considerable evidence that insomnia is a risk factor for suicide. In a recent meta-analysis, insomnia symptoms were significantly associated with suicidal ideation, attempt, or completion (OR = 1.63 to 2.41) after adjusting for comorbidities.[91] Augmenting the insomnia influence, nightmares[92] and hypnotic use[93] are independent risk factors for suicidality.

Anxiety

Insomnia is also closely associated with anxiety disorders, with as many as 64% of patients with generalized anxiety disorder reporting insomnia symptoms.[94] As many as 44% to 73% of patients report their insomnia symptoms started *after* the occurrence of an anxiety disorder, but only 16% to 18% reported their insomnia started *before* the anxiety.[88,89] Studies indicate that anxiety disorders are a significant risk factor for the development of insomnia in both adults (RR = 1.39 to 4.24) and adolescents (RR = 2.3 to 5.5).[88]

Insomnia is also a risk factor for the development of anxiety disorders. One review of the literature found that people with insomnia are 1.97 to 6.3 times more likely to develop an anxiety disorder than people without insomnia.[85] More recently, Morphy and colleagues[44] found that having insomnia at baseline increased the risk for developing incident anxiety disorders by 1.43 to 3.64 times, after controlling for age, sex, socioeconomic status, and baseline depression and pain.

Posttraumatic Stress Disorder

Approximately 41% to 47% of people with posttraumatic stress disorder (PTSD) report symptoms of insomnia.[95] To date, very few longitudinal studies have examined the relation between insomnia and PTSD. Two recent prospective studies produced similar results. Among Palestinians living amid violent political turmoil[96] and deployed U.S. soldiers,[97] PTSD did not predict insomnia, but insomnia did predict PTSD.

Substance Use and Abuse

One early review found that people with insomnia symptoms at baseline were 2.35 times more likely to develop alcohol abuse or dependence disorders ($d = .02$ to $.05$), and 7.18 times more likely to develop drug abuse or dependence disorders ($d = .07$) than people without insomnia symptoms at baseline.[85] These results were not replicated in a study examining insomnia risk in adolescents for predicting substance use as

young adults, after controlling for the higher report of substance use in adolescents with insomnia than adolescents without insomnia (OR = 1.26 to 2.85).[98] A more recent review found that the insomnia symptom of difficulty falling asleep was a consistent predictor of relapse among persons recovering from alcohol addiction.[99]

Medical Comorbidities

Medical comorbidities are ubiquitous in insomnia. One could argue that even though this is based mostly on cross-sectional data, when insomnia prevalence runs two to four times the population prevalence (see Table 81-3), as is the case in many severe medical conditions, these conditions are likely a risk factor for insomnia. The few longitudinal studies that have investigated medical status have found that medical disorders in general and having greater than one medical disorder are risk factors for the development of insomnia (OR = 1.3 to 3.8).[28,63,64] One study found that baseline kidney/bladder problems and migraines, and to a lesser extent allergy/asthma and anemia, were risk factors for incident insomnia, but these significant results were lost after controlling for baseline mental health.[63] Other prospective studies of cancer patients assessed before curative surgery and 2 to 18 months later showed that 18.6% and 14.4%, respectively, developed new-onset insomnia.[100,101]

Even fewer studies have investigated whether insomnia is a risk factor for the development of medical disorders. This is surprising considering that people with severe insomnia report more medical problems, have more physician office visits, are hospitalized twice as often, and use more medications compared with good sleepers.[102] Also relevant is the experimental demonstration that short sleep increases the risk for catching a cold.[103]

Cardiovascular Disease

Perhaps the most thoroughly researched area of insomnia as a risk factor is cardiovascular disease. Studies of the risk of insomnia on hypertension have been mixed. One study (n = 4913) of subjects stratified by age found that baseline insomnia was predictive of hypertension at follow-up (hazard ratio [HR] = 1.01 to 1.09), and these authors hypothesized that insomnia might mediate the relationship between depression and subsequent hypertension diagnosis.[104] A similar study of 4794 Japanese male workers found that sleep-onset and maintenance insomnia symptoms were significant risk factors for development of hypertension (OR = 1.42 to 2.70 and 1.45 to 2.45, respectively).[105] These results were replicated in an even larger study of 8757 participants without hypertension at baseline, which found that difficulty initiating or maintaining sleep predicted slightly increased risk for hypertension (OR = 1.03 to 1.3).[106] However, another relatively smaller cohort of 1419 older adults (≥64 years old) who were not hypertensive at baseline found that insomnia symptoms did not predict hypertension at follow-up, but instead appeared to be a protective factor in white men.[107]

Studies of cardiovascular disease have more consistently found that insomnia symptoms are risk factors. One study of Taiwanese adults found insomnia frequency, as a proxy for severity, was a significant risk for the development of cardiovascular disease and all-cause death (OR = 1.03 to 3.08 and 1.16 to 2.49, respectively).[108] Another study found that sleep-onset insomnia symptoms were a risk for coronary artery disease death in males (RR = 1.5 to 6.3), but not in females.[109]

In a study of 11,863 participants without cardiovascular disease at baseline, only the combination of three sleep complaints (difficulty initiating and maintaining sleep and awakening fatigue) predicted a slight increase of cardiovascular disease (OR = 1.1 to 2).[106]

The largest study to date followed 52,610 participants over 11.4 years to determine whether insomnia increased the risk for an acute myocardial infarction (AMI) identified at hospitals or by the National Cause of Death Registry. After adjusting for numerous health risk factors, there was a moderately increased risk for AMI for people with difficulties initiating (HR = 1.18 to 1.80) or maintaining (HR = 1.01 to 1.68) sleep almost every night compared with people who never experienced these sleep difficulties.[110]

Stress and Life Events

Models of insomnia commonly assume that high stress is a risk factor for the onset of insomnia, but this area has actually received mixed support. Stressful life events have been shown to be only mildly related to the incidence of insomnia.[111,112] Several studies indicate that reactivity to life events is the more important factor in the development of insomnia.[64,112,113] For example, one study found that baseline arousal predisposition, but not baseline perceived stress, predicted incidence of insomnia at follow-up, after controlling for baseline demographic, psychological, and health factors.[64]

Work

Evidence has been steadily growing on the reciprocal relation between work stress and insomnia. Several studies have now identified work stress as a significant predictor of insomnia. For instance, high work demands increased the risk for developing insomnia.[114] People with insomnia also reported poor self-esteem at work, less job satisfaction, and less efficiency at work compared with good sleepers.[115] In turn, insomnia appears to take a considerable toll on worker quality of life and productivity. Insomnia is a significant risk for absenteeism, lost productivity, and disability.[115-117]

CLINICAL PEARLS

- Methodologic faults, primarily inconsistent definitions of insomnia and single-point sleep assessments, corrupt the epidemiology literature on insomnia.
- Advancing age, female gender, and low socioeconomic status are the strongest correlates of insomnia, although it is premature to infer causal influence. Hyperarousal in its many forms and comorbidity are the most common insomnia risk factors, but the mediator or moderator influences associated with this mechanism are poorly understood, as are causal paths.
- Insomnia is itself a risk factor for medical and psychiatric disorders, including depression, hypertension, and substance abuse.

SUMMARY

The prevalence of chronic, clinically significant insomnia is about 10%, but poor sleep complaints occur in as many as 40% of the population. Attributes such as advanced age, female gender, and low socioeconomic status, as well as hyperarousal and psychiatric or medical comorbidity, are the most frequently observed insomnia risk factors, although causal paths are

poorly understood. Transient insomnia and genetic factors may also emerge as important risk factors. Insomnia is itself a risk factor for a wide variety of psychological, psychiatric, and medical disorders, including depression, substance abuse, and hypertension. The highest priorities for future research are a standard definition of insomnia, longitudinal studies, and more attention to monitoring psychological, psychiatric, and medical comorbidity, race, and genetic variables.

Selected Readings

Bonnet MH, Arand DL. Hyperarousal and insomnia: state of the science. *Sleep Med Rev* 2010;**14**:9–15.

Buysse DJ, Ancoli-Israel S, Edinger JD, et al. Recommendations for a standard research assessment of insomnia. *Sleep* 2006;**29**:1155–73.

Cohen S, Doyle WJ, Alper CM, et al. Sleep habits and susceptibility to the common cold. *Arch Intern Med* 2009;**169**:62–7.

Foley DJ, Monjan A, Simonsick EM, et al. Incidence and remission of insomnia among elderly adults: an epidemiologic study of 6,800 persons over three years. *Sleep* 1999;**22**:S366–72.

Jansson-Fröjmark M, Linton SJ. The course of insomnia over one year: a longitudinal study in the general population in Sweden. *Sleep* 2008;**31**: 881–6.

Kamath J, Prpich G, Jillani S. Sleep disturbances in patients with medical conditions. *Psychiatr Clin North Am* 2015;**38**(4):825–41.

Kanady JC, Soehnera AM, Harvey AG. A retrospective examination of sleep disturbance across the course of bipolar disorder. *J Sleep Disord Ther* 2015;**4**(2):pii, 1000193.

Morphy H, Dunn KM, Lewis M, et al. Epidemiology of insomnia: a longitudinal study in a UK population. *Sleep* 2007;**30**:274–80.

Nardo D, Högberg G, Jonsson C, et al. Neurobiology of sleep disturbances in PTSD patients and traumatized controls: MRI and SPECT findings. *Front Psychiatry* 2015;**6**:134.

Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Med Rev* 2002;**6**:97–111.

Roth T, Coulouvrat C, Hajak G, et al. Prevalence and perceived health associated with insomnia based on DSM-IV-TR; International Statistical Classification of Diseases and Related Health Problems, tenth revision; and Research Diagnostic Criteria/International Classification of Sleep Disorders, second edition criteria: results from the America Insomnia Survey. *Biol Psychiatry* 2011;**69**:592–600.

Rumble ME, White KH, Benca RM. Sleep disturbances in mood disorders. *Psychiatr Clin North Am* 2015;**38**(4):743–59.

Taylor DJ, Lichstein KL, Durrence HH. Insomnia as a health risk factor. *Behav Sleep Med* 2003;**1**:227–47.

Taylor DJ, Mallory LJ, Lichstein KL, et al. Comorbidity of chronic insomnia with medical problems. *Sleep* 2007;**30**:213–18.

Zhao L, Wang E, Zhang X, et al. Cortical structural connectivity alterations in primary insomnia: insights from MRI-based morphometric correlation analysis. *Biomed Res Int* 2015;**2015**:817595.

A complete reference list can be found online at ExpertConsult.com.

Chapter 82

Etiology and Pathophysiology of Insomnia

Michael Lloyd Perlis; Jason Gordon Ellis; Jacqueline DeMichele Kloss; Dieter Wilhelm Riemann

Chapter Highlights

- Since the 1990s there has been a proliferation of theoretical perspectives on the etiology of insomnia that now includes nine human models. The central concepts for the nine models include the following:
  - Stress-diathesis
  - Stimulus dyscontrol and classical conditioning
  - The interaction of basal arousal and sleep requirement
  - Sleep extension and the mismatch between sleep opportunity and ability
  - Altered sensory and information processing and an attenuation of the normal mesograde amnesia of sleep
  - Appraisal as a determinant of the patient's perception of disease
  - The concept of "the inhibition of sleep-related dearousal" (vs. hyperarousal)
  - The role of attention, intention, and effort
  - The etiologic importance of daytime deficits, selective attending to sleep-related threats, and safety behaviors
  - Chronic insomnia as a hybrid state that occurs in association with local neuronal wakefulness during non–rapid eye movement and rapid eye movement sleep

Until the late 1990s there were only two models regarding the etiology and pathophysiology of insomnia. The relative lack of theoretical perspectives was due to at least three factors. First, the widespread conceptualization of insomnia as owing directly to hyperarousal (levels of physiologic or central nervous system arousal that are sufficiently high as to directly prohibit sleep) may have made it appear that further explanation was not necessary. Second, the long-time characterization of insomnia as a symptom carried with it the clear implication that insomnia was not itself worth modeling as a disorder or disease state. Third, for those inclined toward theory, the acceptance of the behavioral models (i.e., the three-factor model [3P] and the stimulus control model[1,2]) and the treatments that were derived from them might have had the untoward effect of discouraging the development of alternative or elaborative models. Since the 1990s there has been a proliferation of theoretical perspectives on the etiology and pathophysiology of insomnia that includes both human and animal models. In this chapter, nine of the human models are described and critiqued. The models presented span from the classical behavioral perspectives, to the traditionally cognitively focused frameworks, to the more modern cognitive information–processing perspectives, to an interaction paradigm that takes into account basal arousal and sleep requirement, to the neurocognitive and neurobiologic models that essentially frame insomnia, from a functional and neurophysiologic point of view, as a hybrid state (part wake and part non–rapid eye movement [NREM] sleep).

DEFINITION OF INSOMNIA

The *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5)[3] and *International Classification of Sleep Disorders,* third edition (ICSD3[4]) define *insomnia disorder* as difficulty initiating or maintaining sleep on three or more nights per week for at least 3 months. This definition further stipulates that the diagnosis of insomnia must take into account sleep opportunity, level of daytime impairment and distress, whether symptom presentation (in the case of children and elders) varies with caregiver presence, and the possibility that the insomnia is not better explained by (or does not occur exclusively during the course of) other sleep disorders or medical or psychiatric illnesses.

This definition is different from the DSM-IV-TR and the ICSD2 in several important ways. First, the diagnostic terms primary insomnia and secondary insomnia have been replaced to reflect the change that insomnia is now viewed as a disorder, regardless of whether it is comorbid with other disorders. Second, although quantitative values are not given for insomnia severity (i.e., that sleep latencies or wake after sleep onset durations must be greater than some minimum duration to be of clinical significance), insomnia frequency and chronicity are explicitly stated. The frequency criterion is new, and the

chronicity criterion has been changed from 1 month to 3 months. Third, nonrestorative sleep has been removed as a diagnostic criterion for insomnia. Finally, both the DSM-5 and the ICSD3 allow for the identification of insomnia types (different modes of clinical presentation including initial, middle, and late insomnia), but only the ICSD3 specifies insomnia subtypes, including idiopathic, psychophysiologic, and paradoxical insomnia.

MODELS OF INSOMNIA

Nine models of insomnia are critically reviewed in this chapter. Each of the models described and critiqued presents a view of how insomnia develops, becomes chronic, or comes to be self-perpetuating. Basic and experimental models are not reviewed in this chapter. Information about these may be found in the prior edition of this text (fifth edition). The review is organized chronologically. Following the explication of each of the nine models, a discussion is provided regarding the issues that have not been well integrated into existing models, including how dysregulation of normal sleep-wake regulatory processes may contribute to the development, incidence, or severity of chronic insomnia and why it may be that chronic insomnia occurs disproportionately in women and older adults. The final section of the chapter provides an integrative or transtheoretical perspective represented by a parallel process model that attempts to illustrate how all of the models contribute something unique to our understanding of the etiology of insomnia.

Stimulus Control Model (1972)

Basic Description

Stimulus control, as originally described by Bootzin in 1972,[2,5] is based on the behavioral principle that one stimulus may elicit a variety of responses, depending on the conditioning history. A simple conditioning history, wherein a stimulus is always paired with a single behavior, yields a high probability that the stimulus will yield only one response. A complex conditioning history, wherein a stimulus is paired with a variety of behaviors, yields a low probability that the stimulus will elicit only one response. In individuals with insomnia, the normal cues associated with sleep (e.g., bed, bedroom, bedtime) are frequently paired with behaviors other than sleep. For instance, in an effort to cope with insomnia, the patient may spend a large amount of time in the bed and bedroom awake and engaging in behaviors other than sleep. These coping behaviors appear to the patient to be both reasonable (i.e., staying in bed at least permits the patients to get "rest") and reasonably successful (i.e., engaging in alternative behaviors in the bedroom sometimes appears to result in cessation of the insomnia). These practices, however, set the stage for stimulus dyscontrol, that is, reduced probability that sleep-related stimuli will elicit the desired response of sleepiness and sleep. Figure 82-1 provides a schematic representation of stimulus control and stimulus dyscontrol.

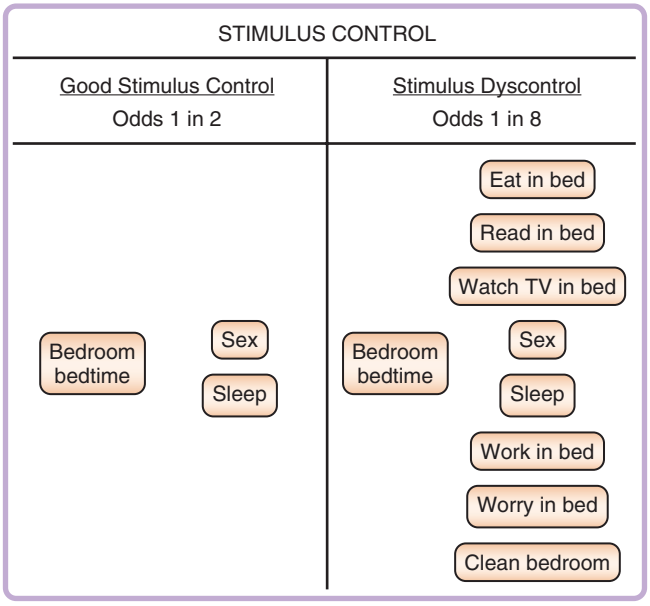


Figure 82-1 The Stimulus Control Model. The schematic represents the instrumental conditioning perspective on stimulus control. In the *left frame* (good stimulus control) the bedroom is tightly coupled with sleep and sex where, given the orthogonality and equal probability of events, the probability of association of bedroom to sleep is 1 in 2. In the right frame (stimulus dyscontrol) the bedroom is no longer a strong associate of sleep and sex where, given the orthogonality and equal probability of events, the probability of association of bedroom to sleep is 1 in 8. The treatment implication of stimulus dyscontrol is that the voluntary elimination (hence instrumental conditioning) of the nonsleep associates except for sex should make it more likely that sleep will occur in the bedroom.

Strengths and Limitations

The treatment derived from stimulus control theory is one of the most widely used behavioral treatments, and its efficacy has been well established.[6-10] The success of the therapy, however, is not sufficient evidence to say that stimulus dyscontrol is responsible for predisposition to, precipitation of, or perpetuation of insomnia.* In fact, one investigation found that the reverse of stimulus control instructions also improved sleep continuity.[11] Another limitation of the stimulus control perspective is that it focuses on instrumental conditioning. That is, there are behaviors that reduce or enhance the probability of the occurrence of sleep. The original model does not explicitly delineate how classical or Pavlovian conditioning may also be an operational factor. Specifically, the regular pairing of the physiology of wakefulness with sleep-related stimuli may lead to sleep-related stimuli becoming conditioned stimuli for wakefulness. This latter possibility, although not part of the classical stimulus control perspective, is clearly consistent with it.

Implications for Current and Future Research and Therapeutics

Given the efficacy of stimulus control therapy, it would be useful to determine how much of the treatment outcomes with cognitive behavior therapy for insomnia (CBT-I) result from the manipulation of this factor. One way to assess the relative importance of stimulus control would be as part of a dismantling study. This alone would not, however, confirm the importance of stimulus dyscontrol as a perpetuating factor. Perhaps what is needed is a dismantling study for stimulus

*The conceptual frame for causality in terms of "predisposition, precipitation, and perpetuation" was first articulated by Spielman as part of the three-factor model. It is used in this context for its general explanatory value.

control itself, given that the treatment contains not only the instruction to limit activities in the bedroom to sleep and sex but also instructions to go to bed only when sleepy, leave the bedroom when awake, get up at the same time every morning irrespective of how much sleep is obtained, and not nap during the day. Any of these components, alone or in combination, may account for the efficacy of stimulus control and may do so through mechanisms other than stimulus dyscontrol, such as sleep homeostasis dysregulation. This is particularly true for the prescription to get up at the same time every morning irrespective of how much sleep is obtained. When patients are compliant with this instruction, it prevents the deleterious effects of excessive time in bed (see Spielman's three-factor model discussed next) and may ensure that sleep loss will prime for better sleep on subsequent nights.[12] In a related vein, the instruction to leave the bedroom when awake may serve as a means of ensuring that patients are fully awake (vs. microsleeping) and thus may also improve sleep through sleep homeostatic or circadian processes.

Three-Factor Model (1987)

Basic Description

This model, alternatively referred to as the Spielman model, 3P model, or behavioral model, delineates how insomnia occurs acutely and how acute insomnia becomes both chronic and self-perpetuating[1] (Figure 82-2). The model is based on the interaction of three factors. The first two factors (predisposing and precipitating factors) represent a stress-diathesis conceptualization of how insomnia comes to be expressed. The third factor (perpetuating factor) represents how behavioral considerations modulate chronicity. Predisposing factors extend across the entire biopsychosocial spectrum. Biologic factors include, for example, the genetic predisposition for insomnia or related etiologic factors, increased basal metabolic rate, hyperreactivity, and fundamental alterations to the neurotransmitter systems associated with sleep and wakefulness. Psychological factors include worry or the tendency to be excessively ruminative. Social factors, although rarely a focus at the theoretical level, include factors such as the bed partner keeping an incompatible sleep schedule or social pressures to sleep according to a nonpreferred sleep schedule (e.g., child rearing). Precipitating factors are acute occurrences that trigger sleep continuity disturbance.[†] The primary "triggers" are thought related to life stress events, including medical and psychiatric illness. Perpetuating factors refer to the behaviors adopted by the individual that are intended to compensate for or cope with sleeplessness, but that actually reinforce the sleep problem. Perpetuating factors include the practice of nonsleep behaviors in the bedroom, staying in bed while awake, and spending excessive amounts of time in bed. Stimulus control speaks to the first two of these (as reviewed earlier). The classic version of the three-factor model focuses primarily on the last of these. Excessive time in bed (or sleep extension) may involve going to bed early, getting out of bed late, and napping as ways of coping with insomnia. Such compensatory behaviors are enacted to increase the opportunity to get more sleep and are likely to be highly self-reinforcing because they allow lost sleep to be "recovered" and the daytime effects of lost sleep to be ameliorated. Extension of sleep opportunity can lead to a mismatch between sleep opportunity and sleep ability.[1,13] The greater the mismatch, the more likely the individual will spend prolonged periods of time awake during the given sleep period, regardless of what factors predisposed to or precipitated the insomnia.

The three-factor model and its graphic representations have periodically been updated by Spielman.[14] As shown in Figure 82-3, one version of the model represents the speed of onset and offset of events in the development of insomnia, and a 4P representation takes into account Pavlovian (classical) conditioning as a perpetuating factor. Classical conditioning refers to the reliable elicitation of specific physiologic responses by what were once neutral stimuli. In the context of insomnia, classical conditioning refers to the elicitation of arousal or wakefulness in response to what were once sleep-related stimuli. This phenomenon corresponds to the common patient report that "it's as if I just walk into the bedroom and I am suddenly wide awake … it's like some switch got flipped from sleepy to wide-awake."

Strengths and Limitations

The three-factor model is conceptually appealing and comports well both with clinical experience and with the two-process model of sleep-wake regulation.[15] The model has good face validity for both patients and clinicians, and the therapy derived from the model (sleep restriction) is efficacious. This said, there have been very few studies evaluating sleep restriction therapy as a monotherapy[16] and no studies evaluating the relative efficacy of sleep restriction therapy as a component of CBT-I (i.e., no dismantling studies). It is therefore difficult to assess the extent to which treatment efficacy supports the model. Further, even if studies could show that sleep restriction therapy accounted for most clinical gains with CBT-I, formal validation of the model would still require a natural history study showing that the transition from acute to chronic insomnia is largely mediated by sleep extension.

Another limitation of the original model is the implication that the predisposition for insomnia varies across individuals

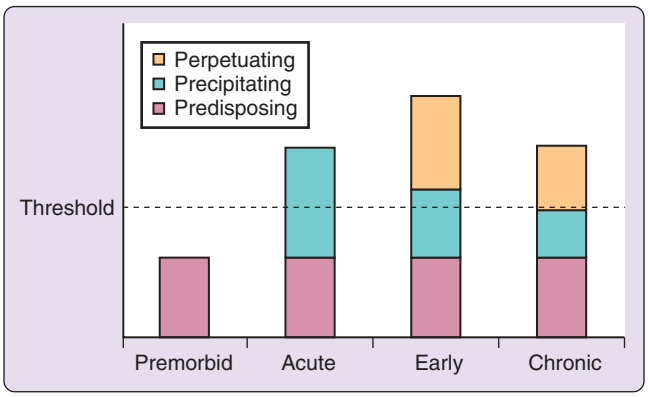


Figure 82-2 The 3P Model. This schematic represents the classic 1987 rendition of the 3P model. There are two more recent representations in Figure 82-3. The reader is encouraged to compare the three versions of the model.

[†]Sleep continuity is meant to denote the class of variables that include sleep latency, number of awakenings, wake after sleep onset, total sleep time, and sleep efficiency (i.e., sleep architecture vs. sleep continuity).

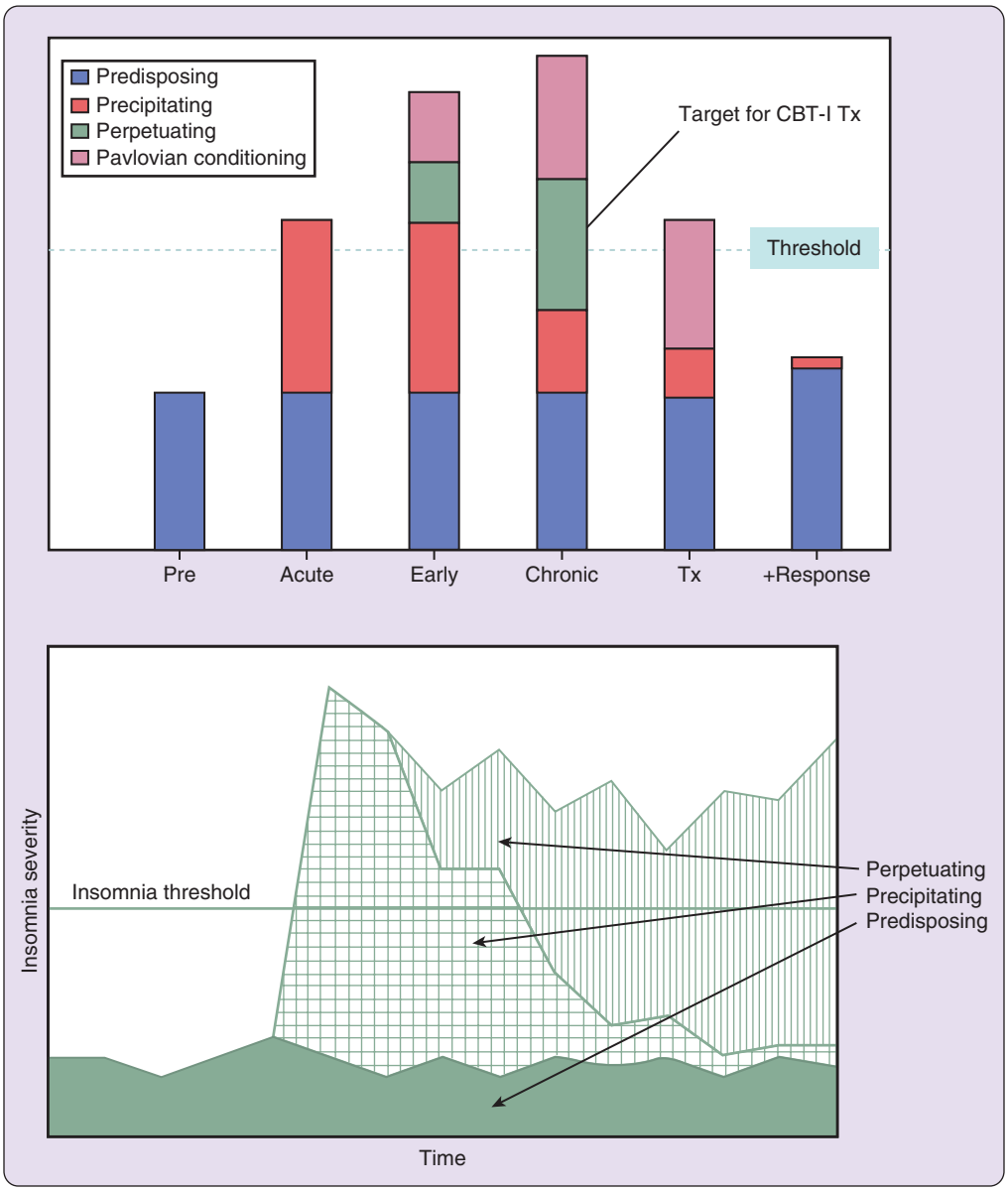


Figure 82-3 The Dynamic 3P and the 4P Models. The dynamic 3P model has the added value (compared with the original model) of illustrating the temporal course of each of the factors. The 4P model has the added value of explicitly incorporating Pavlovian conditioning and how this factor affects the clinical course of insomnia and response to treatment (Tx). CBT-I, Cognitive behavior therapy for insomnia.

but is a trait factor within the individual. With respect to between-subjects variability, presumably this means that some individuals are not prone to insomnia, some are marginally at risk, and still others are at high risk. Although it stands to reason that the vulnerability for insomnia exists on a continuum, it is also plausible that *all individuals* are at risk for insomnia (acute insomnia) and that this may be so to the extent that insomnia represents an adaptive response to stress (i.e., a real or perceived threat, as part of the flight-fight response, triggers a systemic response that overrides the normal homeostatic and circadian imperatives for sleep). Although some predispositions may be indeed be "hard-wired," some predispositions vary over the life span (e.g., new sleep environments or partners, pregnancy or child rearing, altered hormonal status, aging effects, prior insomnia experience). The newer rendition of the Spielman model explicitly allows predisposing factors to vary with time.[14]

Implications for Current and Future Research and Therapeutics

Despite its heuristic and practical value, most tenets of the three-factor model have not been empirically tested. Several avenues for research are possible. Predisposition to insomnia could be, and has recently been, evaluated using molecular and behavioral genetic approaches.[17-20] As a complement to this approach, medical anthropologic studies could be used to assess vulnerability to insomnia at the cultural level (e.g., industrial vs. nonindustrial societies). The precipitation of insomnia, while evaluated with stress induction studies in good sleepers,[21] has not been studied prospectively to

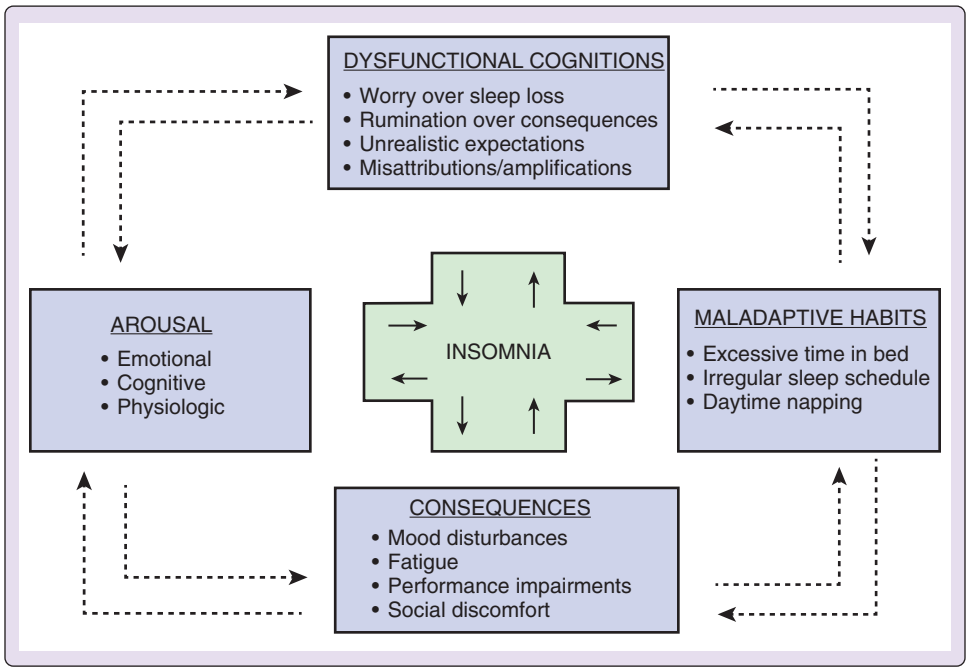


Figure 82-4 The Microanalytic Model. As noted in text, this is primarily a state model that focuses on how insomnia may be self-perpetuating. Note that arousal is arrayed as occurring within three domains: emotional, cognitive, and physiologic.

determine what events reliably trigger acute insomnia. Finally, as noted earlier, natural history studies (with high temporal resolution) are now assessing whether the putative perpetuating factor of sleep extension does indeed mediate the transition from acute to chronic insomnia.

The three-factor model has served as the conceptual basis for one treatment modality in particular: sleep restriction. This therapy, although believed by many to be the single most potent component of CBT-I, was developed to target one particular perpetuating factor, sleep extension. In multicomponent CBT-I, other treatment components may address other perpetuating factors (e.g., stimulus control addresses the engagement of nonsleep behaviors in the bedroom and the tendency to remain in bed when awake; cognitive therapy addresses the problem of catastrophic or dysfunctional thinking about insomnia; sleep hygiene addresses the misuse of counterfatigue measures). The three-factor model may also help to identify alternative treatment targets for insomnia. For instance, this model could help guide the development or adaptation of existing therapies to target predisposing factors. Such treatments could be used to increase treatment response, diminish the risk for recurrence (as an adjuvant to traditional CBT-I), or prophylactically to prevent first episodes of insomnia.

Microanalytic Model (1993)

Basic Description

Morin put forward the microanalytic model in his seminal book, *Insomnia: Psychological Assessment and Management*[22] (Figure 82-4). This model suggests that four bidirectional factors account for the perpetuation of insomnia over time (i.e., how it is that insomnia is self-perpetuating through arousal, dysfunctional cognitions, consequences, and maladaptive habits.) Arousal is conceptualized in terms of emotional, cognitive, and physiologic components. Dysfunctional cognitions are construed in terms of worry, rumination, and unrealistic expectations about sleep and sleep loss. Consequences refer to the negative psychosocial outcomes that occur with insomnia. Maladaptive habits refer to behaviors such as excessive time in bed, irregular sleep-wake schedules, and napping, each of which presumably occurs in relation to the effort to recover lost sleep. Central to this model is the concept that each occurrence of insomnia has consequences, including increased arousal, and results in the engagement of cognitions and behaviors that prolong the index episode or increase the likelihood of additional occurrences of insomnia.

Strengths and Limitations

Major strengths of the microanalytic model are that it posits that insomnia is a self-reinforcing phenomenon that will continue unabated without an adaptive response on the part of the individual or the provision of treatment; construes arousal along multiple dimensions; incorporates the central behavioral concept of excessive time in bed (i.e., the behavioral maladaptation of sleep extension); and implies that adaptive responses or treatment can target any of the four contributory factors. This model, however, is not an etiologic model; that is, it does not delineate how the first episode of acute or chronic insomnia occurs.

Implications for Current and Future Research and Therapeutics

The delineation of the four contributory factors provides a conceptual basis for the assessment of the relative contribution of each to the occurrence and severity of chronic insomnia. To our knowledge, no such study has been conducted on this topic. The clear therapeutic implication of the model is that treatment of insomnia may benefit from adopting a

multicomponent approach. This is appropriate given that the model was introduced as part of a treatment manual (the first of its kind) that delineated a multicomponent treatment approach to insomnia. Given the centrality of dysfunctional cognitions to the model, the form of CBT-I standardized by Morin and colleagues[22] has a cognitive component that is dedicated to the assessment and treatment of dysfunctional beliefs and attitudes about sleep.[23] To date, there is evidence that dysfunctional beliefs and attitudes about sleep vary with treatment outcome with CBT-I[24]; whether monotherapy with this component of CBT-I is effective is not known.

Neurocognitive Model (1997)

Basic Description

The neurocognitive model (Figure 82-5) is based on and is an extension of the 3P and 4P models[25] (see Figure 82-3). The central tenets of neurocognitive model include (1) a pluralistic perspective of hyperarousal (cortical, cognitive, and somatic arousal); (2) the specification that cortical arousal (as opposed to cognitive or somatic arousal) is central to the etiology and pathophysiology of insomnia; (3) the proposition that cortical arousal, in the context of chronic insomnia, occurs as a result of classical conditioning and is permissive of cognitive processes that do not occur with normal sleep; (4) the proposition that sleep initiation and maintenance problems do not occur because of hyperarousal but because of increased sensory and information processing at sleep onset and during NREM sleep; and (5) the suggestion that sleep state misperception derives from increased sensory and information processing during NREM sleep or the attenuation of the normal mesograde amnesia of sleep.

As with the 3P and 4P behavioral models of insomnia, the neurocognitive model posits that acute insomnia occurs in association with predisposing and precipitating factors and that chronic insomnia occurs in association with perpetuating factors. Like the 3P model, chronic insomnia is perpetuated by the instrumental conditioning that occurs with sleep extension. Like the 4P model, the neurocognitive model posits that classical conditioning also serves as a perpetuating factor for chronic insomnia; that is, the repeated pairing of sleep-related stimuli with insomnia-related wakefulness (arousal) ultimately causes sleep-related stimuli to elicit (or maintain) higher than usual levels of cortical arousal at around sleep onset or during the sleep period. This form of arousal is, in the context of chronic insomnia, thought to be independent of somatic arousal; the biological substrate for, and precipitant of, cognitive arousal; and the form of arousal that directly contributes to sleep continuity disturbance and sleep state misperception. In the case of sleep continuity disturbance and sleep state misperception, cortical arousal is not necessarily antithetical to sleep but exerts its deleterious effects through enhanced sensory processing, enhanced information processing, and long-term memory formation.

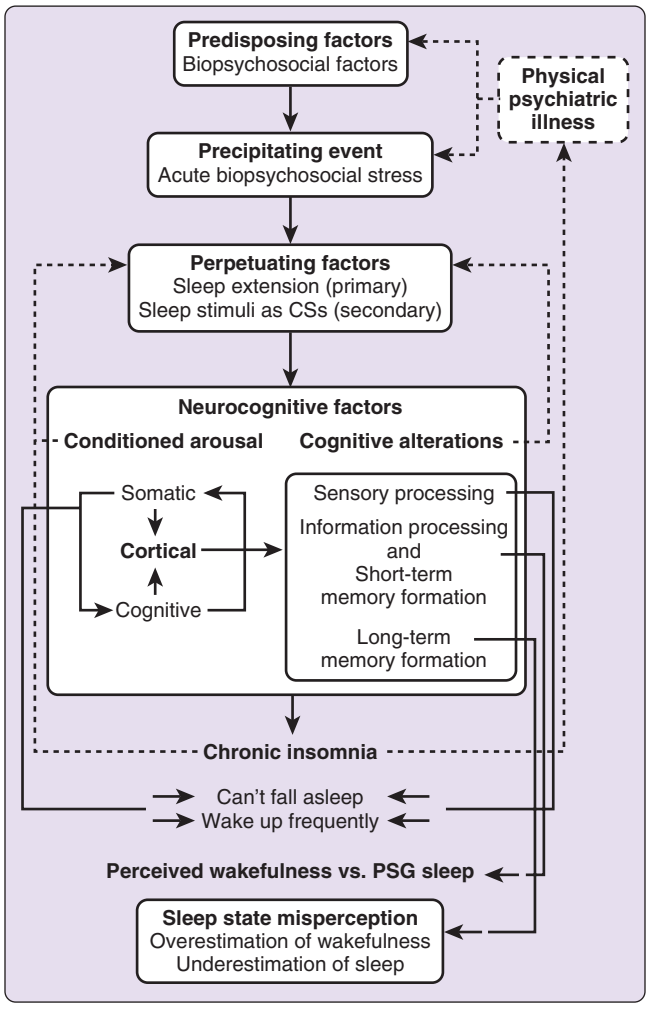


Figure 82-5 The Neurocognitive Model. The schematic is different from prior publications of the neurocognitive model in several ways: (1) *Dotted lines* are provided to highlight feedback loops (*solid lines* represent feed-forward loops). (2) The examples provided for perpetuating factors have been changed. The primary factor is designated as "sleep extension" (previously denoted as increased time in bed and staying awake in bed). The secondary factor is designated as "sleep stimuli as CSs." This is meant to represent when "sleep stimuli" become conditioned stimuli for wakefulness (arousal). The section of the diagram denoted "neurocognitive factors" may well correspond to the "persistence of wakefulness" (when such events occur before sleep onset proper) and the "failure to inhibit wakefulness" (when such events occur during NREM sleep). The latter may correspond to what is characterized by Cano and Saper as a "hybrid state" (not entirely sleep or wakefulness) and may be accounted for by "local neuronal wakefulness" as posited by Buysse and colleagues. PSG, Polysomonography.

Enhanced sensory processing (detection of endogenous or exogenous stimuli and, potentially, the emission of startle or orienting responses) around sleep onset and during NREM sleep is thought to directly interfere with sleep initiation or maintenance.

Enhanced information processing (detection of and discrimination between stimuli and the formation of a short-term memory of the stimulating events) during NREM sleep is thought to blur the perceptual distinction between sleep and wakefulness and thus contribute to sleep state misperception.

Enhanced long-term memory (detection of and discrimination between stimuli and recollection of the stimulating event hours after its occurrence) around sleep onset and during NREM sleep is thought to interfere with the subjective experience of sleep initiation and duration and thus contribute to the discrepancies between subjectively and objectively assessed sleep continuity.

Finally, conditioned cortical arousal is hypothesized to be self-reinforcing and thus, like sleep extension, serves to perpetuate insomnia in the absence of the original precipitants. That is, each time sleep-related stimuli (i.e., the specifics of the sleep environment) elicit cortical arousal, this reinforces the potential of sleep-related stimuli to serve as conditioned stimuli for enhanced sensory and information processing or long-term memory formation.

Strengths and Limitations

In general, the major strengths of the neurocognitive model are that it allows for a pluralistic perspective on the concept of arousal; does not require that hyperarousal be so intense as to directly interfere with sleep initiation and maintenance; delineates a mechanism beyond that of instrumental conditioning (i.e., classical conditioning as a perpetuating factor); specifies how chronic insomnia "takes on a life of its own" (i.e., is self-reinforcing); and is based on hypotheses that are falsifiable.

To date, the evidence for the model derives from observations about patients with insomnia compared with good sleepers exhibiting increased cortical or central nervous system arousal using such measures as quantitative electroencephalography[26-30] and positron emission tomography[31-32]; increased sensory or information processing using such measures as evoked response potentials[33-35]; an attenuation of the normal mesograde amnesia of sleep using such measures as implicit and explicit memory tests for semantic stimuli presented during sleep[36]; and an association between sleep state misperception and objective measures of cortical arousal or evoked response potential abnormalities.[37-39]

The primary limitations of the neurocognitive model are that it does not adequately account for the transition from good sleep to acute insomnia, the importance of circadian and homeostatic influences on sleep, which brain regions or circuits are abnormally activated around sleep onset and during NREM sleep, the likely possibility that abnormal activation may also occur in subcortical regions, and the neurobiologic mechanisms by which insomnia may occur as a hybrid state. Some speculations regarding the functional anatomic substrate of the neurocognitive model have subsequently been published.[40]

Implications for Current and Future Research and Therapeutics

Many of the model's central tenets require further empiric validation. For instance, natural history studies are needed to determine whether the transition from acute insomnia to chronic insomnia is accounted for by sleep extension, and laboratory studies are needed to determine whether neurocognitive processes (sensory and information processing and long-term memory formation) are altered before and during the sleep period in patients with chronic insomnia. Further, it must be shown that altered cognitive processing has clear neurobiologic substrates (e.g., altered metabolic activity in specific brain regions or the occurrence of local neuronal wakefulness) and functional consequences (sleep continuity disturbance and sleep state misperception). In short, novel experimental paradigms need to be developed to test the model's core hypotheses.

The neurocognitive model may provide some insight into the potential mechanisms of action of existing therapies and also some guidance regarding potential targets for new treatments. In the case of existing therapies, pharmacotherapy might be effective to the extent that the various compounds block sensory and information processing or promote amnesia for episodic memories formed during the sleep period. This idea, first espoused by Mendelson,[41-47] seems probable given the effects of benzodiazepines and benzodiazepine receptor agonists on arousal thresholds and memory formation. Sleep restriction therapy might also work through these mechanisms to the extent that this treatment modality serves to deepen sleep, which may augment the endogenous form of sleep-related mesograde amnesia. Potential avenues for new medical treatments include the assessment of compounds that have greater than normal amnestic potential for their efficacy as hypnotics, provided that such effects can be limited to the desired sleep period. Given that this is not possible, potent amnestics may be used experimentally to determine the extent to which amnesia for events occurring during the sleep period influences morning recall about sleep continuity and sleep quality. Alternatively, it may be possible to use stimulants during the day (e.g., modafinil) to promote wake extension and thereby their potential to diminish nocturnal cortical arousal through increased sleep pressure. The latter has been attempted with modafinil alone and in combination with CBT-I. Potential avenues for behavioral treatment include protocols that use more intensive forms of sleep restriction to promote counterconditioning, such as intensive sleep retraining therapy.[48]

Two-Factor Model (1997)

Basic Description

Bonnet and Arand propose a two-factor model to account for the incidence of insomnia and hypersomnia[49] (Figure 82-6). As described by the authors,

> *… each individual has his own sleep requirement determined by his sleep system and each individual has a basal level of arousal determined by his arousal system. Sleep deprivation will eventually override the arousal system, but the arousal system can also mask the sleep system. By thinking of these systems as relatively independent, one can dichotomize their effects. (p. 99)*

High basal arousal and a short sleep requirement are posited to account for idiopathic or psychophysiologic insomnia. Psychophysiologic insomnia occurs when the individual attempts to sleep when sleep is not required and the concurrent high level of basal arousal prohibits the obtention of "optional" sleep. Idiopathic insomnia likely represents the same scenario but as a life-long problem. High basal arousal and a long sleep requirement are posited to account for sleep state misperception (paradoxical insomnia). Presumably, sleep state misperception occurs when individuals sleep "because they can" but the concurrent high level of basal arousal results in shallow sleep, which may be perceived as wakefulness.

Strengths and Limitations

The major strengths of this model are that it has a corresponding program of research that empirically assesses and experimentally models the hyperarousal and sleep continuity disturbance that occurs with insomnia; allows for the distinction between primary insomnia (psychophysiologic insomnia), idiopathic insomnia, and sleep state misperception insomnia

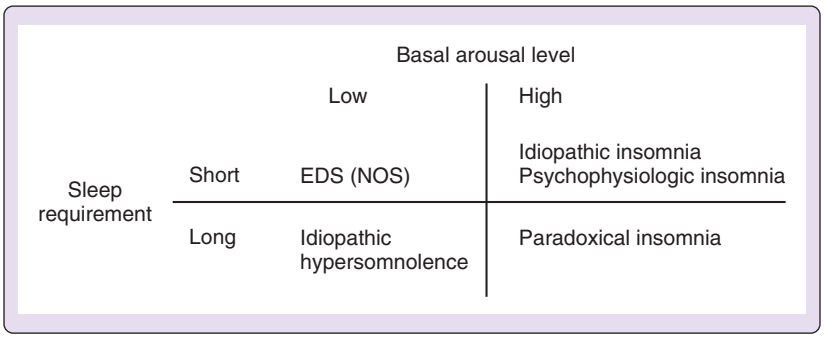


Figure 82-6 The Two-Factor Model. This schematic illustrates how high and low basal arousal and long and short sleep requirement interact to produce six forms of sleep continuity disturbance. EDS, Excessive daytime sleepiness; NOS, not otherwise specified.

(paradoxical insomnia); and takes into account sleep requirement as a moderator of hyperarousal. Programmatic research on this model has included measures of heart rate variability and Vo_2 measures of metabolic rate and experimental studies based on a "yoked insomnia" protocol and a high-dose caffeine paradigm.[50-53] The former provides evidence that patients with primary insomnia do indeed exhibit increased basal metabolic arousal compared with normal controls and patients with sleep state misperception (paradoxical insomnia). The latter provides evidence that many of the symptoms of insomnia can be produced experimentally (i.e., induced hyperarousal or sleep continuity disturbance).

The weaknesses of the model are that it does not directly provide a perspective on the etiology of insomnia (i.e., how basal arousal comes to be or continues to be elevated or how basal arousal varies with time); the concept of sleep requirement appears to combine the concepts of sleep need and sleep ability into a single factor and does not account for the related concept of sleep opportunity (how much sleep individuals attempt to obtain).

Implications for Current and Future Research and Therapeutics

The work related to the definition and measure of arousal has been, and will continue to be, an essential goal for insomnia research. Moving forward, it will be important to determine what type of arousal (e.g., cognitive arousal vs. general metabolic rate vs. global cortical arousal vs. local activation within the central nervous system) and what level of arousal (how much activation) are required to prohibit sleep initiation or maintenance. Further, the delineation of sleep requirement as a potential moderator of insomnia type or subtype will need to be specifically defined, operationalized, and further studied.

The therapeutic implications of the model are clear: arousal and sleep requirement are potential targets for interventions. Currently, basal arousal is targeted pharmacologically with benzodiazepines, benzodiazepine receptor agonists (putatively altering central nervous system arousal through gamma-aminobutyric acid modulation), melatonin agonists (presumably affecting propensity for wakefulness), histamine antagonism (depotentiation of wakefulness), and most recently, with orexin antagonism (depotentiation of wakefulness). Although many directions for research and new therapeutics are possible, one possible direction would be to evaluate the soporific potential of compounds administered at time of bed that attenuate arousal in the periphery (e.g., heart rate) as opposed to within the central nervous system.

Basal arousal is targeted with cognitive behavioral treatments through sleep restriction (diminishes arousal by mild sleep deprivation) and relaxation training. Stimulus control may also serve to diminish basal arousal to the extent that leaving the bedroom and engaging in a nonsleep behavior prevents microsleeps and increases the time awake during the night. This may, in turn, result in mild sleep deprivation and thereby diminish arousal. Newer approaches include the anxiolytic potential of the practice of mindfulness. Currently, there are no therapies that address (modulate) sleep requirement, although CBT-I clearly involves a reset with respect to sleep opportunity and sleep ability.

Sleep Interfering-Interpreting Process Model (2000)

Basic Description

Lundh and Broman propose a two component etiologic model[54] (Figure 82-7). One of the components is identified as sleep interfering and is responsible for arousal and sleep continuity disturbance. The other is identified as sleep interpreting and is responsible for the individual's appraisal of the emergent insomnia (i.e., whether the insomnia is viewed or reported as a problem). The sleep interfering component is represented at two levels of causality.

The first level (distal causation) represents several factors that may lead to or be permissive of hyperarousal. These factors include arousability, stimulus arousal associations, behavioral and cognitive coping strategies, and interpersonal relations. Stimulus arousal associations and behavioral and cognitive strategies are also articulated in other insomnia models, but arousability and interpersonal relations are more novel constructs. Arousability, the magnitude of the individual's tendency to react to and recover from elicited arousal, appears to constitute a predisposing factor for insomnia. Interpersonal relations refer to the potential of social conflict to increase arousal and predispose the individual to insomnia. The second level (proximal causation) is arousal itself, conceptualized in terms of emotional, cognitive, and physiologic activation. These forms of arousal (alone or in combination) are posited to interfere with sleep initiation and maintenance. Finally, the sleep interpreting component is also represented at two levels of causality. The first level includes attributions, perfectionism, and beliefs about sleep and daytime functioning. These factors feed forward and determine the individual's

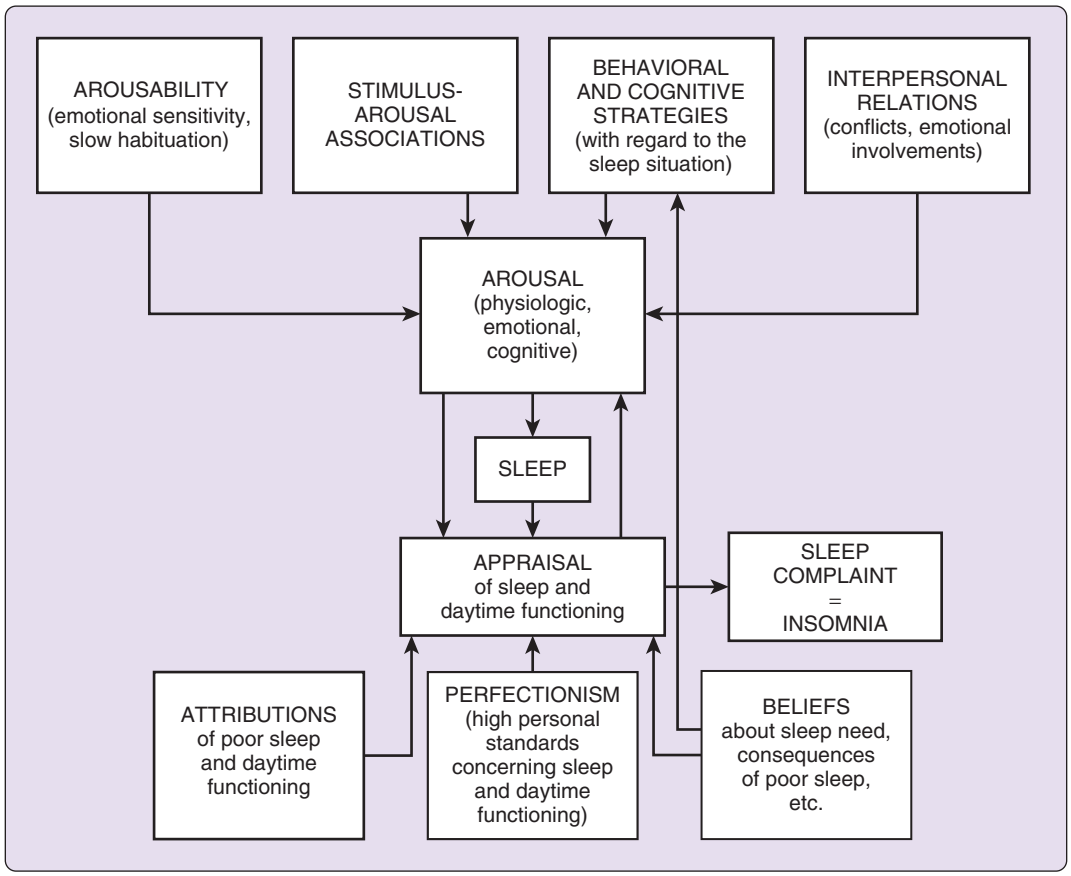


Figure 82-7 Sleep Interfering-Interpreting Process Model. This schematic represents the components of the two factors (sleep interference and sleep interpretation) and how they lead to the physiologic, emotional, and cognitive arousals that impinge on sleep.

appraisal (second level) of whether the insomnia is viewed or reported as a problem.

Strengths and Limitations

The major strengths of the model are that it takes into account the importance of individual factors (arousability and appraisal) and interpersonal relationships as primary triggers for emotional, cognitive, and physiologic arousal. The identification of individual factors appears to be unique to the Lundh and Broman model and is critical because it may explain why, given similar levels of life stress, some develop acute or chronic insomnia, and some do not and why when sleep continuity disturbance occurs, some view the emergent insomnia as a problem whereas others do not. The major limitations of the model are that it does not account for the difference between acute and chronic insomnia and that it assumes that increased arousal necessarily results in sleep continuity disturbance.

Implications for Current and Future Research and Therapeutics

If appraisal and interpretation moderate insomnia severity, these factors may constitute targets for intervention. Morin and colleagues, as part of their formulation of CBT-I, single out at least one aspect of appraisal for assessment and treatment: dysfunctional beliefs and attitudes about sleep.[23] More recently, two additional approaches have been developed to address appraisal and interpretation: behavioral experiments[55] and mindfulness training.[56] Both are likely to influence patients' tolerance of or response to the experience of insomnia or the consequences of insomnia. Empiric studies can evaluate the extent to which each of these approaches produces different outcomes or additive effects when combined with behavioral or pharmacologic interventions.

Psychobiologic Inhibition Model (2002)

Basic Description

The psychobiologic inhibition (PI) model posits that good sleep is ensured by automaticity and plasticity[57,58] (Figure 82-8). Automaticity refers to the involuntary nature of sleep initiation and sleep maintenance, governed by processed such as homeostatic and circadian regulation.[15] Plasticity refers to the ability of the system to accommodate real-world circumstances. Under normal circumstances, sleep occurs passively (without attention, intention, or effort). Within the context of normal sleep, stressful life events precipitate both physiologic and psychological arousal, which can result in inhibition of sleep-related dearousal and the occurrence of selective attending to the life stressors and insomnia symptoms. In acute insomnia, physiologic and psychological arousal interfere with the normal homeostatic and circadian regulation of sleep. Acute insomnia may, in turn, resolve or be perpetuated based on whether the stressor resolves or the individual attends to the insomnia symptoms that occur with the acute insomnia.

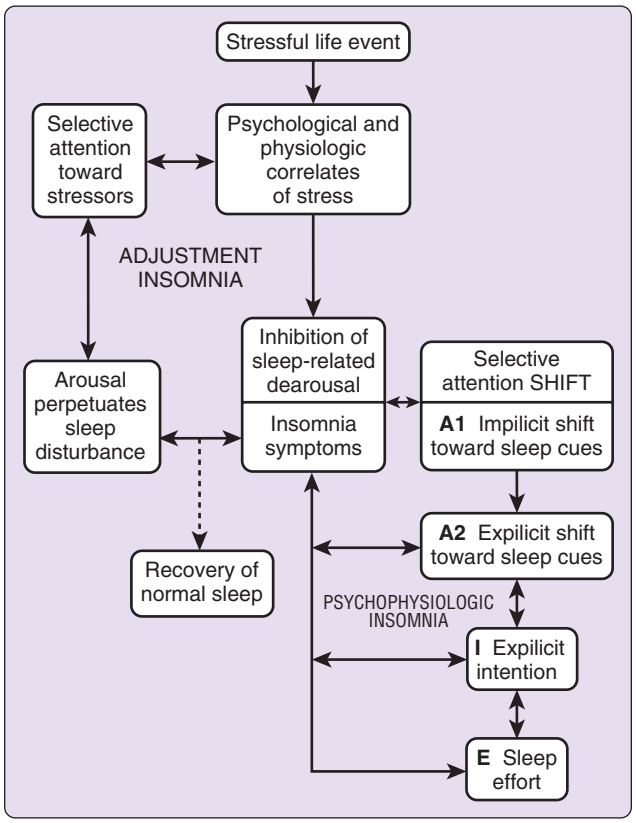


Figure 82-8 The Psychobiologic Inhibition Model. The psychobiologic inhibition model focuses on how insomnia may be perpetuated by (1) the inhibition of sleep-related dearousal and (2) increased sleep-related attention, intention, and effort.

The shift of attention from the life stressor, implicitly or explicitly, to the insomnia symptoms is posited to be the first of three critical events that transition acute insomnia to a form of sleep disturbance that is self-perpetuating. Collectively, the three events (attention, intention, and effort) are referred to as the A-I-E pathway. When individuals are unable to sleep, their attention is drawn to an otherwise automatic process. The very process of attending, in turn, prevents perceptual disengagement and behavioral unresponsiveness (sleep).[59] Because a primary function of attention is to promote action in response to perceived need, an intentional (purposive) process is initiated that acts to further inhibit the normal downregulation of arousal. Finally, the intention to fall asleep triggers sleep effort, and this effort, like enhanced attention and intention, serves only to further inhibit sleep-related dearousal. Ultimately, in chronic insomnia, the inhibition of sleep-related dearousal reflects ongoing or elicited sleep-related attention, intention, and effort.

Strengths and Limitations

Major strengths of the PI model are that it differentiates between acute and chronic insomnia and clearly delineates the mechanisms that are thought to mediate the transition between acute and chronic insomnia. Not only are these mediating variables clearly specified, there is also substantial support for attention bias or selective attention as operational in both in mental illness and insomnia.[60-72] Several studies have assessed the descriptive and predictive utility of the PI model. For instance, sleep-related mental preoccupation appears to be associated with the transition from acute to persistent insomnia in cancer patients.[66] Further, attention bias has been observed in individuals with psychophysiologic insomnia compared with good sleepers and subjects with delayed sleep phase syndrome,[68,69] and sleep-related attentional bias is also associated with self-reported sleep quality and sleepiness.[71] Finally, individuals with psychophysiologic insomnia exhibit effortful preoccupation with sleep.[73]

Another strength of the PI model is that it allows for objective measurement of cognitive processes in insomnia. Insomnia patients commonly complain of mental events interfering with sleep, such as intrusive thoughts, racing thoughts, worry, and inability to disengage from environmental "noise" or bodily sensations. The identification of such mental events relies on self-report. The constructs of the PI model can be operationally defined and tested with objective measures like the computerized emotional Stroop task, the induced-change blindness task, and the dot probe task.

Finally, the PI model poses the novel hypothesis that inhibition of sleep-related dearousal rather than hyperarousal may be responsible for acute and chronic insomnia. In acute insomnia the inhibition of dearousal is engaged by the psychologic and physiologic correlates of stress. In chronic insomnia, the engagement of sleep-related attention, intention, and effort further inhibit dearousal. This shift from hyperarousal as the primary explanatory variable for chronic insomnia to inhibition of sleep-related dearousal represents a potential paradigm shift with regard to the etiology of insomnia.

One limitation of the PI model and the A-I-E pathway is the need for further validation, particularly the intention and effort components. Some conceptual limitations are also present. First, the model focuses on cognitive factors and does account for behavioral mediators or moderators such as sleep extension and stimulus dyscontrol. These factors could be considered forms of sleep effort and thus be accounted for implicitly within the model. This said, explicit inclusion of sleep extension and stimulus dyscontrol would allow the PI model to be more comprehensive and integrative. Second, sleep-related attentional bias tends to be conceptualized as a perpetuating factor, but this factor may also serve as a vulnerability factor for acute or recurrent insomnia.[74] Third, the conceptualization of sleep-related dearousal needs to be explicated in a way that specifically delineates how it is similar to, and different from, the more traditional concept of hyperarousal[75] and the potentially related and alternative concept of the failure to inhibit wakefulness.

Implications for Current and Future Therapeutics and Research

The PI model may help to explain the efficacy of many existing elements of CBT-I and potentially of medical therapies as well. Any behavioral or cognitive intervention that potentiates sleep-related dearousal or promotes the disengagement of attention, intention, and effort should help to restore normal sleep. For example, sleep restriction may help to reinstate sleep automaticity by increasing homeostatic pressure and overcoming the effects of increased attention, intention, or effort. Similarly, stimulus control may strengthen adaptive and automatic bed-sleep dearousal associations. Finally,

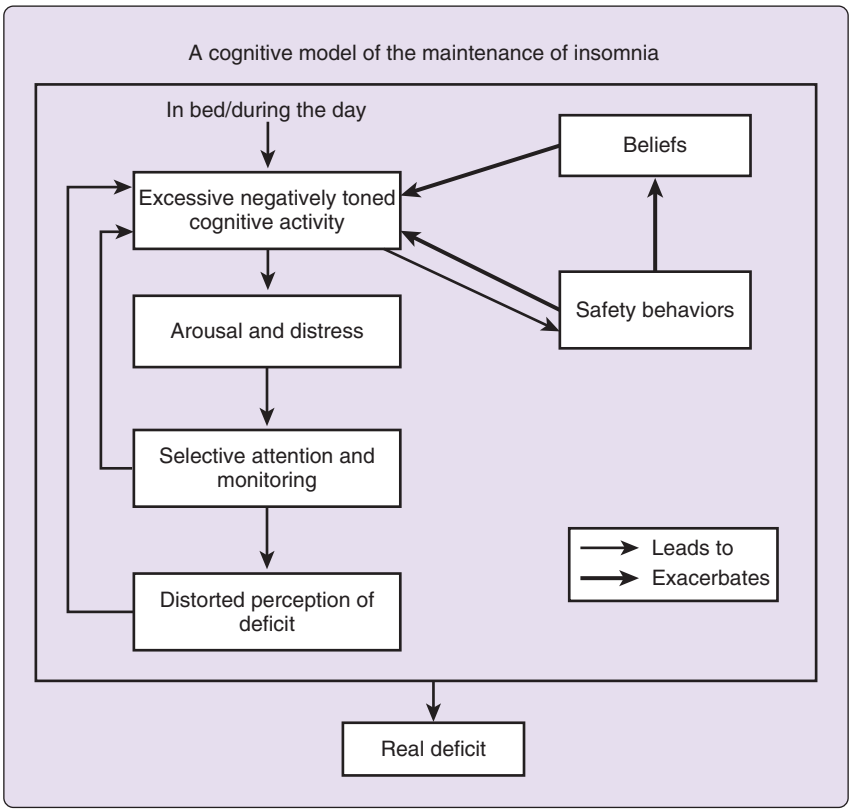


Figure 82-9 **The Cognitive Model.** The cognitive model was originally rendered as a state model that focuses on how insomnia may be perpetuated by (1) selective attention to sleep-related threats and the daytime consequences of insomnia and (2) the engagement of safety behaviors.

relaxation, distraction, and imagery methods may reduce worry about sleep, and paradoxical intention methods may entirely refocus the A-I-E pathway away from sleep preoccupation. The PI model suggests that the mechanisms for existing pharmacotherapies may reside in their capacity to promote relaxation, inhibit exteroception, and reduce sleep-related attention, intention, and effort. Clearly these are features of traditional sedatives (e.g., barbiturates, benzodiazepines, and benzodiazepine receptor agonist therapies). Finally, the model may point to the development of new approaches. For instance, the PI model supports the rationale for sensory gating training and mindfulness therapies. From the pharmacologic point of view, the PI model suggests that it may be productive to antagonize wake-promoting or wake-consolidating systems, for instance, through the modulation of orexin or a histamine.

Cognitive Model (2002)

Basic Description

Harvey's cognitive model adopts a framework that has been applied to anxiety disorders and uses the concepts of selective attention, monitoring, and detection of threat described previously[76] (Figure 82-9). The model posits that in chronic insomnia, sleep-related worry, selective attention and monitoring, and the detection of sleep-related threats perpetuate a level of physiologic arousal that continuously interferes with sleep initiation or sleep maintenance. The transition from acute to chronic occurs when individuals perceive that they have a sleep problem and engage in sleep-related worry, which in turn, prompts them to selectively attend to sleep-related threats and the daytime consequences of insomnia. Enhanced monitoring inevitably results in the increased detection of sleep-related threats and daytime effects of insomnia and the engagement of safety behaviors. Sleep-related worry includes thoughts and ruminations about sleep timing, duration, and quality and the functional and health consequences of sleeplessness. Selective attention to sleep-related threats, including both internal events, such as experiences of alertness or sleepiness, pain or discomfort, and the passage of time, and external events, such as environmental stimuli (e.g., light, sound, temperature) and the passage of time. The detection of sleep-related threats is coupled with the belief that internal sensations or environmental stimuli are interfering with sleep initiation or maintenance. The individual may also detect adverse outcomes during the day (e.g., being late to, missing, or performing badly at work) and attribute them to sleep loss or poor sleep quality. Finally, the engagement of safety behaviors includes both compensatory behaviors (e.g., extending sleep opportunity) and avoidance behaviors (e.g., cancelling social activities) that are intended to mitigate poor sleep and its effects.

Strengths and Limitations

Major strengths of the cognitive model include an explicit focus on the etiologic relevance of the daytime consequences of insomnia, identification of specific factors that perpetuate insomnia (sleep-related worry), and identification of specific mechanisms for how sleep-related worry is perpetuated over

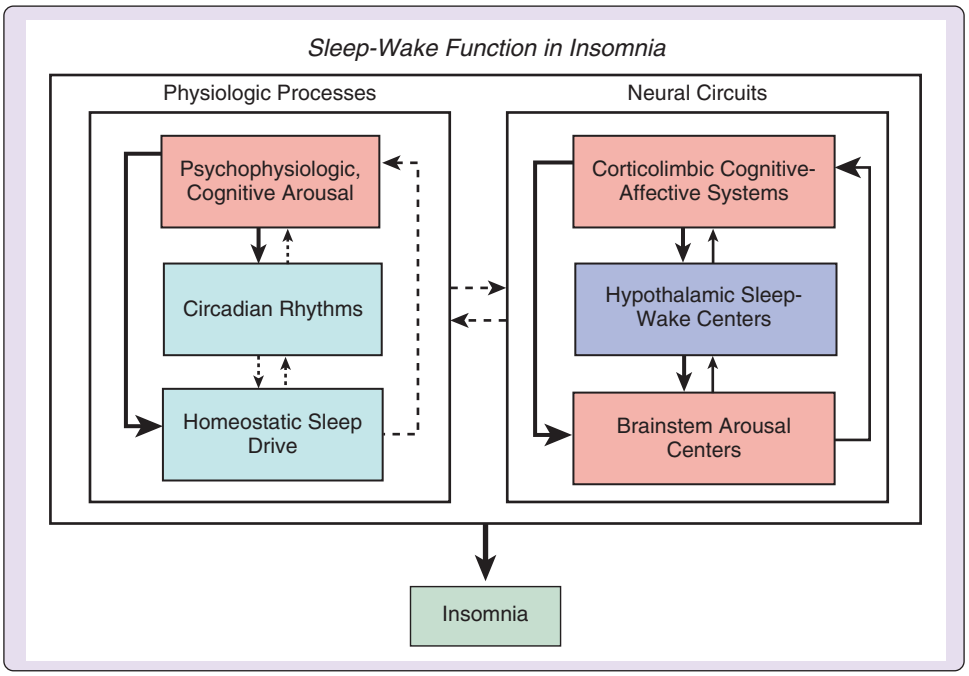


Figure 82-10 The Neurobiologic Model. The neurobiologic model is primarily a state model that focuses on how insomnia may be a hybrid state that has as its neurobiologic substrate "local neuronal wakefulness" during NREM sleep.

time through selective attention and detection of sleep-related threats and consequences. Multiple studies have been conducted on several of the central tenets of the model, including experimental assessments of the relative roles of worry, selective attention, and safety behaviors. For example, experimental manipulations designed to increase worry among good sleepers increase sleep-onset insomnia,[77,78] and experimental manipulations designed to decrease worry in insomnia patients shortened sleep-onset insomnia.[79,80] A range of methodologies, including daily diaries, interviews, questionnaires, and experimental manipulations of monitoring,[81-85] support the prediction that attention to internal and external sleep-related threat is higher among individuals with insomnia relative to good sleepers and contributes to the vicious cycle of insomnia. In addition, one study provided evidence for the predicted association between monitoring and increased negative thoughts and use of safety behaviors at night and during the day. Furthermore, safety behaviors among patients with insomnia have been documented.[86] In addition, one study supported the predicted relationship between unhelpful beliefs about sleep and use of safety behaviors. Specifically, unhelpful beliefs about sleep predicted the use of daily safety behaviors.[86]

Implications for Current and Future Research and Therapeutics

The cognitive model and related empiric work challenge the predominant behavioral perspective on insomnia. The model reasserts the relevance and centrality of cognition as an etiologic factor for insomnia and in so doing suggests that cognitive approaches, long deemphasized in favor of sleep restriction and stimulus control therapies, may deserve further study and clinical use. Several approaches to the management of sleep-related worry have been evaluated, including disputation of dysfunctional beliefs, decatastrophization exercises, mindfulness training to evoke moment-to-moment, nonjudgmental awareness, and behavioral experiments to invalidate worry-related thoughts and beliefs. Dismantling studies could be useful for evaluating the efficacy of cognitive versus behavioral approaches. To date, one such study[87] has been attempted that concluded that, "full CBT is the treatment of choice. Both BT and CT are effective, with a more rapid effect for BT and a delayed action for CT. These different trajectories of changes provide unique insights into the process of behavior change via behavioral versus cognitive routes" (p. 670). Beyond this, the emphasis on centrality of cognition may also provide an insight into the common use of antipsychotics in the management of chronic and severe or treatment-resistant insomnia. Although such treatments may or may not directly alter dysfunctional beliefs, catastrophization, sleep-related worry, or selective attention and monitoring, it remains possible that the sedating effects of these agents (and benzodiazepines and benzodiazepine receptor agonists as well) alter cognition in a manner that contributes to their efficacy.

Neurobiologic Model (2011)

Basic Description

The neurobiologic (NB) model of insomnia primarily focuses on the changes in brain activity and function that may account for insomnia (Figure 82-10). Specifically, Buysse and colleagues[88] posit that insomnia is "a disorder of sleep–wake regulation characterized by persistent wake-like activity in neural structures during NREM sleep, resulting in simultaneous and regionally specific waking and sleeping neuronal activity patterns" (p. 133). Wakelike levels of activity during cortically defined sleep (NREM sleep) are specified as occurring in the prefrontal and parietal cortices, the paralimbic

cortex, the thalamus, and the hypothalamic-brainstem arousal centers. Localized activation within these regions (local wakefulness) during what is otherwise more globally sleep can be expected to be associated with "persistent awareness of the environment" (p. 133). Put differently, coactivation of this sort may directly result in an altered or attenuated capacity to initiate and maintain sleep (hypothalamic-brainstem) but also in abnormal levels of sensory and information processing (thalamus and parietal cortex), emotional processing (paralimbic cortex), and formation of perceptual representations that are evaluated for their appropriateness for action or nonaction (prefrontal cortex) during polysomnographically defined sleep.

Strengths and Limitations

The NB model attempts to be an integrative model that proposes a more specific mechanism for insomnia than provided by the general concept of hyperarousal or the inhibition of "sleep-on" systems. This model defines insomnia as a hybrid state (part sleep and part wakefulness) that occurs with local neuronal variations in sleep depth and may help to explain clinical features of insomnia. The model is informed by the neurocognitive model (described previously), the neuronal transition probability model,[89] the two-process model of normal sleep-wake regulation,[15] and recent findings within neuroscience regarding both the sleep switch[90] and the phenomenon of local neuronal sleep.[91] Further, the NB model echoes the concept of status dissociatus as propounded by Mahowald and Scheck,[92] which suggests that hybrid states of consciousness (coactivations of wake, NREM, and rapid eye movement [REM] sleep) may occur in a variety of the sleep disorders, including narcolepsy, REM sleep behavior disorder, and confusional arousals. The concept of insomnia as a hybrid state was first suggested by Cano and Saper in conceptualizing findings from their rodent model of insomnia.[93]

The proposal that insomnia represents an aberrant state of persistent awareness that occurs as a result of local neuronal wakefulness adds to the existing literature in two ways. First, the model is explicit about what other models only imply: insomnia is, in part, a disorder of persistent wakefulness that may occur globally (objective insomnia) or more locally (subjective insomnia [i.e., sleep state misperception]). Second, the application of the concept of local sleep provides a mechanistic explanation for the proposition that insomnia entails aberrant levels of sensory and information processing or memory formation during the sleep onset period or during sleep. The NB model provides a framework for understanding the phenomena of shallow sleep or sleep state misperception and the paradox that small objective treatment gains are regularly paralleled by larger subjective effects. With respect to paradoxical treatment gains, small treatment-related changes in the polysomnogram may be associated with larger subjective improvements given reduced local waking neural activity in critical regions or circuits, such as the default mode network or the thalamocortical system.

The primary limitation of the NB model is that it is not an etiologic model. It does not focus on how good sleep transitions to insomnia nor on how acute insomnia transitions to the chronic form of the disorder. Like Morin's microanalytic model, the NB model adopts a state perspective. Future elaborations of this model could address how local wakefulness develops and how the functional and physiologic abnormalities that occur with this phenomenon map onto the symptoms of insomnia (difficulties initiating and maintaining sleep).

Implications for Current and Future Research and Therapeutics

Future investigations of the NB model are likely to rely heavily on neuroimaging studies before and during sleep to document regional and circuit-level brain dysregulation in patents with insomnia. Extending such paradigms cross-sectionally and longitudinally could address the transition between acute and chronic insomnia. The therapeutic implications of the NB model are varied and include the exploration of whether present medical and cognitive behavioral approaches minimize or eliminate neuronal local wakefulness. CBT-I may accomplish this by increasing homeostatic pressure for sleep and medical treatment with benzodiazepines, and benzodiazepine receptor agonists may do this through modulation of central nervous system gamma-aminobutyric acid activity. Techniques to increase regional brain activity during wakefulness or to decrease such activity during sleep, such as transcranial magnetic stimulation, may also warrant further study.

LIMITATIONS OF THE EXISTING MODELS

In aggregate, the models presented in this chapter help to explain many of the clinical features and treatment effects commonly observed in chronic insomnia. However, several important aspects of insomnia are not well accounted for.

Most Models Do Not Explicitly Take into Account the Role of Classical Conditioning

Most of the etiologic models do not explicitly address the issue of classical or Pavlovian conditioning, focusing instead on the instrumental side of behavioral processes, that is, on the behaviors that maintain insomnia. Being awake in bed may directly elicit arousal responses or wakefulness through classical conditioning, and such conditioning may contribute to the self-perpetuating nature of insomnia. Classical conditioning as a perpetuating factor can help to explain two reliable findings from the treatment outcome literature. First, CBT-I produces about a 50% reduction in symptoms during the acute treatment phase,[94] which is less than might be expected if only instrumental behavioral factors were responsible for chronic insomnia (see Figure 82-3). Second, patients treated with CBT-I continue to improve over follow-up periods as long as 12 months.[95,96] If only instrumental factors were responsible for chronic insomnia, no additional improvements would be expected beyond the acute treatment phase. From a classical conditioning perspective, successful treatment with CBT-I in the short term may result in counterconditioning over the longer term: Repeated pairing of sleep-related cues with sleep over time may extinguish conditioned arousal.

Most Models Do Not Take into Account Normal Sleep-Wake Sleep Regulation

Although most of the human models of insomnia are compatible with Borbély's two-process model, only the NB and PI models explicitly embrace this perspective on sleep-wake

regulation and how these abnormalities within these systems may serve to predispose, precipitate, or perpetuate insomnia. This is unfortunate because it leaves out the likely possibility that at some point abnormalities within these arenas (process S and process C) contribute to the etiology of insomnia. For example, from the 3P point of view, sleep expansion by retiring to bed earlier likely leads to dysregulation of sleep homeostasis and possibly to a phase advance. Retiring to bed earlier in the evening (while keeping rise time constant or delaying this as well) not only creates a potential mismatch between sleep opportunity and ability but also likely diminishes the homeostatic prime for good sleep continuity and normal sleep architecture and may promote shallow sleep if not local wakefulness. Phase-advancing the sleep period (and the resultant alterations to light exposure and melatonin secretion) may prompt a fundamental shift in all the physiologic parameters that have circadian or ultradian rhythms.

Most Models Do Not Explicitly Differentiate between Acute and Chronic Insomnia

All of the models presented directly or indirectly address how acute insomnia transitions to chronic insomnia or how chronic insomnia is maintained over time. Significantly less attention is paid to what precipitates acute episodes of insomnia, whether acute insomnia is a distinct entity from chronic insomnia, and what may characterize the differences between acute and chronic forms of the disorder.

Within the current nosologies (e.g., the DSM-5[3] and ICSD3[4]), acute and chronic insomnia are defined temporally, with a threshold of 3 months. Individuals who meet all criteria for chronic insomnia except duration are diagnosed with acute insomnia (adjustment insomnia, short-term insomnia disorder, or transient insomnia). It is unclear, however, whether other clinical or physiologic factors distinguish acute and chronic insomnia, such as precipitating and perpetuating factors, symptoms, and polysomnographic features. On a more basic level, it is not clear whether acute insomnia should even be considered a pathologic state. It could be argued that acute insomnia is a normative adaptive phenomenon, part of the fight-or-flight response to threat that overrides the normal homeostatic and circadian imperatives for sleep. Put differently, "it may be the case that we live with insomnia today, because at some point in our evolutionary history insomnia allowed us to live."[97]

Most Models Do Not Account for Gender and Age Differences with Respect to Chronic Insomnia

Women experience insomnia at a rate nearly double that of men,[98-99] and older adults report chronic insomnia at a rate that is approximately three times that of general population,[100] but the reasons for these differences are not well-defined. Although none of the models presented here explicitly address sex or age discrepancies, several of them provide a framework for doing so. For example, within the 3P framework, specific predisposing factors for women may include physiologic and psychosocial concomitants of menstrual cycles, childbirth, and menopause,[101-105] and predisposing factors in older adults may include chronic medical and psychiatric conditions, other sleep disorders, medication effects, and age-related changes in sleep behaviors and physiology (e.g., advanced time to bed, weakened homeostatic sleep drive). Precipitating factors specific to women may again include physiologic and psychosocial aspects of reproductive function, as well as a host of more general interpersonal and social factors; precipitating factors in older adults include acute illnesses, new medications, and psychosocial stressors such as retirement, bereavement, and loss of independence. Finally, although many perpetuating factors may be common across sex and age, worry, rumination, and anxiety are more often exhibited by women,[106,107] which may help explain why acute insomnia occurs with about equal prevalence in men and women[108] but chronic insomnia appears more often in women. Older adults may be particularly vulnerable to increased time in bed as a perpetuating factor.

Most Models Do Not Account for the Types and Subtypes of Insomnia

Most of the models reviewed in this chapter do not account for different clinical presentations such as initial insomnia, middle or late insomnia, or a combination of these symptoms. The models also do not consistently address other phenomena, such as varying degrees of subjective-objective sleep discrepancy or varying age of onset. Finally, most models do not account for differences in clinical presentation that are characterized as idiopathic, psychophysiologic, or paradoxical insomnia. More research is needed to determine whether these states differ with respect to etiology, pathophysiology, or their responsiveness to treatment.

Most Models Posit that Insomnia Occurs in Association with Elevated Arousal or Hyperarousal

At present, no theoretical distinctions have been made (or studies conducted) showing that the hyperarousal of acute insomnia is the same as or different from the hyperarousal that is present with subchronic and chronic insomnia. In the absence of data, it seems reasonable to hypothesize that the arousal (at least somatic hyperarousal) that occurs with the hypothalamic-pituitary axis–related fight-or-flight response far exceeds the arousal that occurs with chronic insomnia. If this is the case, then it follows that the persistent inability to initiate or maintain sleep occurs, in part, as a result of other factors. The NC model focuses on functional changes with respect to perceptual disengagement. The PI focuses on a weakening of the sleep system in terms of the inhibition of sleep-related dearousal. The NB model focuses on the persistent wakelike activity in neural structures during NREM sleep. Taken together, along with basic findings from the Cano-Saper rodent model,[93] this suggests not only that chronic insomnia may be maintained by unique factors but also that chronic insomnia may well be a hybrid state in which there is either a persistence of wakefulness (at sleep onset) or a failure to inhibit wakefulness (following nocturnal awakenings). Moving forward such concepts should be evaluated for how they differ from and interact with the concepts of basal arousal and hyperarousal. Further, these concepts should serve to spur the development of novel interventions.

CONCLUSION

Each of the models presented in this chapter provides a unique perspective, and for the most part, none are mutually exclusive. In recognition of this, we provide in Figure 82-11 an integrative perspective, parallel process model. This model

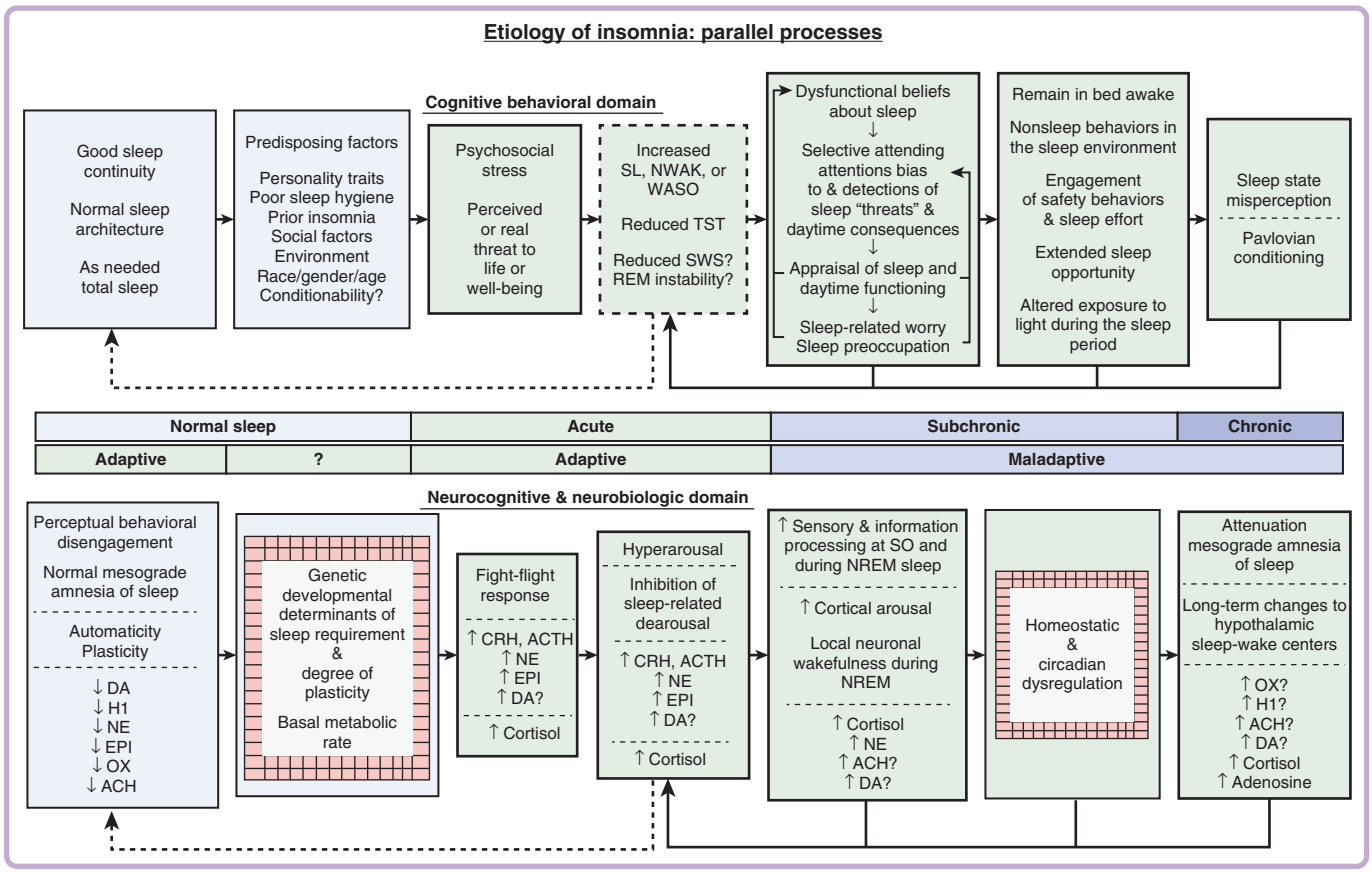


Figure 82-11 The Parallel Process Model. The parallel process model is provided to illustrate how (1) all of the identified factors may be contributory and (2) the cognitive and behavioral domains may be viewed as parallel processes to the neurocognitive and neurobiologic domains. ACH, Acetylcholine; ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; DA, dopamine; EPI, epinephrine; H1, histamine-1 receptor antagonist; NE; norepinephrine; NWAK, number of awakenings; OX, orexin; SL, sleep latency; SO, sleep onset; TST, total sleep time; WASO, wake after sleep onset.

is intended to represent each of the core components from the nine models within one framework, the perspective that the cognitive-behavioral and the neurocognitive-neurobiologic domains represent two sides of the same phenomena, and the possibility that acute insomnia is adaptive. In framing the various factors in this manner we hope to stimulate new ideas for both research and possible interventions.

CLINICAL PEARL

Although many consider theory to be largely an academic enterprise, the models presented in this chapter provide a framework for understanding how insomnia becomes chronic, why the disorder presents as it does, and how or why different treatments may work. Such frameworks, although inevitably imperfect and incomplete, help in the conceptualization of both individual cases and the directions for future research.

SUMMARY

Since the 1990s there has been a proliferation of theoretical perspectives on the etiology of insomnia that now includes nine human models. Each is summarized, reviewed for its strengths and weaknesses, and evaluated for its potential to generate new research and therapeutics. Following the summaries, limitations of the present models are considered and an integrative perspective provided.

Selected Readings

Bonnet MH, Arand DL. Hyperarousal and insomnia. *Sleep Med Rev* 1997;**1**:97–108.

Bootzin RR. Stimulus control treatment for Insomnia. In: *Proceedings of the 80th Annual Convention of American Psychological Association*. 1972. p. 395–6.

Buysse D, Germain A, Hall M, et al. Neurobiological model of insomnia. *Drug Discov Today Dis Mod* 2011;**8**:129–37.

Cano G, Mochizuki T, Saper CB. Neural circuitry of stress-induced insomnia in rats. *J Neurosci* 2008;**28**:10167–84.

Ellis JG, Gehrman P, Espie CA, et al. Acute insomnia: current conceptualizations and future directions. *Sleep Med Rev* 2012;**16**:5–14.

Espie CA, Broomfield NM, MacMahon KMA, et al. The attention-intention-effort pathway in the development of psychophysiologic insomnia: an invited theoretical review. *Sleep Med Rev* 2006;**10**:215–45.

Harvey AG. A cognitive model of insomnia. *Behav Res Ther* 2002;**40**:869–93.

Lundh LG, Broman JE. Insomnia as an interaction between sleep-interfering and sleep-interpreting processes. *J Psychosom Res* 2000;**49**:299–310.

Merica H, Fortune RD. A neuronal transition probability model for the evolution of power in the sigma and delta frequency bands of sleep EEG. *Physiol Behav* 1997;**62**:585–9.

Morin CM, Hauri PJ, Espie CA, et al. Nonpharmacologic treatment of chronic insomnia: an American Academy of Sleep Medicine review. *Sleep* 1999;**22**:1134–56.

Perlis ML, Giles DE, Mendelson WB, et al. Psychophysiological insomnia: the behavioural model and a neurocognitive perspective. *J Sleep Res* 1997;**6**:179–88.

Riemann D, Nissen C, Palagini L, et al. The neurobiology, investigation and treatment of chronic insomnia. *Lancet Neurol* 2015;**14**:547–58.

Spielman A, Caruso L, Glovinsky P. A behavioral perspective on insomnia treatment. *Psychiatr Clin North Am* 1987;**10**:541–53.

Winkelman JW. Clinical practice. Insomnia disorder. *N Engl J Med* 2015;**373**(15):1437–44.

Zhao L, Wang E, Zhang X, et al. Cortical structural connectivity alterations in primary insomnia: insights from MRI-based morphometric correlation analysis. *Biomed Res Int* 2015;**2015**:817595.

A complete reference list can be found online at ExpertConsult.com.

Insomnia Diagnosis, Assessment, and Evaluation

Chapter 83

Jason C. Ong; J. Todd Arnedt; Philip R. Gehrman

Chapter Highlights

- This chapter provides an update on the *Diagnostic and Statistical Manual of Mental Disorders (DSM),* fifth edition and *International Classification of Sleep Disorders (ICSD),* third edition diagnostic criteria for insomnia disorders.
- The tools for conducting a clinical assessment of insomnia are described. The essential tools include the clinical interview, sleep diaries, and self-report measures. Optional tools include polysomnography and actigraphy, which are not routinely used but might be indicated to rule out other sleep disorders.
- Considerations for case formulation, assessing treatment progress, and variations in clinical setting are discussed. Also, the potential use of mobile technology is discussed with regard to a need for further research.

This chapter provides an overview of the diagnosis of insomnia disorders and procedures for conducting clinical assessments of insomnia disorders. The first section reviews the diagnosis of insomnia with an update on the current diagnostic criteria. The second section describes the components for conducting a clinical assessment with insomnia patients, including the clinical interview, the use of self-reported measures, and the role of objective measures. The chapter concludes with a discussion on formulating a case conceptualization, assessment of treatment progress, and contextual considerations for conducting the assessment of insomnia.

DIAGNOSIS OF INSOMNIA

At a surface level, the diagnosis of insomnia seems straightforward and can be as simple as self-reported "trouble sleeping." Even if this complaint is expanded to include difficulty falling asleep, difficulty maintaining sleep, or awakening earlier than desired, these symptoms lack specificity and do not sufficiently capture the complexities of the sleep problem or the impact on daytime functioning. The main classification systems of sleep disorders include the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* and the *International Classification of Sleep Disorders (ICSD),* and each iteration of these systems has sought to improve the value of a diagnosis of insomnia based on new research findings. Beyond the sleep-specific symptoms, the first diagnostic consideration is the requirement that an insomnia disorder only exists if there is clinically significant distress or impairment. Each of the diagnostic systems requires some degree of distress or impairment, such as fatigue, irritability, difficulty with concentration or memory, and worry about inadequate sleep that is perceived by the patient to be associated with the sleep disturbance.[1] There are some patients who present with insomnia symptoms but on subsequent clinical interview fail to experience this criterion and thus are not diagnosed with an insomnia disorder.

A second issue in the diagnosis of insomnia that has been more controversial is whether the insomnia is a symptom of another condition or a disorder in its own right.[2] This distinction stems from the fact that a wide range of mental and physical disorders are known to cause insomnia, including chronic pain, depression, and anxiety disorders. Also, approximately 80% of patients with insomnia have one or more comorbidities.[3] Early concepts posited that when insomnia occurs in conjunction with these comorbidities, it is a secondary symptom rather than an independent condition. In the absence of comorbidities the insomnia could be considered to be a primary disorder. This distinction is reflected in the DSM-IV-TR[4] and ICSD2[5] criteria for insomnia. The DSM-IV-TR[4] included diagnoses of Primary Insomnia and several secondary insomnia diagnoses including Insomnia Related to Another Mental Disorder, Substance-Induced Sleep Disorder—Insomnia Type, and Sleep Disorder Due to a General Medical Condition. Similarly, the ICSD2[5] included insomnia subtypes that could be divided between primary (Psychophysiological Insomnia, Paradoxical Insomnia, Idiopathic Insomnia) and secondary (e.g., Insomnia Due to a Mental Disorder, Insomnia Due to a Medical Condition, and Insomnia Due to a Drug or Substance) insomnia diagnoses.

The distinction between primary and secondary insomnia diagnoses has not held up over time. At an empiric level, most of the diagnoses fail to demonstrate adequate validity and reliability.[6] At a conceptual level, secondary insomnia would be expected to resolve with successful treatment of the primary disorder, but this is often not the case. Insomnia tends to persist over time, even with remission of comorbid conditions. As a result, the current versions of the DSM and ICSD dramatically reduced the number of insomnia diagnoses. The DSM-5 abandoned the primary versus secondary distinction and now has a single diagnosis of Insomnia Disorder with specifiers to indicate comorbidities when present.[7] Similarly, the ICSD3 has included two diagnoses, Chronic Insomnia Disorder and Short-Term Insomnia Disorder, that are distinguished by the duration of illness.[8] Also, nonrestorative sleep is no longer included in the diagnostic criteria, and the diagnosis of insomnia disorder can be applied to both children and adults in DSM-5 and ICSD3. The essential elements for an

Box 83-1 DIAGNOSTIC CRITERIA FOR INSOMNIA DISORDERS (307.42/F51.01) IN THE DSM-5 AND ICSD3

	DSM-5* Insomnia Disorder	ICSD3† Chronic Insomnia Disorder
A. Sleep disturbance/complaint	Difficulty initiating sleep, maintaining sleep (frequent awakenings or difficulty returning to sleep after awakenings), waking up earlier than desired with an inability to return to sleep	
B. Associated consequence	Clinically significant distress or impairment in social, occupational, educational, academic, behavioral, or other important areas of functioning	One or more of the following related to the nighttime sleep difficulty: 1. Fatigue/malaise 2. Attention, concentration, or memory impairment 3. Impaired social, family, occupational, or academic performance 4. Mood disturbance, irritability 5. Daytime sleepiness 6. Behavioral problems (e.g., hyperactivity, impulsivity, aggression) 7. Reduced motivation, energy, initiative 8. Proneness for errors, accidents 9. Concerns about or dissatisfaction with sleep
C. Frequency	Sleep difficulty and associated consequence occurs at least 3 nights per week	
D. Duration	Sleep difficulty and associated consequence is present for at least 3 months	
E. Adequate opportunity	Sleep difficulty occurs despite adequate opportunity and circumstances for sleep	
F. Relationship to another condition	Sleep-wake problems are not better explained by another sleep disorder, a coexisting mental disorder, or coexisting medical condition and are not attributed to the physiologic effects of a substance	

*American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. Arlington, VA: APA; 2013.
† American Academy of Sleep Medicine (AASM). *International Classification of Sleep Disorders.* 3rd ed. Darien, IL: AASM; 2014.
Note: Additional specifiers are made in DSM-5: (1) with non–sleep-disorder mental comorbidity, including substance use disorders; (2) with other medical comorbidity; (3) with other sleep disorder and for *episodic* (symptoms last at least 1 month but less than 3 months), *persistent* (symptoms last 3 months or longer), or *recurrent* (two or more episodes within 1 year).

insomnia disorder that are common between the DSM-5 and ICSD3 are summarized in Box 83-1 and include the following patient-reported criteria:

A. Difficulty initiating sleep, difficulty maintaining sleep, or early morning awakenings
B. Distress or impairment that is caused by the insomnia
C. Frequency of at least three nights per week for at least 3 months' duration
D. Duration of at least 3 months
E. Adequate opportunity for sleep
F. Insomnia that is not better explained by another disorder or a substance

Criterion F retains the option of deciding that the insomnia is secondary to another condition, but in this case no insomnia diagnosis would be given, as opposed to the prior practice of assigning a secondary insomnia diagnosis. Thus, the DSM-5 and ICSD3 took a lumping approach to the diagnosis of insomnia instead of the traditional splitting approach because of the lack of evidence to support multiple diagnoses. As our understanding of the pathophysiology of insomnia advances it may become possible to reintroduce diagnostic subtypes with greater empiric support.

ASSESSMENT OF INSOMNIA

The clinical assessment of insomnia should include, at a minimum, a clinical interview and self-report measures of global insomnia symptoms and sleep diaries. Objective measures, such as polysomnography (PSG) and actigraphy, are not routinely recommended but may be indicated to rule out other sleep disorders as part of the evaluation process. When conducting an insomnia assessment, conceptual models for the development and maintenance of insomnia (see Chapter 82) can provide a useful framework for organizing assessment data from multiple sources. For example, the biobehavioral model, or 3 P model,[9] can be useful when assessing a patient's predisposing factors (e.g., family history of insomnia or trait tendencies toward worry that preceded the onset of insomnia), precipitating events that initially caused the onset of insomnia (e.g., loss of a job or other psychosocial stressors), and perpetuating factors that appear to be maintaining the insomnia (e.g., napping during the day or worry about sleep). Also, etiologic models of hyperarousal[10] and the role of cognitions[11] can be useful to inform the case conceptualization and treatment planning for an individual patient. The following section provides a guide to conducting a clinical assessment for insomnia.

The Clinical Interview

The clinical interview remains the cornerstone of insomnia assessment, and a number of semi-structured[10,12] and structured[13,14] interviews for sleep disorders and insomnia have been developed. A primary advantage of the clinical interview is that it enables the clinician to be flexible in approach,

fostering the clinician-patient relationship and allowing tailoring of questions to collect the most pertinent information. For the experienced sleep practitioner, the clinical interview yields a comprehensive differential diagnosis list and logical treatment plan. Insomnia patients often seek treatment because of the daytime consequences of their nighttime symptoms; thus focusing on the 24-hour impact of insomnia for the patient is critical and provides treatment targets beyond nighttime symptoms. An assessment outline with general category areas and key specific areas of inquiry can be found in Table 83-1.

Characterization of the Nocturnal Insomnia Complaint

The clinical interview often begins by eliciting the patient's chief insomnia complaint. Specific quotes vary widely, but the nature of the complaints falls along a limited number of dimensions: inability to fall asleep, inability to stay asleep, waking too early in the morning (or a combination of these symptoms), poor quality sleep, too little sleep, work or lifestyle interference with sleep, or inability to sleep without medications.[15] As indicated previously, the weekly frequency of the nighttime complaint and its duration are required for the DSM-5 and ICSD3. It is additionally helpful to gauge the perceived severity of the nighttime sleep problem and its effects on next-day functioning, which can be accomplished with a simple Likert scale ranging from 1 (no severity) to 10 (maximum severity). It should be noted that patients commonly report during the clinical interview that the insomnia symptoms occur nightly, but further assessment with a daily sleep diary identifies that the sleep pattern is more variable, with evidence of both good and bad nights. This inconsistency highlights the importance of a multimethod approach to insomnia assessment.

An important line of inquiry involves identifying a precipitant, or precipitating factor, for the initiation of the sleep problem. For many, the onset of insomnia is intimately linked with a specific event, but for others the precipitant is less clear. When an obvious insomnia trigger is identified, the clinician can explore premorbid sleep patterns to identify potential predisposing factors. Commonly, patients will report being "light sleepers" before the onset of a more extended bout of insomnia. Previous episodes of transient insomnia may serve as a precipitant to more chronic insomnia, although few studies have evaluated this empirically. Evidence does suggest that transient episodes of insomnia can be elicited in vulnerable individuals.[16,17]

Daytime Consequences of Nighttime Symptoms

Integral to the ICSD3 and DSM-5 diagnostic criteria is the presence of reported daytime impairments related to the insomnia complaints. Most patients report insomnia-related daytime consequences, most commonly in the domains of daytime fatigue, mood disturbances, cognitive deficits, and physical illness.[10,18,19] Reported daytime consequences can seem excessive relative to the magnitude of the nighttime sleep symptoms, which may reflect the contribution of psychological processes to the perception of daytime impairment.[20] Objective confirmation of daytime deficits among insomnia patients has yielded mixed results,[21] but it is important to maintain the perspective that perceived daytime impairment from insomnia is usually what prompts patients to seek treatment.

Current Sleep Pattern

Information about the patient's typical sleep pattern during a recent week under normal life circumstances should be elicited early in the interview. Patients are often better able to describe a "typical" night than an "average" night. Specific information to collect includes bedtime, rise time, time to fall asleep (sleep latency), number and duration of awakenings during the night, final wake-up time, frequency and duration of daytime naps, estimated total sleep time, perceived sleep quality, and morning restfulness. It is also important to determine how this typical pattern differs on weekdays or workdays versus weekends or nonworkdays, on and off sleep medication, and on good nights versus bad nights. The clinician should additionally inquire about presleep activity, perceived causes for awakenings during the night, the patient's coping response when unable to sleep (e.g., remain in bed, watch television), and factors that improve or exacerbate sleep. A functional analysis examining how patients make various decisions related to sleep can be informative. For example, understanding a patient's responses to insomnia symptoms (e.g., level of distress on awakening during the night) and daytime consequences (e.g., napping due to fatigue or sleepiness) can be helpful for case formulation and for the identification of treatment targets. It is common for many insomnia patients to report their sleep to be invariable, unpredictable, and resistant to environmental circumstances, but the patient should nevertheless be challenged to consider differences between bad and good nights.

History of Insomnia Treatments

The most common form of insomnia treatment is prescription hypnotic medication. However, up to 15% of the population has used alcohol as a sleep aid,[22] and insomnia patients commonly initiate self-help treatments before seeking more formal treatment from a sleep specialist.[23] The evaluation of current and past treatments should therefore include any over-the-counter remedies, nonmedication strategies, and alcohol in addition to any sleep-related prescription medications. Current and past treatments should be assessed for type, dosage, frequency of use, time of administration, typical response, and the conditions surrounding deviations from this pattern. It is important to establish in each case whether the specific treatment was given an adequate trial because past failures can result from suboptimal treatment as well as partial or no response. It is additionally important to gauge patients' attitudes toward pharmacologic and nonpharmacologic therapies because these attitudes will likely influence adherence with treatment recommendations and success with treatment outcome.[24] A thorough evaluation of all treatment efforts should enable the clinician to suggest modifications to previously attempted treatments, while providing plausible reasons for previous failed trials.

Sleep-Incompatible Thoughts and Behaviors

Insomnia patients often engage in patterns of behavior and thinking that precipitate or perpetuate the insomnia episode.[10,25] Moreover, there are many daily activities that interfere with nighttime sleep. A careful assessment of compensatory behavioral strategies used to increase nighttime sleep (e.g., variable bedtimes and rise times, spending excessive amounts of time in bed awake) or to deal with the daytime

Table 83-1 Assessment Outline

Category	Specifics
A. Chief complaint	Difficulty falling asleep, staying asleep, waking up too early, poor quality sleep, too little sleep, work or lifestyle interferes with sleep, inability to sleep without medications Define frequency (weekly, monthly), duration, severity, and course (episodic, seasonal variation)
B. Circumstances surrounding onset	Age of onset, precipitating events, sudden or gradual onset, premorbid sleep pattern and quality, previous insomnia episodes
C. Daytime consequences	Fatigue vs. sleepiness, napping, cognition, performance, and mood
D. Current sleep-wake schedule	Bedtime, wake time, rise time, sleep latency, frequency and duration of nighttime awakenings, estimated total sleep time Define average night and variability (e.g., weekdays vs. weekends, medication vs. nonmedication), identify factors that ameliorate and exacerbate sleep pattern Identify prebedtime activities and environment
E. Current and past treatments, adequacy of trial, efficacy	Type (prescription, over-the-counter, behavioral), dosage, efficacy, adequacy of trial
F. Perpetuating factors	
Behavioral	Practices intended to improve sleep (e.g., going to bed early, staying in bed late, increasing time in bed, falling asleep with TV or radio on, staying in bed during awakenings, delayed rise time on weekends) Practices intended to counter fatigue (e.g., napping vs. dozing vs. resting, increasing caffeine ingestion, decreasing physical activity)
Cognitive	Worry about (1) consequences of insomnia (fatigue, performance deficits, health, appearance), (2) sleeplessness ("I'll never get to sleep tonight"), (3) self-perception ("I'm not together, I can't even sleep like a normal person," "My sleep is out of control—just like my life") *False beliefs:* Misconceptions about sleep ("Everybody sleeps 8 hours," "I can't function on less than 7 hours of sleep"), catastrophizing ("Insomnia is ruining my life"), requirements for sleep ("I cannot sleep without sleeping pills," "I must sleep alone")
Other	Alcohol, noise, pets sleeping in the same bed, clock watching during the night, getting home late without enough time to wind down
G. Work	Stress, work schedule incompatible with sleep schedule or contributing to insomnia
H. Family/social	Childhood habit of staying up late with a parent Family history of sleep and psychiatric disorders Stressful life events (past stressful events may be precipitants and present stressful events may be perpetuators)
I. Medical factors	Pain and other medical conditions that can interfere with sleep
J. Pharmacologic considerations	Activating and sedating drugs that interfere with the sleep-wake cycle (consider type, dosage, frequency, and timing), side effects, history of hypnotic use
K. Psychiatric factors	Key features of depression: • *Mood*: sadness, crying, inability to enjoy oneself, irritability, reduced motivation, diurnal mood variation (worse in the morning) • *Cognitive:* reduced attention and concentration, reduced speed of processing, preoccupations, guilt, hopelessness, helplessness, worthlessness, catastrophic thinking • *Physiologic:* reduced (or increased) appetite, diminished sex drive, poor sleep, early morning awakening, psychomotor agitation or lethargy, increased fatigue • *Other:* social withdrawal, suicidal thoughts, intentions, plans, and attempts Key features of anxiety: • Exaggerated worry, tension and irritability; restlessness, headaches, trembling, sweating, and sleep disturbance Key features of substance use disorders: • Excessive and persistent use of substance with evidence of impaired control, use in risky circumstances, or significant social impairment • Consider relationship with current insomnia complaint and status of treatment
L. Other sleep disorders	Circadian rhythm sleep-wake disorders (mismatch between intrinsic circadian rhythm and external environment), restless legs syndrome (sensory discomfort or restlessness in limbs when sedentary), periodic limb movement disorder (repetitive stereotypical leg movements during sleep), sleep-related breathing disorder (snoring, gasping, witnessed apneas, daytime sleepiness)
M. Significant other report	Obtain collateral information from bed partner when possible about sleep pattern and other sleep disorders

consequences of poor sleep (e.g., napping, increased caffeine intake), as well as a thorough evaluation of all "sleep hygiene" behaviors, will uncover maladaptive behaviors. Insomnia patients may exhibit more adverse sleep hygiene practices relative to good sleepers,[26] but most insomnia patients have already attempted to remedy many of these before seeking formal treatment. However, many patients fail to appreciate that altering only a few of these behaviors is unlikely to benefit sleep over the long-term and that good sleep hygiene practices need to be implemented over an extended period of time to reveal their impact on sleep. It is equally important to understand patients' beliefs and attitudes about sleep in general as well as any cognitive responses and reactions to nighttime and daytime sleep-related symptoms. Recent evidence highlights the importance of modifying unhelpful sleep-related cognitions to produce positive and sustained treatment gains.[27,28]

Work, Family, and Social History

Survey studies suggest that patients with insomnia have a higher incidence of first-degree relatives with sleep disturbances compared with controls, with the mother being the most commonly affected family member.[29-31] Specific inquiry regarding a family history of insomnia and psychiatric disorders can provide insight into potential predisposing factors to the insomnia disorder. Assessing the patient's psychosocial history, including occupational or school performance and typical work hours, perceived quantity and quality of interpersonal support, and presence of psychosocial stressors, can also elicit factors that can be treatment targets. Finally, assessment of work hours and timing (e.g., multiple jobs, night shifts) and caregiver demands can determine whether there is an adequate opportunity to achieve sufficient sleep.

Evaluation of Comorbid Conditions

A significant portion of the clinical interview should focus on the evaluation of medical disorders, psychiatric conditions, other sleep disorders, and substance use that are commonly comorbid with insomnia.[32,33] The general approach to this part of the evaluation is to establish the nature of the relationship between the comorbid condition and the insomnia disorder, including temporal precedence. It should not be assumed that any condition co-occurring with insomnia should be the sole or initial focus of treatment.[34] Insomnia frequently persists despite resolution of other conditions and requires independent treatment.[35,36] Studies indicate that cotreatment of insomnia and cooccurring conditions yields more rapid improvement of both conditions compared with sole treatment of only the cooccurring condition.[37,38]

Evaluation of specific medical conditions can be accomplished during the course of the clinical interview through the use of self-report questionnaires or with physical examinations and laboratory testing. Medical comorbidity is high in insomnia disorders,[39] and accompanying medical conditions can often exacerbate insomnia symptoms, impair the patient's ability to cope with subsequent stressors, and directly interfere with treatment progress. (See Table 83-2 for common medical conditions that are comorbid with insomnia.)

Like medical conditions, insomnia is commonly associated with psychiatric disorders, with rates ranging from 40% to more than 50%.[34] Insomnia is listed as a symptom in several disorders in the DSM-5, the most common of which are indicated in Table 83-2. Thus it is important to assess for undiagnosed or inadequately treated psychiatric conditions that may be causing or contributing to the insomnia complaint. Insomnia symptoms are reported by as many as 80% of outpatients with major depressive disorder and more than 90% of inpatients, and up to 20% of outpatients report suffering from hypersomnia.[40,41] Patients in a hypomanic or manic state may present with complaints of intermittent insomnia, which might arise from a decreased need for sleep during a manic phase. A history of manic episodes should also be considered as part of treatment planning because sleep deprivation resulting from behavioral strategies such as sleep restriction could precipitate a manic episode. Complaints of insomnia are also common in patients with panic disorder, generalized anxiety disorder, and posttraumatic stress disorder; they are less likely to be significant features of social phobia, obsessive compulsive disorder, or simple phobias.[42,43] Differentiating an anxiety disorder from an insomnia disorder can present a diagnostic challenge because patients with insomnia are frequently anxious. However, patients with generalized anxiety disorder will describe worry that encompasses

Table 83-2 Medical and Psychiatric Conditions Commonly Comorbid with Insomnia

Medical Conditions	Psychiatric Conditions
Neurologic (e.g., headache, stroke, seizure disorders, brain injury, dementia, Parkinson disease)	Mood disorders (e.g., major depressive disorder, dysthymic disorder, bipolar disorder, seasonal affective disorder)
Cardiovascular (e.g., angina, congestive heart failure)	Anxiety disorders (e.g., generalized anxiety disorder, posttraumatic stress disorder)
Pulmonary (e.g., chronic obstructive pulmonary disease, asthma)	Psychotic disorders (e.g., schizophrenia)
Digestive (e.g., irritable bowel syndrome, reflux, peptic ulcer)	Eating disorders: bulimia nervosa, anorexia nervosa
Endocrine (e.g., hyperthyroidism, diabetes mellitus)	Attention deficit/ hyperactivity disorder
Musculoskeletal (e.g., rheumatoid arthritis, chronic pain disorders)	Adjustment disorder
Reproductive (e.g., pregnancy, menopause)	Personality disorders (e.g., borderline)
Genitourinary (e.g., incontinence, enuresis, benign prostatic hypertrophy)	
Other sleep disorders (e.g., sleep apnea, restless legs syndrome, periodic limb movement disorder)	

Modified from Schutte-Rodin S, Broch L, Buysse D, et al. Clinical guideline for the evaluation and management of chronic insomnia in adults. *J Clin Sleep Med* 2008;4(5):487–504.

multiple daily events and activities outside of the sleep domain, whereas insomnia patients tend to have sleep-focused worry, such as the ability to fall asleep or the consequences of insufficient sleep. One third to more than three fourths of patients with a substance use disorder complain of insomnia.[44,45] Subjective sleep complaints in alcohol-dependent patients continue for a minimum of several weeks after withdrawal, but sleep abnormalities can persist much longer.[46] Insomnia complaints also predict relapse to alcohol during early abstinence after controlling for other confounding variables.[45,47,48] Consideration should be given to the pattern of substance use to determine whether the substance use is restricted to the sleep period or extends into the daytime with attendant adverse consequences to daily functioning. In addition to alcohol, specific attention should be given to licit prescription substances that may be abused (e.g., methylphenidate) and to illicit stimulants (e.g., cocaine, MDMA or "ecstasy") and depressants (e.g., cannabis, opioids).

Medications and Substances

A wide variety of medications and substances can contribute to insomnia complaints and their daytime consequences. A thorough assessment of prescription and nonprescription medications, herbal remedies, and licit (caffeine, nicotine, and alcohol) substances can determine their role in precipitating and perpetuating the insomnia disorder. The assessment should include type, dosage, frequency, and timing of use. For hypnotics, understanding how and when the patient decides to take the medication can provide important therapeutic material. Medication use can be collected during the course of the clinical interview or with locally developed checklists. Patients can also be asked to have available an updated list of their current medications for the initial evaluation. The clinical evaluation will need to consider the timing of the insomnia disorder relative to the initiation of medications or substances and alterations in sleep as a function of use, extended exposure, and discontinuation.

Other Sleep Disorders

Insomnia symptoms can occur in the context of another sleep disorder, and the assessment should evaluate whether these symptoms are better accounted for by another sleep disorder. Importantly, the current nosology systems allow for diagnosis of an insomnia disorder in addition to the presence of another sleep disorder, as long as the criteria are met for both. Insomnia complaints with daytime sleepiness that involve a misalignment between the patient's endogenous circadian rhythm and the external environment may reflect a circadian rhythm sleep-wake disorder (CRSWD) (see Chapter 40). The distinction between an insomnia disorder and a CRSWD can be informed by asking whether the patient's insomnia complaints are resolved when the patient is able to choose a preferred schedule, such as on weekends or days off work. Those with delayed sleep phase disorder will report difficulty with sleep onset, whereas those with advanced sleep phase disorder will report waking up earlier than their planned times. Furthermore, patterns of sleepiness or alertness (e.g., "second wind") across a 24-hour period can help to distinguish an insomnia disorder from CRSWD, in which the diurnal pattern is related to the timing of the sleep phase. For insomnia related to shift work, the clinician should evaluate the relationship between poor sleep and work hour distribution when making a differential diagnosis. Restless legs syndrome (RLS) is another sleep disorder that could include insomnia symptoms (see Chapter 95). RLS is characterized by a complaint of a strong, irresistible, urge to move the legs, usually accompanied by abnormal sensations. The symptoms are aggravated at rest, relieved with movement or activity, and worse in the evening.[49] RLS patients can present with insomnia complaints because the symptoms become worse in the evening and can disturb a patient's ability to fall asleep or return to sleep after awakening. When these symptoms are reported, laboratory tests to examine blood serum ferritin levels (less than 50 mcg/L) can aid in differential diagnosis. Periodic limb movement disorder is a disorder in which periodic limb movements during sleep are repetitive, highly stereotyped movements of the lower extremities. Approximately 80% to 90% of patients with RLS also have periodic limb movements during sleep,[50] and these movements may occur in up to 15% of insomnia patients. The frequent nighttime limb movements are often disruptive of bed partner's sleep, highlighting the importance of including the significant other in the evaluation. More than 40% of patients with obstructive sleep apnea (OSA) have at least one insomnia symptom,[51] and insomnia symptoms are more likely to be reported by women than men with OSA (see Chapters 113 and 114).[52] Men with OSA are more likely to report snoring, gasping in sleep, and excessive daytime sleepiness. Patients with comorbid OSA and insomnia report sustained awakenings in the night with difficulty returning to sleep. Moreover, the presence of insomnia can affect adherence to positive airway pressure, the first-line treatment of OSA.[53] Therefore assessment of comorbid sleep disorders is important for diagnosis and treatment planning.

Self-Report Measures

To complement the clinical interview, administering self-report measures can provide important additional information or corroborate the symptoms reported during the clinical interview. Depending on the clinical setting and available time for the assessment, these measures might be collected at the time of the assessment or sent to the patient before the assessment for review by the clinician during the initial clinic visit. The selection of these measures should be based on the goals of the assessment, with respondent burden taken into consideration. Table 83-3 provides a summary of the most common self-report measures.

Global Measures of Insomnia Symptoms

Global measures of insomnia can provide an index of the severity of the insomnia symptoms and can also be administered longitudinally to inform response to treatment. A number of these measures have been recommended in the standard assessment of insomnia,[54] and three measures are particularly useful in a clinical setting. The Pittsburgh Sleep Quality Index (PSQI) is one of the most widely used sleep questionnaires.[55] It consists of 10 questions that assess a number of aspects of sleep quality. An advantage of this instrument is its widespread use in a variety of settings and translation into more than 50 languages. The primary disadvantage of its use as an insomnia assessment tool is that it is designed to assess the broader construct of sleep quality and not insomnia per se. As such, an individual can have a high score on the PSQI for reasons other than insomnia. A more specific insomnia questionnaire is the Insomnia Severity

Table 83-3 Summary of Self-Report Measures

Domain	Measure	Utility/Purpose
Global measures of insomnia	Insomnia Severity Index Pittsburgh Sleep Quality Index Patient Reported Outcomes Information System	Assess severity of insomnia symptoms and change in symptoms during follow-up
Sleep-wake patterns	Consensus Sleep Diary	Prospective estimates of sleep parameters, bedtimes, sleep quality ratings, and daytime naps
Other measures	Epworth Sleepiness Scale Fatigue Severity Scale Beck Depression Inventory State-Trait Anxiety Inventory Beliefs and Attitudes about Sleep Pre-Sleep Arousal Scale Glasgow Sleep Effort Scale Morningness-Eveningness Questionnaire	Assess daytime functioning, presence of psychiatric symptoms, or etiologic factors

Index (ISI).[56] The ISI consists of seven items on the nighttime and daytime symptoms of insomnia. The scoring is straightforward and yields a total score with standard cutoffs for clinically significant insomnia. It has been widely used as a screening tool and as an outcome measure in insomnia clinical trials. Lastly, there are the Sleep Disturbance and Sleep-Related Impairment scales from the Patient Reported Outcomes Measurement Information System initiative that measure the nighttime and daytime aspects of sleep disturbance, respectively.[57] Although these scales have not yet been widely used, they are noteworthy in that they were developed using item response theory, a more sophisticated psychometric approach than used in the development of other scales, which allows for computerized adaptive testing. Similar to the PSQI, the items do not pertain just to insomnia, although most items are insomnia specific.

Sleep Diaries

The clinical interview and global measures of insomnia are based on retrospective estimates of sleep, which can be heavily influenced by memory and cognitive biases. Sleep diaries are a prospective form of assessment in which an individual completes a series of questions each morning pertaining to the previous night of sleep. A standard sleep diary is now available[58] that includes items for recording estimated sleep parameters such as bedtime, sleep latency, number of nighttime awakenings, and rise time. Sleep diaries are typically kept for 1 to 2 weeks to capture a representative sample of data, and then the average for each parameter is computed. An advantage of sleep diaries is their prospective nature, which is less subject to bias (e.g., primacy, recency effects). They also yield a series of quantitative values that can more precisely describe an individual's sleep patterns and can be useful in delivering behavioral treatments or measuring treatment-related changes. A disadvantage of sleep diaries is that patient nonadherence to completing diaries as prescribed each morning might result in inaccuracies or missing data. Another disadvantage is that some patients become more anxious or vigilant about sleep when monitoring their sleep patterns. Sleep diaries are an essential ingredient to the assessment of insomnia and are typically kept during treatment as part of cognitive behavioral therapy.

Other Measures

Depending on the particular goals of the assessment process, other measures that assess daytime functioning, psychiatric symptoms, or etiologic factors might be considered to provide a more complete clinical profile.[54] If there is a desire to have a better understanding of daytime functioning, it may be useful to include measures of sleepiness or fatigue, such as the Epworth Sleepiness Scale[59] or the Fatigue Severity Scale.[60] Given the high rates of psychiatric comorbidities in insomnia, self-report measures of depression (e.g., Beck Depression Inventory) and anxiety (e.g., State-Trait Anxiety Inventory) are often included to inform the presence of psychiatric symptoms. A number of questionnaires have been developed to assess various features that are related to etiologic factors. A commonly used measure to assess cognitive features is the Dysfunctional Beliefs and Attitudes about Sleep Scale,[61] which assesses the degree to which an individual endorses particular beliefs about sleep that could contribute to the maintenance of insomnia. Measures also exist to assess pre-sleep arousal (Pre-Sleep Arousal Scale) and sleep effort (Glasgow Sleep Effort Scale[62]), as well as circadian tendencies (e.g., Morningness-Eveningness Questionnaire[63]) and sleep hygiene practices.

Objective Measures

Polysomnography

The use of PSG is not indicated for routine evaluation of a transient or chronic insomnia disorder, which places primacy on the subjective complaint for diagnosis.[64] It is indicated to rule out the presence of a sleep-related breathing disorder (e.g., OSA) or a periodic limb movement disorder, or when the initial insomnia treatment fails.[64] However, emerging evidence suggests that sleep-related breathing disorders in particular may be underrecognized causes of insomnia complaints, even among individuals who deny cardinal sleep-related breathing disorder symptoms on initial presentation.[65] PSG may also be used in patients suspected of paradoxical insomnia, a specific subtype of an insomnia disorder characterized by subjective insomnia complaints without objective evidence of insomnia. Although insomnia patients often complain of daytime sleepiness, the Multiple Sleep Latency Test, the standard assessment tool for objectively measuring daytime sleepiness,[66] is not indicated for the evaluation of sleepiness in insomnia.[67] Patients with insomnia may show longer sleep latencies on the Multiple Sleep Latency Test than normal

controls, which is attributed to their heightened physiologic arousal.[68]

A major limitation of PSG is that 1 to 2 nights of PSG may not adequately capture a representative sleep period for the patient, thus it is less useful for assessing the pattern of sleep that is important in the context of insomnia. Moreover, laboratory-based studies can exaggerate sleep problems (the first-night effect), or paradoxically, some insomnia patients may sleep better in the laboratory than they do typically in the home environment, a phenomenon termed the *reverse first-night effect*.[69]

Actigraphy

Actigraphs are small motion sensor detectors (accelerometers) that are encased in a unit about the size of a wristwatch and can be worn continuously for days to months. Actigraphy does not measure sleep per se, but rather level of activity, which is highly correlated with sleep. Sleep can be reliably inferred when low activity occurs in the presence of other indicators of sleep, such as self-report sleep diaries.[70] Dedicated software provides summary measures of behavioral sleep-wake activity and circadian rhythm parameters. Current practice guidelines suggest that actigraphy is not indicated for the routine diagnosis of insomnia but is a useful adjunct to the sleep history or other sleep assessment, such as a sleep diary, particularly for the evaluation of circadian rhythm sleep disorders or paradoxical insomnia, and in the evaluation of treatment response after insomnia interventions.[71] Compared with PSG, actigraphy is more cost-effective and is capable of measuring sleep-wake patterns across several days in the patient's home environment. However, actigraphy is less accurate at detecting sleep compared with PSG, with the accuracy diminishing with increasing amounts of nighttime wakefulness.[72] Another notable issue with the use of actigraphy is the lack of standardized scoring algorithms for sleep and wakefulness across the different devices and manufacturers. Thus it is difficult to compare results between devices that use different scoring algorithms.

Case Formulation

At the conclusion of the assessment, the clinician should have generated a list of potential differential diagnoses based on a synthesis of the information collected from the clinical interview, self-reported measures, and objective measures, if indicated. These initial diagnostic impressions should be discussed with the patient, emphasizing that they are working hypotheses and subject to change. Preliminary treatment options should be outlined and decided on with patient input, including a discussion of referrals for further assessment or treatment of other identified conditions. It is critically important to emphasize a collaborative approach with the patient as the evaluation continues and treatment strategies are initiated.

Assessment of Treatment Progress

In addition to the diagnostic assessment of insomnia, ongoing assessment should include an evaluation of treatment progress. After initiating treatment, clinicians should continue to assess insomnia symptoms at each follow-up visit. This includes an assessment of changes in nocturnal symptoms, the frequency of insomnia symptoms (e.g., how many times per week), changes in conditioned arousal or perpetuating factors, and changes in daytime functioning. The most common approach to monitoring treatment progress is to have patients complete sleep diaries during treatment. Another approach is to have patients complete an insomnia questionnaire, such as the ISI, at every session to track progress over time. Actigraphy can also be used to monitor treatment progress or adherence to treatment recommendations, if self-report alone is insufficient. In addition to the symptoms of insomnia, evaluation of treatment should include assessment of side effects. For medications, these might include amnestic episodes at night, residual daytime sleepiness, or nighttime falls for older adults. For cognitive-behavioral therapy, side effects are not well defined but might include potential daytime sleepiness associated with sleep restriction therapy.[73] Ongoing evaluation of treatment benefits and side effects can aid the overall clinical management and quality of care.

Other Considerations for Conducting the Assessment of Insomnia

This chapter described the basic tools for conducting an assessment of insomnia. Beyond these nuts and bolts, there are several considerations for conducting the assessment. First, the clinical setting will provide certain parameters for the length of the assessment and the use of self-report or objective measures. For example, if the assessment occurs in a primary care clinic or mental health clinic, the assessment will likely need to be conducted in a short amount of time and thus should focus on making a clear diagnosis of an insomnia disorder and the most appropriate treatment plan or need for further sleep evaluation. If the assessment occurs in the context of a sleep clinic, the focus should be on a comprehensive sleep evaluation for insomnia and differential diagnosis from other sleep disorders, medical conditions, and psychiatric conditions. Second, the clinician's training background and discipline will play a role in the assessment process. Physicians who conduct physical examinations can further inform the need for a PSG as part of the assessment. Evidence of a crowded upper airway, large tonsils, or enlarged turbinates may indicate the presence of sleep-disordered breathing and the need for a PSG. An assessment conducted by a psychologist might focus more on a functional analysis of the sleep schedule or differential diagnoses between insomnia and psychiatric disorders. Ideally, a collaborative care model that includes a sleep physician and a sleep psychologist can cover the entire biopsychosocial spectrum of causes and contributors to insomnia. Third, considerations for specific patient populations are important. For example, assessing children or those who might not be able to provide accurate self-reports (e.g., cognitive impairment) might rely more on objective measures than the clinical interview or questionnaires. For patients who have a bed partner, gathering information from the bed partner's observations is frequently a source of valuable information in the insomnia evaluation and should be included whenever possible. Bed partners can provide collateral information about the nature of the insomnia disorder, including symptom frequency, severity, and duration as well as the nature and degree of daytime impairment. Recent evidence suggests associations between bed partner sleep quality and daytime relationship functioning,[74] and some researchers have proposed models for engaging bed partners of insomnia patients into the treatment plan.[75] Clinical experience suggests that any patient-partner discrepancies in the insomnia

evaluation can be both important and useful therapeutically. The significant other can provide insight into important differential diagnoses, such as the presence of symptoms of sleep-related breathing disorders, periodic limb movements in sleep, or parasomnias.

With the rapid growth in mobile technology, consumer devices for monitoring sleep and daytime activity levels are becoming increasingly popular and could potentially play a role in the assessment of insomnia. These devices typically use accelerometers worn on the wrist, similar to actigraphy, which can connect with a smartphone, computer, or other mobile device and can track sleep-wake patterns. The data are processed through an "app" that the consumer can use to download the data and review progress. Currently, there is scant scientific evaluation of these consumer devices. One study compared the Fitbit device (Fitbit, Inc., San Francisco, Calif.) with actigraphy and PSG on a small sample of healthy adults.[76] Similar to actigraphy, the Fitbit device overestimated total sleep time and sleep efficiency by misidentifying wake as sleep. However, the low cost, accessibility, and convenience of these devices provide an opportunity to gather naturalistic data from patients that could be useful in a clinical assessment. Further scientific research on the use of these devices in clinical populations is needed, and the role of clinicians in interpreting these data should be clarified before these consumer devices can be recommended for the assessment of insomnia.

CLINICAL PEARLS

- The DSM-5 and ICSD3 provide better convergence on the diagnostic criteria for insomnia disorders compared with previous versions by collapsing primary versus secondary distinctions and subtypes of insomnia.
- A comprehensive clinical assessment using a clinical interview, self-report questionnaires, and sleep diaries should be able to characterize the nocturnal sleep disturbance, daytime distress or impairment related to the sleep disturbance, the temporal pattern of these symptoms, and etiologic factors related to the insomnia.

SUMMARY

The current diagnostic criteria in the DSM-5 and ICSD3 include more convergence on the essential elements of an insomnia disorder and less emphasis on distinguishing primary versus secondary insomnia or subtyping insomnia based on presumed etiology. As a result, the clinical assessment of insomnia should aim to evaluate these elements using a comprehensive clinical interview, sleep diaries, and self-report measures. These tools should enable the clinician to characterize the nocturnal sleep disturbance, identify daytime distress or impairment related to the sleep disturbance, and determine the course or temporal pattern of these symptoms. In addition, etiologic factors that predispose, precipitate, or perpetuate insomnia should be assessed, including sleep-related cognitions, circadian preferences, sleep-related arousal, and comorbid medical or psychiatric conditions. PSG and actigraphy are not routinely used in the assessment of insomnia but may be indicated to rule out other sleep disorders. The data gathered from these tools should inform case formulation and treatment planning. Ongoing assessment during treatment should be conducted to monitor progress and potential side effects.

Selected Readings

American Psychiatric Association (APA). *Diagnostic and statistical manual of mental disorders*. 5th ed. Arlington, VA: APA; 2013.

American Academy of Sleep Medicine (AASM). *International classification of sleep disorders*. 3rd ed. Darien, IL: AASM; 2014.

Buysse DJ, Ancoli-Israel S, Edinger JD, et al. Recommendations for a standard research assessment of insomnia. *Sleep* 2006;**29**(9):1155–73.

Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep* 2012;**35**(2):287–302.

Edinger JD, Buysse DJ, Deriy L, et al. Quality measures for the care of patients with insomnia. *J Clin Sleep Med* 2015;**11**:311–34.

Edinger JD, Wyatt JK, Stepanski EJ, et al. Testing the reliability and validity of DSM-IV-TR and ICSD-2 insomnia diagnoses: results of a multitrait-multimethod analysis. *Arch Gen Psychiatry* 2011;**68**(10):992–1002.

Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep* 2007;**30**(4):519–29.

Natale V, Léger D, Bayon V, et al. The consensus sleep diary: quantitative criteria for primary insomnia diagnosis. *Psychosom Med* 2015;**77**:413–18.

Schutte-Rodin S, Broch L, Buysse D, et al. Clinical guideline for the evaluation and management of chronic insomnia in adults. *J Clin Sleep Med* 2008;**4**(5):487–504.

Spielman A, Caruso L, Glovinsky P. A behavioral perspective on insomnia treatment. *Psychiatr Clin North Am* 1987;**10**:541–53.

A complete reference list can be found online at ExpertConsult.com.

Chapter
84

Insomnia and Health

Martica H. Hall; Julio Fernandez-Mendoza; Christopher E. Kline; Alexandros N. Vgontzas

Chapter Highlights

- This chapter is focused on insomnia and three specific health outcomes: cardiovascular disease and diabetes (hereafter referred to as cardiometabolic disease) and all-cause mortality.
- Epidemiologic and experimental studies support an association of insomnia with clinical (e.g., hypertension) and subclinical (e.g., heart rate variability [HRV]) cardiovascular risk factors as well as cardiovascular disease per se. Emerging data suggest that the association of insomnia with cardiovascular morbidity is more pronounced when insomnia is associated with objectively measured short sleep duration or other measures of physiologic arousal, including elevated heart rate, blunted HRV, and longer latencies in MSLT.
- Available evidence—especially epidemiologic research—has linked insomnia to poor metabolic function, ranging from overt metabolic disease (i.e., type 2 diabetes) to subclinical manifestations (e.g., metabolic syndrome, insulin resistance) that are risk factors for future diabetes. These studies also suggest that the metabolic risk is greatest among individuals with insomnia and short sleep duration. However, clinical studies involving small samples of adults diagnosed with chronic insomnia have largely failed to corroborate the epidemiologic findings, and studies that have experimentally induced poor sleep quality may generalize to other sleep disorders (e.g., sleep-disordered breathing) as well as insomnia.
- The association between insomnia and mortality is rather modest based on studies that have used insomnia symptoms not associated with any severity or chronicity criteria. Furthermore, in these studies, the lack of objective polysomnographic measures raises the question of whether this association is confounded by the presence of other sleep pathology, such as sleep apnea. More recent studies that have used more stringent criteria and polysomnographic measures suggest that mortality in insomnia is marked and significant, particularly in men and in those with short sleep duration.
- Candidate pathways through which insomnia adversely affects health include changes in physiology (e.g., increased inflammation, endocrine dysregulation, autonomic imbalance) and poor health behaviors (e.g., physical inactivity and caffeine, alcohol, and nicotine use) observed in association with chronic insomnia. More research is needed to determine whether interventions that reduce symptoms of insomnia are associated with concomitant improvements in physiology (e.g., decreased inflammation, normalization of cortisol and adrenocorticotropic hormone, increased heart failure or HRV) or health behaviors (e.g., increased physical activity, improved diet, decreased alcohol, nicotine, and caffeine use). Also unclear is whether any improvements in physiology and health behaviors associated with improvements in insomnia would attenuate the long-term risk associated with chronic insomnia disorder.

If sleep is essential to health, it stands to reason that sleep disorders may have adverse downstream consequences to health. Mounting evidence suggests that insomnia and sleep apnea, the two most common sleep disorders, are associated with significant morbidity and risk for mortality. The health consequences of sleep apnea are covered elsewhere in this text (see Chapters 14 and 126). This chapter focuses on insomnia and three specific health outcomes: cardiovascular disease and diabetes (hereafter referred to as cardiometabolic disease) and all-cause mortality. Our focus on cardiometabolic disease and mortality is driven by the current state of science. This chapter also presents a discussion of putative mechanisms through which insomnia may influence these outcomes. As research on the health consequences of insomnia becomes more established, the number of insomnia-related health outcomes may grow. Although insomnia has also been linked to mental health, these associations are not covered here because they are covered in other chapters (see Section 17, Psychiatric Disorders). We conclude with a discussion of important next steps needed to advance the state of science linking insomnia

and health, including promising areas of inquiry. This research question is relevant to basic scientific questions about the functions of sleep and definitions of sleep health,[1] as well as public health campaigns to disseminate effective and durable interventions to treat insomnia and reduce its downstream consequences to adverse health outcomes including cardiovascular disease, diabetes, and increased risk for all-cause mortality.

INSOMNIA AND CARDIOVASCULAR HEALTH

Although the association of insomnia with cardiovascular health had been noted for several years,[2,3] its association with significant cardiovascular morbidity such as clinical (e.g., hypertension) and subclinical (e.g., heart rate) risk factors and cardiovascular disease (CVD) has remained largely unexplored until recently.[4-6] This renewed interest in the association of insomnia with cardiovascular health is highlighted by the fact that only in the past few years have several systematic reviews and meta-analyses been published on the association of insomnia with hypertension, CVD, or heart rate variability (HRV).[7-10]

Several questionnaire-based studies have shown a significant relationship between insomnia, either defined as a symptom or as a disorder, and hypertension (Table 84-1).[11-20] Based on findings from large longitudinal studies, the estimated risk for incident hypertension associated with insomnia symptoms ranges between 5% and 20%.[7] Furthermore epidemiologic studies have also examined whether insomnia is a risk factor for CVD such as myocardial infarction (MI), coronary heart disease (CHD), heart failure, or, to a lesser extent, stroke (Table 84-2).[21-37] Among these cohort studies, those

Table 84-1 Cohort Studies on the Association of Insomnia Symptoms or Disorder with Hypertension

Study (Design)	*n* (Men, Age)	Insomnia	HTN
Suka et al., 2003[11] (Longitudinal, 4-yr follow-up)	4794 (100%, 40–55 yr)	Persistent DIS Persistent DMS	OR = 1.96* OR = 1.88*
Phillips & Mannino, 2007[12] (Longitudinal, 6-yr follow-up)	8757 (45%, 44–64 yr)	DIS, DMS, NRS	OR = 1.2*
Phillips et al., 2009[13] (Longitudinal, 6-yr follow-up)	1419 (41%, 64–91 yr)	DIS, DMS, EMA	RR = 0.50–1.60
Vgontzas et al., 2009[14] (Cross-sectional)	1741 (48%, 20–88 yr)	Poor sleep (DIS, DMS, EMA, NRS) Insomnia (at least 1 yr) Poor sleep + PSG 5–6 h sleep duration Insomnia + PSG 5–6 h sleep duration Poor sleep + PSG <5 h sleep duration Insomnia + PSG <5 h sleep duration	OR = 1.23 OR = 2.41* OR = 1.48 OR = 3.53* OR = 2.43* OR = 5.12*
Gangwisch et al., 2010[15] (Longitudinal, 9-yr follow-up)	4913 (36%, 32–86 yr)	Insomnia score (DIS, DMS, EMA)	HR = 1.05*/unit (32–59 yr) Mediator
Bansil et al., 2011[16] (Cross-sectional)	10,308 (49%, 18–60+ yr)	Sleep disorders† Sleep disorders + Subjective <7 h sleep duration	OR = 1.65 OR = 2.30*
Rod et al., 2011[114] (Longitudinal, 19-yr follow-up)	16,989 (50%, 36–52 yr)	DIS, DMS, EMA, NRS	RR = 1.05–1.21*
Fernandez-Mendoza et al., 2012[18] (Longitudinal, 7.5-yr follow-up)	786 (49%, 20–84 yr)	Poor sleep (DIS, DMS, EMA, NRS) Insomnia (at least 1 yr) Poor sleep + PSG <6 h sleep duration Insomnia + PSG <6 h sleep duration	OR = 1.01 OR = 2.21* OR = 1.34 OR = 3.75*
Vozoris, 2013[19] (Cross-sectional)	12,643 (51%, ≥16 yr)	DIS, DMS, EMA + Subjective <6 h sleep duration DIS + Subjective <6 h sleep duration DMS + Subjective <6 h sleep duration EMA + Subjective <6 h sleep duration	OR = 1.47* OR = 1.58* OR = 1.50* OR = 1.67*
Vozoris, 2014[20] (Cross-sectional)	12,643 (51%, ≥16 yr)	DIS, DMS, EMA 1–4 times + Subjective <6 h sleep duration DIS, DMS, EMA 5–15 times + Subjective <6 h sleep duration DIS, DMS, EMA 16–30 times + Subjective <6 h sleep duration	OR = 1.31 OR = 1.44 OR = 1.87*

*$P < .05$

†This study included a wide range of sleep disorders in the insomnia definition (e.g., sleep apnea).

DIS, Difficulty initiating sleep; DMS, difficulty maintaining sleep; EMA, early morning awakening; HR, hazard ratio; HTN, prevalent or incident hypertension in cross-sectional and longitudinal studies, respectively; NRS, nonrestorative sleep; OR, odds ratio; PSG, polysomnography; RR, relative risk.

Table 84-2 Cohort Studies on the Association of Insomnia Symptoms or Disorder with Cardiovascular Diseases

Study (Design)	*n* (Men, Age)	Insomnia	CVD
Appels et al., 1987[21] (Longitudinal, 4-yr follow-up)	3269 (100%, 39–65 yr)	DIS	RR = 1.60$^{(MI)}$
Eaker et al., 1992[22] (Longitudinal, 20-yr follow-up)	749 (0%, 45–64 yr)	DIS	RR = 3.9*$^{(MI\ or\ CHD)\dagger}$
Schwartz et al., 1998[23] (Longitudinal, 3-yr follow-up)	2960 (34%, 65–101 yr)	Restless sleep DIS DMS EMA	RR = 1.15$^{(MI)}$ RR = 1.22$^{(MI)}$ RR = 0.91$^{(MI)}$ RR = 1.08$^{(MI)}$
Elwood et al., 2006[24] (Longitudinal, 5-yr follow-up)	1874 (100%, 55–69 yr)	DIS and/or DMS (≥1–2/wk)	RR = 1.47$^{(CHD)}$ RR = 1.75*$^{(Stroke)}$
Phillips & Mannino, 2007[12] (Longitudinal, 6-yr follow-up)	11,863 (45%, 44–64 yr)	DIS, DMS, and/or NRS	RR = 1.5*$^{(CHD)}$
Meisinger et al., 2007[25] (Longitudinal, 10-yr follow-up)	6896 (51%, 45–75 yr)	DIS	RR = 1.16$^{(MI\ men)}$ RR = 1.30$^{(MI\ women)}$
Chien et al., 2010[26] (Longitudinal, 16-yr follow-up)	3430 (47%, ≥35 yr)	DIS, DMS and/or NRS >3 times/wk	RR = 1.78*†
Loponen et al., 2010[27] (Longitudinal, 7-yr follow-up)	2753 (100%, 40–55 yr)	5–6 of DIS, DMS, EMA, NRS, or others	RR = 1.42$^{(CHD)}$
Chandola et al., 2010[28] (Longitudinal, 15-yr follow-up)	10,308 (67%, 35–55 yr)	Lost sleep over worry Restless, disturbed nights Restless, disturbed nights + Subjective <6 h sleep duration	RR = 1.02† RR = 1.36*† RR = 1.28*†
Laugsand et al., 2011[29] (Longitudinal, 11-yr follow-up)	51,982 (45%, 20–89 yr)	1 of DIS, DMS, or NRS 2 of DIS, DMS, or NRS 3 of DIS, DMS, or NRS	HR = 1.30*$^{(MI)}$ HR = 1.47*$^{(MI)}$ HR = 2.12*$^{(MI)}$
Westerlund et al., 2013[30] (Longitudinal, 13-yr follow-up)	41,192 (36%, NA)	DIS DMS EMA NRS	HR = 0.91$^{(CVD)\dagger}$ HR = 0.97$^{(CVD)\dagger}$ HR = 0.89$^{(CVD)\dagger}$ HR = 0.95$^{(MI/CHD)\dagger}$
Jaussent et al., 2013[31] (Longitudinal, 6-yr follow-up)	5494 (43%, 65–95 yr)	1–2 of DIS, DMS, and EMA 3–4 of DIS, DMS, and EMA	HR = 0.94$^{(MI/CHD)}$ HR = 0.94$^{(MI/CHD)}$
Sands-Lincoln et al., 2013[32] (Longitudinal, 10-yr follow-up)	86,329 (0%, 50–79 yr)	3 ≤<6 6 ≤ WHIIRS <9 WHIIRS ≥9 WHIIRS ≥9 + Subjective >10 h sleep duration	HR = 1.11* $^{(MI/CHD)\dagger}$ HR = 1.09 $^{(MI/CHD)\dagger}$ HR = 1.19* $^{(MI/CHD)\dagger}$ HR = 1.93* $^{(MI/CHD)\dagger}$
Laugsand et al., 2014[33] (Longitudinal, 11-year follow-up)	44,047 (46%, 20–89 yr)	1 of DIS, DMS, or NRS 2 of DIS, DMS, or NRS 3 of DIS, DMS, or NRS	HR = 0.95$^{(HF)}$ HR = 1.43$^{(HF)}$ HR = 5.25*$^{(HF)}$
Michal et al., 2014[34] (Cross-sectional)	10,000 (48%, 35–74 yr)	DIS, DMS, or sleeping too much several days DIS, DMS, or sleeping too much more than half the days DIS, DMS, or sleeping too much nearly every day	OR = 1.22$^{(CHD)}$ OR = 1.59*$^{(CHD)}$ OR = 1.64*$^{(CHD)}$
Michal et al., 2014[34] (Cross-sectional)	10,000 (48%, 35–74 yr)	DIS, DMS, or sleeping too much several days DIS, DMS, or sleeping too much more than half the days DIS, DMS, or sleeping too much nearly every day	OR = 0.95$^{(MI)}$ OR = 1.57*$^{(MI)}$ OR = 1.73*$^{(MI)}$
Michal et al., 2014[34] (Cross-sectional)	10,000 (48%, 35–74 yr)	DIS, DMS, or sleeping too much several days DIS, DMS, or sleeping too much more than half the days DIS, DMS, or sleeping too much nearly every day	OR = 1.23$^{(HF)}$ OR = 1.56$^{(HF)}$ OR = 1.55$^{(HF)}$

Table 84-2 Cohort Studies on the Association of Insomnia Symptoms or Disorder with Cardiovascular Diseases—cont'd

Study (Design)	*n* (Men, Age)	Insomnia	CVD
Sivertsen et al., 2014[35] (Longitudinal, 11-yr follow-up)	24,715 (43%, 32–66 yr)	DSM-IV Insomnia	OR = 1.46*(MI)
Canivet et al., 2014[36] (Longitudinal, 13-yr follow-up)	13,617 (43%, 45–64 yr)	DIS, DMS, EMA, and/or NRS	HR = 1.00(men)†
		DIS, DMS, EMA, and/or NRS + Subjective <6 h sleep duration	HR = 1.40*(women)†
		DIS, DMS, EMA, and/or NRS + Subjective >9 h sleep duration	HR = 1.50*(women)†
			HR = 2.10*(women)†
Wu et al., 2014[37] (Longitudinal, 4-year follow-up)	85,752 (38%, 18–65+ yr)	ICD-9 Insomnia	HR = 1.54*
		Remitted ICD9 Insomnia	HR = 1.57*
		Relapsed ICD9 Insomnia	HR = 1.52*
		Persistent ICD9 Insomnia	HR = 1.55*

*$P < .05$.
†These studies included death from CVD in the definition of incident CVD (see section on Insomnia and Mortality).
CVD, Prevalent or incident cardiovascular diseases in cross-sectional and longitudinal studies, respectively, which primarily included myocardial infarction (MI), coronary heart disease (CHD), heart failure (HF), and, in a few studies, stroke; DIS, difficulty initiating sleep; DMS, difficulty maintaining sleep; DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition; HR, hazard ratio; ICD, *International Classification of Diseases,* ninth edition; NRS, nonrestorative sleep; OR, odds ratio; RR, relative risk: WHIIRS, Women's Health Initiative Insomnia Rating Scale score.

using a longitudinal design have estimated that the risk for incident CVD associated with insomnia symptoms is about 45% higher than among individuals without insomnia symptoms.[8] More specifically, the relative risk associated with insomnia compared with those without insomnia is 41% for incident MI and 55% for incident CHD or stroke.[9]

Although these large cohort studies have shown a significant association between self-reported insomnia symptoms and hypertension or CVD, most did not include polysomnography (PSG) evaluation and thus could not control for other sleep pathology such as sleep-disordered breathing (SDB). Furthermore, although most studies reported significant associations, the effect sizes reported are modest; in fact, the degree of association found in meta-analyses[2,7-9] had been dismissed as artifact or flawed by some clinicians and researchers.[12,38] However, more recent work has identified a specific high-risk phenotype of individuals with insomnia and PSG-assessed short sleep duration.[5] Two studies from the Penn State Adult Cohort have shown a synergistic effect between insomnia (i.e., a chronic complaint) and PSG-measured short sleep duration (i.e., <6 hours) on the risk for hypertension even after controlling for multiple confounders, including SDB (see Table 84-1).[14,18] Compared with good sleepers, patients with insomnia who slept objectively between 5 and 6 hours and patients with insomnia who slept less than 5 hours had 3.5-fold and 5.1-fold increased cross-sectional odds, respectively, of prevalent hypertension—a degree of association about as high as that between SDB and hypertension. In contrast, insomnia with more than 6 hours of sleep was not significantly associated with prevalent hypertension (odds ratio [OR] = 1.3).[14] Longitudinally, patients with insomnia who slept less than 6 hours had a 3.8-fold odds of incident hypertension compared with good sleepers, whereas insomnia with more than 6 hours of sleep was not significantly associated with incident hypertension (OR = 0.85).[18] The combined effects of short sleep duration and sleep disturbance have been examined in other cohort studies with the limitation of using subjective sleep measures, which may account for some of the inconsistent findings (see Tables 84-1 and 84-2).[16,19,28,32,36]

The association of insomnia with subclinical risk factors for CVD, such as blood pressure dipping, resting HR, or HRV, has been examined primarily in laboratory-based studies. An elevated nighttime systolic blood pressure and a lower day-to-night dipping in systolic blood pressure have been found in a study of normotensive primary insomnia patients.[39] Furthermore nighttime systolic blood pressure was positively correlated with beta activity in the electroencephalogram (EEG) during the night, a marker of cortical hyperarousal.[39] Another study has shown that morning-evening and day-to-day ambulatory blood pressure variability is significantly higher in patients with persistent insomnia compared with subjects without insomnia and that insomnia combined with short sleep duration was associated with even greater variability.[40] A recent study using the Multiple Sleep Latency Test (MSLT) as a measure of physiologic hyperarousal has shown that patients with insomnia who had an MSLT score higher than 14 minutes and those who had an MSLT score higher than 17 minutes had 3.3-fold and 4.3-fold odds, respectively, of high blood pressure, whereas insomnia with a MSLT of 8 to 14 minutes was not significantly associated with high blood pressure (OR = 1.17).[41]

Several studies have shown an elevated resting HR or blunted HRV in patients with insomnia (Table 84-3),[40,42-48] with the latter suggesting a shift of the sympathovagal balance toward a predominance of sympathetic modulation during both wake and nighttime.[10] However, these findings have been inconsistent.[49-53] In those studies reporting a positive association, insomnia patients were carefully defined and presented with objective sleep disturbances. For example, the association between insomnia and alterations in nighttime HR and HRV was examined in a recent study of 58 subjectively defined patients with primary insomnia and 46 healthy controls.[53] Differences between patients with primary insomnia and healthy controls in resting HR were not found, and previous results of blunted HRV[43] could not be replicated. However,

Table 84-3 Experimental and Cohort Studies on the Association of Insomnia Symptoms or Disorder with Heart Rate

Study (Design)	*n* (Men, Age)	Insomnia	Elevated HR	Impaired HRV
Monroe, 1967[49] (In-lab)	32 (100%, 20–42 yr)	DIS, DMS	No	
Haynes et al., 1981[42] (In-lab)	21 (19%, 18–21 yr)	DIS, DMS	Yes	
Freedman & Sattler, 1982[50] (In-lab)	24 (21%, 19–56 yr)	DIS, DMS	No	
Bonnet & Arand, 1998 [43] (In-lab)	24 (50%, 30.0 ± 6.0 yr)	Psychophysiologic insomnia	Yes	Yes (nighttime)
Nilsson et al., 2001[44] (Cross-longitudinal)	22,933 (59%, 46.7 ± 7.0 yr)	DIS, DMS	Yes	
Varkevisser et al., 2005[51] (In-lab constant routine)	24 (54%, 44.4 ± 8.3 yr)	Primary insomnia	No	No (daytime)
Fang et al., 2008[45] (In-lab)	39 (33%, 31.0 ± 10.0 yr)	Primary insomnia	No	Trend (daytime)
Jurysta et al., 2009[52] (In-lab)	28 (100%, 41.0 ± 11.0 yr)	Primary insomnia	No	No (nighttime)
Spiegelhalder et al., 2011[53] (In-lab)	104 (38%, 39.5 ± 11.8 yr)	Primary insomnia Insomnia + PSG <6 h sleep duration	No Yes	No Yes (nighttime)
Johansson et al., 2011[40] (Cross-sectional)	1908 (44%, 41–74 yr)	Persistent insomnia Insomnia + Subjective <6 h sleep duration		Yes (daytime) Yes (daytime)
Yang et al., 2011[46] (In-lab)	187 (37%, 43.9 ± 10.4 yr)	Primary insomnia	Yes (daytime) Yes (nighttime)	Yes (daytime) Yes (nighttime)
de Zambotti et al., 2013[47] (In-lab)	18 (44%, 19–28 yr)	Primary insomnia		Yes (nighttime)
Farina et al., 2014[48] (In-lab)	140 (44%, 53.2 ± 13.6 yr)	Primary insomnia	Yes	Yes (nighttime)

DIS, Difficulty initiating sleep; DMS, difficulty maintaining sleep; HR, resting heart rate; HRV, heart rate variability; PSG, polysomnography.

when the authors used the criterion put forward in previous epidemiologic studies (i.e., PSG-defined short sleep duration),[14] patients with insomnia who slept objectively less than 6 hours had reduced parasympathetic activity compared with healthy controls.[53]

In summary, epidemiologic and experimental studies support an association of insomnia with clinical (e.g., hypertension) and subclinical (e.g., HRV) cardiovascular risk factors as well as CVD per se. A growing literature suggests that the association of insomnia with cardiovascular morbidity is more pronounced when insomnia is associated with objectively measured short sleep duration or other measures of physiologic arousal such as elevated HR, blunted HRV, and longer latencies in MSLT.

INSOMNIA AND DIABETES OR METABOLIC SYNDROME

Relative to CVD morbidity, less research has been conducted on the relationship between insomnia and metabolic dysfunction. Nevertheless, available evidence—especially epidemiologic research—has linked insomnia to poor metabolic function, ranging from overt metabolic disease (i.e., type 2 diabetes) to subclinical manifestations (e.g., metabolic syndrome, insulin resistance) that are risk factors for future diabetes.

Multiple epidemiologic studies have examined the association between insomnia symptoms and type 2 diabetes risk. Although conflicting findings have been observed,[54,55] most cross-sectional and longitudinal studies have reported significant associations between the presence of insomnia symptoms and prevalent or incident diabetes, respectively (Table 84-4).[56-59] For instance, a 2010 meta-analysis of five longitudinal studies found that difficulty initiating sleep and difficulty maintaining sleep were associated with significantly greater risk for developing type 2 diabetes (RR = 1.57 and RR = 1.84, respectively) relative to adults without these complaints.[60] Notably, the diabetes risk associated with these sleep complaints was greater in magnitude than the diabetes risk associated with short (≤5 to 6 hour) or long (>8 to 9 hour) sleep duration (RR = 1.28 and RR = 1.48, respectively).[60]

As previously noted in regard to CVD risk, the risk for type 2 diabetes may be most pronounced among those with insomnia and short sleep duration. In the Penn State Adult Cohort, adults with insomnia along with PSG-assessed short sleep duration (<6 hours) were significantly more likely to have diabetes relative to adults without sleep complaints and 6 hours or more sleep duration (OR = 2.95 for ≤5 hours, OR = 2.07 for 5 to 6 hours).[61] Notably, the odds of having diabetes among adults without sleep complaints but 6 hours or less sleep duration and adults with insomnia and normal sleep duration was not significantly elevated (OR ≤ 1.45 and OR = 1.10, respectively).[61] Conflicting results have been observed in the few other studies that have examined the combined effect of insomnia symptoms and short sleep duration on diabetes risk, perhaps owing to a reliance on self-reported sleep duration.[62,63]

The metabolic syndrome (MetSyn) is a cluster of adverse cardiometabolic factors—obesity, dyslipidemia, insulin resistance or glucose dysregulation, and elevated blood

Table 84-4 Studies on the Association of Insomnia Symptoms or Disorder with Type 2 Diabetes

Study (Design)	*n* (% Men, Age)	Insomnia	T2D
Kawakami et al., 2004[58] (Longitudinal, 8-yr follow-up)	2265 (100%, unknown)	DIS DMS	HR = 2.98* HR = 2.23*
Nilsson et al., 2004[115] (Longitudinal, 7- to 22-yr follow-up)	6599 (100%, 35–51 yr)	DIS	OR = 1.61*
Bjorkelund et al., 2005[54] (Longitudinal, 32-yr follow-up)	1462 (0%, 38–60 yr)	Sleep complaints	RR = 1.06
Mallon et al., 2005[56] (Longitudinal, 12-yr follow-up)	1170 (47%, 45–65 yr)	DIS (men only) DMS	RR = 2.4 RR = 4.8*, 1.8†
Meisinger et al., 2005[57] (Longitudinal, 7.5-yr follow-up)	8269 (50%, 25–74 yr)	DIS DMS	HR = 1.10, 1.42† HR = 1.60*, 1.98*†
Hayashino et al., 2007[59] (Longitudinal, 4-yr follow-up)	6509 (74%, 19–69 yr)	DIS DMS	HR = 1.63* HR = 1.34
Vgontzas et al., 2009[61] (Cross-sectional)	1741 (48%, 20–88 yr)	Insomnia (≥1 year) Insomnia + PSG 5–6 h sleep duration Insomnia + PSG <5 h sleep duration	OR = 1.69 OR = 2.07 OR = 2.95*
Rod et al., 2011[17] (Longitudinal, 19-yr follow-up	16,989 (74%, 36–52 yr)	DIS DMS EMA Poor sleep quality	HR = 1.04, 1.92*† HR = 1.16, 1.54*† HR = 1.10, 1.28† HR = 1.49*, 1.86*†
Kita et al., 2012[116] (Longitudinal, 4-yr follow-up)	3570 (79%, 35–55 yr)	Total insomnia score‡ DIS‡ DMS‡ EMA‡ Sleep quality‡	OR = 0.95, 1.16* OR = 0.97, 0.66 OR = 1.69, 5.03* OR = 0.32, 1.74 OR = 0.60, 3.71*
Lou et al., 2012[63] (Cross-sectional)	16,893 (46%, 18–75 yr)	Poor sleep (DIS or DMS ≥ 8 days/month) Poor sleep + Subjective ≤6 hr sleep duration	OR = 1.81* OR = 1.41*
Zhang et al., 2012[55] (Longitudinal, 4-yr follow-up)	2316 (48%, 41 yr [mean])	Insomnia symptoms (DIS, DMS, or EMA ≥ 3 days/wk for ≥1 yr) Insomnia disorder (insomnia symptoms + daytime impairment)	OR = 0.34 OR = 0.49
Lai et al., 2013[117] (Longitudinal, 9-yr follow-up)	136,806 (36%, 51 yr [mean])	Insomnia (ICD-9) Sleep disturbance (ICD-9)	HR = 1.02 HR = 1.11*
Liu et al., 2013[118] (Cross-sectional)	3668 (48%, 58 yr [mean])	Sleep disturbance§ Sleep disorder§	OR = 1.37* OR = 1.38

*$P < .05$
†Ratios provided separately for men and women, respectively.
‡Odd ratios given are for those with a family history of diabetes and those without a family history of diabetes, respectively.
§Based on self-report to a health professional (sleep disturbance) or being told by a health professional (sleep disorder).
DIS, Difficulty initiating sleep; DMS, difficulty maintaining sleep; EMA, early morning awakening; HR, hazard ratio; ICD-9, *International Classification of Diseases,* ninth edition; NRS, nonrestorative sleep; OR, odds ratio; PSG, polysomnography; RR, relative risk; T2D, prevalent or incident type 2 diabetes in cross-sectional and longitudinal studies, respectively.

pressure—that is closely linked to risk for type 2 diabetes. In cross-sectional epidemiologic studies, poor subjective sleep quality has been associated with a greater likelihood of having MetSyn,[64,65] although some studies have failed to demonstrate this relationship (Table 84-5).[66,67] Hall and colleagues observed greater odds for MetSyn with lighter and more fragmented sleep, as assessed by PSG indexes of beta EEG power and sleep efficiency, respectively.[66] Other cross-sectional research has found significantly higher odds of MetSyn among adults who report any insomnia symptoms (i.e., difficulty initiating sleep, difficulty maintaining sleep, or early morning awakening), although individual symptoms were not predictive of MetSyn.[68] However, the only prospective study on the topic found that subjective sleep complaints of difficulty initiating sleep and unrefreshing sleep were associated with incident MetSyn over a 3-year follow-up, although a clinical diagnosis of insomnia was not.[69] No research has evaluated the combination of insomnia symptoms and short sleep duration on MetSyn risk.

Subjective sleep disturbance has also been associated with key preclinical indicators of early progression toward type 2 diabetes, such as impaired fasting glucose levels and insulin

Table 84-5 Studies on the Association of Insomnia Symptoms or Disorder with Metabolic Syndrome

Study (Design)	*n* (% Male, Age)	Insomnia	MetSyn
Jennings et al., 2007[64] (Cross-sectional)	210 (57%, 30–54 yr)	Sleep quality (PSQI)	OR = 1.44*
Troxel et al., 2010[69] (Longitudinal, 3-year follow-up)	812 (33%, 45–74 yr)	DIS DMS NRS Insomnia syndrome (insomnia symptoms + daytime impairment)	OR = 1.18* OR = 1.04 OR = 1.71* OR = 1.60
Hall et al., 2012[66] (Cross-sectional)	340 (0%, 46–57 yr)	Low PSG sleep efficiency PSG NREM beta-frequency EEG activity	OR = 1.40* OR = 1.45*
Kazman et al., 2012[67] (Cross-sectional)	248 (37%, 18–60 yr)	Poor sleep quality (PSQI ≥6)	OR = 0.67
Sakura et al., 2014[119] (Cross-sectional)	3936 (40%, 57 yr [mean])	DIS, DMS, or EMA	OR = 1.23 *
Okubo et al., 2014[65] (Cross-sectional)	1481 (37%, 20–80 yr)	Poor sleep quality (PSQI ≥6)	OR = 2.37*, 2.71*†

*$P < .05$.
†Ratios provided separately for men and women, respectively.
DIS, Difficulty initiating sleep; DMS, difficulty maintaining sleep; EMA, early morning awakening; MetSyn, prevalent or incident metabolic syndrome in cross-sectional and longitudinal studies, respectively; NREM, non–rapid eye movement sleep; NRS, nonrestorative sleep; OR, odds ratio; PSG, polysomnography; PSQI, Pittsburgh Sleep Quality Index.

resistance. Adults with difficulty maintaining sleep or early morning awakening are more likely to have impaired fasting glucose levels.[70] Studies have noted greater insulin resistance among adults with poor sleep quality,[64,71-73] insomnia symptoms,[74] or insomnia symptoms in conjunction with hypnotic use.[75] These studies have found little evidence to suggest β-cell dysfunction in those with insomnia.[71,74,75] In the sole longitudinal study, adults with difficulty maintaining sleep were significantly more likely to develop impaired fasting glucose over a 2-year follow-up.[76] Once again, data suggest that short sleep duration may augment the risk associated with insomnia on glucose control: the greatest odds for impaired glucose tolerance were observed with poor sleep quality and less than 6 hours self-reported sleep duration (OR = 6.37).[77]

Overall, epidemiologic research suggests that insomnia symptoms are related to type 2 diabetes and its metabolic precursors. However, limitations in this research cast suspicion on the independent association between insomnia and the risk for metabolic dysfunction. Studies varied widely on their definitions of sleep disturbance or operationalization of insomnia symptoms, and rarely did studies attempt to account for insomnia symptom severity or chronicity. Moreover, many studies neglected or were unable to account for important factors associated with diabetes risk (e.g., depression, sleep medications, SDB). Importantly, because most studies did not use PSG in their assessment of sleep disturbance, subjective sleep complaints may have been due to other sleep disorders rather than insomnia. Therefore studies with well-characterized samples of adults with diagnosed insomnia and objective measures of glucose metabolism have the potential to clarify the commonly observed epidemiologic associations.

Studies incorporating small samples ($n \leq 28$) of adults carefully screened for insomnia without comorbid conditions have failed to replicate the associations seen in most of the epidemiologic studies. Two studies compared adults who met diagnostic criteria for primary insomnia with matched controls on indexes of glucose metabolism derived from laboratory assessments (i.e., oral glucose tolerance test,[78] hyperinsulinemic-euglycemic clamp[79]). Glucose metabolism did not differ between the insomnia and control groups in either study.[78,79] Moreover, in a sample of adults with primary insomnia, those who also had PSG-assessed short sleep duration (≤6 hours) responded to an oral glucose tolerance test with lower insulin secretion in conjunction with greater insulin sensitivity and no difference in glycemic control compared with those with insomnia and more than 6 hours sleep duration.[80] Despite being in opposition to many of the epidemiologic findings, these results are consistent with an epidemiologic study that found that insomnia symptoms combined with poor actigraphic sleep efficiency were associated with lower insulin levels and greater insulin sensitivity among adults without diagnosed diabetes; however, among those with diabetes, these sleep symptoms were associated with worse insulin sensitivity.[81]

Several studies have used experimental sleep disruption to partially model the effects of insomnia on metabolic dysfunction. Although these studies differ from insomnia disorder in several ways (i.e., exogenous stimuli, acute disruption), they are able to increase sleep fragmentation or reduce time spent in slow wave sleep without reducing total sleep time. In one study, two nights of experimental sleep fragmentation across all sleep stages led to reduced insulin sensitivity and reduced glucose effectiveness.[82] Two other studies found that one to three nights of selective slow wave sleep suppression significantly impaired glucose tolerance and insulin sensitivity.[83,84] Despite differences between both the source and duration of sleep disruption, these experimental studies that partially model insomnia provide compelling evidence that experimental sleep disruption causes significant metabolic alterations, independent of sleep duration. Causal associations among insomnia and metabolic function could also be examined in the context of treatment. Yet, we are not aware of any

studies that have examined whether treatment of insomnia, either pharmacologic or behavioral, alters metabolic function in these patients.

In summary, most epidemiologic research suggests a significant association between insomnia and indexes of diabetes risk. These studies also suggest that the metabolic risk associated with insomnia may be augmented further when accompanied by short sleep duration. However, clinical studies involving small samples of adults diagnosed with chronic insomnia have largely failed to corroborate the epidemiologic findings, and studies that have experimentally induced poor sleep quality may generalize to other sleep disorders (e.g., SDB) as well as insomnia. Further research involving better characterization of insomnia samples, assessment of insomnia severity and chronicity, and measurement of potential confounders are needed. Moreover, examining the effects of successful insomnia treatment is necessary to better understand the actual causal role of insomnia in the development and worsening of metabolic dysfunction.

INSOMNIA AND MORTALITY

It has been a decade since Phillips and Mannino evaluated whether insomnia "kills" in a large prospective study in North America; the authors concluded that neither insomnia complaints nor the use of hypnotics were associated with increased risk for death.[85] In this section, we will examine the literature that has been published on this important question. First, we will review studies of large cohorts with subjective sleep data. Then we will assess studies that have used objective sleep data. Finally, we will examine the association of hypnotics and mortality.

Many large epidemiologic studies have examined the association of insomnia with mortality, including all-cause or cardiovascular. Most of these studies were not designed specifically to address the insomnia-mortality link; that is, their main focus was mortality associated with major medical problems, such as cardiovascular disorders or cancer. Thus it is not surprising that the definition of insomnia varies across studies, objective sleep data are not available, length of follow-up is variable, and the number of confounding factors is often limited and differs from study to study. The early studies either failed to find an association between insomnia and mortality[85] or reported a protective effect of insomnia on mortality risk.[86]

Two recent meta-analyses have been published including large cohort studies from the past 20 years. Li and colleagues included 13 studies and evaluated the association between insomnia and risk for cardiovascular mortality. Insomnia overall significantly increased the risk for cardiovascular mortality by 33% compared with individuals without insomnia.[9] Another meta-analysis included 10 published studies that evaluated the association between individual insomnia symptoms and total mortality. The pooled hazard ratios of total mortality were 1.14 (95% confidence interval [CI] = 1.04 to 1.24) for difficulty initiating sleep, 1.08 (95% CI = 0.96 to 1.22) for difficulty maintaining sleep, 1.00 (95% CI = 0.94 to 1.06) for early morning awakenings, and 1.17 (95% CI = 1.01 to 1.36) for nonrestorative sleep relative to individuals without those symptoms.[87] The same authors conducted a meta-analysis using data from six published studies that examined the associations between insomnia symptoms and cardiovascular mortality. The pooled hazard ratios of cardiovascular mortality were 1.45 (95% CI = 1.09 to 1.93) for difficulty initiating sleep, whereas there was no significant association between difficulty maintaining sleep and early morning awakening and mortality. A recent prospective study of 23,447 U.S. adults participating in the Health Professionals Follow-Up Study showed a significant association between difficulty initiating sleep (55% increased risk) and nonrestorative sleep (32% increased risk) and cardiovascular mortality.[5] Collectively, individual studies and meta-analyses provide somewhat equivocal evidence that insomnia is associated with increased risk. Inconsistencies in the assessment of insomnia across studies may contribute to the variability in findings, suggesting that more work is needed to reliably estimate associations among insomnia and mortality. Also warranted is the assessment of sleep apnea, which is also highly prevalent, with known consequences to morbidity and cardiovascular mortality.[88]

In the past few years, following the findings that insomnia with short sleep duration is associated with elevated indexes of physiologic arousal, a series of publications have examined the synergistic effect of insomnia with short sleep duration and cardiometabolic morbidity.[5] As described earlier, insomnia with short sleep duration is associated with a marked, significant, and dose-response risk for hypertension and diabetes. The insomnia–short sleeper phenotype has also been examined in relation to mortality in a random general population sample of 1741 adults who were studied in the sleep laboratory and followed after 10 years (women) and 14 years (men). Insomnia was defined by a complaint of insomnia with duration of at least 1 year. In men, mortality risk was significantly increased in insomniacs who slept less than 6 hours compared with the normal sleep duration and no insomnia group (about 400% increased risk). This risk was even higher in patients with diabetes or hypertension (about 700% increased risk) who suffered from this type of insomnia. No significant association between mortality and insomnia with short sleep duration was found in women.[89] One of the limitations of this study was that the number of deaths was too small to allow analysis for a specific cause, that is, cardiovascular mortality; furthermore the percentage of women who died was only 5% at the time of follow-up, leaving open the question of a gender-specific effect. In a recent reanalysis of the same cohort, but with a longer follow-up period (18 years for men, 14 years for women) and with a mortality rate of about 16% in women, the results remained unchanged. Specifically, in men, there was a significant association of this type of insomnia with mortality, whereas again no such association was found in women. The authors concluded that although insomnia is more prevalent in women, it appears that its biologic effect is more consequential in men.[90]

A recent large study from Norway examined the association of midlife insomnia and subsequent all-cause mortality in more than 6000 participants aged 40 to 45 years. These subjects provided information on insomnia using the Karolinska Sleep Questionnaire (which follows the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*), various sociodemographic and health behaviors, as well as obstructive sleep apnea symptoms and sleep duration. Insomnia was associated with a 3-fold risk of mortality over 13 to 15 years follow-up. The risk seemed even higher in males or when insomnia was combined with short sleep duration. The

authors concluded that the insomnia-mortality association may be confined to subgroups: males and short sleepers.[91] Although based on subjective data, this study is different from the other similar studies in that the criteria for insomnia were stringent and the presence of obstructive sleep apnea was assessed based on symptoms such as snoring and breathing cessation during sleep reported by the subjects or their bed partners. These differences in methodology may potentially explain the consistencies between this subjective study and studies using objective sleep measures.

Hypnotics are frequently prescribed from physicians to treat insomnia or insomnia symptoms, either as a primary or as an adjunct therapy. In the late 1970s, a paper by Kripke and colleagues, based on data from a large registry of the American Cancer Society, reported that frequent use of sleeping pills was associated with 1.5 times higher mortality compared with that in individuals who had never used sleeping pills.[92] These results were replicated in subsequent studies by the same group with hazard ratios of 1.35 for men and 1.22 for women.[93] Similar results were reported more recently based on data from a large group of community-dwelling older adults in Taiwan; specifically, frequent hypnotic use had an increased mortality risk of about 37%.[94] Another study using an insurance database in Pennsylvania also reported higher mortality risk; specifically, the hazard ratios ranged from 3.6 to 5.3 for several of the common hypnotics, such as zolpidem and temazepam.[95] However, in this study there was no control for confounding factors such as sleep disorders and psychiatric pathology. These limitations seem to be important; a study from France in an elderly general population reported that the increased mortality associated with hypnotics became nonsignificant after adjusting for sleep and psychiatric disorders.[96] The authors concluded that underlying psychiatric disorders appeared to be the principal confounders of the observed association. Overall, it appears that, based on recent published studies, the association of hypnotic use with mortality is rather small but warrants further examination given the widespread use of these medications, particularly in the older, frail population.

The association between insomnia and mortality is rather modest based on studies that have used insomnia symptoms not associated with any severity or chronicity criteria. Furthermore, in these studies, the lack of objective PSG measures raises the question of whether this association is confounded by the presence of other sleep pathology, such as sleep apnea. More recent studies that have used more stringent criteria and PSG measures suggest that mortality in insomnia is marked and significant, particularly in men and in those with short sleep duration. Further studies are needed with larger cohorts, uniform criteria for insomnia disorder, and objective sleep data to confirm these more recent observations.

MECHANISMS LINKING INSOMNIA WITH CARDIOMETABOLIC DISEASE

As summarized earlier, mounting evidence suggests that insomnia is prospectively linked to both the risk for and clinical course of cardiometabolic disease. The evidence is strongest for chronic insomnia disorder, although symptoms of insomnia, too, have been linked to increased risk. Proposed physiologic mechanisms include factors known to change in response to experimental sleep deprivation and restriction, including endocrine dysregulation, increased inflammation, and autonomic imbalance. Early studies in this area focused on stress hormones in light of the hypothesized role of psychological stress in the pathogenesis of insomnia as well as evidence of psychophysiologic hyperarousal in insomnia first documented by Monroe in 1967.[49,97] Most studies reported increased hypothalamic-pituitary-adrenal axis activation in patients with insomnia, including increased cortisol secretion and alterations in the diurnal cortisol profile,[79,98-101] although some have failed to observe group differences.[102,103]

Other studies have focused on inflammation as a pathway through which chronic insomnia may contribute to cardiometabolic disease. Two large-scale epidemiologic studies reported no significant association among C-reactive protein and symptoms of insomnia in community-dwelling adults.[101,104] Two other studies have documented increased inflammation in patients with insomnia compared with good sleepers, as measured by the secretion[105] and diurnal profiles of interleukin-6.[106] At present, differences across studies, including indexes of inflammation and severity and chronicity of insomnia, preclude conclusions about inflammation as a mechanism linking insomnia to cardiometabolic disease, or the extent to which insomnia is a cause, or effect, of inflammation. The insomnia-inflammation pathway is certainly plausible given the effects of experimental sleep deprivation on circulating markers of inflammation and their cellular and molecular upstream pathways.[107,108]

More studies have evaluated autonomic tone in relation in insomnia, perhaps because it can be measured continuously and noninvasively during sleep, and HRV-derived indexes of autonomic tone exhibit greater short-term stability than do measures of the hypothalamic-pituitary-adrenal axis or inflammation.[109] Although non–rapid eye movement sleep is a predominantly parasympathetic state, most studies report decreased parasympathetic activity as indexed by high-frequency HRV during sleep in patients with insomnia compared with controls.[43,45,46,48,52,109] Autonomic imbalance in insomnia may be further compounded by alterations in dynamic indexes of high-frequency HRV[52,110] and increased sympathetic nervous system activity, as measured by impedance cardiography and the ratio of low- to high-frequency HRV (LF/HF ratio), also known as *sympathovagal balance*.[47,53,111] The link between insomnia and the LF/HF ratio has been further instantiated by the demonstration that intradermal, but not sham, acupuncture was associated with improvements in both symptoms of insomnia and the LF/HF ratio in hospitalized stroke patients with insomnia.[112]

Insomnia may also predispose to greater cardiometabolic risk through poor health behaviors. Although the relationships between insomnia symptoms and health behaviors are likely bidirectional,[1,2] adults with insomnia are more likely to use alcohol,[1,3] smoke,[3] and be physically inactive[1] than adults without insomnia; each of these behaviors is associated with increased cardiometabolic risk. Adults with insomnia have also been found to have lower cardiorespiratory fitness,[4] which is a strong independent predictor of cardiovascular morbidity and mortality.[5] Finally, insomnia symptoms may predispose to cardiometabolic risk through poor diet; poor sleep quality, fragmented sleep, and altered sleep architecture (e.g., low slow wave sleep) are associated with dietary macronutrient and micronutrient composition[6,7] and satiety.[8,9] More work is needed to establish whether these putative mechanisms do, in

fact, lie on the causal pathway linking insomnia to cardiometabolic disease.

It is currently unclear whether interventions that reduce symptoms of insomnia are associated with concomitant improvements in physiology (e.g., decreased inflammation, normalization of cortisol and adrenocorticotropic hormone, increased high-frequency HRV) or health behaviors (e.g., increased physical activity, improved diet, decreased alcohol, nicotine, and caffeine use). Also unclear is whether any improvements in physiology and health behaviors associated with improvements in insomnia would attenuate the long-term risk associated with chronic insomnia disorder. Moreover, little is known about factors that may moderate associations among insomnia, insomnia mechanisms, and cardiometabolic risk. Do age and gender matter? Do other comorbidities such as major depression and sleep apnea have an additive or synergistic effect on cardiometabolic risk in patients with insomnia? What is clear is that identification of mechanisms and effect moderators is critical to treatment and prevention of the cardiometabolic consequences of insomnia.

CLINICAL PEARLS

- Several large cohort studies have demonstrated that reports of insomnia, particularly in those with short sleep duration, are associated with incident cardiovascular disease and events.
- Insomnia disorder is associated with increased risk for mortality, particularly in men and in those with short sleep duration. Further studies are needed with larger cohorts, uniform criteria for insomnia disorder, and objective sleep data to confirm these observations.
- Symptoms of insomnia, including difficulty initiating and maintaining sleep, are prospectively associated with incident type 2 diabetes and metabolic syndrome. This effect appears to be stronger in those with objective short sleep duration.

SUMMARY

Mounting evidence suggests that insomnia is associated with subclinical and prognostic indicators of cardiovascular disease, diabetes and metabolic syndrome, and all-cause mortality.[113] Chronicity and severity of symptoms appear to play an important role in risk, although risk profiles differ for CVD and diabetes. The insomnia–short sleeper phenotype, in particular, appears to confer greater risk than insomnia alone. Although data provide several promising hints about possible mechanisms through which insomnia may influence cardiometabolic risk, the cross-sectional nature of this literature limits testing and refinement of mechanistic models. One of the promising aspects of the work on insomnia and cardiometabolic risk is that these data may apply to other disorders and disease outcomes that have been linked to similar mechanisms (e.g., cancer and arthritis with inflammation and hormone dysregulation; cognitive decline and impairment with neuroinflammation and autonomic imbalance). Finally, the advent of ambulatory technologies and Web-based intervention and assessment platforms will extend this work into underserved populations in whom health disparities are most prevalent and where scalable sleep health interventions may do the most good.

Selected Readings

Burgos I, Richter L, Klein T, et al. Increased nocturnal interleukin-6 excretion in patients with primary insomnia: a pilot study. *Brain Behav Immun* 2006;**20**:246–53.

Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* 2010;**33**:414–20.

Chandola T, Ferrie JE, Perski A, et al. The effect of short sleep duration on coronary heart disease risk is greatest among those with sleep disturbance: a prospective study from the Whitehall II cohort. *Sleep* 2010;**33**:739–44.

Jackson CL, Redline S, Emmons KM. Sleep as a potential fundamental contributor to disparities in cardiovascular health. *Annu Rev Public Health* 2015;**36**:417–40.

Li Y, Vgontzas AN, Fernandez-Mendoza J, et al. Insomnia with physiological hyperarousal is associated with hypertension. *Hypertension* 2015;**65**(3):644–50.

Meng L, Zheng Y, Hui R. The relationship of sleep duration and insomnia to risk of hypertension incidence: a meta-analysis of prospective cohort studies. *Hypertens Res* 2013;**36**:985–95.

Sofi F, Cesari F, Casini A, et al. Insomnia and risk of cardiovascular disease: a meta-analysis. *Eur J Prev Cardiol* 2014;**21**:57–64.

Vasisht KP, Kessler LE, Booth JN III, et al. Differences in insulin secretion and sensitivity in short-sleep insomnia. *Sleep* 2013;**36**:955–7.

Vgontzas AN, Fernandez-Mendoza J, Liao D, Bixler EO. Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Med Rev* 2013;**17**:241–54.

A complete reference list can be found online at ExpertConsult.com.

Chapter
85

Cognitive Behavior Therapies for Insomnia I: Approaches and Efficacy

Charles M. Morin; Judith R. Davidson; Simon Beaulieu-Bonneau

Chapter Highlights

- Cognitive behavior therapies (CBTs) target psychological, behavioral, and cognitive factors that perpetuate or exacerbate sleep difficulties. The main therapeutic components of CBT include sleep restriction, stimulus control, relaxation training, cognitive therapy, sleep hygiene, or a combination of these methods.
- Evidence from randomized controlled trials, meta-analyses, and systematic reviews indicate that CBT is an effective therapeutic approach to improve sleep continuity parameters (e.g., sleep-onset latency, time awake after sleep onset) as well as improve global measures of insomnia severity and sleep quality. These clinical benefits are well sustained over time.
- There is also increasing evidence showing that CBT is effective for insomnia disorder presenting with comorbid psychiatric (e.g., depression) or medical (e.g., cancer, chronic pain) conditions, in older adults, and in patients who are chronic users of hypnotics.
- Despite its well-documented short- and long-term benefits, as well as broad acceptability by patients, CBT remains underused in the management of insomnia. An important challenge for the future will be to develop more efficient strategies and incentives for health practitioners to integrate CBT in their day-to-day clinical practices.

Insomnia can be triggered by a variety of precipitating events and when it becomes a persistent problem, psychological, behavioral, and cognitive factors are instrumental in perpetuating or exacerbating sleep disturbances over time. The effective management of persistent insomnia must target these factors, which involve sleep-scheduling issues, poor sleep habits, conditioning, hyperarousal, excessive worrying, and erroneous beliefs about sleep. Cognitive behavior therapies (CBTs) for persistent insomnia include sleep restriction, stimulus control, relaxation training, cognitive therapy, mindfulness-based interventions, sleep hygiene education, or any combination of those methods. Evidence from controlled clinical trials and systematic reviews indicates that between 70% and 80% of patients with persistent insomnia benefit from CBT, and approximately half of them achieve clinical remission. Treatment produces significant improvements of specific sleep parameters such as sleep-onset latency, wake after sleep onset, sleep efficiency, sleep quality, and reduction of global insomnia severity. These benefits are paralleled by reductions of daytime fatigue, psychological symptoms, and improvements in quality of life. Changes in sleep patterns are well sustained after therapy has ended. Treatment outcomes have been documented primarily with sleep diaries and other self-report measures; studies using polysomnography and actigraphy have also shown sleep improvements but of smaller magnitude. CBT should be the first-line therapy for insomnia disorder. Specific indications include insomnia with or without comorbid medical or psychiatric disorders, insomnia in older adults, and insomnia associated with chronic hypnotic usage. Despite strong evidence supporting its efficacy and effectiveness, CBT remains underused by health care practitioners. An important challenge for the future will be to disseminate more effectively these evidence-based therapies and increase their use in routine clinical practice.

CURRENT TREATMENT PRACTICES

Insomnia is a common condition and carries a significant psychosocial, medical, and economic burden.[1,2] Despite its high prevalence and negative effects, insomnia often remains unrecognized and untreated. Most patients who initiate treatment do so without professional consultation and often resort to alternative remedies (herbal or dietary supplements) of unknown risks and benefits.[3] When insomnia is brought to professional attention, typically to a primary care physician, treatment is often limited to medication. Although hypnotic medications are clinically indicated and useful in selected situations, cognitive and behavior factors are almost always involved in perpetuating sleep disturbances,[4-7] and these factors must be addressed for effective management of chronic insomnia.

With solid evidence supporting its short- and long-term efficacy and clinical utility, as well as its acceptability by patients, CBT is increasingly recognized as the treatment of choice for persistent insomnia.[8,9] This chapter describes clinical procedures that are typically included in CBT for

Box 85-1 COGNITIVE BEHAVIOR THERAPIES FOR PERSISTENT INSOMNIA

Sleep Restriction

A method designed to restrict time spent in bed (i.e., the sleep window) as close as possible to the actual sleep time, thereby strengthening the homeostatic sleep drive. This sleep window is then gradually increased over a period of a few days or weeks until optimal sleep duration is achieved.

Stimulus Control

A set of instructions designed to reinforce the association between the bed and bedroom with sleep and to reestablish a consistent sleep-wake schedule:

- Go to bed only when sleepy
- Get out of bed when unable to sleep
- Use the bed/bedroom for sleep only (no reading, watching TV, etc.)
- Arise at the same time every morning
- No napping

Relaxation Training

Clinical procedures (e.g., progressive muscle relaxation) aimed at reducing autonomic arousal, muscle tension, and intrusive thoughts interfering with sleep. Most relaxation procedures require some professional guidance initially and daily practice over a period of a few weeks.

Cognitive Therapy

Psychological approach using Socratic questioning and behavioral experiments to reduce excessive worrying about sleep and reframe unhelpful beliefs about insomnia and its daytime consequences. Usually requires a trained and skilled clinician. Additional cognitive strategies may involve paradoxical intention technique to alleviate performance anxiety associated with the attempt to fall asleep.

Mindfulness-Based Interventions

The core principle of mindfulness-based interventions is nonjudgmental awareness in the present moment. It involves meditation practice and its most common variant is mindfulness-based stress reduction (MBSR).

Sleep Hygiene Education

General guidelines about health practices (e.g., diet, exercise, substance use) and environmental factors (e.g., light, noise, excessive temperature) that may promote or interfere with sleep. This may also include some basic information about normal sleep and changes in sleep patterns with aging.

Cognitive Behavior Therapy

A multimodal intervention combining some of the above cognitive and behavior (e.g., stimulus control, sleep restriction, relaxation) procedures. Multicomponent behavior therapy would include more than one behavioral procedure but without a cognitive component.

insomnia and summarizes the evidence regarding their efficacy and generalizability. Clinical issues related to feasibility and treatment implementation are addressed only briefly because these issues are the focus of another chapter.

TREATMENTS

Treatment options for insomnia can be grouped in three broad classes of interventions including CBTs, pharmacotherapy, and complementary and alternative therapies. This chapter is about CBTs that have been validated in controlled clinical trials for persistent insomnia. These methods include sleep restriction, stimulus control therapy, relaxation, cognitive strategies, mindfulness-based stress reduction, sleep hygiene education, or a combination of these. CBT is a generic label used to describe a combination of therapeutic methods that include at least one cognitive and one behavioral component. The label *multicomponent behavior therapies* is also used to describe a similar approach involving different components but without cognitive therapy. Sleep hygiene education (i.e., recommendations about health practices and environmental factors that may positively or negatively impact sleep) by itself, although useful, must be distinguished from more formal CBT because it is often insufficient to treat chronic insomnia.[10] A summary of these interventions is provided in Box 85-1; more extensive descriptions are available in other sources.[6,11,12]

Rationale and Indications

The main targets of CBT include psychological, behavioral, and cognitive factors that perpetuate or exacerbate sleep disturbances. Such features may involve counterproductive sleep scheduling, sleep habits (e.g., spending excessive amounts of time in bed), conditioning, hyperarousal, dysfunctional beliefs and excessive worry about sleep, and inadequate sleep hygiene practices.[4,6,13-16] Although numerous factors (e.g., life events, medical illness) can precipitate insomnia, when it becomes a persistent problem, these factors are almost always involved in perpetuating insomnia over time; hence the need to target them directly in treatment.

The main indication for CBT is persistent insomnia, whether it is occurring on its own or in the context of another medical or psychiatric disorder. It is indicated for insomnia in younger and older adults and for patients with prolonged use of hypnotics, although chronic use of hypnotics can interfere with CBT's objectives. There is no evidence for using CBT with acute insomnia and very little evidence for its use in children and adolescents.

There are few contraindications. Sleep restriction should be avoided in patients with a history of seizures, certain parasomnias (e.g., sleepwalking), or manic or hypomanic episodes (as in bipolar disorder) because it may lower the threshold for, or exacerbate, these conditions. Sleep restriction should also be avoided or used very cautiously with patients who need to drive or operate heavy equipment and with those who are already excessively sleepy during the day. Some stimulus control procedures (e.g., getting out of bed when unable to sleep) should be used with caution with frail elderly people, who may be at risk for falls when getting out of bed.[17]

Sleep Restriction

There is a natural tendency among individuals with insomnia to increase the time they spend in bed simply to rest or provide

more opportunity for sleep. Although this strategy may be effective in the short term, in the long run it is more likely to result in fragmented and poor-quality sleep.

Sleep restriction therapy consists of curtailing the amount of time spent in bed as close as possible to the actual amount of time asleep.[18] For example, if a person reports sleeping an average of 6 hours per night out of 8 hours spent in bed, the initial prescribed sleep window (i.e., from initial bedtime to final arising time) would be 6 hours. Time in bed is subsequently adjusted on the basis of sleep efficiency (total sleep time over time in bed × 100%) for a given period, usually 1 week. The sleep window is increased by about 15 to 20 minutes for a given week when sleep efficiency exceeds 85%, it is decreased by the same amount when sleep efficiency is lower than 80%, and it is kept stable when sleep efficiency falls between 80% and 85%. Adjustments are made weekly until optimal sleep duration is achieved. Changes to the prescribed sleep window can be made at the beginning of the night (i.e., postponing bedtime), at the end of the sleep period (i.e., advancing arising time), or at both ends. Some variations in implementation might involve changing the time in bed on the basis of a moving average of the sleep efficiency (e.g., the past 3 to 5 days) or changing it on a weekly basis regardless of changes in sleep efficiency.[19]

Sleep restriction improves sleep through two complementary mechanisms: It strengthens the homeostatic sleep drive through a mild sleep deprivation, and it alleviates some of the sleep anticipatory anxiety by changing the patient's focus of attention (i.e., trying to stay up later rather than going to bed early). Side effects of this intervention include daytime sleepiness, reduced vigilance, and slowed reaction time.[20] To prevent excessive daytime sleepiness, the prescribed time in bed should not be less than 5 hours per night, regardless of reported initial sleep duration. This caution is particularly important for those whose jobs require operating motor vehicles or who have duties in which drowsiness may be a danger to the patient or others.

Stimulus Control Therapy

Stimulus control therapy[21] involves five instructions designed to reassociate temporal (bedtime) and environmental (bed and bedroom) stimuli with rapid sleep onset and to establish a regular circadian sleep-wake rhythm:

- Go to bed only when sleepy—not just fatigued, but sleepy.
- Get out of bed when unable to sleep (e.g., after 20 minutes), go to another room, and return to bed only when sleep is imminent.
- Curtail all sleep-incompatible activities: no eating, watching television, using electronic devices, or planning or problem solving in bed.
- Arise at a regular time every morning regardless of the amount of sleep achieved.
- Avoid daytime napping.

Persons with insomnia often develop apprehension around bedtime and the bedroom and come to associate this particular time of the day and environment with the frustration of being unable to sleep. Over time, the presleep rituals usually associated with relaxation and sleep become cues or stimuli for worrying and wakefulness. In addition, many insomnia patients display counterproductive sleep habits that emerge as a means of coping with sleep disturbances. For example, poor sleep at night can lead to daytime napping or sleeping late into the morning on weekends in an effort to catch up on lost sleep. Individuals might lie in bed for prolonged periods trying to force sleep, only to find themselves becoming more awake. Stimulus control procedures are designed to re-create a positive association between presleep rituals and the bedroom environment.

Stimulus control instructions appear quite simple; however, the challenge for clinicians is to foster strict adherence to these instructions. Several consultation visits held on a weekly or biweekly basis are often necessary to assist patients in implementing these behavioral changes. Stimulus control instructions can be combined with sleep restriction, for example, by having patients stay up until they are sleepy and it is at least their prescribed bedtime ("threshold time for bed").[6] Also, the stimulus control instruction to avoid napping can be modified to allow a short midafternoon nap (e.g., less than 60 minutes starting no later than 3 PM) in the early phase of sleep restriction, particularly in older adults or those who are overly sleepy with the nighttime sleep restriction.

Relaxation-Based Interventions

Because stress, tension, and anxiety are often contributing factors to sleep disturbances, relaxation is probably the most commonly used intervention for insomnia. The goal of this treatment is to reduce arousal at bedtime or on nighttime awakening. Among the different relaxation interventions, some methods (e.g., progressive muscle relaxation) focus primarily on reducing somatic arousal, whereas attention-focusing procedures (e.g., imagery training) target mental arousal in the form of worries, intrusive thoughts, or a racing mind. Biofeedback is designed to train patients to control some physiologic parameters (e.g., tension) through visual or auditory feedback; despite its popularity in the 1980s, this method is not commonly used today.

Selection of a particular method should be based on the patient's preference or the predominant subtype of arousal (somatic versus mental) interfering with sleep. There is no formal contraindication to using relaxation, but some patients might have a paradoxical response and become more anxious when trying to relax. The most critical issue is to encourage daily practice of the selected method for at least 2 to 4 weeks. The focus should be on reducing arousal rather than on inducing sleep, and the relaxation method should be done out of bed, especially if the patient is also using stimulus control therapy. Professional guidance is helpful during initial training, especially to demonstrate the relaxation technique during an in-office session and to support the patient with daily practice.

Cognitive Therapy

Cognitive therapy for insomnia seeks to alter sleep-disruptive cognitions (e.g., beliefs, expectations) and maladaptive cognitive processes (e.g., excessive self-monitoring, worrying) through Socratic questioning and behavioral experiments.[7,22-24] The basic premise of this approach is that appraisal of a given situation (sleeplessness) can trigger negative thoughts and emotions (fear, anxiety) that are incompatible with sleep. For example, worries about the consequences of sleep loss on the next day's performance can lead to further arousal and more sleep disturbance, thus forming a vicious cycle. Likewise, on night waking, a person may engage in self-monitoring (e.g., clock watching to check how many hours are left in the night)

and safety activities (e.g., trying to stop thinking), which can prolong the nocturnal awakenings.[7]

Cognitive therapy is designed to short-circuit the self-fulfilling nature of this vicious cycle through verbal interventions and behavioral homework. Some therapeutic targets for cognitive restructuring include unrealistic expectations ("I must get 8 hours of sleep every night"), faulty causal attributions ("My insomnia is entirely caused by a biochemical imbalance"), and amplification of the consequences of insomnia ("After a poor night's sleep, I can't function the next day"). Key messages to communicate to patients in the context of cognitive therapy include the following:

- Keep realistic expectations with regard to sleep requirements and daytime energy.
- Do not blame insomnia for all daytime impairments because there may be other explanations (worries about family, conflicts with coworkers) for these deficits.
- Never *try* to sleep because it is likely to exacerbate sleep difficulties.
- Do not give too much importance to sleep. Although sleep should be a priority, it should not become the central point of life.
- Do not catastrophize after a poor night's sleep. Insomnia is very unpleasant, but it is not necessarily dangerous to health, at least not in the short term.
- Develop some tolerance to the effects of insomnia. If you are predisposed to insomnia, it is likely that you will remain vulnerable to sleep disturbances even after treatment, and you can develop strategies to cope with these occasional nights of poor sleep.

In addition to these verbal interventions, behavioral experiments can be helpful to change a person's beliefs about sleep and insomnia. For example, if a patient is convinced that bed rest is a good strategy to conserve energy, a behavioral experiment is designed to test the validity of this belief: The patient is instructed to engage specifically in this strategy (bed rest) on a day following insomnia and, on another day, to engage in the opposite behavior, such as performing a series of activities designed to generate energy after a poor night's sleep (e.g., exercising, meeting with friends). Such homework can be quite effective to change a person's belief about ways to preserve or generate energy in the context of insomnia.[24]

Additional cognitive strategies may be useful. For instance, paradoxical intention is a procedure designed to eliminate performance anxiety. In the context of insomnia, any attempt to control or induce sleep voluntarily is likely to generate performance anxiety and to delay sleep onset. With paradoxical intention, the patient is instructed to remain passively awake and to give up any attempt (intention) to fall asleep. To minimize worrying and mental activity interfering with sleep, it is also helpful to instruct patients to set aside a time and a place (other than bedtime and the bedroom) to write down thoughts or worries and plans for the next day. Imagery can also be useful to block out unwanted presleep thoughts.[25]

Mindfulness-Based Interventions

Developed through meditation, mindfulness is based on awareness in the present moment, without judgment. One of the most common mindfulness programs used in health care today is mindfulness-based stress reduction (MBSR). It includes eight weekly group sessions and a 1-day silent retreat. Participants learn various meditation practices, including breathing meditation, the body scan, sitting meditation, Hatha yoga, and walking meditation. Daily personal meditation is expected. The program teaches principles of nonjudging, beginner's mind, trust, nonstriving, acceptance, and letting go. MBSR has been used with various health conditions to reduce symptoms, enhance calmness, and improve quality of life.[26] As used in the context of insomnia, the goal is to reduce stress and psychophysiologic arousal.

Mindfulness-based therapy for insomnia is a group program that integrates behavioral techniques for insomnia, namely sleep restriction, stimulus control therapy, and sleep hygiene, into MBSR.[27,28] The mindfulness principles are believed to complement the conventional behavioral sleep techniques and to replace the need for cognitive therapy. For example, nonstriving is consistent with relinquishing the counterproductive desire to actively control sleep. Also, mindfulness encourages a nonjudgmental observation of one's thoughts as mental events, and this is believed to improve clarity of thought and calmness, thereby reducing arousal.[29]

Several other complementary and alternative therapies, within the nondrug domain, have been used in the treatment of chronic insomnia. These include acupuncture, tai chi, hypnosis, exercise, and electrosleep therapy. Although potentially useful in clinical practice, these methods have not been evaluated extensively in controlled studies.

Sleep Hygiene Education

Sleep hygiene education is intended to provide information about lifestyle (diet, exercise, substance use) and environmental factors (light, noise, temperature) that might either interfere with or promote better sleep.[30] It could also include general sleep-facilitating recommendations, such as allowing enough time to relax before bedtime, avoiding clock watching, and maintaining a regular sleep schedule. Some of these instructions overlap with other behavioral procedures. Sleep hygiene guidelines include the following:

- Avoid stimulants (e.g., caffeine, nicotine) for several hours before bedtime.
- Avoid alcohol around bedtime because it fragments sleep.
- Exercise regularly (especially in late afternoon or early evening).
- Allow at least a 1-hour period to unwind before bedtime.
- Keep the bedroom environment quiet, dark, and comfortable.
- Maintain a regular sleep schedule.
- Remove electronic devices from the bedroom.

Although sleep-hygiene education may be helpful for mild insomnia, it is rarely sufficient for more severe insomnia, which requires more directive and potent behavioral interventions.[31] A didactic approach can also be used to provide basic information about normal sleep, individual differences in sleep need, and changes in sleep physiology with aging. This information is very useful to help some patients distinguish clinical insomnia from normal (age-related) sleep disturbances. Such knowledge can prevent excessive worry and concern, which can also lead to clinical insomnia.

Cognitive Behavior Therapy and Multicomponent Behavior Therapy

Single interventions described to this point can be combined effectively. Multicomponent therapy is becoming the preferred approach to treating insomnia. In a systematic review

of the literature, 26 of 37 clinical studies conducted between 1999 and 2004 had evaluated a multicomponent approach to persistent insomnia.[31] This approach typically included a behavioral (stimulus control, sleep restriction, and, sometimes, relaxation), a cognitive (cognitive restructuring), and an educational component (sleep hygiene); hence the term CBT. This multimodal approach is appealing because it addresses different insomnia features with different therapeutic recommendations, which is consistent with a multidimensional etiologic model of insomnia (see Chapter 82).[4-7,14]

TREATMENT OUTCOME EVIDENCE

Evidence for Efficacy

Several meta-analyses[32-36] and systematic reviews[31,37,38] have summarized the findings from clinical trials evaluating the efficacy of CBT for insomnia. Evidence from these sources shows that treatment produces reliable changes in several sleep parameters (Figure 85-1), including sleep latency (effect sizes, 0.41 to 1.05), wake after sleep onset (0.61 to 1.03), number of awakenings (0.25 to 0.83), total sleep time (0.15 to 0.49), and sleep quality ratings (0.94 to 1.14). These effect sizes are considered large (Cohen's effect size $d > 0.8$) for sleep latency, wake after sleep onset, and sleep quality and moderate ($d > 0.5$) for other sleep parameters. When transformed into a percentile rank metric, these data indicate that approximately 70% to 80% of patients with insomnia benefit from treatment. These effect sizes are similar to those obtained for benzodiazepine-receptor agonists,[34,39] with a slight advantage for CBT on measures of sleep-onset latency and sleep quality and for pharmacotherapy on total sleep time. A more recent meta-analysis of 73 randomized controlled trials found that CBT produced a large effect on the Insomnia Severity Index (ISI) and moderate effects on most sleep diary parameters, apart from total sleep time, which showed a small effect.[36]

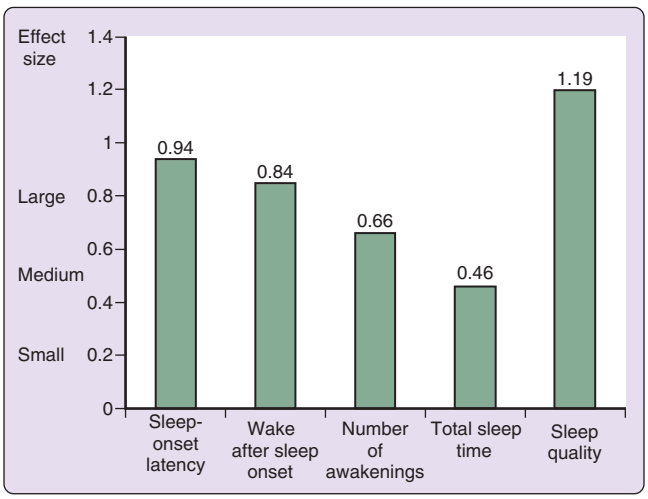


Figure 85-1 Mean effect size (reported as a standardized *z*-score) across sleep parameters. These effect sizes are pooled across three meta-analyses. (Data from Morin CM, Culbert JP, Schwartz SM. Nonpharmacological interventions for insomnia: a meta-analysis of treatment efficacy. *Am J Psychiatry* 1994;151:1172–80; Murtagh DR, Greenwood KM. Identifying effective psychological treatments for insomnia: a meta-analysis. *J Consult Clin Psychol* 1995;63:79–89; and Smith MT, Perlis ML, Park A, et al. Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. *Am J Psychiatry* 2002;159:5–11.)

In terms of absolute changes on sleep parameters, treatment reduces subjective (sleep diary) sleep latency and time awake after sleep onset from an average of 60 to 70 minutes at baseline to about 35 minutes after treatment. Total sleep time is increased by an average of 30 minutes, from 6 to 6.5 hours after treatment, with additional gains made after treatment is completed. Thus, for the average patient with persistent insomnia, often exceeding 10 years in duration, CBT may be expected to reduce sleep latency and wake after sleep onset by an average of about 50% and to bring the absolute values of those sleep parameters below or near the 30-minute cutoff criterion initially used to define sleep-onset or sleep-maintenance insomnia. Treatment effects are generally equivalent for sleep-onset and sleep-maintenance problems, although very few studies have targeted specifically the early-morning awakening problems. Overall, findings from meta-analyses represent fairly conservative estimates of treatment effects because they are based on group averages pooled across all CBT interventions, insomnia diagnoses (i.e., primary and comorbid), and age groups. Nonetheless, although most patients benefit from treatment, only about half of them achieve full remission, and a significant number continue to experience residual sleep disturbances after treatment.[31]

Treatment outcome has been documented primarily with prospective daily sleep diaries, although several studies have also complemented those findings with polysomnography (PSG)[40-42] and actigraphy.[43-45] In general, the magnitude of improvements is smaller on PSG measures, but those changes tend to parallel sleep improvements reported in daily sleep diaries. For example, in a study of older adults with sleep-maintenance insomnia,[40] average baseline values for wake after sleep onset were 62 minutes for diaries and 73 minutes for PSG measures. Posttreatment values were 29 minutes for the diary and 35 minutes for PSG, yielding improvement rates of 54% and 51%, respectively, for the two assessment methods. In another study,[44] sleep efficiency increases of 8% and 12% were obtained for PSG and diary measures, respectively, after CBT. Collectively, these findings indicate that CBT not only alters sleep perception on daily diaries but also produces objective changes on EEG-defined sleep-continuity measures. Except for a modest increase in stages 3 and 4 after sleep restriction, there are few changes in sleep stages with CBT.

Clinical Significance

In addition to showing robust changes on sleep-wake parameters, some studies have also shown that CBT produced improvements on secondary end points, including global measures of insomnia severity, sleep quality, fatigue, quality of life, sleep-related beliefs, and psychological symptoms.[46-53] Despite frequent reports of cognitive impairments among insomnia patients, there is no evidence that treatment improves performance-based measures of attention, concentration, and memory. It is unclear whether these negative findings are due to the absence of objective deficits at baseline or to the selection of measures that are not sensitive enough to detect such deficits. Nonetheless, it remains essential to document outcome beyond reduction of insomnia symptoms because it is often the perceived consequences of insomnia, rather than insomnia per se, that prompt treatment seeking. Investigators should incorporate multiple outcome measures, including sleep, functional impairments, psychological symptoms, and quality of life.[54,55]

Generalizability of Treatment Effects to Comorbid Insomnia

Insomnia is often a pervasive problem among patients with medical or psychiatric conditions.[8,56,57] Although it is generally assumed that the sleep disturbance will remit with appropriate treatment of the comorbid condition, often it does not. Insomnia that persists after treatment of the comorbid condition, for instance depression, may actually increase the risk for future relapse of that condition. Thus it is often essential to treat both the sleep and the coexisting disorder. An important study by Manber and colleagues[58] showed that CBT had an augmentation effect when used in combination with antidepressant medication in the treatment of insomnia comorbid with major depression. The addition of CBT for insomnia, relative to antidepressant therapy alone, resulted in higher remission rates of depression (61.5% vs. 33.3%) and insomnia (50% vs. 7.7%) and larger improvements in measures of sleep continuity.

Several additional studies have shown that patients with medical and psychiatric conditions can also benefit from insomnia-specific treatment, even though the outcome with those patients is sometimes more modest compared with insomnia patients without comorbid disorders.[59-66] A systematic review of 16 randomized controlled trials of CBT for insomnia in psychiatric populations found that CBT was effective to improve sleep in patients with comorbid depression, anxiety, posttraumatic stress disorder, and substance abuse disorders. The comorbid disorder also improved after CBT for insomnia in some instances.[67,68] Findings from another review also supports the use of CBT for insomnia in individuals with alcohol-related disorders, although the evidence is more limited for this condition.[69]

Controlled studies have also shown that CBT is effective for treating insomnia comorbid with medical disorders, including chronic pain,[70] fibromyalgia,[71] cancer,[52,72] and various medical conditions in older adults.[35,65,73-75] CBT was generally superior to MBSR for cancer patients with insomnia.[76] Single-case experimental studies also support the use of CBT for insomnia associated with traumatic brain injury.[61] In general, the findings of comorbid insomnia studies indicate that baseline and posttreatment scores on insomnia measures are usually more severe among patients with comorbid disorders relative to those without comorbid conditions, but they can still benefit from treatment. To optimize outcome, it is often necessary to adapt these interventions to disease-specific conditions[56]; for example, the addition of a fatigue management module to standard CBT has proved helpful for insomnia comorbid with cancer[52] and traumatic brain injury.[61]

Treatment of Insomnia in Older Adults

Insomnia is age-related, and its treatment in older adults has received increased attention from clinical investigators (see Chapter 153). Two evidence-based review papers concluded that CBT is effective for managing insomnia in older adults. A meta-analysis[35] showed robust treatment effects in sleep continuity (mean effect sizes, 0.52 for sleep latency and 0.64 for time awake after sleep onset) and sleep quality (0.76) for older adults, with similar effects for CBT, relaxation, and behavioral approaches. Another systematic review reported that multimodal CBT and sleep restriction or sleep compression met criteria for empirically validated therapies, but there was not adequate evidence to support cognitive therapy, relaxation, and sleep hygiene education as stand-alone interventions for insomnia in older adults.[75] A dismantling study comparing stimulus control, sleep restriction, and a multicomponent intervention in older adults found that all three conditions produced significant and sustainable sleep improvements, whereas the multicomponent intervention yielded the highest insomnia remission rate.[77]

Insomnia in elderly people is more likely to be comorbid with another medical condition or even another sleep disorder than to occur on its own. The presence of a comorbid medical disorder, however, is not a contraindication to using CBT for insomnia, unless it is an unstable or advanced illness that precludes the patient from being able to use the techniques. There is increasing evidence that CBT is effective for insomnia in the context of other medical and, to a lesser extent, psychiatric conditions.[65,74,78,79] One study[74] found that older adults with chronic obstructive pulmonary disease, osteoarthritis, or coronary artery disease responded equally well to group CBT relative to control patients. In addition to improving sleep continuity, treatment of insomnia in elderly people is usually associated with enhanced sleep satisfaction and quality of life.

When patients are considered for CBT, they might already have been on hypnotic medication for a long period, which can represent a challenge for clinicians. This clinical situation is more common in older adults, but it can occur with all age groups. Several studies have shown that a supervised, structured, and time-limited withdrawal program, with or without CBT, can facilitate discontinuation of hypnotic medications among patients with prolonged use.[80-84] In a study of 76 older adults who had used benzodiazepines for insomnia for nearly 20 years, treatment was effective in reducing both the quantity (90% reduction) and the frequency (80% reduction) of medication use, and 63% of the patients were drug free within the 10-week intervention.[80] The addition of CBT to the withdrawal program minimized rebound insomnia and reduced worries about drug discontinuation.

Initial Treatment Response versus Long-Term Outcome

Because insomnia is often a recurrent or persistent problem and CBT is a brief intervention, it is important to evaluate outcome beyond initial treatment. A robust finding across clinical trials is that sleep improvements achieved with CBT are well sustained after treatment is completed.[31,33,37] Although sleep restriction may yield only modest increases (and even a reduction) of sleep time during initial therapy, there is generally an increase at follow-up. Long-term outcome must be interpreted cautiously, however, because few studies report follow-up data later than 6 months after treatment and, among those that do, attrition rates increase over time. In addition, it is important to keep in mind that patients with chronic insomnia, even those who benefit from short-term therapy, might remain vulnerable to recurrent episodes of insomnia in the long term.[85,86] The addition of maintenance therapy in the form of booster therapy sessions might facilitate better integration of newly learned self-management skills and contribute to improved long-term outcome.[87]

Comparative Efficacy of Single Therapies

Although multimodal CBT is becoming the standard approach in the field, there is little information on the relative

efficacy and unique contribution of each therapeutic component. Nonetheless, studies comparing the relative efficacy of single therapies have found that stimulus control and sleep restriction therapies are more effective than relaxation alone, which, in turn, is more effective than sleep hygiene education.[31,33,37] Sleep restriction tends to produce better outcomes than stimulus control in terms of sleep efficiency and continuity, but it also decreases sleep time during initial treatment. Relaxation methods focusing on cognitive arousal (e.g., reducing intrusive thoughts) yield slightly greater improvements than those targeting somatic arousal. Sleep hygiene education, when used alone, produces little effect on sleep and insomnia symptoms; this didactic approach should be seen as a minimal intervention. Cognitive therapy, although increasingly used as one therapeutic component of multimodal interventions,[31] has been evaluated as a single treatment in two studies.[24,88] In a dismantling study investigating the relative efficacy and unique contribution of behavior therapy (sleep restriction and stimulus control), cognitive therapy, and combined CBT for insomnia, all three therapies produced significant reductions of insomnia severity symptoms.[88] CBT produced the best outcomes overall, and although behavior therapy produced immediate benefits after treatment, these were not as well sustained 6 months after treatment, although the benefits associated with cognitive therapy were slower to accrue but were more sustained 6 months after treatment. Figure 85-2 illustrates some of these results: all three conditions showed significant improvements on the ISI, and these were well maintained at the 6-month follow-up. Depending on the condition, rates of treatment response (i.e., reduction in ISI score of at least 8) were between 42% and 67% after treatment and between 44% and 67% at follow-up, whereas rates of insomnia remission (i.e., ISI score below 8) were between 31% and 57% after treatment and between 37% and 52% at follow-up.

Based on criteria set forth by the American Psychological Association, there are currently five interventions that have sufficient evidence to meet criteria for well-established psychological treatment for insomnia: CBT, sleep restriction, stimulus control, relaxation, and paradoxical intention.[31] When the evidence is evaluated specifically with older adults, CBT, sleep restriction, and stimulus control meet similar efficacy standards[75] (Table 85-1).

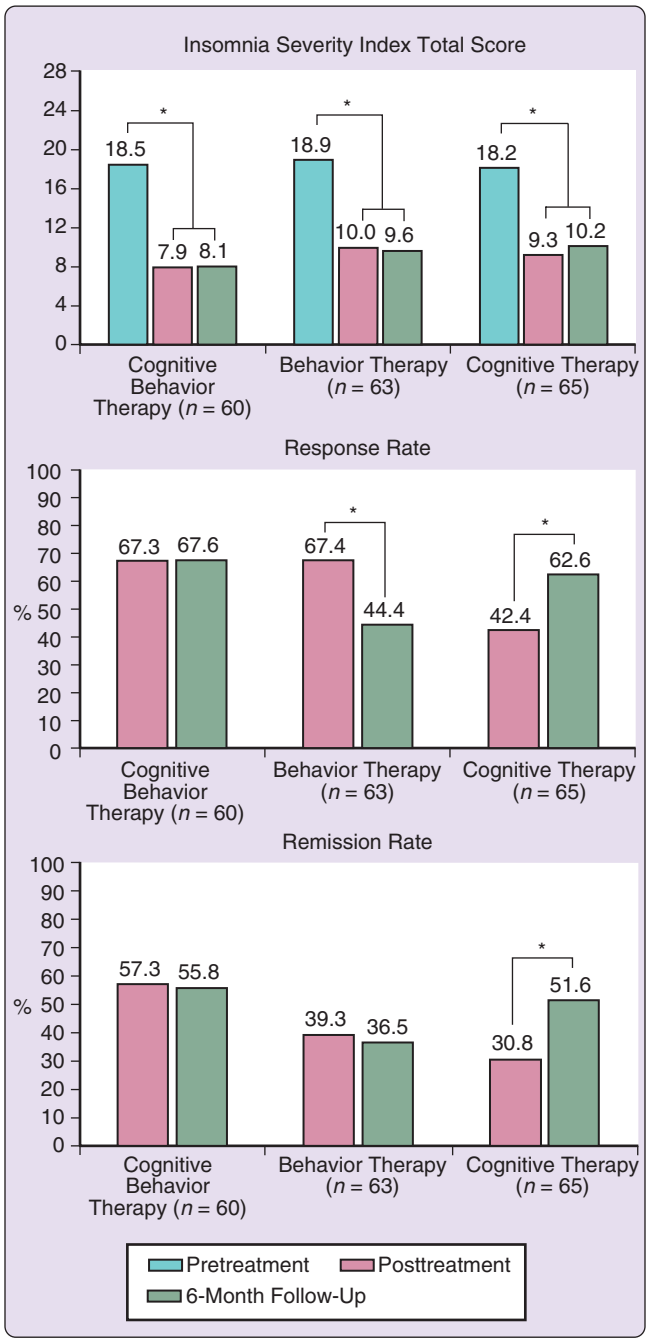


Figure 85-2 Changes in the Insomnia Severity Index (ISI) following cognitive behavior therapy (CBT), behavior therapy (BT), or cognitive therapy (CT). Response was defined as a reduction in total ISI score of at least 8 points from baseline and remission was defined as an ISI score below 8 at posttreatment or 6-month follow-up. Flagged comparisons (*) are significant at $P < .05$. Also significant: greater response rates at posttreatment for CBT and BT compared with CT, greater response rate at follow-up for CBT compared with BT, greater remission rate at posttreatment for CBT compared with CT, and greater remission rate at follow-up for CBT compared with BT. (Data from Harvey AG, Belanger L, Talbot L, et al. Comparative efficacy of behavior therapy, cognitive therapy, and cognitive behavior therapy for chronic insomnia: a randomized controlled trial. *J Consult Clin Psychol* 2014;82:670–83.)

Treatment Specificity and Mechanisms of Changes

Although there is strong evidence supporting the efficacy of CBT, there is little information about the mechanisms responsible for sleep improvements. In a review of 21 randomized controlled trials and 11 secondary analyses, several potential mediators of CBT were examined, including behavioral (time spent in bed, napping), cognitive (sleep effort, self-efficacy), and hyperarousal-related mediators (physiologic or cognitive arousal). It was concluded that although CBT elicits changes in the processes involved in the maintenance of insomnia, it remains unclear whether these changes are responsible for sleep improvements.[89]

With a few notable exceptions, which have used attention-placebo conditions,[19,44,90] most CBT trials have used waitlist control groups, precluding the unequivocal attribution of treatment effects to any specific therapeutic ingredient. The lack of a pill-placebo control equivalent in psychological treatment studies makes it difficult to determine what percentage of the variance in outcomes is the result of specific clinical procedures (e.g., restriction of time in bed, cognitive restructuring), the measurement process (e.g., self-monitoring), or nonspecific factors (e.g., therapist attention, patients' expectations). A recent trial comparing online CBT, attention-placebo

Table 85-1 Evidence Level for Cognitive Behavior Therapies for Persistent Insomnia

Intervention	Level of Evidence*	Key Studies
Sleep restriction	Guideline	Epstein et al, 2012[77] Fernando et al, 2013[114]
Stimulus control	Standard	Epstein et al, 2012[77] Riedel et al, 1998[115]
Relaxation training	Standard	Lichstein et al, 2001[19] Means et al, 2000[116]
Cognitive therapy	Insufficient evidence	Harvey et al, 2007[24] Harvey et al, 2014[88]
Mindfulness-based interventions	Insufficient evidence	Gross et al, 2011[117] Ong et al, 2014[27]
Sleep hygiene education	The evidence indicates that sleep hygiene education alone is not effective	Morin et al, 1999[31]; 2006[37]
Multicomponent cognitive behavior therapy (CBT)	Standard (for CBT with or without relaxation training) Guideline (for CBT without cognitive therapy)	Bothelius et al, 2013[104] Harvey et al, 2014[88] Edinger et al, 2009[66] Morin et al, 1999[40]; 2009[87] Morin et al, 2006[31]

*Level of evidence according to the Practice Parameters for the psychological and behavioral treatment of insomnia,[10] except for mindfulness-based interventions, which were not evaluated in these practice parameters.

(imagery relief therapy), and treatment as usual showed that CBT successfully altered sleep-related attributions, thought content, and depression and anxiety symptoms, and those changes partly mediated sleep improvements.[90]

This issue of mechanism of change has been addressed indirectly by examining predictors of changes and by surveying patients about their perception of the most effective treatment components. In a follow-up study of patients who had completed CBT for insomnia,[91] the most critical ingredients associated with long-term improvements were the use of stimulus control and sleep restriction, followed by cognitive restructuring. Relaxation was the most commonly cited component (79% of respondents), but it did not predict improvement in any of the sleep variables. Other studies have examined the association between changes in sleep beliefs and attitudes and sleep improvements. Reductions of dysfunctional sleep cognitions during treatment were correlated with sleep improvements after treatment, and fewer dysfunctional cognitions after treatment were associated with better maintenance of sleep changes over time.[92] In contrast to these findings, one study found that reductions of dysfunctional sleep beliefs following CBT did not mediate improvements of insomnia symptoms.[93] Thus it is likely that factors yet to be identified are critical in mediating treatment response. With increasing evidence that central nervous system hyperarousal is implicated in insomnia,[94,95] more attention is needed to identify the biologic, as well as the psychological, mechanisms responsible for sleep changes.

Combined Cognitive Behavior Therapy and Medication

CBT and medication can play complementary roles in the management of insomnia.[96] Medication produces fast symptomatic relief, and CBT yields durable benefits over time. Thus combined approaches should, in principle, optimize outcome by capitalizing on the more immediate and potent effects of medication and the more sustained effects of CBT.

Few studies have directly evaluated the combined and differential effects of behavioral and pharmacologic therapies for insomnia.[30,40,41,50,97-101] Three studies compared triazolam with relaxation or sleep hygiene,[30,97,98] and six other studies compared CBT with temazepam,[40,41] zolpidem,[50,87] or zopiclone.[100,101] Collectively, the evidence from those studies indicates that both treatment modalities, when used singly, are effective in the short term. Medication produces rapid improvements, but these benefits are also quickly lost after the drug is discontinued. Sleep improvements obtained with CBT take longer, but they are also well sustained even after treatment ends. There is a slight additive effect of combined approaches relative to medication alone, but not necessarily relative to CBT alone.

The long-term effect of combined versus single approaches is more equivocal. Some studies[97,98] indicate that a combined intervention (e.g., triazolam plus relaxation) produces more sustained benefits than medication alone, and others[30,40,41] report more variable long-term outcomes. Treatment responses vary across individual patients; some patients retain their initial sleep improvements, but others return to their baseline values. Because of the mediating role of psychological factors in chronic insomnia, behavioral and attitudinal changes seem to be essential to sustaining improvements in sleep patterns. When behavioral and drug therapies are combined, patients' attributions of the initial benefits may be critical in determining long-term outcomes. Attribution of therapeutic benefits to the medication alone, without integration of self-management skills, can place a patient at greater risk for insomnia recurrence after the drug is discontinued. In addition, there is a risk that the availability of medications might undermine the patient's effort and motivation to implement behavioral changes in sleep habits and sleep scheduling. Thus,

despite the intuitive appeal of combining therapies, it is not entirely clear when, how, and for whom combining CBT and medication is indicated. Additional research is needed to examine optimal methods for combining or integrating these therapies.[102]

Rather than initiating and discontinuing both treatments simultaneously, it may be preferable to introduce them sequentially to optimize outcome.[103] One study has examined the effect of different treatment sequences when combining CBT with medication.[100] The optimal sequence appeared to be the introduction of CBT early in treatment, either before or concurrently with medication. In another study, medication had a modest augmentation effect on CBT, primarily in terms of improving total sleep time, and this added benefit may be important given that the sleep-restriction component of CBT reduces total sleep time during the initial course of therapy and could lead some patients to discontinue therapy prematurely.[87] Furthermore, CBT combined with medication produced faster sleep improvements, with optimal improvement reached after only 1 week of treatment compared with 2 to 3 weeks for CBT alone.[96] Overall, the best sequence may be a combined approach initially, followed by CBT alone while discontinuing medication.[87]

Efficacy versus Effectiveness

Because the main entry point for professional insomnia treatment is typically in primary care medicine, it is important to examine whether treatment validated in academic research centers is also effective in various clinical settings. Several pragmatic trials[43,49,65,84,104] in which CBT was implemented in the context of primary care practices or sleep clinics have yielded results comparable to those obtained with patients treated within more controlled research protocols. Brief consultation models involving one or two consultations have also been used successfully in primary care.[105] Finally, the effectiveness of insomnia treatment provided in behavioral sleep medicine clinics has been shown in several clinical replication series.[46,59,60]

Methods of Treatment Delivery

Group therapy can be a cost-effective method to deliver CBT for insomnia in clinical settings.[106] Although it might not provide the same flexibility, group therapy is generally as effective as individual therapy.[47] Self-help approaches using printed materials or DVDs can also be useful in insomnia treatment.[51,107,108] For example, studies[51,81] have shown that a self-help treatment program consisting of six printed booklets mailed weekly to participants was effective to treat insomnia and to reduce use of hypnotics; the addition of professional guidance in the form of brief telephone consultation added to the initial outcomes. There is also increasing evidence of the benefits of Internet-based intervention for treating insomnia.[109,110] Collectively, the evidence indicates that CBT for insomnia is effective not only in the context of controlled research studies but also in various clinical settings such as in primary care and in behavioral sleep medicine clinics. Self-help materials, brief consultations, group therapy, and even the Internet can enhance treatment access and facilitate its implementation in clinical practice. However, despite their increasing popularity,[111] self-help therapies should be seen as a complement rather than a replacement for direct face-to-face therapy.

Clinical and Practical Considerations

CBT approaches to treating primary insomnia are time limited, structured, and sleep focused. For typical patients with chronic insomnia, direct consultation time takes between 4 and 6 hours per patient, which is usually spread over a treatment period of 6 to 8 weeks.[44] The amount of direct clinical contact and the number of follow-up visits are likely to vary as a function of several factors, including insomnia severity, comorbidity, use of hypnotic medications, and the patient's motivation and education. This treatment period is considerably shorter than for other forms of psychotherapy, and it is shorter than behavioral approaches applied to other health and psychological problems (e.g., chronic pain, anxiety, depression).

Despite ample evidence supporting the efficacy of CBT for insomnia, there is a dearth of data on the effect of adherence on outcomes. Although there is a need for adding standardized measures of treatment adherence in clinical studies,[112] the success of CBT depends largely on the patient's compliance with the recommended changes. For this reason, it is particularly important to schedule follow-up visits after the initial evaluation to address adherence. Although treatment is generally well accepted by patients, it is more time-consuming and requires more effort than drug therapy for both clinicians and patients. Thus it is crucial to find ways to optimize adherence by using motivational interviewing strategies, soliciting the bed partner's collaboration, and providing feedback and guidance early in treatment.[113] Given the multifactorial nature of insomnia, it is often necessary to combine several clinical procedures rather than to rely on a single treatment modality. Sleep restriction is almost always relevant for an insomnia patient, and the use of this procedure early in therapy is likely to produce rapid therapeutic benefits. Some stimulus control procedures (e.g., getting out of bed when unable to sleep) might not be necessary early on when the initial sleep window is restricted to 5 to 6 hours per night, but those procedures become particularly relevant as the sleep window is increased. Relaxation is helpful for persons with elevated tension or anxiety; however, caution is needed because some patients have a paradoxical response and become more anxious. Cognitive therapy is particularly helpful to challenge some misconceptions about sleep and daytime impairments, to alleviate emotional distress, and to minimize recurrence of sleep disturbances over time. For milder forms of insomnia, basic sleep hygiene education may be sufficient, whereas more severe insomnia often requires multimodal interventions and more frequent follow-up visits.

CLINICAL PEARL

Sleep scheduling factors, conditioning, hyperarousal, and misconceptions about sleep are important contributing factors to persistent insomnia. Effective clinical management of insomnia must address these perpetuating factors. CBT is an effective and well-accepted therapeutic approach for persistent insomnia, and sleep improvements are well sustained over time.

SUMMARY

Significant advances have been made in the development and validation of CBT for insomnia. With increasing

research-based evidence supporting its efficacy and acceptability, CBT is often recognized as the treatment of choice for persistent insomnia. However, there is still a major gap between research evidence and actual clinical practice in that few clinicians, other than behavioral sleep medicine specialists, use this therapeutic approach in their clinical practice. Important challenges remain for professional sleep organizations and public health decision makers to promote wider dissemination and use of CBT for managing insomnia.

ACKNOWLEDGMENT

Preparation of this chapter was supported in part by grants from the National Institute of Mental Health (MH-091053) and the Canadian Institutes of Health Research (MOP-42504).

Selected Readings

Ashworth DK, Sletten TL, Junge M, et al. A randomized controlled trial of cognitive behavioral therapy for insomnia: an effective treatment for comorbid insomnia and depression. *J Couns Psychol* 2015;**62**(2):115–23.

Conley S, Redeker NS. Cognitive behavioral therapy for insomnia in the context of cardiovascular conditions. *Curr Sleep Med Rep* 2015;**1**(3): 157–65.

Edinger JD, Carney CE. *Overcoming insomnia: a cognitive-behavioral therapy approach, therapist guide*. 2nd ed. New York: Oxford University Press; 2014.

Edinger JD, Wohlgemuth WK, Radtke RA, et al. Cognitive behavioral therapy for treatment of chronic primary insomnia: a randomized controlled trial. *JAMA* 2001;**285**:1856–64.

Espie CA, Inglis SJ, Tessier S, et al. The clinical effectiveness of cognitive behaviour therapy for chronic insomnia: Implementation and evaluation of a sleep clinic in general medical practice. *Behav Res Ther* 2001;**39**: 45–60.

Harvey AG. A cognitive theory and therapy for chronic insomnia. *J Cogn Psychother* 2005;**19**:41–59.

Harvey AG, Belanger L, Talbot L, et al. Comparative efficacy of behavior therapy, cognitive therapy, and cognitive behavior therapy for chronic insomnia: a randomized controlled trial. *J Consult Clin Psychol* 2014; **82**:670–83.

Irwin MR, Cole JC, Nicassio PM. Comparative meta-analysis of behavioral interventions for insomnia and their efficacy in middle-aged adults and in older adults 55+ years of age. *Health Psychol* 2006;**25**: 3–14.

Johnson JA, Rash JA, Campbell TS, et al. A systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy for insomnia (CBT-I) in cancer survivors. *Sleep Med Rev* 2015;**27**:20–8.

Lichstein KL, Nau SD, Wilson NM, et al. Psychological treatment of hypnotic-dependent insomnia in a primarily older adult sample. *Behav Res Ther* 2013;**51**:787–96.

Morgenthaler T, Kramer M, Alessi C, et al. Practice parameters for the psychological and behavioral treatment of insomnia: an update. An American Academy of Sleep Medicine Report. *Sleep* 2006;**29**:1415–19.

Morin CM, Espie CA. *Insomnia: a clinical guide to assessment and treatment*. New York: Kluwer Academic/Plenum; 2003.

Morin CM, Vallières A, Guay B, et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. *JAMA* 2009;**301**:2005–15.

Ong JC, Sholtes D. A mindfulness-based approach to the treatment of insomnia. *J Clin Psychol* 2010;**66**:1175–84.

Smith MT, Perlis ML, Park A, et al. Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. *Am J Psychiatry* 2002;**159**:5–11.

Taylor DJ, Pruiksma KE. Cognitive and behavioural therapy for insomnia (CBT-I) in psychiatric populations: a systematic review. *Int Rev Psychiatry* 2014;**26**:205–13.

A complete reference list can be found online at ExpertConsult.com.

Chapter
86

Psychological and Behavioral Treatments for Insomnia II: Implementation and Specific Populations

Jack D. Edinger; Melanie K. Leggett; Colleen E. Carney; Rachel Manber

Chapter Highlights

- Psychological and behavioral insomnia therapies have a strong legacy of research support showing their effectiveness for insomnia management. A number of monotherapies including stimulus control, relaxation training, and sleep restriction have shown their value for insomnia management, whereas multicomponent cognitive-behavioral insomnia therapy is now considered a first-line therapy for insomnia complaints.
- Implementation of the psychological and behavioral therapies usually entails a number of tools, such as sleep diaries and actigraphy for insomnia assessment and treatment monitoring, as well as handout materials for patients to reinforce treatment instructions received during clinical visits.
- Psychological and behavioral insomnia therapies can be provided in individual, group, and various self-help formats, and they can be delivered effectively by specialist and nonspecialist providers. The reach of these therapies can be extended by use of abridged treatment protocols in primary care settings or through sophisticated Internet protocols designed to mimic in-person treatment methods.
- The psychological and behavioral insomnia therapies have proved to be effective with various age groups as well as with patients with sleep-disruptive psychiatric and medical comorbidities. This chapter reviews important considerations in the application of these therapies with various types of patients.

As evidenced by the preceding chapter, psychological and behavioral insomnia therapies have a strong legacy of research support. Many of the stand-alone (e.g., stimulus control, sleep restriction) and multicomponent (e.g., cognitive-behavioral therapy for insomnia [CBT-I]) treatments are now regarded as well-established therapies that can and should be considered first-line or adjunctive therapies for many treatment-seeking insomnia patients. This chapter provides a description of some tools commonly employed in implementing psychological therapies for insomnia and a discussion of several issues clinicians may wish to consider when implementing these treatments across clinical settings. The chapter first describes some tools that are useful in delivering these treatments and monitoring their outcomes. The chapter then provides a discussion of treatment formats, the types of providers who might deliver these treatments, treatment dosing, and patients' acceptance and adherence to these therapies. Subsequently, the chapter provides a discussion of implementation issues and the efficacy and potential value of psychological therapies in the management of insomnia among patients with comorbid psychiatric and medical conditions. The chapter concludes by considering the applicability of the psychological and behavioral insomnia therapies with various age groups.

BASIC INTERVENTION TOOLS

The psychological and behavioral insomnia therapies include a diverse group of treatments (e.g., relaxation training, stimulus control, cognitive therapy, sleep restriction).[1,2] Research and clinical experience with these treatments have produced a number of helpful tools for identifying treatment targets, monitoring adherence and response, and delivering and reinforcing treatment instructions. In this section, we discuss some of the commonly used clinical tools for the implementation of these therapies.

Sleep Diaries

A sleep diary is an assessment and intervention tool that provides a daily recording of sleep from the patient's perspective. Sleep diaries are a mainstay of insomnia assessment and treatment as they quantify the severity of the sleep disorder, aid in diagnostic discriminations and case conceptualization, guide the implementation of behavioral interventions by tracking progress and adjustments of the behavioral instructions, identify problems with treatment adherence (e.g., varying from a prescribed sleep schedule), and measure treatment outcomes.[3]

Sleep diaries provide a more accurate and detailed depiction of sleep patterns compared with an individual's retrospective report.[4] They are inexpensive and easily administered as a paper questionnaire. There are many published versions available, resulting in variability among clinicians and researchers in type and amount of information obtained, definitions and calculations of sleep parameters, and response formats.[3] In efforts to standardize insomnia research methodologies, a Consensus Sleep Diary was published in 2012.[5] This sleep diary represents the consensus of a group of insomnia experts in collaboration with focus groups of individuals with and without sleep disorders. The Core version elicits information about the respondent's sleep-onset latency, number of awakenings, wake time during the night, total sleep time, sleep efficiency, and sleep quality in addition to a "comments" section (Figure 86-1). Two expanded versions provide queries about early morning awakenings, napping, feelings of restedness, estimated total sleep time, and use of alcohol or caffeine and medications.

Alternative versions have been developed to meet the needs of different clinical populations, such as children and adolescents, and are widely available on the Internet. For individuals who may have difficulty in accurately completing a traditional diary, such as those with highly erratic sleep patterns or shift workers, an analog or raster-style sleep diary captures a richer picture of sleep continuity. In an analog version, such as that published by the American Academy of Sleep Medicine (Figure 86-2; http://yoursleep.aasmnet.org/pdf/sleepdiary.pdf),[6] individuals are instructed to fill in blocks of time slept during 24-hour cycles.

Disadvantages of the traditional paper sleep diary include inaccurate or incomplete entries, inability to monitor or to confirm daily completion, clinician time required for hand scoring, and potential for scoring errors. These drawbacks can be mitigated with technologies such as electronic sleep diaries on hand-held devices,[7,8] computerized scoring programs,[9] and interactive voice recording systems.[10-12] CBT-i Coach,[13] which can be used in conjunction with face-to-face CBT-I, is a mobile phone application available to anyone with a smartphone (http://www.ptsd.va.gov/professional/materials/apps/cbticoach_app_pro.asp). It includes an interactive sleep diary and sleep calculations for time-in-bed prescriptions. Thus, technology is providing several sleep diary alternatives designed to improve on the traditional paper diary.

Actigraphy

Actigraphy is an objective measurement methodology that assesses limb movement activity by a small recording device usually worn on the wrist (see Chapter 171). Typically, 24-hour activity is recorded for periods of 1 to 2 weeks and is accompanied by sleep diaries so that bedtimes, rise times, and times the device was removed can be determined. Recorded data are subjected to a proprietary algorithm that produces estimates of sleep-wake variables. Current standards of practice outline the primary roles of actigraphy in insomnia as the characterization of circadian or sleep disturbance and

	Sample							
Today's date	4/5/08							
1. What time did you get into bed?	10:15 p.m							
2. What time did you try to go to sleep?	11:30 p.m							
3. How long did it take you to fall asleep?	1 hour 15 min.							
4. How many times did you wake up, not counting your final awakening?	3 times							
5. In total, how long did these awakenings last?	1 hour 10 min.							
6. What time was your final awakening?	6:35 a.m							
7. What time did you get out of bed for the day?	7:20 a.m							
8. How would you rate the quality of your sleep?	☐Very poor ☑Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good	☐Very poor ☐Poor ☐Fair ☐Good ☐Very good
9. Comments (if applicable)	I have a cold							

Figure 86-1 Consensus Sleep Diary. Copyright 2012 Consensus Sleep Diary. Permission for use is granted solely for clinical, nonprofit use.[5]

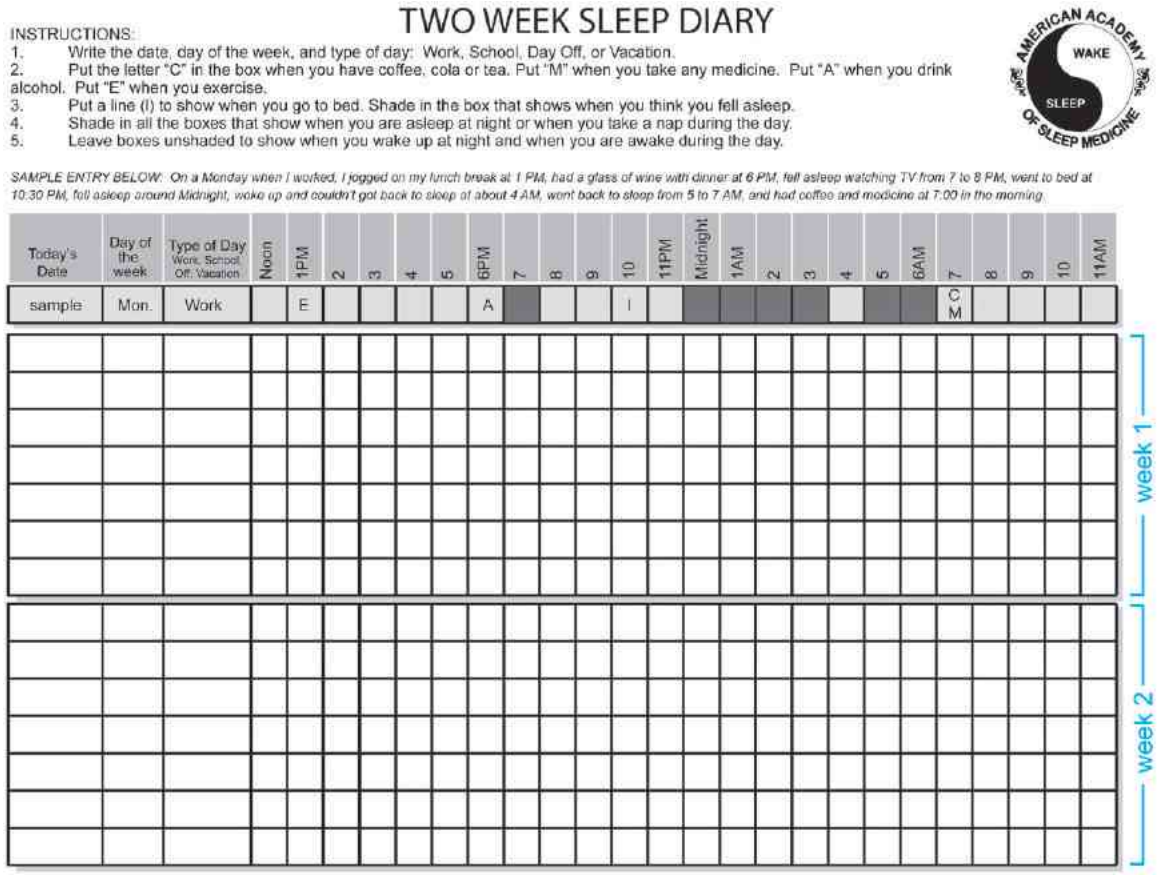


Figure 86-2 Analogue Sleep Diary offered by the American Academy of Sleep Medicine,[6] a raster or analog-style sleep diary made available by the American Academy of Sleep Medicine at http://yoursleep.aasmnet.org/pdf/sleepdiary.pdf.

the evaluation of treatment response.[14] In general, studies have shown that actigraphy can reliably differentiate individuals with and without insomnia.[15-17] However, because actigraphy tends to interpret wakeful stillness in bed as sleep, its validity in insomnia patients is less robust than in sleepers without insomnia.[3,15]

Actigraphy is unobtrusive and simple to use, requiring minimal effort from the patient. It can be used in the home environment for extended periods to capture night-to-night variability in sleep patterns characteristic of insomnia. It can also be used to measure sleep in those who may be unable to complete sleep diaries. Because actigraphy records continuously, information on 24-hour rest and activity patterns can be examined to detect unreported daytime naps or day-to-day variations in the sleep schedule. Actigraphy may also be useful in monitoring adherence to behavioral treatment recommendations and for identifying when violations of those recommendations occur.[18] Furthermore, actigraphy provides streamlined electronic data collection and analysis. Although it is more costly than sleep diaries, actigraphy provides an objective estimate of sleep patterns that is much less expensive than polysomnography.

The major disadvantages of actigraphy revolve around a lack of standardization in instrumentation, methodologies, sampling procedures, and analytic algorithms.[15,16] For example, quantitative criteria used to differentiate between insomnia and normal sleepers vary by type of device.[19] Also, data loss may occur if the individual does not follow instructions for using the device (e.g., removes the device or does not depress the device button for documenting events).[20]

Written Behavioral Prescriptions for Home Use

The success of psychological and behavioral insomnia therapies depends on the patient's ability to implement therapeutic techniques in the home environment. Patients may be more successful in adhering to treatment recommendations if they understand the rationale behind the therapy instructions. This knowledge is particularly salient for treatment recommendations that are counterintuitive, such as directives to limit time in bed and to leave the bedroom if not sleeping. Indeed, as has been demonstrated in cognitive-behavioral therapy (CBT) for depression, a patient's acceptance of the treatment rationale enhances the likelihood of a positive treatment outcome.[21] Furthermore, providing a written summary of behavioral prescriptions helps the patient review and remember the information discussed during the therapy session.

Boxes 86-1 to 86-3 exemplify handouts that facilitate the implementation of behavioral treatments. Sleep restriction and stimulus control are efficacious, stand-alone therapies but are more commonly folded into a multicomponent CBT-I. Discussions of how these various treatment components are used in combination are provided in several therapist guidebooks.[1,2,22,23] The following text provides descriptions of the individual CBT-I components.

Behavior Therapy Tools

Sleep Restriction

In its original version,[24] sleep restriction requires the patient to initially limit the time in bed to the average total sleep time shown on 2 weeks of pretherapy sleep diaries. However, 4.5 hours was set as the minimum time in bed. Patients were instructed to go to bed 15 minutes earlier if, after 5 days, sleep efficiency (based on subjective reports) was 90% or more. If sleep efficiency was less than 85%, time in bed was decreased; no change was made if sleep efficiency fell between 85% and 89%.

Although the general principle of sleep restriction (that of inducing mild sleep deprivation to consolidate sleep) endures today, many clinicians relax the original sleep restriction protocol to promote the patient's adherence.[1,2,25-27] Morin[1] provides guidelines for implementing sleep restriction that include flexibility in restricting time in bed to what the patient can tolerate, increasing time in bed (after the initial restriction) at predetermined intervals rather than in response to sleep efficiency improvements in some situations, adjusting the frequency of time in bed alterations on the basis of the patient's comfort, giving patients control in selecting bed and rise times, and reassuring the patient about the transient side effect of daytime sleepiness. Recommending reduction of time in bed may seem counterintuitive to patients and contradict strategies they have tried on their own to manage insomnia. Thus, a clear explanation of the therapy's rationale as shown in Box 86-1 is particularly important to facilitate adherence.

Stimulus Control

The stimulus control instructions, originally published by Bootzin in the 1970s, essentially have remained unchanged since then.[28] The instructions to the patient, along with the rationale for each instruction, are reproduced in Box 86-2.[29] Stimulus control and sleep restriction often are used together in multicomponent insomnia treatments. In such applications, it is necessary to slightly alter the stimulus control instruction to go to bed at night when sleepy, particularly when sleepiness occurs too early to allow adherence to the sleep schedule required for sleep restriction. In these cases, patients are instructed to go to bed when sleepy but not before the earliest bedtime recommended by the sleep restriction schedule. Thus, the bedtime recommended as part of sleep restriction can be viewed as the earliest allowed bedtime, but the patient may choose to stay up later on a given night when not sufficiently sleepy at that time.

Box 86-1 SLEEP RESTRICTION RATIONALE

The following rationale is provided for a patient whose sleep logs show an average total sleep time of 6 hours, with an average time in bed of 8 hours.

Restrict the time you spend in bed to the actual amount of sleep you get. Since you average only 6 hours of sleep per night out of 8 hours spent in bed, your task consists of curtailing the time you spend in bed to these 6 hours. There is no reason for staying in bed any longer than that, since the remaining time is spent awake anyway. Spending excessive amounts of time lying down attempting to relax, rest, nap, or simply find a comfortable body position fragments rather than consolidates sleep. Although it may have been a useful coping strategy early on, it is most likely contributing to maintaining the sleep problem currently. As your sleep becomes more efficient, you will gradually be allowed to spend more time in bed.

From Morin CM. *Insomnia: psychological assessment and management*. New York: Guilford Press; 1993. p. 114.

Sleep Hygiene

Sleep hygiene practices are broadly defined and vary among clinicians, but most recommendations cover four core areas: keeping regular bedtimes and rise times, maintaining a comfortable sleep environment, curbing substance use (e.g., alcohol, nicotine, caffeine), and exercising.[30] As part of a multicomponent treatment package, recommendations about bedtimes and rise times might not be included because they are already subsumed under stimulus control and sleep restriction instructions. Box 86-3 shows a sleep hygiene handout that includes the rationale for each recommendation provided. Sleep hygiene practices can be assessed with instruments such as the Sleep Hygiene Index,[31] the Sleep Hygiene Awareness and Practice Scale,[32] or the Sleep Hygiene Checklist developed by Riedel.[33]

Sleep Education

Including a component of sleep education can cultivate positive patient expectations and maximize treatment adherence. Sleep education may include information about normative sleep patterns, circadian functioning, homeostatic sleep drive, hyperarousal, effects of sleep loss, and realistic sleep expectations.[2,9]

Additional Educational Resources

Patient-oriented educational resources in the public domain (e.g., mobile applications, websites, pamphlets, self-help books) are widely available at minimal cost and can reinforce or supplement nonpharmacologic insomnia treatments. For example, the American Academy of Sleep Medicine (www.sleepeducation.com) and the National Sleep Foundation (www.sleepfoundation.org) provide information about insomnia.

Cognitive Therapy Tools

Psychological and behavioral insomnia therapies also address cognitions that interfere with sleep, mostly by increasing sleep-disruptive cognitive arousal. Cognitive therapy for insomnia is typically combined with the behavioral methods discussed before. A recent trial confirmed that the most effective cognitive therapy outcomes are achieved by combining cognitive therapy with established behavioral therapy methods.[34] The following are a few cognitive therapy tools that help in implementing this therapy.

Worry Control and Structured Problem Solving

Worry in the presleep period is one of the strongest cognitive predictors of delayed sleep-onset latency.[35] For patients who tend to worry in bed, a variety of intervention strategies have proved useful, including stimulus control, relaxation, and techniques more directly focused on sleep-disruptive cognitions.[36-39] One of the original problem-solving cognitive strategies is a presleep writing procedure developed by Espie and Lindsay[38]; this technique is similar to a procedure called constructive worry.[36] The premise of such techniques is that unresolved

Box 86-2 STIMULUS CONTROL RATIONALE AND TREATMENT INSTRUCTIONS

1. Lie down intending to go to sleep only when you are sleepy.

   The goal of this rule is to help the patients become more sensitive to internal cues of sleepiness so that they will be more likely to fall asleep quickly when they go to bed.

2. Do not use your bed for anything except sleep; that is, do not read, watch television, eat, or worry in bed. Sexual activity is the only exception to this rule. On such occasions, the instructions are to be followed afterward when you intend to go to sleep.

   The goals here are to have activities that are associated with arousal occur elsewhere and to break up patterns that are associated with disturbed sleep. If bedtime is the only time patients have for thinking about the day's events and planning the next day, they should spend some quiet time doing that in another room before they go to bed. Many people who do not have insomnia read or listen to music in bed without problem. This is not the case for insomniacs, however. This instruction is used to help those who have sleep problems establish new routines to facilitate sleep onset.

3. If you find yourself unable to fall asleep, get up and go into another room. Stay up as long as you wish and then return to the bedroom to sleep. Although we do not want you to watch the clock, we want you to get out of bed if you do not fall asleep immediately. Remember the goal is to associate your bed with falling asleep *quickly*! If you are in bed more than about 10 minutes without falling asleep and have not gotten up, you are not following this instruction.

   In order to associate the bed with sleep and disassociate it from the frustration and arousal of not being able to sleep, the patients are instructed to get out of bed after about 10 minutes (20 minutes for those over 60 years old). This is also a means of coping with insomnia. By getting out of bed and engaging in other activities, patients are taking control of their problem. Consequently, the problem becomes more manageable and the patient is likely to experience less distress.

4. If you still cannot fall asleep, repeat Step 3. Do this as often as is necessary throughout the night.

   See rationale for earlier.

5. Set your alarm and get up at the same time every morning irrespective of how much sleep you got during the night. This will help your body acquire a consistent sleep rhythm.

   Insomniacs often have irregular sleep rhythms because they try to make up for poor sleep by sleeping late or by napping the next day. Keeping consistent wake times helps patients develop consistent sleep rhythms. In addition, the set wake times mean that the patients will be somewhat sleep deprived after a night of insomnia. This will make it more likely that they will fall asleep quickly the following night, strengthening the cues of the bed and bedroom for sleep. Often insomniacs will want to follow a different sleeping schedule on weekends or nights off than they do during the work week. It is important to have as consistent a schedule as possible, seven nights a week. In our experience, a deviation of no more than one hour in the wake time on days off does not produce problems in establishing a consistent rhythm.

6. Do not nap during the day.

   The goals of this rule are to keep insomniacs from disrupting their sleep patterns by irregular napping and to prevent them from losing the advantage of the sleep loss of the previous night for increasing the likelihood of faster sleep onset the following night. A nap that takes place seven days a week at the same time would be permissible. For those elderly insomniacs who feel that they need to nap, a short daily afternoon nap of 30 to 45 minutes or the use of 20 to 30 minutes of relaxation as a nap substitute is recommended.

From Bootzin RR. Cognitive-behavioral treatment of insomnia: knitting up the ravell'd sleave of care. In: Kenny DT, Carlson JG, McGuigan FJ, Sheppard JL, editors. *Stress and health: research and clinical applications*. Amsterdam: Harwood Academic Publishers; 2000. p. 243-66.

problems lead to worry about such matters and interfere with sleep onset by increased arousal. However, creating a plan to address the problem decreases the likelihood that such worries will follow the person to bed. Within the constructive worry approach, the patient sets aside some time several hours before bedtime (i.e., usually in the early evening) to complete a worksheet containing one column labeled Concerns and a second column labeled Solutions. In the Concerns column, patients list worries that have the greatest likelihood of keeping them awake at night. In the Solutions column, patients generate the next, immediate step they can take toward the goal of solving the problem. Patients are instructed not to write down the ultimate end solution as they may view this solution to be out of reach or difficult to implement. They are refocused on the most immediate step in the sequence. When there are no apparent solutions to solve a problem, patients are encouraged to write down an acceptance-based or self-care strategy. For example, the problem may be that their terminally ill pet will likely die tonight. In such a case, it is helpful to write down plans to implement coping strategies, such as calling a friend for moral support.

Constructive worry was also formally tested as an adjunct to behavioral therapy techniques such as stimulus control and sleep restriction, and the addition of the constructive worry procedure reduced insomnia symptom severity as well as worry to a greater degree than did behavioral therapy alone.[40] Formal problem-solving training (described by Malouff et al[41]) is also a helpful adjunct to behavioral therapy as the effects of combining problem-solving training with behavioral therapy were comparably large to those resulting from combined behavioral therapy and constructive worry.[42] A similar intervention is a writing technique called the Pennebaker writing task.[43] Harvey and Farrell[44] found that poorly sleeping young adults who used this strategy and wrote about their thoughts and negative emotions before going to bed decreased their sleep-onset latencies. A replication of this technique in a clinical sample did not find evidence for a sleep improvement; however, the writing strategy led to greater improvements in self-reported cognitive arousal.[45] Indeed, worry management techniques are likely most helpful to reduce presleep cognitive arousal and worry,[36] rather than producing sleep improvements, when tested alone.

Thought Records

Thought records are a classic cognitive therapy tool for modifying the maladaptive thinking[46] styles implicated in a variety of disorders including insomnia.[47-50] Negative sleep-related thoughts are rated as more compelling or more "believable" by

Box 86-3 SLEEP HYGIENE INSTRUCTIONS AND RATIONALE

The sorts of daytime activities in which you engage, the foods and beverages you consume, and the surroundings in which you sleep may all influence how well you sleep at night and how you feel in the daytime. Thus, you may benefit from making some changes to your lifestyle and bedroom to promote better sleep.

- Limit your use of caffeinated foods and beverages, such as coffee, tea, soft drinks, and chocolate. Caffeine is a stimulant that may make it harder for you to sleep well at night. Caffeine stays in your system for several hours after you consume it. Therefore, we recommend that you limit caffeine to the equivalent of no more than 3 cups of coffee per day and that you not consume caffeine in the late afternoon or evening hours.
- Limit your use of alcohol. Alcoholic beverages may make you fall asleep more easily. However, alcohol also causes sleep to be much more broken and far less refreshing than normal. Therefore, we recommend against using much alcohol in the evening or using alcohol as a sleep aid.
- Try some regular moderate exercise, such as walking, swimming, or bike riding. Generally, such exercise performed in the late afternoon or early evening leads to deeper sleep at night. Also, improving your fitness level will likely improve the quality of your sleep. However, avoid exercise right before bed because it may make it harder to get to sleep quickly.
- Make sure your bedroom is quiet and dark. Noise and even dim light may interrupt or shorten your sleep.
- Make sure the temperature in your bedroom is comfortable. Generally, temperatures above 75 degrees Fahrenheit cause unwanted awakenings from sleep. During hot weather, we suggest you use an air conditioner to control the bedroom temperature.
- Try a light bedtime snack that includes such items as cheese, milk, or nuts. These foods contain chemicals that your body uses to produce sleep and may help bring on drowsiness.

Modified from Edinger J, Carney C. *Overcoming insomnia: a cognitive-behavioral therapy approach—therapist guide.* New York: Oxford University Press; 2015.

those with insomnia[47-51] relative to their "good sleeper" counterparts. More specifically, people with insomnia have cognitions that (1) overestimate the negative effects of insomnia, (2) focus on the belief that sleep is out of one's control, (3) are rigid or unrealistic about how much sleep is needed for functioning, and (4) are misconceptions about the causes of insomnia.[47-50] Moreover, such thoughts seem equally present among insomnia sufferers with and without sleep-disruptive comorbidities.[51] Negative thoughts and moods can exert reciprocal negative influences on each other, and both can perpetuate insomnia. The purpose of the thought record is to observe the connections between negative thoughts and distress and to challenge or to interrupt this process.

This connection of negative thinking and distress is achieved through prospective monitoring using a specially designed thought record worksheet divided into columns. Several versions of thought records are available, including a version specifically designed for use with insomnia patients.[2] The first three columns of an insomnia-focused thought record are labeled (1) Situations, (2) Mood and Mood Intensity Rating (0% to100%), and (3) Thoughts. Observing which situations (Column 1) tend to give rise to distress (Column 2) and subsequent negative sleep-related thoughts (Column 3) can provide insight into a seemingly automatic cognitive-emotional process. Although observing these connections can be helpful in initiating cognitive change, the main use of the thought record is to challenge unhelpful thoughts. This is accomplished by having the patient list some experiences or perspectives that support the thought (Column 4) as well as some experiences or perspectives that refute the thought (Column 5). The patient then constructs a more balanced appraisal of the sleep-wake difficulties that combines the evidence supporting and refuting the original negative thought and writes that in a separate column (Column 6).

Patients often initially have difficulty completing all aspects of the thought record and particularly the task of providing evidence that refutes the unhelpful thought. For example, after a poor night's sleep, an insomnia sufferer may struggle with refuting the negative thought "after a poor night's sleep, I just can't function." For such patients, a series of Socratic questions may prove helpful. Has this thought been true 100% of the time? Is it guaranteed that a poor night's sleep will prevent them from functioning? Will he or she be able to function albeit with somewhat more than usual effort? Are there times when they had a good day even after a poor night's sleep, or a poor day after a good night's sleep? Are there strategies they have used in the past to be able to cope with the negative effects of their insomnia? Could they see how this thought becomes a self-fulfilling prophesy? Through this questioning, the patient may find evidence to refute the thought and develop a more balanced statement, such as "I don't feel well now after my poor night's sleep, but I've always made it through, and things tend to get better as the day goes on anyway." Worry management through the use of thought records or constructive worry is often a helpful adjunct to behavioral therapy. Also, given that patients view mental arousal as one of the most important factors in their insomnia, it is a worthwhile clinical endeavor.[52]

Acceptance-Based Approaches

These approaches, which include acceptance and commitment therapy and mindfulness-based cognitive therapy, represent a relatively new approach to insomnia management and focus on reducing emotional and behavioral reactivity to these anxiety-provoking thoughts about sleep, thus reducing arousal and sleep effort. Specific techniques of this nature are described in a conceptual paper by Ong, Ulmer, and Manber[53] and supported by a recent randomized clinical trial.[54]

SELECTION OF TREATMENT COMPONENTS AND DELIVERY METHODS

As noted thus far, the psychological and behavioral insomnia therapies comprise a diverse group of techniques that can be used individually or in combination to address insomnia patients' sleep-wake complaints. Admittedly, multicomponent CBT-I that combines one or more of the cognitive therapy techniques described before or in the prior chapter with stimulus control therapy, sleep restriction, sleep hygiene education, and perhaps also relaxation training has emerged as the "gold standard" nonmedicinal insomnia therapy. In fact, this form of therapy is regarded as a front-line intervention for insomnia by both the National Institutes of Health in the United States and the British Association of Psychopharmacology.[55,56] Of course, in practice, the specific psychological and behavioral

treatment components selected may be guided by the preferences and biases of practitioners who employ them as well as by patients' acceptance of those techniques. Many other factors also may dictate which treatment components are selected for use and how those treatments are actually delivered. The expertise of providers in the settings wherein the insomnia patient presents, the amount of treatment time that is practical and available in that setting, and some more general considerations about the cost-effectiveness of the treatment may all influence such choices. These considerations, in turn, argue for flexibility in choosing the components actually employed and how they are delivered so as to make these treatments broadly applicable and practical. Fortunately, both research and clinical experience suggest that these treatments afford the practitioner and patient considerable flexibility in regard to their actual composition and method of delivery. Indeed, as the following discussion demonstrates, the treatment components employed, the therapeutic modality used (i.e., individual vs. group vs. self-help), the type of provider delivering treatment, and the means by which treatment instructions are communicated can all vary such that a multitude of effective treatment delivery options can be considered for these therapies.

Individual Treatment Format

Perhaps the best tested and most commonly used method of delivering psychological and behavioral insomnia therapies has been by individualized treatment consisting of one-on-one sessions between a therapist and single patient. Most published studies using this treatment format have employed doctoral-level or graduate student psychologists as therapists primarily because the psychological and behavioral insomnia therapies have, for the most part, been developed by psychologists. Several meta-analyses and systematic reviews suggest that individualized treatment is efficacious for short-term and longer insomnia management.[57-62] Moreover, this literature suggests that individual therapy is effective for the stand-alone therapies, including relaxation training, sleep restriction, paradoxical intention, and stimulus control, as well as for the multicomponent CBT-I protocols. Thus, individualized therapy can be considered a well-established and viable method for delivering these treatments.

As evidenced by previous critical reviews[58,59] as well as meta-analyses,[57,60,61] the literature contains many descriptions of treatment protocols for both the stand-alone and multicomponent psychological and behavioral insomnia treatments provided in individual treatment formats. Publications during the past 25 years suggest that CBT-I and other multicomponent renditions of the behavioral insomnia therapies have become increasingly popular. This trend likely results from the presumption that multicomponent treatments are best able to address the range of cognitive and behavioral mechanisms that perpetuate insomnia.[59] Most available descriptions of these multicomponent interventions come from clinical trials that outline a fixed treatment package, usually in four to eight sessions occurring at predetermined intervals. However, in clinical practice, patients' treatment needs vary, so pretreatment case formulations often guide the specific treatment components and duration of treatment provided. Although no consensus guidelines exist for matching treatment components to patients, some previous clinic-based studies[63,64] describe flexible treatment packages that allow tailoring of therapies to patients.

In deciding on the attractiveness and suitability of individualized therapy for a particular venue, consideration of the advantages and disadvantages of this treatment method is useful. As is the case for individualized treatment in general, this form of insomnia treatment delivery fits well with traditional mental health outpatient and private practice settings wherein one-on-one meetings between providers and patients are the norm. Individualized treatment offers reasonable scheduling flexibility, subject to the common availability of the provider and patient, and it requires only the limited office space to accommodate the patient-provider dyad. It also allows maximum flexibility in tailoring of treatment to best address each patient's problem sleep-related cognitions and behaviors. However, individualized treatment is at the high-intensity end of the treatment continuum. It is relatively costly to the patient and arguably is an inefficient use of the limited number of providers (largely psychologists and other doctoral-level sleep specialists) who have expertise with behavioral insomnia therapies. Yet, some patients prefer individual therapy to other forms of treatment, given the level of confidentiality and treatment tailoring it affords them, despite its relatively high cost. Thus, the patient's preferences may dictate the choice of this approach when knowledgeable providers are available.

Group Treatment Format

To address the noted disadvantages of individualized insomnia therapy, various alternative, more cost-efficient delivery methods have been developed. One common albeit lower intensity treatment delivery method entails the use of a group therapy format. Most previously described group treatment protocols have used CBT-I or other multicomponent packages that are typically provided in four to eight group treatment sessions. Table 86-1 provides a listing and brief descriptions of a sampling of these group protocols. As suggested by this information, there is no one standard brand of group therapy that has been employed universally for insomnia treatment. However, the descriptions presented show that group therapy protocols characteristically include an amalgam of such well-established, stand-alone insomnia treatments as stimulus control, relaxation training, and sleep restriction, and these components are often supplemented by sleep education, sleep hygiene instructions, and cognitive therapy. Thus, group protocols are typically omnibus and include components with well-established efficacy for treating chronic insomnia.

Results of the studies listed in Table 86-1 as well as previous meta-analyses[57,60,65] and critical reviews[58,59] suggest that group therapy is an efficacious delivery format for behavioral insomnia therapy. As shown by Figure 86-3, recent meta-analytic findings suggest that group CBT-I produces medium to large effect sizes for most sleep variables. Both comparisons conducted within patient cohorts receiving group CBT-I and comparisons between cohorts assigned to group CBT-I or a control condition support this contention. As is also apparent from the data presented in Figure 86-3, group CBT-I has durable treatment effects because the effect sizes noted at follow-up assessment conducted 3 to 12 months after treatment cessation indicate reasonable retention of treatment gains. Thus, group delivery of psychological and behavioral insomnia therapy seems a reasonably efficacious treatment methodology.

Whether group and individual insomnia treatments are equally efficacious is not clear. On the one hand, conclusions

Table 86-1 Sampling of CBT-I Group Therapy Approaches for Insomnia

Citation	No. of Sessions	Treatment Components	Types of Patients	No. of Patients per Group	Type of Providers
Backhaus et al[208]	6	Sleep education Sleep hygiene Progress relaxation Cognitive relaxation Stimulus control Sleep restriction Cognitive therapy	Primary insomnia	4–8	PhD clinical psychologists
Morin et al[112]	8	Sleep education Sleep hygiene Stimulus control Sleep restriction Cognitive therapy	Primary insomnia	4–6	PhD clinical psychologists
Espie et al[72]	6	Sleep information Sleep hygiene Stimulus control Sleep restriction Relaxation Cognitive therapy	Medical clinic patients with primary insomnia	4–6	Health visitors (nurses)
Rybarczyk et al[209]	8	Sleep education Sleep hygiene Stimulus control Sleep restriction Cognitive therapy Relaxation	Older mixed medical patients with insomnia	5–6	PhD clinical psychologists
Epstein and Dirksen[71]	4 + 2 phone follow-ups	Sleep education Sleep hygiene Stimulus control Sleep restriction	Breast cancer survivors	4–8	Master's-level clinical nurse specialist
Bootzin and Stevens[210]	6	Stimulus control Bright light Sleep hygiene Cognitive therapy Mindfulness-based stress reduction	Teenage substance abusers	2–6	Advanced psychology graduate students
Verbeek et al[67]	6	Psychoeducation Stimulus control Sleep restriction Relaxation Cognitive therapy Medication withdrawal	Clinically referred primary insomnia	5–7	PhD-level therapists
Ong et al[120]	7	Sleep education Deep breathing and mental imagery Stimulus control Sleep restriction Cognitive restructuring	Clinically referred mixed insomnia subtypes	8–15	Licensed PhD clinical psychologist
Vitiello et al[211]	6	Sleep hygiene Sleep restriction Stimulus control Relaxation/guided imagery Cognitive restructuring Pain education Physical activation Pacing	Patients with osteoarthritis pain and insomnia	5–12	Master's-level family counselor and PhD psychologist

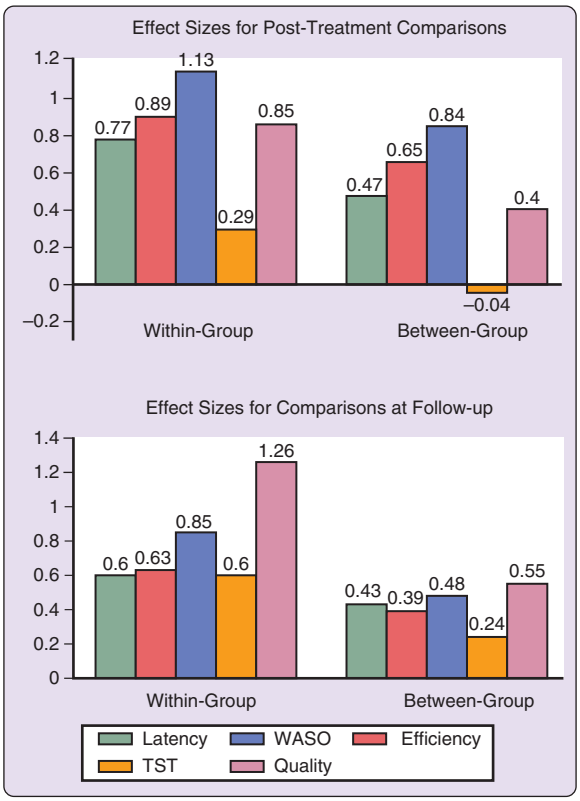


Figure 86-3 Group cognitive-behavioral therapy for insomnia (CBT-I) treatment effect sizes. Data taken from Koffel et al.[65] All data are based on sleep diary assessments. Within-group comparisons are based on pretreatment to posttreatment change within cohorts receiving group CBT-I. Between-group data are based on comparisons between cohorts assigned to group CBT-I and a control group. Latency, Sleep-onset latency; WASO, wake time after sleep onset; Efficiency, sleep efficiency; TST, total sleep time; Quality, subjective sleep quality ratings.

derived from previous meta-analyses conflict; one such report[57] concluded that individually administered treatments are slightly superior to group interventions, whereas another report[60] concluded that the two methods of insomnia treatment produce comparable results. However, conclusions based on meta-analyses should be accepted cautiously because they are based on comparisons of studies that employed different therapists, different patient samples, and often different treatment packages. More trustworthy are randomized trials that provide direct comparisons of group and individual therapy within the same study. Unfortunately, only three such studies have been conducted. Two of these studies[66,67] concluded that the two methods of CBT-I delivery are equally effective, whereas the third reported that individual CBT-I resulted in greater improvement in sleep-onset latency and overall sleep quality than did group CBT-I.[68] Thus, more comparisons of this nature are needed before conclusions can be drawn.

What is clear is that group therapy represents a reasonably efficacious form of insomnia treatment delivery with several advantages that make it an attractive treatment medium. First, as there currently are a limited number of providers competent in behavioral insomnia therapies, group treatment allows multiple patients access to a single trained provider. As such, the group format entails a relatively efficient use of the provider's time. Because of this efficiency, group treatment is generally a less expensive treatment alternative for the patient and third-party payers. Furthermore, the group milieu, composed of many patients with the same sorts of sleep difficulties, offers individual patients some sense of validation and social support that they would likely not derive from individualized treatment. Arguably, these nonspecific factors enhance treatment outcomes for at least a portion of those who attend group insomnia therapies. Yet group treatment offers less scheduling flexibility than one-on-one therapy, and it may afford patients less individualized attention. It may be difficult for all patients to attend all sessions at prescheduled times, and if patients miss group sessions, individual "make-up" sessions may be needed to help patients catch up on missed material. Such occurrences, of course, raise the cost and reduce the efficiency of group intervention. Group treatment may be optimized by the individual prescreening of participants to identify those with special treatment needs (e.g., uncontrolled comorbid psychiatric or medical disorders) or who might be disruptive to the group process. Furthermore, group treatment requires a meeting room sufficiently large to accommodate at least one therapist and several patients. The space needed and the desirability of minimizing the time patients wait for a new group to form may render group treatment unfeasible in some settings because of limited space and patient volume. Nonetheless, group therapy should be considered a viable treatment alternative that has a place in many clinical venues.

Delivery by Nonspecialist Providers

As noted recently,[69] the ranks of sleep specialists who are expert in the behavioral insomnia therapies are currently limited. In efforts to facilitate broad dissemination of behavioral insomnia treatments, some investigators have examined whether nontraditional "therapists" can effectively deliver these interventions. Because insomnia sufferers typically present first in primary care settings, it seems reasonable to consider using health care professionals (e.g., nurses, physician's assistants, general practitioners) who are not sleep specialists to provide behavioral interventions. Several studies designed to test the efficacy of using nontraditional therapists have demonstrated that various general health care staff, including family physicians,[70] nurses,[71-73] nondoctoral rural mental health clinic staff such as mental health counselors and social workers,[74] and primary care counselors,[75] can effectively administer treatments such as stimulus control and multicomponent CBT-I in general practice settings. In several of these trials, the nontraditional therapists received training and supervision from an experienced specialist (i.e., sleep psychologist) such that the therapists employed could be considered specialist provider "extenders." This model wherein the CBT-I specialist assumes a trainer or consultant role may represent a reasonable alternative to higher priced specialty care and prove useful for optimal dissemination of the behavioral insomnia therapies in primary care and other medical settings.

Perhaps the most ambitious effort to train nonspecialists in CBT-I delivery has been the training dissemination program effected by the Department of Veterans Affairs (VA).[9,76-78] This program was devised to train available licensed mental health workers (psychologists, social workers, nurses, and psychiatrists) at VA medical centers across the United States in the delivery of CBT-I to VA patients with insomnia complaints. The training itself requires providers to read a treatment manual and to attend a 3-day workshop as well as 4 months of weekly telephone consultation or supervision

with an expert consultant who provides case consultation and reviews the audio recordings of CBT-I therapy sessions they conduct. The most recent report[78] indicates that this program has successfully trained more than 300 VA providers across the nation, and most of these trainees have gone on to use their CBT-I skills in treating almost 700 insomnia patients at their respective medical centers. Furthermore, the program evaluation data suggest that these providers show acceptable CBT-I proficiency levels as measured by systematic ratings of their recorded therapy sessions and also by patient-reported outcome measures assessing their insomnia symptoms, mood statuses, and quality of life perceptions. Admittedly, this training program was time-consuming to develop and costly to effect. Yet it provides a promising model for optimizing the use of a limited number of CBT-I experts in a consultant role to expand their reach to a much larger population of insomnia patients than they themselves could manage.

Collaboration between a specialist consultant and the nonspecialist health care providers who deliver treatment has obvious appeal. This delivery model represents an efficient use of behavioral sleep medicine specialists and ensures that their expertise is made available, albeit indirectly, in those medical settings where most treatment-seeking insomnia patients present.[79] The training of existing health care personnel to deliver behavioral insomnia therapies also seems reasonable, given the relevant training most such professionals have in the provision of related health care activities. At least one study[73] using this delivery method noted treatment effect sizes that fall short of those usually reported for efficacy studies in which more traditional therapists (PhD or graduate-student psychologists) were employed. In another study,[80] primary care health providers who received CBT-I training solely from a manual helped 47% of the insomnia patients they treated achieve clinically significant improvement after treatment. However, these patients also showed significant deterioration in their symptoms by a subsequent follow-up assessment, a finding contrary to many studies that employed more expert therapists. Whether such differences in treatment outcome are due to "therapist effects" or the presence of more difficult to treat patients encountered by the CBT-I-trained nonspecialists is difficult to determine. Only through studies that directly compare specialist and nonspecialist CBT-I therapists will this question be answered. When licensed health care providers who are not CBT-I specialists can be properly trained and given access to an experienced CBT-I specialist or consultant, this sort of delivery method seems a reasonable alternative to consider. However, the more complex insomnia cases seen in such settings may ultimately benefit by referral to an experienced behavioral sleep medicine specialist for insomnia management.[69]

Abridged Protocols Designed for Primary Care Settings

A large percentage of insomnia patients first seek treatment in primary care settings for their sleep problems. Yet the high patient flow and consequent restrictions on the face-to-face time providers have with their patients obviate use of the relativity time-consuming CBT-I protocols in such venues. Hence, some investigators have developed and tested abridged CBT-I interventions specifically designed for use in primary care settings. In an early study, Edinger and Sampson[81] developed a treatment composed of brief sleep education on the individual differences in sleep needs, the effects of aging on sleep, and the influences of sleep drive and circadian rhythms on sleep along with a condensed set of behavioral treatment recommendations, including (1) eliminate sleep-incompatible activities (television watching, reading, planning, and worrying) in the bed and bedroom, (2) avoid all daytime napping, and (3) follow a consistent sleep-wake schedule by adhering to agreed-on bedtimes and rising times. This treatment was delivered to military veterans with sleep complaints in two biweekly 25-minute sessions with a psychologist. A comparison of this intervention against a sleep hygiene control intervention showed highly favorable results for the abridged treatment.

In two subsequent studies, Germain et al[82] and Buysse et al[83] tested a similar brief behavioral therapy for insomnia (BBTI) against an information control condition for treating the sleep problems of older primary care patients with insomnia. BBTI consisted of brief sleep education about homeostatic and circadian sleep regulation as well as a set of behavioral recommendations, including (1) reduce time in bed, (2) get up at the same time every day, regardless of sleep duration, (3) do not go to bed unless sleepy, and (4) do not stay in bed unless asleep. Napping was also discouraged. BBTI and information control interventions were delivered by a master's-level nurse practitioner who had no prior experience with behavioral insomnia therapies. Both treatments were delivered in one 45- to 60-minute initial face-to-face session followed by a 30-minute visit 2 weeks later and then by additional 20-minute phone calls scheduled 1 and 3 weeks after the latter in-person session. The study results suggest that BBTI was significantly more effective than the control intervention and resulted in a 55% insomnia remission rate. Figure 86-4 shows the treatment effect sizes for selected sleep diary variables in the comparisons conducted between BBTI and information control. Comparisons of these findings with previous studies suggest that BBTI produced more modest treatment effects than found in previous studies using unabridged treatments in middle-aged cohorts but similar effects to such treatments applied to older adults. Nonetheless, considering that the promising effects were obtained with the sort of nurse interventionist that would be found in many primary care settings, this type of intervention holds promise for broad dissemination in primary care venues. However, how best to broadly disseminate the training necessary to the health care providers

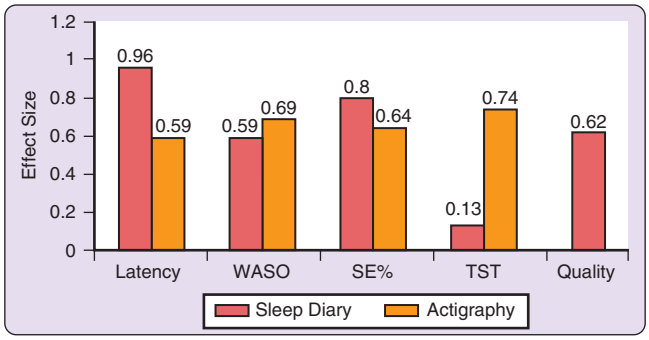


Figure 86-4 Treatment effect sizes for brief behavioral insomnia therapy.[83] Findings taken from Buysse et al.[83] Effect sizes shown are for a brief behavioral insomnia therapy relative to an information control therapy. Latency, Sleep-onset latency; WASO, wake time after sleep onset; SE%, sleep efficiency percentage; TST, total sleep time; Quality, subjective sleep quality rating (rated on a scale of 0 to 100) from sleep diary only.

who might deliver this treatment remains a challenge to be addressed in making this sort of entry-level intervention more generally available.

Self-Help and Remote Treatment Delivery

The various approaches discussed thus far provide a number of delivery methods for disseminating behavioral insomnia treatments. However, these options considered collectively may not meet the needs of all insomnia sufferers. Given the 10% to 22% population prevalence[84-87] of chronic insomnia and the limited number of skilled providers, many patients do not have easy access to such treatment, especially those far outside urban areas where most such providers are found. Others with limited financial resources (e.g., the poor and uninsured) may not have sufficient financial resources to seek professional help for their insomnia. Finally, some insomnia sufferers initially may prefer to use self-help approaches with limited or no provider contact. Thus, they may resist presenting to clinical settings for insomnia treatment.

Given these considerations, several investigators have tested less expensive treatment delivery methods that can be largely initiated in the home by patients themselves. For example, Mimeault and Morin[88] tested a booklet of self-help CBT-I instructions used independently or with therapist guidance provided by telephone against a wait-list control group. Compared with the control, those treated with the self-help therapy showed substantially greater sleep improvements, and the improvements were maintained at a 3-month follow-up. The addition of therapist phone consultations conferred some advantage over the self-help booklet alone, but these benefits disappeared by follow-up. In another study, Currie et al[89] compared individual behavioral insomnia therapy, treatment with a self-help manual, and a waiting list condition for ameliorating insomnia in a group of recovering alcoholics. Results showed that treated patients achieved significantly greater improvements than controls, but no significant differences were noted between the in-person individual therapy and the home-administered self-help program. Bastien et al[66] found comparable effectiveness of cognitive-behavioral insomnia therapy provided in individual, group, and telephone consultation formats. Finally, Rybarczyk et al[90] tested a home-based video CBT-I program for treating older insomnia patients with comorbid medical conditions. This treatment produced greater improvements in sleep and waking functioning than did a wait-list condition, but the improvements fell short of those achieved with CBT-I delivered face to face in a classroom setting. Considered collectively, these findings suggest that self-administered behavioral insomnia treatments are promising, although contact with a therapist by phone consultation may be needed to improve outcome.

To determine whether behavioral insomnia treatments can be delivered effectively without direct provider-patient interaction, some investigators have tested treatment protocols provided through mass media dissemination. Oosterhuis and Klip[91] tested a behavioral insomnia therapy provided through a series of eight 15-minute educational programs broadcast on radio and television in the Netherlands. More than 23,000 people ordered the accompanying course material. Data from a random subset of these showed that sleep improvements and reductions in hypnotic use, medical visits, and physical complaints were achieved among those who took part in this program. However, given the single-group nature of their design, it is difficult to discern how these results might compare with more conventional treatment. More recently, Gehrman[92] described a group CBT-I program for veterans with posttraumatic stress disorder (PTSD) delivered by a remotely located expert therapist by video telehealth connection. The therapist was located at one VA medical center in Philadelphia, whereas the patients were located elsewhere at various VA facilities around the country. The program entailed six group therapy sessions delivered by video conferencing to the remotely located patients. Early results reported improvements in both insomnia and PTSD symptoms, although endpoint data suggest that the treatment effects may fall short of what has been reported for other group CBT-I interventions. Nonetheless, the reported results appear promising, particularly in considering the difficult sleep problems typically presented by the types of patients enrolled in this treatment.

Given its widespread availability and extensive societal reach, the Internet has also been tapped for dissemination of CBT-I. Various websites (particularly those developed by medical organizations and professional sleep societies) offer their readers general sleep hygiene information as well as sleep improvement guidelines derived from interventions such as stimulus control and sleep restriction therapy. These sorts of "interventions" arguably represent a relatively low intensity treatment perhaps best suited for a subset of insomnia sufferers who are highly self-motivated. However, at the other end of the Internet delivery continuum are a number of recently developed, automated and interactive CBT-I interventions designed to mimic therapist-delivered interventions. Several investigators have developed and tested these sorts of interventions with generally positive results.[93-99] In one well-designed trial, Ritterband et al[95] tested an Internet CBT-I intervention composed of video-based material from expert therapists and automated interactive components designed to tailor treatment to individual patients (www.shuti.net). Study results showed that the Internet CBT-I produced significantly greater reductions in overall insomnia severity and greater improvements in sleep efficiency than did a wait-list control. Likewise, Espie et al[99] tested an Internet-based intervention (www.sleepio.com) in the largest ($n = 164$) and only placebo-controlled trial to date against both an Internet placebo therapy and treatment as usual in primary care patients. The intervention is delivered by an engaging animated character, "The Prof," who interacts with program users and provides treatment advice and guidance based on information entered into the program by the user-patient. Study findings showed that this intervention produced significantly greater improvements than did the other two conditions for a range of outcome measures.

Overall, the various self-help and remote delivery methods for CBT-I have several advantages that give them appeal. First, they greatly extend the reach of these insomnia treatments to remote users who would otherwise have no access to such treatments. They also can be considered efficient because they allow access to treatment for a large number of patients at one time. Furthermore, the more sophisticated online delivery programs mimic what occurs in the typical psychological and behavioral insomnia therapy yet do so at a reasonable cost. Although online interventions do charge users for their programs, the costs are notably lower that the costs for in-person interventions. Yet the majority of these interventions require significant independence and self-motivation on the patient's

part to derive benefits. As such, they may not be suited for every patient. Furthermore, they may not be quite as effective as in-person treatment delivery methods. Recent meta-analyses[100,101] that evaluated self-help CBT-I show that treatment effect sizes appear larger when patients are afforded some additional therapist support (e.g., phone consultations) and when they lack sleep-disruptive comorbidities. Thus, further exploration of patient characteristics that best fit these approaches is needed.

Deciding When to Use Each of the CBT-I Treatment Options

The preceding discussion documents the myriad treatment protocols and delivery options that can be considered for behavioral insomnia therapies. It should be understood that the best-tested and well-established forms of delivery are the individual and group therapy protocols directed by those with extensive training backgrounds in the provision of psychotherapy in general and CBT-I specifically. Treatment delivery approaches that diverge from that common paradigm through use of alternative providers, abridged treatments, or no direct provider contact at all currently have relatively less empiric support. This does not mean they are not effective, at least for a subset of insomnia sufferers.

What is clear is that there is a wide array of treatment delivery options for these therapies ranging from the relatively low intensity yet highly accessible self-help interventions to the high-intensity yet more scarcely available individual therapist-directed protocols. Given this range of delivery methods, several insomnia experts[102-104] have suggested use of the stepped care, pyramid-shaped model for treatment allocation shown in Figure 86-5. At the bottom of this pyramid are the lowest intensity albeit most widely available self-help protocols delivered through the Internet and other mass media. A "step up" from these treatments would be the group or individually administered entry-level treatments designed specifically for those patients who actively seek intervention in primary care settings. At the next higher step are individual and group treatments offered by health care providers who have some mental health or psychotherapy expertise and journeyman training in CBT-I yet lack behavioral sleep medicine specialty. At yet a higher and less accessible level of the pyramid are the treatments offered by those with expertise in behavioral sleep medicine. Finally, at the top of the pyramid are the treatments available at sleep centers that employ behavioral sleep medicine specialists and can manage patients who have insomnia comorbid to other primary sleep disorders (e.g., obstructive sleep apnea). In theory, this model would allocate patients to the various levels and intensities of treatment to fit their specific treatment needs. The most motivated and self-sufficient might be best served by treatments at the base of the pyramid, whereas those more complex cases are assigned to higher intensity treatments toward the upper portion of the pyramid. Moreover, those who fail to respond to a lower intensity treatment could be "stepped up" to a higher intensity intervention.

Obviously this treatment model has significant heuristic appeal as it balances treatment intensity and cost against the needs of specific patients, thus resulting in the efficient use of the limited number of providers with a high level of expertise. Unfortunately, to date, there has been limited testing of this model to determine its usefulness. However, a recent study conducted by Vincent and Walsh[104] tested online with therapist-directed CBT interventions in a large medical outpatient setting and provided some insight into the sorts of patients who might need to be stepped up to more intensive in-person therapies. Results of this study showed that patients who were older, unemployed, and presented with poorer sleep quality responded better to a higher intensity therapist-delivered intervention than they did to online CBT-I.

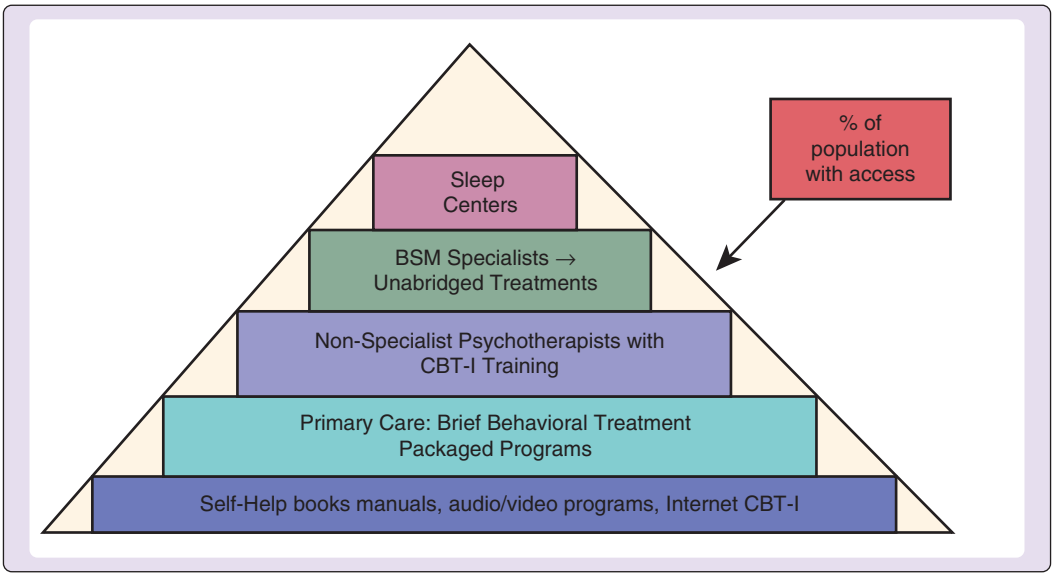


Figure 86-5 Stepped care model for allocation of insomnia patients to treatment options. Stepped care model adapted from Edinger and Carney[102] and Espie.[103] This model shows the varying levels of treatment intensity with provider expertise. The treatments toward the base of the pyramid are most widely accessible to patients, whereas those toward the top of the pyramid represent less accessible levels of care to which the more challenging and complex patients should be referred. BSM, Behavioral sleep medicine; CBT-I, cognitive-behavioral therapy for insomnia.

As noted by Edinger,[105] a number of issues would need to be addressed in placing the stepped care model into general practice. First, we have to decide which treatments should be included in the model at the different levels of delivery to optimize the overall efficacy of treatment allocation. We also need to answer questions such as these: (1) Which patients are likely to benefit from self-help approaches and which are not? (2) Does treatment failure with one of the nontraditional or self-help approaches adversely affect a patient's treatment expectations and subsequent response to a more intensive traditional behavioral insomnia treatment? and (3) In what settings can this model actually be implemented? In regard to the last question, it would admittedly be difficult to effect a stepped care model at a broader societal level in a country such as the United States, given its complex free market health care system. It would seem to better fit within closed health maintenance organizations wherein a finite patient population of health plan enrollees is closely tracked by computer technology. Likewise, it may be possible to implement it in countries that have a socialized, government managed, and centralized health care system. However, even in such settings, the allocation of patients to proper treatment represents a massive triage endeavor requiring systematic patient screenings to optimize the treatment matching process. Admittedly, much has to be learned about patient characteristics that predict treatment response to effectively use this stepped care model.

TREATMENT DOSING

In addition to considering the various methods and providers that can be used to deliver CBT-I, it is also important to consider what constitutes the optimal "dose" of such treatments. Previous studies with multicomponent CBT-I protocols have shown efficacy for interventions delivered in one, two, four, six, and eight individualized sessions and between five and eight group sessions. Yet little is known about how many individual and group sessions are needed on average to achieve optimal treatment outcomes. Despite the ample number of exemplary dose-response studies extant in the pharmacologic literature,[106] research specifically designed to determine the optimal dosing for any of the psychological and behavioral insomnia therapies has generally been lacking.

One exception to this oversight is a study conducted by Edinger et al.[107] The goal was to determine the optimal dosing of their CBT-I model delivered in an individual therapy format. In this study, patients with primary insomnia were randomly assigned to a no-treatment control (wait list) or to active treatment composed of one, two (weeks 1 and 5), four (every other week), or eight (weekly) CBT-I sessions delivered during an 8-week period. Results showed that subjective (sleep diary) and objective (actigraphic) sleep measures favored the one- and four-session CBT-I doses over the other doses and wait-list control. However, comparisons of pretreatment data with data acquired at the 6-month follow-up showed that only the four-session model maintained significant improvements in objective wake time and sleep efficiency measures over time. Moreover, a notably greater percentage of those assigned to the four-session condition achieved 50% or more reduction in the study's primary sleep outcome measure, subjective wake time after sleep onset, than did those in the other conditions. Considered collectively, this study's findings imply that for the specific CBT model tested, a dose of four sessions delivered on an every-other-week schedule appears optimal. However, a more recent study[108] showed that those with insomnia comorbid to a psychiatric condition had less impressive treatment outcomes than did insomnia patients without comorbidities in response to the four-session CBT model shown to be optimal in the aforementioned dose-response study.[107] Thus, concurrent comorbidities may be important factors that affect the optimal dosing of CBT-I.

In addition to the number of treatment sessions provided, an alternative important but largely ignored dosing consideration is the number and types of treatment components composing an insomnia therapy. However, Harvey et al[34] recently investigated the relative effectiveness of treatments composed solely of behavioral therapy (e.g., stimulus control, sleep restriction, and sleep hygiene) or cognitive therapy techniques compared with a multicomponent CBT-I. Cognitive therapy included empiric testing of unhelpful beliefs about sleep through behavioral experiments in addition to using direct verbal challenge by Socratic testing or thought records. Results of this trial showed that behavioral therapy produced more immediate sleep improvements than cognitive therapy but that cognitive therapy provided longer term benefits than behavioral therapy as measured at a follow-up. However, the best outcomes were obtained with the full CBT-I intervention, supporting the notion that combining behavioral therapy and cognitive therapy provides the optimal approach. Of course, the combined treatment entailed longer treatment sessions on average than did the other two treatments (75 minutes vs. 40 to 60 minutes).

TREATMENT ACCEPTABILITY AND ADHERENCE ISSUES

The ultimate success of an insomnia therapy relies on patients' willingness to accept and to engage in treatment as well as their ability and motivation to implement the treatment components at home. Factors related to treatment acceptability and adherence are reviewed in this section.

Treatment Acceptability

Patients' preferences and values play a critical role in treatment response.[109] The most efficacious treatments will have little clinical utility if they are not palatable to the patients. To assess patients' attitudes toward insomnia treatments, Morin and colleagues[1,110] developed the Insomnia Treatment Acceptability Scale. This scale asks respondents to rate their opinions of CBT-I versus a pharmacologic intervention for insomnia. Two different studies conducted with this questionnaire showed that community-dwelling adults with insomnia rated a cognitive-behavioral intervention as more acceptable, more suitable for sleep difficulties, more likely to be effective long term, less likely to produce side effects, and more likely to improve daytime functioning than a pharmacologic intervention for insomnia.[110,111] In a study of older adults with insomnia, those randomized to receive CBT-I reported greater treatment satisfaction than those receiving pharmacotherapy.[112] A survey of hospital patients taking benzodiazepines for insomnia showed that 67% of those surveyed reported they would try an alternative treatment if available, and most (82%) viewed these alternatives as "healthier" than benzodiazepines.[113]

A few studies have compared patient preferences for the various therapeutic components typically composing CBT-I

approaches. In one such study, insomnia patients rated their satisfaction with sleep hygiene, stimulus control, sleep restriction, cognitive therapy, relaxation training, and medication tapering after participation in a group CBT-I treatment.[111] Sleep hygiene was rated the most liked and most useful element, whereas sleep restriction was rated lowest. Interestingly, sleep restriction but not sleep hygiene was related to symptom improvement. In another investigation comparing sleep hygiene, meditation, and stimulus control as treatments for chronic insomnia, participants in the sleep hygiene group were the least satisfied with their treatment despite similar sleep improvements in all three groups.[114] Hospitalized patients rating various insomnia therapies report the strongest preferences for sleep hygiene and relaxation, with sleep restriction and stimulus control holding the least appeal.[113]

Collectively, these findings imply that most patients prefer psychological treatments to medications. Some prefer treatment components that are not necessarily the most efficacious ones. Sleep restriction seems to be unpopular with patients, perhaps because of its counterintuitive nature and the enhanced sleepiness it may cause, particularly during its initial stages.[115] Implementation of sleep restriction may require a particular focus on the therapeutic rationale to foster patients' acceptance. It also may be that original sleep restriction approaches, which limit time in bed to the average total sleep time derived from a baseline sleep diary, may be too restrictive, especially for those patients who grossly underestimate their actual sleep time. For such patients, more lenient approaches proposed by some authors may be more acceptable and effective.[2]

Treatment Adherence

It is not difficult to appreciate the central importance of treatment adherence in ensuring the success of therapies like CBT-I. These treatments require considerable commitment from the patient, and adherence is recognized as one of the most critical factors affecting treatment success.[78,116-118] It has been estimated that approximately 15% of patients who enroll in CBT-I treatment studies fail to complete a full course of such treatment[119]; however, this rate may approach 40% among those who seek treatment in clinical settings.[119,120] Factors linked to nonadherence and attrition include greater sleep impairment, poorer perceived general health, higher levels of depression, less favorable ratings of the treatment, lower levels of support from spouses or partners, and viewing the therapist as critical or confrontive.[111,119-124] Most findings have been based on adherence indicators derived from sleep diaries, although other tools (e.g., actigraphy, patient checklists) can provide supplemental information and sometimes have been used for this purpose.

Studies reporting adherence to CBT-I in research participants are encouraging,[58,125-127] but less favorable outcomes are reported in clinical samples.[117] The percentage of patients reporting continued adherence to treatment recommendations at long-term follow-up ranges from 20% to 74%.[128,129] Higher adherence is related to better treatment outcomes.[126]

Unfortunately, there are few evidence-based data on promoting adherence. One study showed a benefit of modafinil in adherence to CBT-I[130] mainly through the role of this medication in offsetting the enhanced daytime sleepiness resulting from sleep restriction. Suggestions for promoting adherence have included providing the treatment rationale, discussing adherence directly with the patient, enlisting the support of the patient's family, providing encouragement, soliciting the patient's commitment through a verbal or written contract, encouraging questions from the patient, understanding the patient's expectations, and setting realistic goals.[1,117,131] Adherence strategies targeting specific treatment components are described elsewhere.[1,2]

APPLICATIONS TO PATIENTS WITH COMORBID PSYCHIATRIC CONDITIONS

Insomnia is highly prevalent as a coexisting complaint among those with various psychiatric disorders.[132,133] Yet most studies testing the efficacy of CBT-I among patients with psychiatric comorbidities have been published in the past decade or so. This is true because traditionally such insomnia was viewed as a symptom of the comorbid condition. Hence, its management required successfully treating the suspected "primary" psychiatric condition with the expectation that insomnia remission would follow. However, this view ignores findings suggesting that (1) psychiatric patients presenting with chronic insomnia are more difficult to treat than those presenting without insomnia[134,135]; (2) a sizable group of people continue to suffer from insomnia as a clinically significant residual symptom after remission from their so-called primary psychiatric disorder[136-140]; and (3) there is also little evidence to regard insomnia that co-occurs with a psychiatric disorder as categorically distinct from insomnia without a co-occurring disorder because they have the same purported insomnia-maintaining factors targeted by the cognitive and behavioral interventions composing CBT-I as do those with patients presenting solely with an insomnia disorder (e.g.,[141,142]). Hence, insomnia can be regarded as a comorbid disorder when it is observed in patients with concurrent psychiatric disorders.[55] In fact, this change in thinking is reflected in both the newest diagnostic nosologies[143,144] of the American Academy of Sleep Medicine and the American Psychiatric Association, which eliminate the distinction between "primary" and "secondary" insomnia in favor of a more global insomnia disorder diagnosis that can be assigned to patients with and without sleep-disruptive comorbidities.

Given that those with an isolated insomnia disorder and those suffering from insomnia comorbid to a psychiatric disorder have many shared perpetuating factors, we might expect a cognitive-behavioral conceptualization of comorbid insomnia to look something like the model in Figure 86-6. In this conceptualization, stress may be an endogenous event (e.g., a change in neurochemical activity) or a life event (e.g., the birth of a baby). Whether the stressor is an external life event or a physiologic event, it can directly cause either sleep disruption or the comorbid disorder. Alternatively, the stressor may operate through a moderator variable (i.e., a shared vulnerability trait) of insomnia and the comorbid condition. Several potential moderator variables exist, including ruminative tendency[145,146] and specific maladaptive beliefs.[47,51] Once sleep disruption occurs, attempts at coping or compensating for the sleep loss result in behaviors such as sleeping in, increasing time in bed, using alcohol or sleep aids, or avoiding obligations or attending social events. Such behaviors may be linked to circadian or homeostatic disruption, increased arousal, or reinforcement of beliefs that one is helpless to cope with sleep loss. These factors have been posited as perpetuating etiologic

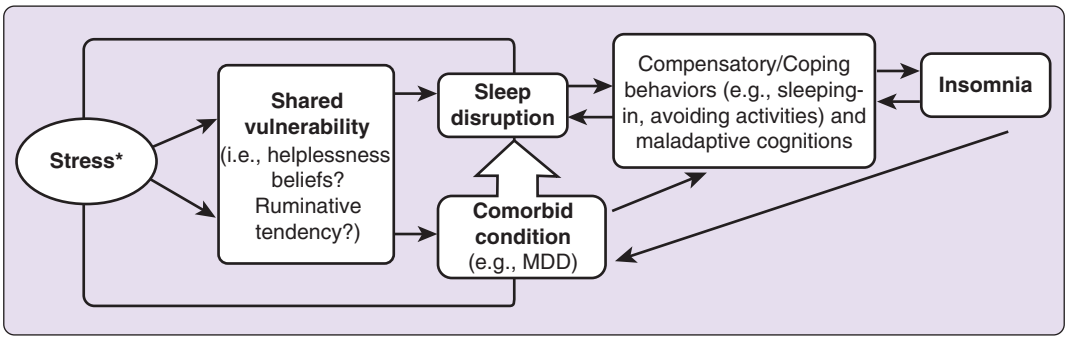


Figure 86-6 Possible etiologic pathways for comorbid insomnia. *Stress may be in the form of an internal (e.g., hormonal change) or external (e.g., interpersonal conflict) event that can directly cause either sleep disruption or a comorbid condition. Compensatory behaviors or arousing cognitions are thought to play a role in causing or maintaining the insomnia. The stressor may also operate through a moderator variable (i.e., a shared vulnerability trait) of insomnia and the comorbid illness. MDD, Major depressive disorder.

factors in insomnia.[1,29,47,147-149] Insomnia would be expected to exert negative influence on the comorbid disorder and in some cases could cause the disorder. Thus, even in cases in which a comorbid disorder may have initially caused the sleep disruption, cognitive and behavioral factors can become perpetuating factors for insomnia that remains after the comorbid condition remits.

To date, there have been a limited number of studies investigating CBT-I among those with comorbid psychiatric conditions.[64,132,150] In several studies, CBT-I has been tested in those with insomnia and depression, and results showed improvements in both sleep and depression symptoms.[151-153] Moreover, some studies have reported significant and clinically meaningful improvements in depression resulting from CBT-I alone.[152-155] Although this is a somewhat surprising finding, these studies tended to enroll small participant samples, so large-scale studies are needed to further investigate whether insomnia therapy alone can markedly improve major depressive disorder. Manber et al[156] reported that augmentation of antidepressant medication with CBT-I produced a far greater remission rate (depression remission rate, 61.5%) than antidepressant therapy alone (depression remission rate, 33%); Carney et al[155] reported the same results in a replication of the randomized controlled trial of Manber and colleagues. Thus, CBT-I is an effective treatment for insomnia in those with comorbid depression; there also appear to be significant benefits to mood by treating insomnia as well.[157]

In addition to the promising results found with depressed patients, CBT-I applications to patients with other psychiatric comorbidities have been promising as well. For example, CBT-I has been used effectively in those with PTSD[158-161] and in those in recovery from alcohol abuse.[89,162] Elements of CBT-I have also been incorporated into an effective nocturnal panic disorder treatment protocol.[163] Given the promising results thus far, much more research is needed to test the applicability and efficacy of CBT-I in patients with other psychiatric conditions.

Most previous studies support the use of standard CBT-I protocols among patients with psychiatric comorbidities. However, as noted by Smith et al,[164] there may be unique delivery issues among comorbid populations that require special alterations of or additions to CBT-I to optimize insomnia treatment outcomes. To date, there has been an absence of studies conducted to test this assumption, and available evidence has only partially supported this idea thus far (e.g.,[120]). Nonetheless, clinical observations suggest that certain subgroups of psychiatric patients present special challenges that may call for some alterations or augmentations of typical CBT-I protocols. For example, it has been noted that depressed patients often show little motivation to face each new day. As such, they have difficulties in getting out of bed in the morning at a scheduled time as required by the stimulus control instructions.[2] For such patients, it may be useful to construct a morning routine that includes attractive breakfast items or comforting activities that invite the patient out of bed at the preplanned time. In contrast, many patients with PTSD, particularly those with military combat experiences, show heightened safety concerns that cause them to delay bedtime by repeatedly checking all the doors and windows in their homes to ensure they are locked and secure before they retire to bed. For such patients, the addition of strategies such as response prevention or placing an agreed on limit on checking behaviors may be useful to prevent this compulsive behavior from interfering with sleep onset. Whereas the sort of CBT-I alterations required may seem obvious from a clinical viewpoint, admittedly more research is warranted to better elucidate the sorts of therapeutic alterations needed to produce optimal responses in patients with various types of psychiatric comorbidities.

APPLICATIONS TO PATIENTS WITH COMORBID MEDICAL CONDITIONS

There is growing evidence that CBT-I is effective for people with comorbid medical conditions including cancer, chronic pain, autoimmune diseases, and human immunodeficiency virus infection.[164] Adaptations may include adding components to address disease-specific symptoms, such as fatigue and pain, or using psychological and behavioral principles to address comorbidity factors that could interfere with adherence to the insomnia treatment.[164] In some studies, researchers adapted standard CBT-I to the specific population studied. For instance, in the context of breast cancer, some investigators have tested protocols that include cognitive and behavioral strategies to directly address fatigue.[165,166] However,

others tested unaltered versions of CBT-I with cancer patients[167] and reported insomnia symptom improvements. Research concerning use of CBT-I for insomnia in patients with medical comorbidities has not directly tested the utility of specific treatment adaptations. However, based on phenomenologic studies and our collective clinical experience, we discuss several issues that are relevant to the implementation of CBT-I for insomnia occurring in the context of medical diseases.

The Impact of Disease-Specific Symptoms on Insomnia

The sleep of patients with medical disorders can be disrupted by physical discomfort, such as pain, night sweats, cough, dyspnea, nocturia, and diarrhea. When such symptoms are disrupting sleep, the clinician needs to include in the treatment plan consultation with the medical team that is involved in the treatment of the comorbid medical condition. The most common symptom encountered among comorbid medical patients is pain. In addition to the direct negative impact of pain on sleep, patients' reactions to pain can sometimes contribute to insomnia indirectly. Examples of reactions to pain that can be detrimental to insomnia include spending much more time in bed at night than is actually needed for sleep, using the bed during the day for resting and napping, and using the bed and bedroom as a "headquarters" for daily activities. These reactions can contribute to insomnia by weakening the sleep drive or by rendering the bed a much weakened cue for sleep. These behaviors are also common reactions to the experience of fatigue, which like pain is common in several medical disorders. Some patients use large amounts of caffeine as a measure to counter fatigue,[168,169] a practice that might also interfere with sleep at night, particularly when the consumption is late in the day. In addition, because impaired sleep quality magnifies pain[170,171] and is associated with reduced energy the next day, patients often are concerned that a night of poor sleep will be followed by increased pain or fatigue on the following day. These concerns are often exaggerated (catastrophic cognitions) and contribute to insomnia by creating sleep-focused anticipatory anxiety.

Addressing Pain and Fatigue

Treatment components that are not included in standard CBT-I protocols for insomnia can be used to modify sleep-interfering behaviors associated with pain and fatigue. These include the following.

Increase Activities

Behavioral activation is an intervention that encourages increased engagement in life's activity. The treatment provider helps the patient identify obstacles to daytime activities and discusses how to overcome these obstacles. Behavioral activation could be easily integrated into CBT-I when pain and fatigue are associated with significant reduction in activities. Used in such situations, activation can help break the vicious circle connecting low level of activity to physical deconditioning, which further reduces activity level.[172] Exercise is a specific form of behavioral activation that has shown efficacy for pain management[173] and reduction of daytime fatigue.[174] Moderate exercise (55% to 75% of heart rate maximum) is safe and well tolerated by many patients with medical comorbidities, including cancer survivors. It effectively reduces cancer-related fatigue during and after various forms of cancer treatment (surgery, transplantation, chemotherapy, radiation therapy, and hormone therapy),[175] although effect sizes among cancer patients are small[176] to moderate.[177] In the context of pain, paced behavioral activation and moderate exercise are often included for pain management.[178]

Schedule Short Naps

Both stimulus control and sleep restriction instruct patients to avoid napping. This is to ensure a strong homeostatic sleep drive at night. However, for patients who find daytime fatigue and sleepiness overwhelming, careful timing of short daytime naps should be considered. Short (10-minute) afternoon naps can improve alertness and performance and do not seem to lead to a meaningful reduction in delta wave activity (an index of the homeostatic drive) in the subsequent night.[179] One advantage of scheduling a nap over unintentional dozing is that it reduces the sense of helplessness. Another advantage is that scheduled naps can reduce the need to cancel scheduled daytime activities because of overwhelming fatigue. Thus, taking short naps can be an effective strategy to combat fatigue and to improve quality of life. However, emphasizing the importance of keeping the naps short and earlier in the day (preferably before 3 PM) is important to ensure strong homeostatic pressure to sleep at bedtime.

Other Considerations for Patients with Chronic Pain

Hypervigilance to environmental stimuli, such as noise and uncomfortable temperatures, is common among chronic pain patients and predicts poor sleep quality.[164] A white noise–generating device can mask extraneous sleep-disruptive sounds and help pain patients maintain sleep. Likewise, encouraging proper temperature regulation to maintain a comfortable sleeping environment is often useful. In addition to these environment-focused strategies, person-focused techniques, such as relaxation training and cognitive distraction, are also useful for pain management in many patients.[178]

The Effects of Treatment for Comorbid Diseases on Insomnia

Treatments including certain antiretroviral medications, cancer chemotherapy, and radiation therapy can disrupt sleep and lead to increased fatigue. When disturbed sleep arises as a treatment-related side effect, development of chronic insomnia can be prevented by educating and preparing patients for this eventuality and discouraging the use of compensatory coping strategies that could perpetuate sleep problems. However, when disease-specific factors affect treatment adherence, they need to be addressed as part of the overall insomnia management strategy. As noted in the following discussion, unhelpful beliefs may need to be addressed to promote treatment adherence, or treatment strategies may need to be altered to prevent exacerbation of the comorbid medical condition.

Unhelpful Sleep-Related Beliefs

Research suggests that patients who have insomnia comorbid with fibromyalgia and other medical or mental disorders have higher scores on measures of dysfunctional beliefs and attitudes about sleep.[51] These beliefs are presumed to contribute to or to perpetuate the insomnia disorder of these patients. Thus, cognitive therapy might need to be emphasized and play a large role in managing insomnia with these patients.

Admittedly, the importance of targeting sleep-interfering thoughts when applying CBT-I with these populations has not been specifically evaluated. Nonetheless, as reductions in sleep-disruptive beliefs have been linked to sleep improvements in insomnia patients who lack sleep-disruptive comorbidities,[47,180,181] such cognitive changes would be expected to be critical to these comorbid insomnia patients as well.

Overconcern that insufficient sleep will worsen a medical disease, impede healing, or lead to a recurrence of the medical problem can increase worry about sleep and sleep-related performance anxiety. These concerns need to be addressed because they may lead to reluctance to restrict time in bed as indicated by the sleep restriction therapy protocol or to get out of bed when unable to sleep as indicated by stimulus control therapy. Helping patients challenge these beliefs by use of cognitive therapy and education can be useful to overcome treatment resistance in many patients with comorbid medical disorders.

Alterations to Standard Behavioral Instructions

Pain and fatigue can contribute to insomnia severity and chronicity and interfere with adherence to prescriptions to curtail time in bed. Because many insomnia patients underestimate their sleep times, their treatment with sleep restriction could result in actual sleep loss, which in turn could produce a hyperalgesic response and impair the functioning of their endogenous pain-inhibitory systems.[182] Such patients may respond better to sleep compression therapy,[183] which gradually shortens sleep opportunity during several weeks instead of drastically curtailing sleep at the onset of therapy. However, it has not been shown that sleep restriction therapy worsens pain or that sleep compression produces better outcomes with pain sufferers. Given the state of the science with respect to this issue, the best approach is to be cognizant of this potential difficulty and to use clinical judgment when implementing this component of treatment.

Addressing Psychological Reactions to the Comorbid Medical Disease

Patients' psychological reactions to their diseases, particularly depression and anxiety (including panic), are common[184-186] and can contribute to sleep disturbance in patients with medical comorbidities. It is therefore important to assess and to treat or refer patients who show such psychiatric symptoms when indicated.

USE WITH YOUNGER AND OLDER AGE GROUPS

Normal sleep and the experience of insomnia are affected by maturational and lifespan issues. Psychological and behavioral treatments for insomnia during childhood actively involve the parents and are based on principles of learning and behavior change, such as reinforcement, extinction, and shaping. There are many unique problems common in childhood sleep difficulties, including nighttime fears,[187] scary dreams,[187] resisting appropriate bedtimes, and needs for parental presence and special circumstances to aid sleep onset. For older children and teenagers, cell phones, computer tablets, and other forms of technology may provide sufficient late-night distraction to cause a delay in bedtimes and adherence to healthy sleep schedules. Such problems require special treatments that lie outside the focus of this chapter. For information about the management of such cases, the reader should consult the writings of Mindell et al[188] or Lewin.[189]

Older Adults

Reviews and meta-analyses have supported the efficacy of cognitive-behavioral therapies for insomnia in older adults.[58,62,190-194] Recent studies suggest that CBT-I is effective for older adults even when they have hypnotic dependence[195] and when the intervention is very brief (two sessions)[82] or delivered by telehealth.[196] Telehealth is a particularly attractive treatment modality for older adults who do not have easy access to in-person care. Several factors specific to older adults need to be considered in implementing CBT-I. These include comorbid medical diseases and life changes that can weaken the homeostatic and circadian regulation of sleep.[197] The latter set of factors includes decreased physical activity, social isolation, and reduced exposure to light, which are particularly relevant to residents of assisted living facilities and people with dementia, including Alzheimer disease. Effective treatment of sleep difficulties in older adults with dementia includes decreasing in-bed time during the day, increasing physical activity and social activation, ensuring daily sunlight (or artificial light) exposure and structured bedtime routines, and, in the case of institutional settings, decreasing nighttime noise and light and involving the caregivers.[198-200] Current evidence shows that increasing social activities improves nocturnal sleep[200] and introducing structured social activities increases slow wave sleep and improved memory.[197] On the other hand, a review of the literature on bright light therapy for older adults with dementia concluded that there is no adequate evidence of the effectiveness of this therapy in managing sleep.[201] Low-impact aerobics, brisk walking, and tai chi all seem to improve sleep of sedentary older adults with moderate sleep disturbance.[202,203] Yet little is known about exercise or tai chi as a single-component intervention for older adults who meet criteria for insomnia disorder. One small study of 17 individuals found that aerobic activity combined with sleep hygiene education is more effective than sleep hygiene alone for patients 55 years or older.[204] Finally, passive body heating (immersing the body in 40°C water for 30 minutes 1.5 to 2 hours before bedtime) has also been used successfully to improve sleep in older women[205,206] and in patients with vascular dementia.[207] This procedure is hypothesized to operate through its action on the thermoregulatory, circadian, and autonomic systems. Nonetheless, it remains to be determined whether such findings can be translated into practical suggestions (e.g., taking a warm bath in the evening) that can be used as a reliable method for managing insomnia complaints in older adults.

CLINICAL PEARL

Psychological and behavioral insomnia therapies have as their heritage a range of tools that aid in their implementation. These treatments can be delivered in a variety of formats ranging from high-intensity individual therapy to low-intensity self-help paradigms. They are effective with a variety of young and older insomnia sufferers who present with isolated and comorbid forms of insomnia disorders. However, alterations or augmentations of these treatments may be indicated for certain age groups and those patients with sleep-disruptive comorbidities.

SUMMARY

Psychological and behavioral insomnia therapies have a strong research legacy that supports their widespread use in clinical practice. Clinicians wishing to use these treatments have a number of useful tools that aid in their implementation, including sleep diaries, written behavioral prescriptions, structured cognitive exercises, and, for selected cases, actigraphy. These treatments can be delivered in different formats, in limited "doses," by various provider types, and in various self-help formats, but the involvement of a specialist in behavioral sleep medicine as either a consultant or therapist appears to enhance treatment outcome. The reach of these therapies can be expanded by telehealth and Internet delivery protocols. The treatments are useful not only for those who present solely with insomnia complaints but also for patients who present with a range of psychiatric and medical comorbidities. Moreover, these therapies have efficacy for the range of younger and older adult insomnia patients encountered in clinical practice. Future research will likely show how best to package these treatments for various clinical settings, varying types of therapy providers, and patients with comorbid conditions. Nonetheless, the guidance provided in this chapter gives clinicians a better understanding of tools that aid in the delivery of treatment and issues that merit consideration in their implementation.

Selected Readings

Ashworth DK, Sletten TL, Junge M, et al. A randomized controlled trial of cognitive behavioral therapy for insomnia: an effective treatment for comorbid insomnia and depression. *J Couns Psychol* 2015;**62**:115–23.

Buysse DJ, Germain A, Moul DE, et al. Efficacy of brief behavioral treatment for chronic insomnia in older adults. *Arch Intern Med* 2011;**171**:887–95.

Carney CE, Buysse DJ, Ancoli-Israel S, et al. The Consensus Sleep Diary: standardizing prospective sleep self-monitoring. *Sleep* 2012;**35**:287–302.

Conley S, Redeker NS. Cognitive behavioral therapy for insomnia in the context of cardiovascular conditions. *Curr Sleep Med Rep* 2015;**1**(3): 157–65.

Edinger JD, Carney CE. *Overcoming insomnia: a cognitive-behavioral approach—therapist guide*. New York: Oxford University Press; 2015.

Espie CA. "Stepped care": a health technology solution for delivering cognitive behavioral therapy as a first line insomnia treatment. *Sleep* 2009;**32**:1549–58.

Espie CA, Kyle SD, Williams C, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. *Sleep* 2012;**35**: 769–81.

Harvey AG, Belanger L, Talbot L, et al. Comparative efficacy of behavior therapy, cognitive therapy, and cognitive behavior therapy for chronic insomnia: a randomized controlled trial. *J Consult Clin Psychol* 2014;**82**: 670–83.

Ho FY, Chan CS, Tang KN. Cognitive-behavioral therapy for sleep disturbances in treating posttraumatic stress disorder symptoms: a meta-analysis of randomized controlled trials. *Clin Psychol Rev* 2015 Sep 25. pii: S0272-7358(15)00123-3.

Koffel EA, Koffel JB, Gehrman PR. A meta-analysis of group cognitive behavioral therapy for insomnia. *Sleep Med Rev* 2015;**19**:6–16.

Manber R, Edinger JD, Gress JL, et al. Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disorder and insomnia. *Sleep* 2008;**31**:489–95.

McCrae CS, Lichstein KL. Secondary insomnia: a heuristic model and behavioral approaches to assessment, treatment, and prevention. *Appl Prev Psychol* 2001;**10**:107–23.

van Straten A, Cuijpers P. Self-help therapy for insomnia: a meta-analysis. *Sleep Med Rev* 2009;**13**:61–71.

A complete reference list can be found online at ExpertConsult.com.

Chapter
87

Pharmacologic Treatment of Insomnia: Benzodiazepine Receptor Agonists

James K. Walsh; Thomas Roth

Chapter Highlights

- Considerable evidence indicates that the benzodiazepine receptor agonists (BzRAs) are efficacious in the treatment of insomnia, and somewhat limited data suggest a substantial degree of effectiveness.
- BzRAs are also generally safe as commonly used in clinical practice; common concerns regarding tolerance, misuse, and abuse by insomnia patients are not typical.
- BzRAs produce improvement in daytime symptoms as well as nighttime symptoms of insomnia and may reduce the severity of comorbid illnesses.
- Sufficient differences in pharmacokinetic profiles exist among the BzRAs to provide physicians with therapeutic options that have differing durations of action.
- Individual differences in metabolism and elimination of BzRAs have led to a lowering of recommended starting doses for some BzRAs.

A 2005 National Institutes of Health State of the Science Conference concluded that the benzodiazepine receptor agonists (BzRAs) were the only drugs for which adequate scientific evidence exists to support their use for chronic insomnia.[1] Subsequently, the U.S. Food and Drug Administration (FDA) has approved low-dose doxepin (Silenor) for the treatment of insomnia, and an orexin antagonist (Suvorexant) has recently been approved (see Chapters 42 and 88). The efficacy, effectiveness, and safety of the BzRAs approved for the treatment of insomnia by the FDA is the focus of this chapter. A number of other BzRAs are approved for the treatment of insomnia in other countries, and research with those drugs is highly consistent with the data reviewed here. In addition, implementation and assessment strategies for the pharmacologic treatment of insomnia are discussed.

The group of drugs referred to as "benzodiazepine receptor agonists" are categorized according to the well-described common mechanism of action of these medications. All BzRAs act as allosteric modulators of gamma-aminobutyric acid (GABA) activity by binding to inotropic benzodiazepine receptors at the $GABA_A$ receptor complex. BzRAs serve to increase GABA binding and thus the frequency of chloride ion channel openings, facilitating inhibitory activity. Some of these drugs (Table 87-1) have a benzodiazepine chemical structure (i.e., estazolam, flurazepam, quazepam, temazepam, triazolam), and others do not (i.e., eszopiclone, zaleplon, zolpidem). BzRAs demonstrate affinity for four benzodiazepine receptor subtypes (referred to as α_1, α_2, α_3, and α_5) located at the $GABA_A$ pentameric complex. Benzodiazepines tend toward comparable affinity for the four receptor subtypes, whereas the affinities of zolpidem and zaleplon for the α_1 subtype are higher than for other subtypes. Eszopiclone shows less preference for α_1, having some activity at α_2 or α_3 (or both). The functional significance of these binding differences for effects on sleep or other actions in humans remains unclear. In contrast, preclinical investigations of knock-in rodent strains suggest different functional characteristics of the benzodiazepine receptor subtypes.[2]

The most clinically relevant differences among BzRAs are associated with pharmacokinetic properties of the drugs, particularly terminal elimination half-life, which is the most important determinant of duration of drug action. Distribution characteristics (for single doses) and absorption also contribute to duration of action, although to a much lesser degree. Of course as with all biologic systems, substantial individual differences exist in the metabolism and elimination of BzRAs, with about a 30% intersubject variability that translates into individual differences in duration of action among patients and hence associated efficacy and safety (see recent related FDA dosing recommendations discussed later in the section titled Safety). Finally, factors other than pharmacokinetics also influence duration of action, in particular drug dose and formulation (e.g., extended release, sublingual).

Of the available BzRA hypnotics in the United States, the most rapidly eliminated drugs used at recommended doses in healthy adults have estimated durations of action in the range of 2 to 5 hours (zaleplon, sublingual zolpidem lozenge) and 4 to 7 hours (triazolam, zolpidem), whereas the slowly eliminated drugs (including active metabolites) have durations of action of 24 hours or more (flurazepam, quazepam). Estazolam, temazepam, and eszopiclone have intermediate durations of action.

A number of BzRAs are available in other countries for the treatment of insomnia. They include the short-acting midazolam and brotizolam, intermediate-acting zopiclone, loprazolam, lormetazepam, and long-acting flunitrazepam and nitrazepam. All are benzodiazepines except zopiclone, a cyclopyrrolone; eszopiclone, the active stereoisomer, is available in the United States.

Table 87-1 Characteristics of Benzodiazepine Receptor Agonists with an FDA Indication for Insomnia

Generic Name	Receptor Binding Specificity	Dose Range (mg)	Elimination Half-Life (hr)	Metabolism
Estazolam	Nonspecific	1–2	10–24	CYP3A
Flurazepam	Nonspecific	15–30	48–120*	CYP3A4
Temazepam	Nonspecific	15–30	8–20	None
Triazolam	Nonspecific	0.125–0.25	2–6	CYP3A4
Quazepam	Nonspecific	7.5–15	39–73*	Not available
Eszopiclone	$GABA_A$, $\alpha_{1,2,3}$	1–3	6	CYP3A4, CYP2E1
Zaleplon	$GABA_A$, α_1	5–20	1	Minor: CYP3A4
Zolpidem	$GABA_A$, α_1	5–10	1.5–2.4	CYP3A4, CYP2C9
Zolpidem extended-release	$GABA_A$, α_1	6.25–12.5	1.6–4.5	CYP3A4, CYP2C9
Zolpidem sublingual tablet	$GABA_A$, α_1	5–10	1.5–2.4	CYP3A4, CYP2C9
Zolpidem oral spray	$GABA_A$, α_1	5–10	1.5–2.4	CYP3A4, CYP2C9
Zolpidem sublingual lozenge	$GABA_A$, α_1	1.75–3.5	1.5–2.4	CYP3A4, CYP2C9

*Refers to elimination half-life of active metabolite.
CYP, Cytochrome P-450 (letters and numbers refer to specific CYP enzymes); $GABA_A$, gamma-aminobutyric acid, type A receptor complex.

The BzRAs are generally recommended as first-line hypnotics for several reasons. All BzRAs have been shown to be efficacious, although there are some differences among drugs, associated with pharmacokinetic properties, and marketed dose. Compared with other central nervous system (CNS) drug classes, the margin of safety or therapeutic index (i.e., the effective dose relative to lethal dose) is wide.[3] For example, barbiturates have margins of safety on the order of two to four times the effective dose, whereas for BzRAs, the margin of safety can be as great as 100. Abuse of or dependence on BzRAs is uncommon in the therapeutic context, most likely accounted for by the relatively mild reinforcing effects and self-administration patterns. Summaries of the efficacy, effectiveness, and safety characteristics of BzRAs follow. Unless otherwise stated, the summary material refers to the FDA-approved dosages of the drugs.

EFFICACY AND EFFECTIVENESS

Traditional measures of hypnotic efficacy include polysomnographic (PSG) or patient estimates of induction and maintenance of sleep. Sleep latency (or latency to persistent sleep), whether given by PSG or self-report, is the standard sleep induction variable and is thought to be relevant to a patient report of difficulty falling asleep. Number of awakenings and wake after sleep onset (WASO) are the most common sleep maintenance measures, corresponding to patients' reports of difficulty staying asleep. Total sleep time and sleep efficiency reflect both sleep induction and sleep maintenance properties. Qualitative measures of efficacy such as morning ratings of sleep quality, sleep depth, or global impression ratings performed by either the clinician or the patient are also used in efficacy trials. Given the variability of sleep across individual nights, efficacy is evaluated by averaging across nights and in many cases across a week. Importantly, patient-reported data are typically collected using diaries or questionnaires with an electronic time stamp to determine the time the evaluation was completed.

Given the desire to maximize objective data and to minimize costs, the monitoring of movements or activity with a wrist-worn sensor (actigraphy) has also been used to evaluate hypnotic efficacy. However, to date, actigraphy has been shown to be sensitive only to timing and duration of sleep. This makes actigraphy more valuable in evaluating efficacy in insomnia that is comorbid with circadian rhythm disorders (e.g., phase-delay syndrome).

Primary Insomnia

Most insomnia clinical trials have documented the efficacy of hypnotics using patients' reports, PSG, or both in patients with primary insomnia. A meta-analysis of clinical trials with benzodiazepines and zolpidem found that these drugs in aggregate produce reliable improvements in the sleep of persons with chronic insomnia. Importantly, the median duration of the studies included in the analysis was only 1 week.[4] Other meta-analyses[5-8] largely concur regarding short-term efficacy. However, the utility of meta-analyses to assist the clinician is fairly low because they typically combine data from several different drugs with widely different pharmacokinetic properties, which can affect the outcome variables examined in the meta-analysis. Additionally, two or more doses of a drug may be included in the analysis. It is much more instructive to examine the strengths and weaknesses of individual drugs as opposed to generalizing across all drugs in a category. Simply put, meta-analyses tell us about the pharmacology of a group of drugs, but they are not informative about the properties of individual drugs or what the best drug is for a given patient. Below, important differences among drugs will be emphasized, followed by a discussion of similarities among BzRAs.

At appropriate doses all BzRA hypnotics reduce sleep latency, and most increase total sleep time. The exception is zaleplon, which does not reliably increase total sleep time. The reduction of sleep latency is attributable to a rapid onset of hypnotic effect. The longer the drug's duration of action (i.e., the longer the half-life or the higher the dose, or both), the

more likely it is that the drug will show efficacy on sleep maintenance properties such as number of awakenings or WASO. At doses within the therapeutic range, a dose-response effect is observed on efficacy measures, and little additional effect is generally apparent at supratherapeutic doses. As might be expected, sleep latency shows a relatively flat dose-response curve, and WASO and total sleep time show more pronounced dose-response curves, related to increased duration of action with increased doses.

Tolerance is defined as a reduction of a drug's effect with repeated administration of a constant dose or the need to increase the dose to sustain a specific level of effect. Despite frequent speculations in the medical literature, most studies do not show tolerance to the hypnotic effects of BzRAs in most subjects, at least for therapeutic doses and for the periods of time that have been studied. Investigations that are often cited as evidence for developing tolerance[9] show gradual improvement in sleep over time in the placebo group in the face of a constant effect in the drug group, resulting in loss of statistical significance between groups. Yet the within-group comparison of an early drug effect with a late drug effect fails to show any change in drug activity with repeated use. Thus it cannot be concluded that tolerance has developed, but rather that unspecified changes occur over time with placebo, such as spontaneous remission, regression to the mean (if there are sleep-disruption criteria for entering the study), sleep habit influences inherent in protocol adherence, Hawthorne effects, and true placebo effects.[10] This is further evidenced by the fact that patients rarely escalate the nightly use of hypnotics, even with long-term use.[11]

Nearly 30 years ago, Oswald and colleagues[12] reported that lormetazepam and nitrazepam, two benzodiazepines available in some European countries, retained their effect by some patient estimates of hypnotic efficacy during 24 weeks of nightly use. More recently, in rigorous PSG studies, zolpidem 10 mg and zaleplon 10 mg retained efficacy for 5 weeks of nightly use.[13,14] A 2003 landmark study of several hundred patients with primary insomnia showed continued hypnotic efficacy of eszopiclone for 6 months of nightly use.[15] Patient reports of sleep latency and WASO were significantly reduced with eszopiclone 3 mg compared with placebo at each monthly time point. Total sleep time, number of awakenings, and sleep quality were also better than with placebo at each monthly time point. These 6-month findings have been replicated by Walsh and colleagues[16] (Figure 87-1). Examination of open-label extension data of the 2003 6-month eszopiclone study suggested sustained efficacy for a total of 12 months.[17] Other evidence for sustained efficacy of nightly BzRA use includes a 3-month trial with indiplon, a nonmarketed hypnotic.[18]

Efficacy of non-nightly use of zolpidem 10 mg has been investigated for up to 12 weeks.[19,20] In these studies, ratings of sleep latency, total sleep time, number of awakenings, and sleep quality were all improved on nights when zolpidem was taken (an average of three or four nights per week) compared with placebo (also an average of three or four nights per week). Additionally, investigator global ratings, which considered both medication nights and no-medication nights, indicated reduced insomnia severity with zolpidem. Total sleep time data on nights when a pill (either zolpidem or placebo) was taken or not taken during an 8-week study are shown in Figure 87-2. Importantly, an examination of sleep on nights when no drug was taken immediately after a zolpidem night failed to reveal evidence of rebound insomnia.[14] A 6-month study with an extended-release formulation of zolpidem used three to seven nights per week has generated similar results over the duration of the study.[21]

More recently, a sublingual lozenge formulation (Intermezzo) was approved for the novel indication of middle-of-the-night awakening and difficulty returning to sleep.[22]

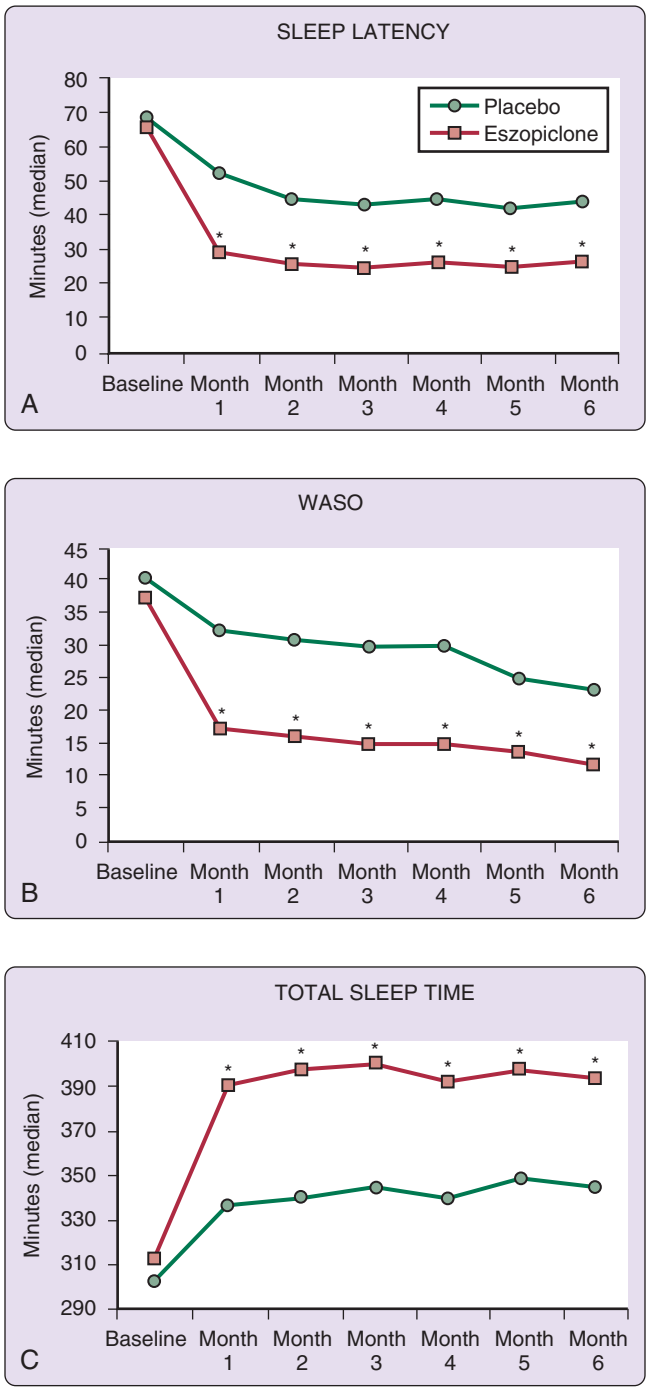


Figure 87-1 Median values for key hypnotic efficacy measures at baseline and during months 1 to 6 of double-blind treatment for eszopiclone (*red*) and placebo (*green*) groups. **A,** Sleep latency. **B,** Wake after sleep onset (WASO). **C,** Total sleep time. *, Change from baseline *P* value versus placebo < .0001. (Modified from Walsh JK, Krystal AD, Amato DA, et al. Nightly treatment of primary insomnia with eszopiclone for six months: effect on sleep, quality of life and work limitations. *Sleep* 2007;30:959–68.)

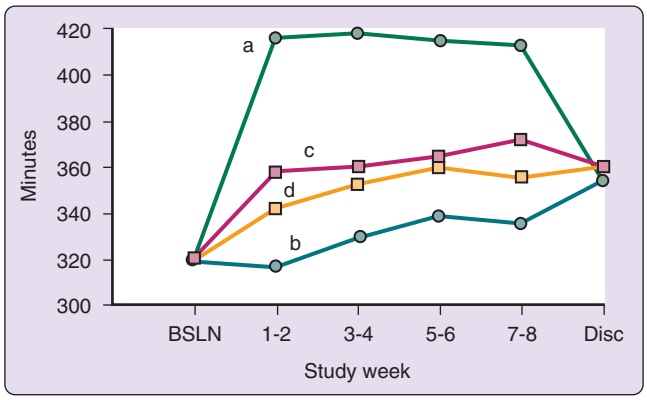


Figure 87-2 Mean subjective total sleep time for 8 weeks: (a) Zolpidem pill nights (*green line* and *green circles*); (b) Zolpidem no-pill nights (*blue line* and *blue circles*); (c) Placebo pill nights (*red line* and *red squares*); and (d) Placebo no-pill nights (*yellow line* and *yellow squares*). $P < .001$ for a versus c and for a versus b, each at study periods 1-2, 3-4, 5-6, and 7-8 weeks. BSLN, baseline (both groups receive placebo, double-blind); Disc, discontinuation week during which both groups received placebo, double-blind. (Modified from Walsh JK, Roth T, Randazzo AC, et al. Eight weeks of non-nightly use of zolpidem for primary insomnia. *Sleep* 2000;23:1087–96.)

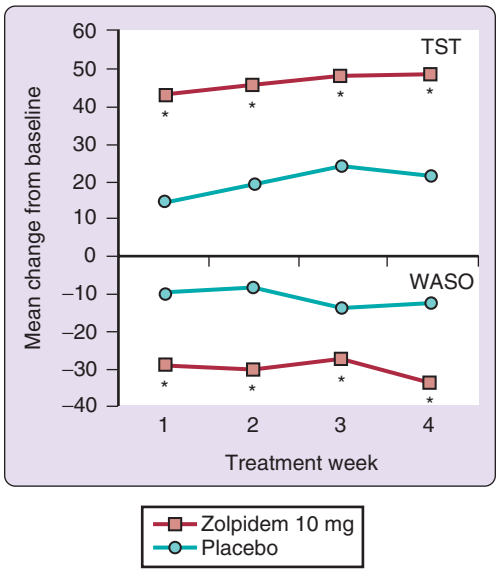


Figure 87-3 Mean weekly change from baseline values for reported total sleep time (TST) (*upper panel*) and wake after sleep onset (WASO) (*lower panel*) in patients with clinically significant insomnia despite effectively treated depressive disorders (treated with fluoxetine, sertraline, or paroxetine). Patients were randomized in double-blind fashion to either zolpidem 10 mg or placebo. *, Indicates the weekly zolpidem value is significantly different from the corresponding placebo value; $P < .05$. (Modified from Asnis GM, Chakraburtty A, DuBoff EA, et al. Zolpidem for persistent insomnia in SSRI-treated depressed patients. *J Clin Psychiatry* 1999;60:68–76.)

Dosages are 1.75 mg for nonelderly women, 3.5 mg for nonelderly men, and 1.75 mg for elderly people. Other zolpidem formulations intended for bedtime use with indications for sleep-onset difficulty include both sublingual (Edular) and oral spray (Zolpimist) forms.

Comorbid Insomnia

Comorbid insomnia refers to insomnia coexistent with other medical or psychiatric conditions. Although BzRAs have not been extensively studied in patients with comorbid insomnia, published data indicate improvements in sleep similar to those seen with primary insomnia. One of the earliest studies relevant to BzRA efficacy in comorbid insomnia examined rheumatoid arthritis patients with insomnia using PSG techniques. Triazolam 0.25 mg was found to reduce sleep latency and number of awakenings and to increase total sleep time across six nights of treatment.[23]

Zolpidem 10 mg increased total sleep time and improved sleep quality in a 4-week placebo-controlled study in depressed patients with refractory insomnia despite adequate control of depressive symptoms with selective serotonin reuptake inhibitor (SSRI) medication.[24] Figure 87-3 contains total sleep time data and wake after sleep time values during that 4-week study. Extended-release zolpidem coadministered with escitalopram improved subjective measures of sleep in patients with comorbid depression[25] and in those with comorbid anxiety,[26] but there were no changes in symptoms of depression or anxiety.

A number of studies have been conducted to assess efficacy of eszopiclone in comorbid insomnia. Eszopiclone combination therapy with fluoxetine was superior to fluoxetine and placebo in improving sleep in patients with newly diagnosed major depressive disorder and comorbid insomnia.[27] In addition, the combination-therapy group was found to have a significantly greater reduction in depressive symptoms, a higher rate of remission, and a faster time to the onset of the antidepressant effect. Similarly, eszopiclone co-therapy with escitalopram was superior to escitalopram plus placebo in improving the sleep of patients with insomnia and generalized anxiety disorder,[28] and anxiety and depressive symptoms were reduced further in patients receiving co-therapy. Insomnia patients treated with eszopiclone 3 mg for 4 weeks during perimenopause or early menopause reported significantly improved sleep throughout the trial, but in addition they had significant improvements in mood, quality of life, and menopause-related symptoms (Figure 87-4).[29] Finally, addition of eszopiclone to a standardized naproxen pain regimen significantly improved sleep, pain, and depression in patients with chronic low back pain.[30]

Effectiveness Studies

Studies of treatment efficacy enroll a selected homogenous sample using a variety of exclusion criteria (e.g., for concomitant medications, many medical conditions). Efficacy trials use highly controlled methods, compare treatment effects to placebo or other active treatments, and typically examine a limited number of outcome variables. In contrast, investigations of treatment effectiveness include patient samples that are much more representative of clinical practice (few exclusion criteria), and the degree of experimental control is minimal. Often no control group is used, and the outcome focus is not only on efficacy measures but also on treatment adherence, patient satisfaction, and quality of life. Hypnotic effectiveness trials, per se, have not been carried out. However, some reports provide information on hypnotic use in the general population that might approximate reports of clinical investigations of effectiveness.

Using an epidemiologic methodology, Ohayon and colleagues[31] reported that of 532 patients describing chronic insomnia and long-term use of hypnotic medications to help them sleep, 67% rated their sleep quality as improved "a lot,"

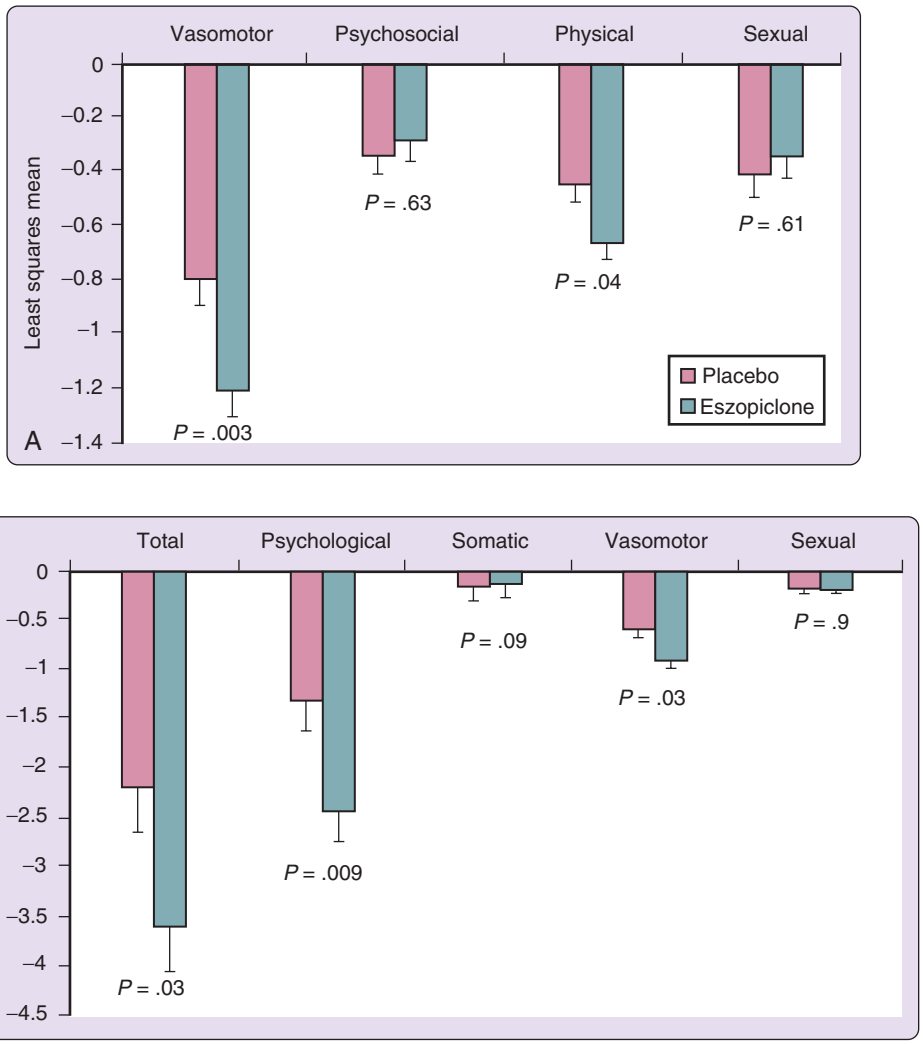


Figure 87-4 Menopause outcomes after 4 weeks of treatment with eszopiclone 3 mg. **A,** Change in menopause-specific quality of life at week 4. **B,** Change in Greene Climacteric Scale at week 4. Data presented are least square means ± standard error of the means. *P* values reflect change from baseline using analysis of covariance. (Modified from Soares CN, Joffe H, Rubens R, et al. Eszopiclone in patients with insomnia during perimenopause and early postmenopause: a randomized controlled trial. *Obstet Gynecol* 2006;108:1402–10).

and only 14.4% reported little or no improvement with medication. Balter and Uhlenhuth[32] interviewed subjects who in the past year either reported significant trouble with insomnia or had taken a medication to help them sleep. They found very high satisfaction rates among users of hypnotics when they asked, "Taking into account both the positive effects on your sleep and daytime functioning and any negative effects you may have experienced, would you take this medication again for the same purpose?" Affirmative responses were received from 84% of those taking triazolam, 82% for flurazepam, 74% for temazepam, and 61% for over-the-counter aids. In summarizing the findings of this intensive interview study, the authors concluded that "the benefits of treatment with prescription hypnotics are substantial and far outweigh the risks as defined in this study." It is important to recognize that in both of these studies, the cited data were obtained from either current users of sleep-promoting medications or those who had used them at some time in the past year. In these studies, the severity or nature of the insomnia was not characterized, nor were the duration, dose, or pattern of hypnotic use or other aspects of the patients. Clearly, effectiveness data in the management of insomnia need to be an important future research priority.

Long-term open-label studies conducted in both adult and elderly subjects in outpatient settings also provide some information about effectiveness.[17,33–36] Zolpidem, zaleplon, and eszopiclone have been evaluated over periods of 6 to 12 months in such studies. In general, patients and physicians report sustained benefit of BzRAs for the duration of the studies, without adverse reactions unique to long-term use. The generalizability of results and the confidence level regarding any conclusions drawn must be tempered by the self-selection of those remaining in these open-label studies.

Impact on Daytime Symptoms of Insomnia and Comorbid Conditions

Diagnostic criteria for insomnia include some form of subjective daytime impairment perceived to be consequent to the sleep disruption, such as fatigue, sleepiness, concentration and memory problems, and mood changes. If the daytime

impairment is mediated by insufficient or disturbed sleep, it should be possible to demonstrate a reduction of impairment with improved sleep. Elimination or reduction of the daytime problem could be viewed as another measure of treatment efficacy. Until recently, few studies have shown daytime consequences of insomnia, especially on laboratory-based tasks of psychomotor and cognitive performance. There is growing evidence, however, that in addition to the perceived daytime dysfunction, insomnia patients have a substantially increased risk for future mood and other psychiatric disorders,[37-40] higher rates of health care use and health care costs,[41-44] elevated risk for falling,[45-47] reduced quality of life,[48,49] and increased absenteeism from work.[41,50,51] Cognitive impairments are reported less often, but recent investigations have differentiated insomniac subjects from normal sleepers.[52,53]

Until recently, the focus of research on the daytime effects of BzRAs used to treat insomnia was the assessment of residual sedation rather than relief of daytime symptoms or other consequences. However, the few studies that have systematically assessed patient-reported daytime measures generally show improvement. For example, in the two 6-month eszopiclone studies of patients with primary insomnia previously discussed,[15,16] patients' reports of daytime alertness, ability to function during the daytime, and physical sense of well-being were all significantly better in the eszopiclone group than in the placebo group.

The waking impact of chronic comorbid insomnia may differ from that of primary insomnia. For example, patients with insomnia and periodic limb movement disorder[54] and those with insomnia and rheumatoid arthritis[55] have been shown to have lower than optimal daytime alertness as determined by Multiple Sleep Latency Test (MSLT) scores, which improve significantly after six nights of treatment with triazolam. Three studies of older adult patients showed increased daytime alertness on the MSLT resulting from increased sleep consolidation and duration with BzRA treatment.[56-58] McCall and colleagues[59] have recently examined health-related quality-of-life metrics in patients with a major depressive episode with insomnia. Combined treatment with fluoxetine and eszopiclone, compared with fluoxetine alone, improved health-related quality-of-life, sleep, and depression severity. In part, the difficulty in finding systematic daytime impairment in chronic insomnia may be the result of differing daytime consequences in subgroups of insomnia patients. More systematic and standardized assessments of daytime function are also needed in hypnotic trials.

SAFETY

In general, BzRAs are well tolerated, with few significant safety concerns. For example, adverse reactions recorded in clinical trials or in clinical practice are seen in the minority of patients, the severity is most commonly rated as mild, and reactions rarely result in abandoning the study or treatment. For inpatient use in a large academic hospital, the median rate of adverse events, across all hypnotics, was found to be about 1 in every 10,000 doses.[60] Data on the incidence of abuse of hypnotic medications, independent of other sedatives, is not available. However, a study in Switzerland in the early 1980s indicated that abuse of all benzodiazepines (not limited to use as a hypnotic) occurred at the rate of 2 per 10,000 prescriptions.[61] A population-based study in the United States indicated that 2.3% of the population reported nonmedical use of sedatives and tranquillizers, and approximately 0.002% of the population met criteria for substance abuse.[62] BzRAs used in COPD patients may increase the risk of respiratory failure.[62a]

The adverse events associated with BzRAs are related to peak plasma concentration (e.g., amnesia, ataxia), which is highly associated with dose, and to duration of action (e.g., residual sedation), which is primarily determined by elimination half-life and dose. Specific safety concerns for hypnotics and their potential mediators are discussed later.

Residual Effects

Many of the side effects associated with BzRAs are mediated by their desired pharmacologic activity, sedation.[63] Residual sedation, which is a prolongation of the hypnotic effect of the drug into the wake period, results in adverse reactions such as drowsy feelings, sleepiness, and impairment in psychomotor performance and driving. The likelihood of residual sedation is determined by the duration of drug activity, which in turn is determined by the elimination half-life and the dose of the drug. Many studies using the MSLT, performance assessments, and simulator or on-road driving tests have shown differences in residual effects between short- and long-acting drugs and between different doses of the same drug.

In the past several years, the FDA has become increasingly concerned about the residual effects of hypnotics,[64] especially as they pertain to driving impairment. Specifically, there is a concern for individuals that metabolize drugs more slowly and hence are at a greater risk for residual sedative effects in the morning. Therefore, the FDA has recommended that clinicians start hypnotic treatment at lower doses and titrate up after a patient has experienced the medication' effects in the morning. As a specific example, the FDA recently released a safety announcement stating: "The U.S. Food and Drug Administration (FDA) is warning that the insomnia drug Lunesta (eszopiclone) can cause next-day impairment of driving and other activities that require alertness. As a result, they have decreased the recommended starting dose of Lunesta to 1 mg at bedtime."

The FDA has also reduced the maximum therapeutic dose of all zolpidem products for women based on sex-related pharmacokinetic differences for zolpidem. At any given dose, women will have higher AUC values than men.[65,66] The mechanism of the gender difference in zolpidem pharmacokinetics is not understood. Importantly, at this time it is not clear whether these pharmacokinetic differences translate into differential efficacy or safety. There are no studies reporting gender differences in efficacy. Other reports of no pharmacodynamic differences[67] require cautious interpretation because these studies were not statistically powered to examine gender differences.

Amnestic Effects

Another adverse effect that is, in part, related to the sedative effects of hypnotics is anterograde amnesia. Anterograde amnesia is memory failure for information presented after administration of the drug. It is associated with all sedative-hypnotics, including all the BzRAs, alcohol, and barbiturates. The extent of amnesia is related to the plasma concentration of the drug. That is, the proximity of information input to peak

plasma concentration determines the degree of amnesia, and higher doses, which increase plasma concentration, are associated with both a greater degree of amnesia and a higher prevalence of amnestic events.[68,69] Furthermore, maintaining wakefulness for 10 to 15 minutes after presentation of memory material, rather than allowing a drug-induced rapid sleep onset, attenuates the amnesia.[70] It is important to recall that sleep itself has amnestic effects. It is not uncommon for people to wake in the middle of the night, and if they fall back to sleep rapidly (within about 5 to 7 minutes), to not remember the awakening. Thus the promotion of rapid sleep onset and a direct drug effect combine to produce the risk for amnesia. There have been reports of global amnesia or traveler's amnesia.[71] These reports are anecdotal and describe persons forgetting long periods of wakefulness after ingestion of a BzRA. These reports tend to be associated with higher doses, drug taken in combination with another CNS depressant drug or alcohol, or drug taken during a period of sleep deprivation.

Discontinuation Effects

The most commonly cited discontinuation effect of BzRA hypnotics is rebound insomnia. Rebound insomnia is defined as a worsening of sleep relative to the patient's status before starting treatment. Unlike other discontinuation effects, rebound insomnia only lasts for one or two nights after the hypnotic is discontinued. Rebound insomnia can occur even after one or two nights' hypnotic use[72] and does not increase in severity with the number of repeated nights of use, at least within the time frame of a few weeks of nightly use. Rebound insomnia is more likely to occur after high doses of short- and intermediate-acting BzRAs. It does not typically occur with long-acting drugs because of the gradual decline in plasma concentration that is inherent in the pharmacology of such drugs. Similarly, it can be minimized with short- and intermediate-acting drugs by gradually tapering the dose over a few nights. A reduction of one clinical dose per week is usually recommended. Importantly, rebound usually can be avoided by using the lowest effective dose.

A critical distinction between rebound and recrudescence is necessary. Recrudescence is a return of symptoms to pretreatment level. This should be expected because BzRAs manage but do not "cure" the symptoms, and hence drug discontinuation is associated with return of symptoms.

One must differentiate rebound insomnia from a withdrawal syndrome. A withdrawal syndrome is the appearance of a new cluster of symptoms (not present before treatment) that are unpleasant and generally last a few days to a few weeks rather than 1 or 2 days. Rebound insomnia is the brief (1 or 2 nights) exacerbation of the original symptom. It does not reflect the presence of a withdrawal syndrome, which would involve the appearance of new symptoms lasting for more than a night or two. Also, rebound insomnia does not increase the likelihood of hypnotic self-administration and behavioral dependence.[72] Withdrawal has been found with both short-acting and long-acting BzRAs, unlike rebound insomnia. Withdrawal phenomena are typically associated with long-term use and importantly can occur with clinical doses. The patient's expectancies can also play a role in the experience of rebound insomnia. Discontinuing placebo pills, that is, stopping pill-taking per se, has been found to produce a sleep disturbance.[73]

Abuse Potential

Dependence on hypnotic medication has continued to be a concern despite only anecdotal evidence to support the concern. Epidemiologic data from representative population samples indicate that most patients do not persist in taking hypnotics for long periods because about 70% use them for 2 weeks or less.[74,75] A percentage of patients use them nightly on a chronic basis (for months or years) but with rare dose escalation.[30] It is unlikely that this pattern of use reflects physical or psychological dependence given the absence of dose escalation and nontherapeutic use of the medications. Although there are reports of physical dependence at therapeutic doses in long-term daytime use of benzodiazepines, none has been reported in 6- to 12-month studies of BzRA hypnotic use.

Daytime studies of the reinforcing effects of these drugs indicate that they have a low to moderate behavioral dependence liability.[76] Studies of their behavioral dependence liability in the context of their use as hypnotics have come to a similar conclusion.[73,77] Among individual patients, the rate of hypnotic administration varies with the severity of the sleep disturbance; within an individual patient, rate of administration depends on the degree of sleep disturbance the prior night.[78] Dose escalation in controlled trials (as in the clinical situation) does not occur, and daytime use is uncommon.[79] The same is true for normal volunteers, the population that best corresponds with patients being treated for transient insomnia. When insomnia patients are given the opportunity in research studies to self-administer hypnotics during the daytime, those that take hypnotic medication during the daytime show the physiologic hyperarousal discussed previously.[79] Hypnotic self-administration in insomniac patients is best explained by therapy-seeking behavior, in that it does not lead to dose escalation, it rarely generalizes to daytime use (i.e., it does not occur outside the therapeutic context), and the rate of self-administration varies as a function of the severity of the sleep problem.

Falls, Cognitive Effects, and Other Considerations for Older Adults

Given the age-related prevalence of insomnia, older adults represent a significant fraction of those using hypnotic medications (see Chapter 153 for more on treatment of insomnia of older patients). The therapeutic index for hypnotics in this population is narrower than for younger adults because of the increased prevalence in older patients of medical, neurologic, and primary sleep disorders, the common use of concomitant medications (particularly other drugs active on the CNS), and changes in drug metabolism and excretion.

Drugs that are primarily metabolized by conjugation are potentially safer for older patients and patients with liver disease. The pharmacokinetics of oxidatively metabolized drugs are altered in these two categories of patients, resulting in increased area under the plasma concentration curve. For some drugs (e.g., triazolam), this alteration is a consequence of increasing the peak plasma concentration, and for others (e.g., flurazepam) it is a consequence of extending the duration of significant blood levels. The lower recommended dose for most hypnotics when treating older patients is related in part to these kinetic changes. However, it is also thought that even with comparable blood levels, elderly people are more sensitive to drug effects.

A number of investigations have demonstrated an increased risk for falls in institutionalized older adults taking benzodiazepines and other psychotropic medications, without regard for the condition being treated or when the medication was taken. Results of early studies of community-dwelling older adults are inconsistent,[80-82] but recent large-scale studies suggest elevated risk. Kelly and colleagues[83] found an elevated risk for an injurious fall for seven categories of medications. When controlling for comorbid illness, the elevated risk remained significant for narcotics, antidepressants, and anticonvulsants but not hypnotics. Ensrud and colleagues[84] reported an elevated risk for falls in older women taking antidepressants (SSRIs or tricyclic antidepressants), anticonvulsants, and benzodiazepines, but not narcotics. Data from the same study, however, showed that risk for a fracture was elevated in those taking narcotics and antidepressants, but it was not affected by use of benzodiazepines and anticonvulsants.[85] Long-acting sedative drugs have been reported to increase risk for falls more than short-acting drugs, but recent data have not borne out this difference.[84] It is important to understand that a variety of CNS-active medications statistically increase fall rates in older adults, and antidepressants often have the highest risks.[86,87]

Investigators have begun to assess the risk for falls in insomniac patients with or without pharmacologic treatment for insomnia. Brassington and coworkers[45] found that reported sleep problems, but not use of psychotropic medication, was an independent risk factor for falls in a large sample of community-dwelling adults aged 64 to 99 years. In another recent study, the risk for falls was statistically significant in patients who had insomnia and did not use hypnotics and in those with insomnia that persisted despite hypnotic use, but not in patients who used hypnotics but did not have insomnia. One interpretation of these findings is that if the hypnotic relieves the insomnia, the hypnotic is not a risk factor for falling.[46] Stone and colleagues studied community-dwelling older women and reported that short sleep duration and fragmented sleep, determined by actigraphy, were associated with an increased risk for falls, independent of benzodiazepine and other BzRA use.[47] Future investigations will need to carefully differentiate between sleep disturbance and its treatment in evaluating contributing factors to falls.

Long-term use of benzodiazepines has been reported to be associated with cognitive decline in older adults,[88] including Alzheimer dementia,[89] although this is certainly not a universal finding.[90] The nature of these investigations (which are predominantly cross-sectional and retrospective) does not allow conclusions regarding causality because it is extremely difficult to implement proper controls for the aging process, the disease being treated, exposure to other drugs, and other factors. The cognitive changes are often subtle,[91] and the clinical significance of these effects has been questioned.[92] Nevertheless, the possibility that cognitive decline might at least in part be attributable to use of hypnotics, or any psychotropic medication, should be kept in mind by the clinician prescribing chronic therapy in elderly patients. However, withholding hypnotic treatment because of the anticipation of cognitive effects does not appear to be warranted on the basis of available information regarding magnitude of risk.

Complex Behavior in Sleep

Reports of idiosyncratic side effects associated with BzRA hypnotics, and other drugs, have appeared periodically in the public press. These reports of "global amnesia," "somnambulism," and "sleep-related eating disorders" are anecdotal. Peer-reviewed case reports describing unusual behavior concurrent with BzRA use also have appeared in the medical literature. Although case reports provide more detail and presumably more accurate information, including assessment of other contributing factors, the level of scientific evidence ascribed to case reports is low. Until a higher level of scientific evidence becomes available through placebo-controlled investigations, the actual risk for complex behavior in sleep and global amnesia will remain unknown. Even if adequate surveillance data were available to assess event frequency, the risk for an event cannot be determined unless the rate of exposure to various medications is known. Moreover, the dose consumed and concomitant medications and substances ingested at the time of the event are unknown.

Somnambulism has been reported with zolpidem and zaleplon.[93,94] These episodes of somnambulism have occurred in persons taking two to three times the clinical doses of the drug, in persons who have a prior history of somnambulism, and in persons who have experienced prior traumatic head injury. Zolpidem-associated somnambulism also has been reported in combination with antidepressant medications and alcohol. In a clinical context, somnambulism is believed to be associated with incomplete arousal from sleep. Although BzRAs increase somnambulism, alcohol and sleep deprivation also produce partial arousals and increase the likelihood of a somnambulistic event.

Sleep-related eating disorder associated with ingestion of BzRAs has also been reported.[95,96] It is disputed whether sleep-related eating disorder is a disorder of partial arousal from sleep with altered levels of consciousness or is the psychiatric disorder of nocturnal eating with awareness and recall. Sleep-related eating disorder is hypothesized to share a common pathophysiology with somnambulism. Zolpidem was reported to exacerbate sleep-related eating disorder and in several cases to induce it de novo. In some of these cases doses of zolpidem greater than 10 mg were being used, and in other cases there was use of sedating antidepressants. Sleep-related eating disorder also has been reported with triazolam. A common thread links much of this case-report information: excessive hypnotic activity or sleep drive. The excessive hypnotic activity can occur as a result of high doses, clinical doses in vulnerable persons (i.e., those who have a past history of sleep disorders or brain injury), the combination of clinical or high doses with prior sleep deprivation caused by stress or illness, or the combination of clinical or high doses with the prior consumption of alcohol.

Mortality

Examination of mortality risk associated with use of medication taken for insomnia and other indications often shows an elevated risk, the explanation for which is unknown.[97-99] Clearly, these findings merit investigation of the mediators and nature of the risk. However, a number of factors limit the conclusions that can be drawn from currently available data.

The epidemiologic nature of the studies (as opposed to controlled experimental trials) prevents any conclusions regarding causality or mediators. It is also important to remember that 90% of insomniac patients have a comorbid condition, and typically they have more than one. Thus the mediators of the mortality are not clear. It would be

important to look at the risk in terms of specific causes of mortality.

The exact medications responsible for the elevated mortality risk are not constant from study to study, and, in fact, they are often unknown. Data for one study were collected in 1959 to 1960,[97] when barbiturates were the most commonly used hypnotics; for the second study[98] data were collected in 1982, when benzodiazepines were the most commonly used hypnotics. Several studies also did not differentiate between prescription and nonprescription drugs. All investigations concluding that an elevated mortality risk exists must include any medication the respondent reports as being taken for sleep, whether a hypnotic or not, and whether the prescribing physician actually intended another indication. The only study that obtained reasonable verification of the actual drugs being taken by respondents found that sedative-hypnotics were not associated with an increased mortality risk. Rather, "other drugs" taken for sleep, which in that study were often analgesics, showed a statistically higher mortality risk.[100]

Long-term, controlled investigations would be needed (although ethical considerations clearly are of concern) to clarify whether any medication used to promote sleep increases mortality risk, or whether the statistical risk is explained by the poor health of the users of sleep-promoting drugs, or by other factors. Alternatively, longitudinal population-based studies aimed at answering this question could better define the risk. Such a study would need to identify the specific hypnotic drug being taken, pattern of drug taking, other drugs being taken, and the nature of the comorbid disorders and should institute adequate experimental and statistical control to allow reasonable inferences about the data. Until such data become available the possibility that hypnotics contribute to mortality risk cannot be discounted.

CONSIDERATIONS FOR PHARMACOTHERAPY

Hypnotic therapy should be considered for insomnia when the patient is significantly distressed by the presence or possibility of disturbed sleep, or when the physician judges the sleep disturbance to be deleterious to the patient's overall safety or health. The 2005 National Institutes of Health State of the Science Conference report on the management of chronic insomnia[1] concluded that BzRAs are the only medications with an established scientific basis (i.e., clearly defined risk and benefit by dose) for use for insomnia. Thus, for most patients with chronic insomnia, pharmacotherapy should be initiated with a BzRA at the lowest effective dose for the shortest clinically necessary period of time. There is also widespread agreement that BzRA hypnotics are an appropriate therapy for short-term insomnia.

The specific BzRA should be carefully chosen, considering the pharmacokinetics of the drug and the patient characteristics (e.g., age, nature of the sleep problem, concurrent illness). The dose and treatment regimen should be individualized, agreed to by the patient, and most importantly monitored by the physician. The hypnotic should be taken only at the time and frequency agreed on. The initial dose should be the lowest clinically indicated dose, and the clinician and patient should jointly determine effectiveness of that dose soon after treatment is initiated (e.g., after three to five nights). The dose can then be adjusted as appropriate based on both efficacy and safety reports from the patient. Patients should generally not be allowed to adjust the dose themselves, particularly during the early portion of therapy, because a fluctuating dose makes interpretation of clinical response very difficult.

Hypnotics can be prescribed to be taken at bedtime nightly, on a predetermined intermittent schedule (e.g., every third night), or as needed. With the exception of zaleplon, all bedtime medications should be taken only when the patient has the opportunity to stay in bed 7 to 8 hours. Zaleplon has been evaluated when administered with 5 hours or less of time in bed remaining and been found to improve sleep without residual effects the next morning.[101,102] As previously mentioned, a sublingual lozenge formulation of zolpidem (Intermezzo) is the only hypnotic approved for nonbedtime use, that is, for middle-of-the-night awakening and difficulty returning to sleep. If nightly hypnotic use is not desirable for individual patients, they can be instructed to attempt sleep without medication and, if unsuccessful, to take the drug later, provided there are 4 to 5 hours remaining before they must rise. Thus, in cases of sleep-maintenance insomnia characterized only by a middle-of-the-night awakening, or in cases of intermittent sleep-onset insomnia, middle-of-the-night administration should be considered.

The physician should specifically inquire about the aspects of sleep and daytime function that were most problematic for the patient before treatment, rather than relying solely on global statements. A sleep diary can be helpful for comparing pretreatment and posttreatment symptoms. In reasonably healthy persons, monthly visits should be sufficient to identify most potential adverse effects. For older adults or others at increased risk for drug interactions, alterations in drug metabolism, incoordination, or cognitive problems, close follow-up is recommended throughout the period that hypnotics are administered. Discontinuation of therapy should be considered if symptoms occur that may be side effects of the medication, to confirm whether they are indeed related to the medication. After long-term use, gradual discontinuation, by dose tapering at a rate of one clinical dose per week, is a reasonable practice.

The primary contraindications to BzRA hypnotic therapy are concomitant illnesses, such as severe COPD, obstructive sleep apnea, substance abuse disorder, or advanced liver disease. All sedative medications have the potential to worsen sleep apnea by blunting arousal from sleep, although central apnea has been reduced by BzRAs in some investigations.[103] The dependence liability of BzRAs and other sedative drugs, although not high, leads to the conclusion that most patients with a history of alcoholism or drug abuse should not receive BzRAs in outpatient settings without close supervision. Caution is advised for moderate users of alcohol because of additive sedative effects with hypnotics, which narrows the wide margin of safety described earlier. Because most BzRAs undergo hepatic metabolism, advanced liver disease requires the use of a lower dose or avoidance of these medications.

Pharmacotherapy for insomnia during pregnancy is also contraindicated; the teratogenic effects of all psychoactive drugs are a matter of concern.[104] People who may be required to awaken and perform duties in the middle of the night should avoid CNS-depressant drugs when on call. All hypnotics have the potential to disrupt alertness and cognitive function for the duration of the sedative activity of the drug, and they can affect motor function.

CLINICAL PEARL

Extensive efficacy and safety evidence supports the use of BzRAs for the treatment of insomnia for as long as it is clinically warranted. Convincing evidence for use of other available medications (e.g., sedating antidepressants) is not currently available. The duration of action between different BzRAs varies greatly, and therefore the desired duration of sleep-promoting action and the need for morning alertness should guide the clinician in the choice of a specific medication and the dose for each person.

SUMMARY

Considerable evidence indicates that the BzRAs are efficacious in the treatment of insomnia, whether primary or comorbid, chronic or transient. Limited data suggest a substantial degree of effectiveness in clinical use. BzRAs are also generally safe as commonly used in clinical practice. Research refutes the common concerns of tolerance, misuse, and abuse by insomnia patients. Sufficient differences in pharmacokinetic profiles exist among the BzRAs to provide physicians with therapeutic options that have differing durations of action. Individual differences in metabolism and elimination of BzRAs have led to a lowering of recommended starting doses for some BzRAs. Several investigations indicate that treatment of insomnia with BzRAs produces improvement in both daytime and nighttime symptoms of insomnia, and some might also reduce the severity of comorbid illnesses. Despite significant advances in understanding the neurobiologic mechanisms of sleep and arousal, which involve multiple neural systems, attempts to produce novel drug treatments for insomnia have not produced clear advances in insomnia pharmacotherapy. BzRAs remain the most effective, safest, and best-understood therapeutic agents for the treatment of insomnia.

Selected Readings

Avidan AY, Fries BE, James ML, et al. Insomnia and hypnotic use, recorded in the minimum data set, as predictors of falls and hip fractures in Michigan nursing homes. *J Am Geriatr Soc* 2005;**53**:955–62.

Becker WC, Fiellin DA, Desai RA. Non-medical use, abuse and dependence on sedatives and tranquilizers among U.S. adults: psychiatric and sociodemographic correlates. *Drug Alcohol Depend* 2007;**90**:280–7.

Chen SJ, Yeh CM, Chao TF, et al. The use of benzodiazepine receptor agonists and risk of respiratory failure in patients with chronic obstructive pulmonary disease: a nationwide population-based case-control study. *Sleep* 2015;**38**(7):1045–50.

Fava M, McCall WV, Krystal A, et al. Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. *Biol Psychiatry* 2006;**59**:1052–60.

Goforth HW, Preud'homme XA, Krystal AD. A randomized, double-blind, placebo-controlled trial of eszopiclone for the treatment of insomnia in patients with chronic low back pain. *Sleep* 2014;**37**(6):1053–60.

National Institutes of Health. State of the Science Conference Statement on Manifestations and Management of Chronic Insomnia in Adults statement. *J Clin Sleep Med* 2005;**1**:412–21.

Okun ML, Ebert R, Saini B. A review of sleep-promoting medications used in pregnancy. *Am J Obstet Gynecol* 2015;**212**(4):428–41.

Pollack M, Kinrys G, Krystal A, et al. Eszopiclone coadministered with escitalopram in patients with insomnia and generalized anxiety disorder. *Arch Gen Psychiatry* 2008;**65**:551–62.

Roth T, Krystal A, Steinberg FJ, et al. Novel sublingual low-dose zolpidem tablet reduces latency to sleep onset following spontaneous middle-of-the-night awakening in insomnia in a randomized, double-blind, placebo-controlled, outpatient study. *Sleep* 2013;**36**:189–96.

Walsh JK, Krystal AD, Amato DA, et al. Nightly treatment of primary insomnia with eszopiclone for six months: effect on sleep, quality of life and work limitations. *Sleep* 2007;**30**:959–68.

Walsh JK, Roth T, Randazzo AC, et al. Eight weeks of non-nightly use of zolpidem for primary insomnia. *Sleep* 2000;**23**:1087–96.

A complete reference list can be found online at ExpertConsult.com.

Chapter 88

Pharmacologic Treatment of Insomnia: Other Medications

Andrew D. Krystal

Chapter Highlights

- Many different types of agents other than benzodiazepine binding site–positive allosteric modulators (also referred to as benzodiazepine receptor agonists) are used in the treatment of insomnia, including antidepressants, antipsychotics, antihistamines, anticonvulsants, and melatonin agonists.
- Each of these agents has unique properties that should be considered in determining which agent should be selected for use in a given patient with insomnia.
- Awareness of the characteristics of the nonbenzodiazepine receptor agonist agents and of the clinical circumstances in which they might be the preferred pharmacotherapy is important for optimizing the pharmacotherapy of insomnia in clinical practice.

The predominant treatments for insomnia in the twentieth century were medications that bind to the benzodiazepine binding site on the gamma-aminobutyric acid (GABA) receptor complex.[1-3] These agents, often referred to as benzodiazepine receptor agonists (BZRAs), are discussed in detail in Chapter 87. Their therapeutic effects in insomnia patients occur through enhancing the inhibitory effects of GABA, which is believed to be the neurotransmitter that is most important for promoting sleep.[1,2,4] Because of the long-standing and widespread use of BZRAs, this mechanism and the clinical properties of these agents have become almost synonymous with insomnia pharmacotherapy. Experience suggests that these insomnia agents have effects that are proportional to their serum level (maximal effect occurs at peak blood level), have some abuse potential, and have daytime sedation and psychomotor impairment as their primary potential side effects.[2,3] However, there are agents used to treat insomnia that improve sleep by mechanisms other than enhancing GABAergic inhibition, and their clinical properties differ from those of the BZRAs. These agents act through a variety of mechanisms and represent at least 11 different medication classes. The mechanisms of their therapeutic effects on insomnia include melatonin receptor agonism, serotonin receptor antagonism, histamine receptor antagonism, hypocretin/orexin receptor antagonism, norepinephrine receptor antagonism, and possibly antagonism of dopamine and acetylcholine receptors.[3] In addition, these agents have a diverse set of other pharmacologic effects that contribute to their side effect profile and are responsible for their effects on conditions other than insomnia.[3] In fact, all of these agents, with the exception of the melatonin receptor agonist ramelteon, the tricyclic drug doxepin, and the hypocretin/orexin antagonist suvorexant, are indicated for use in conditions other than insomnia by the U.S. Food and Drug Administration (FDA). As a result, they are not usually categorized as insomnia therapies but are typically referred to by their other intended actions as antidepressants, anxiolytics, antipsychotics, anticonvulsants, or antihistamines.

The pharmacology of these agents is reviewed in Chapter 43. In this chapter, we discuss these agents, focusing on their clinical effects and their potential utility in the clinical treatment of insomnia (Table 88-1).

MELATONIN AND MELATONIN RECEPTOR AGONISTS

Melatonin

Overview. Melatonin is a hormone primarily produced by the pineal gland. The release of this hormone occurs during the period of darkness from sundown to sunrise. When it is taken orally, melatonin appears to have sleep-enhancing and circadian rhythm–modifying effects. The circadian rhythm effects of melatonin are discussed in Chapter 40. Here we discuss the sleep-enhancing effects.

Melatonin is available as an over-the-counter medication in the United States. It has a time to maximum concentration (T_{max}) of approximately 0.5 hour and an elimination half-life ($T_{1/2}$) of roughly 1 hour.[5] A BZRA with such a pharmacokinetic profile would be expected to have a short duration of effect and be best suited to treatment of sleep-onset problems. However, this cannot be assumed with melatonin, which works by a mechanism that differs from the BZRAs. Melatonin binds to melatonin receptors, predominantly MT_1 and MT_2 receptors, and the mechanism by which these receptors affect sleep is unknown, although it is clear that the sleep enhancement of this agent is divorced from the serum level. A dose-response relationship does not appear to exist for melatonin, and the means to determine optimal dosing are lacking. Further complicating optimization of the use of melatonin, a number of studies have identified that the timing of sleep enhancement that occurs with melatonin depends on the time of day of administration and may not be manifested for as long as 3 hours after dosing, despite achievement of maximal serum concentration in approximately 30 minutes.[6-8] As a result, attention must be given to the time of day that this agent is administered as well as to the relationship of the time

Table 88-1 Medications Other Than Benzodiazepine Receptor Agonists Used in the Treatment of Insomnia

Medication	FDA Indication	Insomnia Dosage (mg)	T_{max} (h)	$T_{1/2}$ (h)*	Efficacy in Placebo-Controlled Insomnia Trials	Most Frequent Side Effects	Relative Contraindications to Use	Most Promising Clinical Use
Amitriptyline	MDD	10–100	2–5	10–100	—	Sedation Dizziness Weight gain Orthostatic hypotension Dry mouth Blurred vision Constipation Urinary retention	History of myocardial infarction, ischemia, or conduction abnormalities Closed-angle glaucoma Decreased gastrointestinal motility Urinary retention Hypotension History suggestive of bipolar disorder Parkinson disease Seizure disorder Hepatic disease	Insomnia comorbid with MDD, chronic pain, or anxiety
Doxepin	MDD Anxiety	1–25 mg	1.5–4	10–50	5/5	Sedation Dizziness Weight gain Orthostatic hypotension Dry mouth Blurred vision Constipation Urinary retention	History of myocardial infarction, ischemia, or conduction abnormalities Closed-angle glaucoma Decreased gastrointestinal motility Urinary retention Hypotension History suggestive of bipolar disorder Parkinson disease Seizure disorder Hepatic disease	Efficacy with relative absence of side effects in 1–6 mg Insomnia comorbid with MDD, chronic pain, or anxiety Treatment of early morning awakenings
Trimipramine	MDD	25–100	2–8	15–40	2/2	Sedation Dizziness Weight gain Orthostatic hypotension Dry mouth Blurred vision Constipation Urinary retention	History of myocardial infarction, ischemia, or conduction abnormalities Closed-angle glaucoma Decreased gastrointestinal motility Urinary retention Hypotension History suggestive of bipolar disorder Parkinson disease Seizure disorder Hepatic disease	Insomnia comorbid with MDD, chronic pain, or anxiety

Continued

Table 88-1 Medications Other Than Benzodiazepine Receptor Agonists Used in the Treatment of Insomnia—cont'd

Medication	FDA Indication	Insomnia Dosage (mg)	T_{max} (h)	$T_{1/2}$ (h)*	Efficacy in Placebo-Controlled Insomnia Trials	Most Frequent Side Effects	Relative Contraindications to Use	Most Promising Clinical Use
Trazodone	MDD	25–150	1–2	7–15	0/1	Sedation Dizziness Headache Dry mouth Blurred vision Orthostatic hypotension Priapism	Hepatic disease Renal disease Suspected bipolar disorder	Insomnia comorbid with MDD Insomnia comorbid with substance abuse
Mirtazapine	MDD	7.5–30	0.25–2	20–40	—	Sedation Dry mouth Increased appetite Weight gain Constipation	Patients with suspected bipolar disorder Hepatic disease Renal disease	Insomnia comorbid with MDD Insomnia comorbid with SDB
Suvorexant	Insomnia	10–20	3	9–13	4/4	Sedation	Narcoleptics Substance abuse–prone individuals	Patients with early morning awakening and difficulty in falling asleep Patients with mild to moderate COPD Patients for whom long-term treatment may be needed
Olanzapine	Schizophrenia Mania	2.5–20	4–6	20–54	—	Sedation Agitation Dizziness Constipation Orthostatic hypotension Akathisia Weight gain Increased incidence of cerebrovascular events in dementia patients	Dementia patients Hepatic disease Prostatic hypertrophy Closed-angle glaucoma Paralytic ileus Urinary retention Hypotension History of myocardial infarction, ischemia, or conduction abnormalities	Insomnia in individuals with psychosis, mania-spectrum, anxiety
Quetiapine	Schizophrenia Mania MDD	25–200	1–2	7	—	Sedation Orthostatic hypotension Dry mouth Tachycardia Weight gain	Hepatic disease Patients with hypotension, cerebrovascular disease, heart failure, history of myocardial infarction or ischemia, or conduction abnormalities	Insomnia in individuals with psychosis, mania-spectrum, anxiety, or comorbid MDD
Gabapentin	Partial seizures Pain	100–900	3–3.5	5–9	0/1	Sedation Dizziness Ataxia Diplopia	Renal impairment	Insomnia in patients with pain, partial seizures, alcohol dependence, or alcohol withdrawal

Pregabalin	Fibromyalgia Pain Partial seizures	50–300	1	4.5–7	1/1	Sedation Dizziness Dry mouth Cognitive impairment Increased appetite Discontinuation effects	Renal impairment Substance abuse risk	Insomnia in patients with pain, fibromyalgia Insomnia occurring in those with partial seizures
Tiagabine	Partial seizures	2–16	1–1.5	8	2/4	Sedation Nausea Dizziness Seizures	Hepatic disease	Insomnia occurring in patients with partial seizures
Chloral hydrate	Insomnia Alcohol withdrawal Anxiety	250–2000	0.5	7–10	1/2	Sedation Nausea Diarrhea Parasomnias Renal damage may occur with chronic use Discontinuation effects	Neonates Those at risk for substance abuse Suicidal individuals Gastritis Intermittent porphyria Hepatic or renal failure	—
Sodium oxybate	Narcolepsy	2000–4500	0.5–0.8	0.3–1.2	—	Headache Nausea Dizziness Drowsiness Discontinuation effects	Those vulnerable to respiratory depression SDB, COPD, or hypoxemia of any cause Those at risk for substance abuse	Insomnia in fibromyalgia patients Sleep disturbance in narcolepsy patients
Melatonin	OTC hormone	0.3–10	0.3.–1	0.6–1	13/14	Headache Sedation	Men and women attempting to conceive	Chronic sleep-onset insomnia in children Alzheimer dementia Children with neurodevelopmental disorders
Ramelteon	Insomnia	8	0.7–0.95	0.8–2	4/4	Drowsiness Fatigue Dizziness	Severe hepatic failure	Sleep-onset insomnia Insomnia comorbid with COPD, SDB
Diphenhydramine	OTC allergy Insomnia	25–50	2–2.5	5–11	2/2	Sedation Dizziness Dry mouth Blurred vision Constipation Urinary retention	Severe hepatic failure Closed-angle glaucoma Asthma COPD Bladder obstruction Gastrointestinal obstruction Ileus Urinary retention	Insomnia in those with allergy or upper respiratory infections
Doxylamine	OTC insomnia	25–50 mg	1.5–2.5	10–12	—	Sedation Dizziness Dry mouth Blurred vision Constipation Urinary retention	Severe hepatic failure Closed-angle glaucoma Asthma COPD Bladder obstruction Gastrointestinal obstruction Ileus Urinary retention	Insomnia in those with allergy or upper respiratory infections

*Includes active metabolites.
FDA, Food and Drug Administration; MDD, major depressive disorder; OTC, over-the-counter; COPD, chronic obstructive pulmonary disease; SDB, sleep disordered breathing.

of administration to desired sleep-enhancing effect. The dosages administered in placebo-controlled trials have ranged from 0.1 to 75 mg; however, the majority of studies used dosages in the range of 2 to 6 mg. The timing of administration varied from 30 minutes to 3 hours before bedtime. There was no clearly predominant administration time. The primary metabolism of melatonin occurs through cytochrome P450 1A2 and 2C19 isoenzymes, so the elimination of melatonin would be expected to be affected by factors that modify the function of these isoenzymes.[9]

Evidence of Efficacy. Placebo-controlled studies of the effects of melatonin on sleep have been carried out in a number of different subject populations, including adults without sleep complaints; children, adults, and elderly insomnia patients; patients with Alzheimer dementia; patients with insomnia comorbid with medical illnesses; mood disorder patients; patients with schizophrenia; children with attention-deficit/hyperactivity disorder; and children with neurodevelopmental disorders.[10-24] However, drawing conclusions from these studies is difficult because they employed widely varying melatonin dosages and different melatonin formulations (immediate and delayed release) and varied substantially in the timing of melatonin administration.[25,26] Despite these limitations, the available self-report, polysomnographic, and actigraphic data suggest that melatonin has a therapeutic effect on sleep onset that is greater and more consistent than its effects on the ability to stay asleep or the duration of sleep.*

Profile of Adverse Effects. No substantive risks of melatonin have been identified. However, no large-scale systematic studies assessing the adverse effects of melatonin have been carried out. The most commonly reported side effect is headache.[12] Slowing of reaction time and sedation have been observed with melatonin after daytime administration, which could impair ability to function.[6,32,33] Melatonin is nonreinforcing, so abuse potential is minimal. Physiologic or psychological dependence does not appear to occur with melatonin therapy, although this has not been systematically evaluated. The longest studies of nightly melatonin therapy have been 8 weeks in duration.[11,24]

Potential Clinical Utility. Melatonin appears to have primary use in addressing problems of sleep onset. It is relatively unique in that data suggest it can be safely administered to children, those with Alzheimer dementia, and children with neurodevelopmental disorders.[10-24] However, the degree of therapeutic effects in these populations is not well established. Given the absence of abuse potential, this medication would also be reasonable to use in substance abuse–prone individuals; however, no studies of melatonin treatment have been carried out in this population.

A number of studies suggest that melatonin regulates reproductive function in both males and females. Some evidence suggests that higher melatonin levels may be associated with reversible inhibition of spermatogenesis and ovulation.[34-37] On this basis, some have suggested that those who are trying to conceive should avoid taking melatonin.[36,37]

*References 10, 12, 14, 18, 19, 27-31.

Ramelteon

Overview. Ramelteon is an agonist at MT_1 and MT_2 melatonin receptors, has a T_{max} of 0.75 to 1 hour, and has a $T_{1/2}$ of 1 to 2.5 hours.[38] Based on this pharmacokinetic profile, ramelteon, like melatonin, would be expected to have effects primarily on sleep onset. Indeed, it is approved by the U.S. FDA with an indicated use in the treatment of sleep-onset difficulty. Ramelteon, like melatonin, has no clear dose-response relationship; however, a therapeutic dosage of 8 mg and timing of dosing 30 minutes before bedtime have been established.[3] Primary metabolism of ramelteon occurs through the cytochrome P450 isoenzymes CYP1A2, CYP2C, and CYP3A4, so that factors affecting these isoenzymes would be expected to alter the metabolism of ramelteon.[39]

Evidence of Efficacy. In placebo-controlled trials, ramelteon consistently shortens sleep latency with an accompanying increase in total sleep time.[3] These effects have been more consistent on polysomnographic than on self-reported measures of sleep. Therapeutic effects on the ability to stay asleep have generally not been noted.

Profile of Adverse Effects. Ramelteon is generally well tolerated; dosages up to 64 mg have been associated with no substantive increase in adverse effects. The most common side effects reported with ramelteon include headache, dizziness, somnolence, fatigue, and nausea.[40,41] There is no evidence for abuse potential or tolerance in studies as long as 5 weeks in duration.[40,41] Statistically significant but not clinically significant elevations of prolactin have been noted with ramelteon compared with placebo in women receiving 6 months of nightly treatment with 16 mg.[42] Several studies suggest that ramelteon does not exacerbate breathing problems when it is used in those with sleep disordered breathing and mild to moderate chronic obstructive pulmonary disease.[43,44]

Potential Clinical Utility. Ramelteon is indicated solely for the treatment of sleep-onset difficulties. The data in patients with sleep disordered breathing and chronic obstructive pulmonary disease suggest that ramelteon can be safely used in these populations of patients when treatment of sleep-onset insomnia is indicated.[43,44] The absence of abuse potential suggests that this medication could also be used in substance abuse–prone individuals, although no studies of the treatment of insomnia with ramelteon have been carried out in this population.

Ramelteon is relatively free of contraindications. The only relative contraindication to use is severe liver failure.

SEDATING ANTIDEPRESSANTS

A number of agents that are approved by the U.S. FDA for the treatment of depression are used in relatively lower dosages for the treatment of insomnia.[3] In fact, for some of these agents, their use in low dosage for insomnia therapy exceeds their use for alleviation of depression.[45] These agents have therapeutic effects on sleep through blocking of the receptors of a number of key wake-promoting neurotransmitters, including serotonin, norepinephrine, and histamine.[3,46] The antidepressants most commonly used to treat insomnia include tricyclic antidepressants, most notably, doxepin,

trimipramine, and amitriptyline; the chlorophenyl piperazine; trazodone; and the tetracyclic antidepressant mirtazapine.[3]

Tricyclic Antidepressants: Doxepin, Amitriptyline, Trimipramine

Overview. Three tricyclic antidepressants—doxepin, amitriptyline, and trimipramine—are used relatively frequently for the treatment of insomnia, presumably because of the sedating effects noted when they have been used to treat major depression.[3,45,47] Interactions with the tricyclic antidepressants may occur through effects on their predominant metabolic pathways, which involve the cytochrome P450 system CYP3A4, CYP2C19, CYP2D6, and CYP2C9 isoenzymes.[48,49] By affecting these isoenzymes or altering hepatic blood flow, a number of factors can increase blood levels of tricyclic antidepressants, including age, antipsychotic medications, methylphenidate, fluoxetine, paroxetine, cimetidine, diltiazem, and grapefruit juice.[50] There is also genetic variation in the capacity to metabolize tricyclic antidepressants in the population, which leads to substantial person-to-person differences in the relationship between dosage and therapeutic and adverse effects.[50] Amitriptyline, doxepin, and trimipramine reach their maximum serum concentration (T_{max}) in 1.5 to 6 hours after oral ingestion and have elimination half-lives ($T_{1/2}$) on the order of 10 to 50 hours.[48-50]

Evidence of Efficacy. These agents have been reported to have therapeutic effects on indices of sleep onset and maintenance in depressed patients.[51] In addition, a number of placebo-controlled trials of the treatment of insomnia have been carried out with doxepin and trimipramine. Trimipramine 50 to 200 mg has been evaluated in two studies of primary insomnia patients, leading to improved sleep quality and polysomnographic sleep efficiency compared with placebo, with no effect on sleep-onset latency.[52,53] Doxepin has been evaluated in three studies at a dosage of 25 to 50 mg and found to improve sleep quality and reported daytime well-being as well as polysomnographic indices of sleep onset and maintenance compared with placebo.[52,54-56] More recently, doxepin has been evaluated in placebo-controlled trials in insomnia patients at dosages of 1 to 6 mg.[57,58] Because the predominant pharmacologic effect of low-dose doxepin is the blockade of H_1 histamine receptors, as the dosage is decreased, doxepin becomes an agent with increasingly specific H_1 antagonist effects.[59] The studies carried out with doxepin 1 to 6 mg have identified evidence of self-reported and polysomnographic efficacy in indices of sleep onset and maintenance with particular effects on maintenance of sleep.[57,58] Notably, the therapeutic effects appear to be largest in the last third of the night, and of particular note, they are greatest in the last hour (hour 8) of the night.[57] This suggests a dissociation of the clinical effects from the serum blood level as the peak blood level occurs between 1.5 and 4 hours after dosing. At the same time, 1 to 6 mg of doxepin was not associated with significant sedation or impairment when subjects were tested after waking 1 hour after the peak sleep-enhancing effect.[57] Daytime somnolence was not noted as an adverse effect more frequently than with placebo in these studies, despite the relatively long half-life of this agent. The absence of daytime impairment with doxepin 1 to 6 mg suggests that an increase in wake-promoting neurotransmitter activity occurs after waking that counteracts the H_1 antagonism.

Profile of Adverse Effects. The side effects of the tricyclic antidepressants primarily derive from their blockade of H_1 histaminergic, M_1 muscarinic cholinergic, serotonergic ($5\text{-}HT_2$), and $alpha_1$- and $alpha_2$-adrenergic receptors. These actions result in potential side effects, such as sedation, orthostatic hypotension (and the attendant risk of falls), weight gain, dry mouth, constipation, urinary retention, cardiac dysrhythmias (increased heart rate and impairment of cardiac electrical conduction, which can lead to heart block in some instances), exacerbation of narrow-angle glaucoma, and, rarely, seizures and anticholinergic delirium.[46,49] These effects are dose dependent. Most published literature relates to their use in dosages used to treat depression and not the lower dosage ranges more commonly used in the treatment of insomnia. The best data on the adverse effects of tricyclic antidepressants when they are used in lower dosages for insomnia therapy exist for doxepin. In dosages of 1, 3, and 6 mg, doxepin has a remarkably favorable side effect profile, in contrast to the side effects noted with its antidepressant use.[57,58] In particular, the available data suggest that doxepin is not associated with significant weight gain, it does not appear to be associated with substantive daytime impairment or sedation, and anticholinergic effects are uncommon.

Potential Clinical Utility. It is natural to consider the use of these antidepressants in patients with insomnia associated with depression. However, there are no data indicating that an antidepressant effect occurs with the dosages of these agents that are generally used to treat insomnia. Low-dose monotherapy with these agents may not have antidepressant effects, but a therapeutic effect on depression might be seen when they are combined with a nonsedating antidepressant (e.g., selective serotonin reuptake inhibitor or serotonin-norepinephrine reuptake inhibitor). At the same time, there may be increased risks when these agents are combined.[60] Tricyclic antidepressants have also been used frequently in the treatment of anxiety and chronic pain and may be useful to treat sleep problems associated with these conditions. However, as with depression, dosage may be an important consideration.[61] Given the lack of abuse potential, the tricyclic antidepressants are a potential option for use in treating the sleep difficulties of individuals with substance abuse history. Doxepin dosed from 1 to 6 mg appears to have unique potential for treating sleep difficulties in the latter part of the night/early morning period. As low-dose doxepin is also a relatively selective H_1 antihistamine, it may be useful in individuals with insomnia occurring in conjunction with allergies.

Relative contraindications to the use of tricyclic antidepressants derive from their risks of cardiac arrhythmias, anticholinergic effects, seizures, and orthostatic hypotension.[46,49] Because of these risks, tricyclic antidepressants should be used with caution in those with a history of myocardial infarction, ischemia, or conduction abnormalities; closed-angle glaucoma; decreased gastrointestinal motility; urinary retention; hypotension; or seizure disorder. Antidepressants should be used with caution in those with bipolar disorder because of the increased risk of precipitating mania.[62] Based on the metabolism of the tricyclic antidepressants, significant hepatic disease is also a relative contraindication to their use. As previously noted, the risks and associated relative contraindications of these agents are dose related and may not be relevant in the lower dosages often used to treat insomnia.

Trazodone

Overview. Trazodone is most frequently used as an "off-label" agent in the treatment of insomnia.[45] Whereas it is generally used at a dosage of 200 to 600 mg in the treatment of depression, the dosage used to treat insomnia typically ranges from 25 to 150 mg. With a T_{max} of 1 to 2 hours and a $T_{1/2}$ of 7 to 15 hours, trazodone has the potential to have therapeutic effects on the initiation and maintenance of sleep.[3] Trazodone is primarily metabolized through CYP3A4 and CYP2D6 pathways and can have significant interactions with other agents that are associated with these liver cytochrome systems.[3] Trazodone has one active metabolite, *m*-chlorophenylpiperazine (mCPP).[63] Significant variation in the effects of trazodone occur because many factors can affect the metabolism of trazodone and mCPP, including genetic polymorphisms that are common in the general population.[64]

Evidence of Efficacy. The primary evidence of a sleep-enhancing effect of trazodone is that sedation was a frequently reported side effect in depression treatment trials. Few studies of the use of trazodone in insomnia patients have been carried out. A small placebo-controlled study in abstinent alcoholics ($n = 16$) found significant improvement in sleep efficiency and awakenings compared with placebo.[65] A double-blind, placebo-controlled, crossover study was carried out in 17 patients with depression who had insomnia while being treated with fluoxetine or bupropion.[66] Compared with placebo, trazodone significantly improved the Pittsburgh Sleep Quality Index and the sleep items of the Yale–New Haven Hospital Depressive Symptom Inventory. One study of the treatment of patients with primary insomnia was also carried out with trazodone.[67] This was a parallel-group, 2-week study comparing trazodone 50 mg, placebo, and zolpidem 10 mg. A number of sleep parameters were significantly improved versus placebo for both trazodone and zolpidem in the first week of treatment. However, by the second week, these differences were not significant because of improvement in the placebo-treated patients.

Profile of Adverse Effects. The most common side effects occurring with trazodone include sedation, dizziness, headache, dry mouth, blurred vision, and orthostatic hypotension. Weight gain may occur in some cases. Individuals who have relatively diminished capacity to metabolize mCPP because of genetic or other factors are likely to experience anxiety or insomnia with trazodone. Priapism, a painful sustained erection, is a relatively uncommon side effect of trazodone that may require surgical intervention and can lead to impotence.[68] There is no evidence that trazodone has significant abuse potential.

Potential Clinical Utility. As with other antidepressants, it is natural to consider the use of trazodone in patients with insomnia associated with depression. One study demonstrated the safe and effective use of trazodone in depressed patients treated with fluoxetine or bupropion.[66] However, trazodone may be associated with increased risks when it is combined with other antidepressants. Further, it is unlikely to have significant antidepressant effects when it is used as monotherapy in the dosing range generally used to treat insomnia (50 to 150 mg), although whether it has antidepressant effects when it is used in combination with other antidepressants is unknown. Given the lack of abuse potential, trazodone is reasonable to consider for use in individuals with substance use disorders. In this regard, one study carried out in abstinent alcoholics reported a therapeutic effect on sleep and a reasonable safety profile.[65]

Trazodone should be used with caution in those at risk for falls because of the risk of orthostatic hypotension. Trazodone should also be administered judiciously in men because of the risk of priapism. Like any antidepressant, trazodone should be used with caution in those with bipolar disorder because of the increased risk of precipitating mania.[62] Based on the metabolism of trazodone, significant hepatic disease and renal disease are also relative contraindications.

Mirtazapine

Overview. Mirtazapine is FDA approved for the treatment of major depression in dosages of 7.5 to 45 mg. The dosages generally used to treat insomnia are 7.5 to 30 mg. The sedating effects of mirtazapine may diminish in dosages larger than 30 mg because of increasing adrenergic effects.[69] The antidepressant, sleep-enhancing, and adverse effects of this agent derive from antagonism of adrenergic ($alpha_1$ and $alpha_2$), serotoninergic ($5\text{-}HT_2$ and $5\text{-}HT_3$), and histaminergic (H_1) receptors.[70] Interactions with mirtazapine can occur through effects on liver cytochrome P450 enzymes CYP2D6, CYP3A4, and CYP1A2.[3] The T_{max} of this agent is 0.25 to 2 hours, and the $T_{1/2}$ is 20 to 40 hours.[3]

Evidence of Efficacy. Mirtazapine has not been evaluated in placebo-controlled trials in insomnia patients. Evidence for effects enhancing sleep onset and sleep maintenance comes from open-label studies of depressed patients and normal sleepers and one double-blind, randomized study comparing the sleep effects of this agent versus fluoxetine.[71-73] However, the *S*-isomer of mirtazapine (*S*-mirtazapine), which has pharmacokinetics comparable to the racemate and like doxepin 3 to 6 mg has a predominant effect of selective H_1 antagonism in the dosages studied, has been evaluated in four placebo-controlled trials in primary insomnia patients. These studies indicate a consistent therapeutic effect on sleep maintenance, with a tendency to improve sleep onset as well, although the onset effect is not as large or consistent and is dose dependent, as are the risks of daytime sedation.[74-77]

Profile of Adverse Effects. The predominant side effects of mirtazapine are increased appetite and weight gain, dry mouth, and constipation.[69] Consistent with the sleep-enhancing effects of this agent, sedation is also a common adverse effect of treatment. Mirtazapine is not associated with significant abuse potential or cardiac or sexual adverse effects.

Potential Clinical Utility. Because of the antidepressant effects of mirtazapine, it is particularly well suited for use in patients with insomnia occurring with major depression. A study suggesting that mirtazapine may decrease the apnea-hypopnea index in those with significant sleep disordered breathing raises the possibility that this agent should be considered in individuals who have insomnia comorbid with sleep disordered breathing.[74] The lack of abuse potential of mirtazapine also suggests its potential use in those with insomnia who are

at risk for substance abuse, although there are no studies of the use of mirtazapine in this population.

Mirtazapine has few relative contraindications. Like other antidepressants, it should be used with caution in those with bipolar disorder because of the risk of precipitating mania.[62] It should also be used with caution in those with obesity, hepatic disease, and renal disease.

HYPOCRETIN/OREXIN ANTAGONISTS

Overview. Orexin A and B (also known as hypocretin A and B) are peptides released by neurons of the lateral hypothalamus that play an important role in mediating wakefulness and arousal.[78] Loss of the neurons releasing orexin is associated with narcolepsy.[78] Because orexin is thought to play an important role in mediating wakefulness, blockade of orexin receptors is expected to enhance sleep and is a mechanism of interest for developing treatments for patients with insomnia. One agent that blocks the receptors for both orexin A and B (as a result, it is referred to as a dual orexin receptor antagonist) is FDA approved for the treatment in dosages from 10 to 20 mg. The T_{max} of suvorexant is 3 hours, and the terminal $T_{1/2}$ is 9 to 13 hours.[79] The primary metabolism of suvorexant occurs through CYP3A and to a lesser degree CYP2C19.

Evidence of Efficacy. Suvorexant has been evaluated in four placebo-controlled trials in insomnia patients, which indicate that this agent has statistically significant dose-dependent therapeutic effects on both sleep onset and maintenance in the range from 10 to 40 mg.[78,80,81] Among these is a study that is unique among insomnia agents, which consisted of a 1-year, placebo-controlled trial of suvorexant 30 to 40 mg followed by a double-blind, randomized, placebo-controlled discontinuation phase.[80] This study demonstrated sustained efficacy over the entire year of dosing. After discontinuation, there was no evidence for an increase in adverse events. Among subjects receiving suvorexant, those continued on suvorexant maintained improvement compared with placebo subjects, whereas those switched to placebo were comparable to those who had been taking placebo throughout the double-blind and discontinuation phases. No statistically significant evidence for rebound insomnia was found after discontinuation of suvorexant, although the proportion of patients who met rebound insomnia criteria was numerically greater in subjects switched from suvorexant to placebo compared with the subjects who stayed on placebo throughout the study. Notably, like doxepin 3 to 6 mg, suvorexant has a therapeutic effect even in the last third of the night. However, unlike with doxepin, there is also evidence for a consistent effect on sleep onset.

Profile of Adverse Effects. The most important side effect of suvorexant is daytime sleepiness and drowsiness. Patients should be cautioned about potential driving impairment with this agent. This agent appears to have no significant risks of weight gain or cardiac or sexual side effects. The use of this agent is contraindicated in those with narcolepsy on the basis of its mechanism of action and the evidence that loss of orexin neurons is associated with this disorder. There is no evidence of dependence occurring during 1 year of regular use.[80] The abuse potential of suvorexant is believed to be comparable to that of zolpidem. One study examined the use of suvorexant 30 to 40 mg in patients with mild to moderate chronic obstructive pulmonary disease and did not find evidence for a respiratory depressant effect in this high-risk population.[82]

Potential Clinical Utility. Suvorexant has therapeutic effects throughout every portion of the night without substantively increasing the risk of morning sedation compared with other agents. This agent shares with doxepin 3 to 6 mg the distinction of being able to improve sleep in the last third of the night at modest or no cost of morning sedation. However, suvorexant, unlike doxepin, also has a consistent sleep-onset effect. As a result, this agent may be particularly well suited for treating patients with early morning awakening who also have difficulty falling asleep. The study in patients with chronic obstructive pulmonary disease also suggests that suvorexant might be particularly useful in this poplation.[82] Suvorexant is also the only drug for which we have systematic, randomized discontinuation data after a year of nightly use.[80] These data support the efficacy and safety of using this agent in those in whom long-term therapy may be needed. Further refinement of the role of suvorexant in insomnia pharmacotherapy will follow more widespread clinical use.

ANTIPSYCHOTICS

Overview. Antipsychotic agents are sometimes used to treat insomnia in clinical practice.[3] In general, these agents are used in dosages that are lower than typically used in their FDA-approved indications, which include the treatment of psychotic disorders (including schizophrenia and schizoaffective disorder), mania, and, in some cases, bipolar and unipolar major depression. The antipsychotic agents most commonly used to treat insomnia include quetiapine and olanzapine.[45,83] When it is used to treat insomnia, quetiapine is generally dosed from 25 to 250 mg; olanzapine is typically administered at a dosage of 2.5 to 20 mg. The primary effects of these agents are antagonism of the following receptors: dopamine, histamine (H_1), serotonin, muscarinic cholinergic, and adrenergic ($alpha_1$).[3] Olanzapine has a T_{max} of 4 to 6 hours and $T_{1/2}$ of 20 to 54 hours, which makes it relatively unlikely to be effective for sleep-onset difficulty when it is taken near bedtime. However, it is likely to have a prolonged sleep-enhancing effect.[3] Olanzapine elimination may be affected by factors that alter the liver enzymes CYP1A2 and CYP2D6. This means that there is slower elimination among females, tobacco smokers, and individuals of Japanese descent. Quetiapine, with a T_{max} of 1 to 2 hours and a $T_{1/2}$ of approximately 7 hours, has the potential to improve sleep onset as well as sleep maintenance on the basis of pharmacokinetics. The metabolism of quetiapine may be affected by factors that alter the CYP3A4 and CYP2D6 liver enzymes.[3]

Evidence of Efficacy. There are no placebo-controlled trials of any antipsychotic agent for the treatment of primary insomnia patients.[3] However, quetiapine 25 to 75 mg improved self-reported and polysomnographic indices of sleep in an open-label study in primary insomnia patients.[84] Both quetiapine and olanzapine have been found to improve sleep in small studies of normal sleepers.[83] These agents have also been found to improve sleep in open-label and placebo-controlled

studies in patients with schizophrenia, mania, bipolar depression, and unipolar depression.[83,85,86]

Profile of Adverse Effects. The primary side effects of the antipsychotic agents used to treat insomnia include orthostatic hypotension, dizziness, dry mouth, constipation, blurred vision, urinary retention, increased appetite, weight gain, and sedation.[82] Because antipsychotics antagonize dopamine receptors, they may lead to extrapyramidal side effects, such as parkinsonism, acute dystonic reactions, akathisia, and tardive dyskinesia.[84] However, these side effects are relatively uncommon in the newer generation antipsychotic agents, which includes those most commonly used to treat insomnia. Olanzapine has also been associated with cognitive impairment, glucose intolerance, and an elevated risk of mortality among patients with dementia.[87] Antipsychotic agents do not have significant abuse potential.

Potential Clinical Utility. Antipsychotic agents are best suited for use in the treatment of insomnia occurring in individuals with psychosis or bipolar disorder. Quetiapine is also FDA approved for use in the treatment of bipolar and unipolar major depression, suggesting its potential use for the treatment of insomnia occurring in patients with these disorders. Given the lack of abuse potential, antipsychotics have potential for use in patients with insomnia and substance abuse.

Because of their adverse effect profiles, antipsychotic agents commonly used to treat insomnia should be used with caution in those with a history of myocardial infarction, ischemia, or conduction abnormalities; closed-angle glaucoma; decreased gastrointestinal motility; urinary retention; or hypotension. These agents should also be used with caution in those with hepatic disease or when weight gain would be of significant concern. Olanzapine is also relatively contraindicated in those with dementia because of a reported increase in risk of mortality in this population.[87]

ANTICONVULSANTS

Gabapentin and Pregabalin

Overview. Gabapentin and pregabalin are structural analogues of GABA but are thought to exert their primary central nervous system effects by binding to the alpha$_2$-delta subunit of N-type voltage-gated calcium channels, thereby diminishing the release of glutamate and norepinephrine.[88,89] Gabapentin and pregabalin are approved by the FDA with indicated use in the treatment of partial seizures and pain. Pregabalin is also indicated for the treatment of fibromyalgia. These agents are also used to treat insomnia, periodic limb movement disorder, restless legs syndrome, and bipolar disorder. Gabapentin has a T_{max} of 3 to 3.5 hours and a $T_{1/2}$ of 5 to 9 hours and is generally used at a dosage of 100 to 900 mg for insomnia therapy. Because of the relatively slow absorption, gabapentin is relatively less likely than other agents to improve sleep onset when it is dosed at bedtime. Pregabalin is more rapidly absorbed with a T_{max} of 1 hour and $T_{1/2}$ of 4.5 to 7 hours. Both agents are eliminated by renal excretion, so factors that affect renal function can affect the elimination of these agents.[88,89]

Evidence of Efficacy. Whereas neither of these agents has been evaluated in a placebo-controlled trial of the treatment of primary insomnia, evidence of a sleep-enhancing effect has been noted in studies of the treatment of normal sleepers, pain patients, patients with restless legs syndrome, patients with generalized anxiety disorder, and epilepsy patients.[90-94] Gabapentin also improved sleep in a small placebo-controlled trial of the treatment of insomnia occurring in alcohol-dependent individuals.[95] A small open-label pilot study also suggests that gabapentin may have utility in treating alcohol withdrawal.[96]

Profile of Adverse Effects. The most common adverse effects of gabapentin are sedation, dizziness, ataxia, and diplopia. The most common side effects of pregabalin include sedation, dizziness, dry mouth, cognitive impairment, and increased appetite. Rarely, leukopenia has been noted with these agents. Gabapentin appears to have minimal abuse potential; however, abuse potential may be a consideration with pregabalin.[97]

Potential Clinical Utility. Gabapentin and pregabalin should be considered for treatment of insomnia occurring in patients with pain or partial seizures because of their efficacy in the treatment of those conditions. Pregabalin, an FDA-approved treatment for fibromyalgia, should be considered in treatment of patients with insomnia occurring with fibromyalgia. Both agents may be useful in treating patients with insomnia occurring with restless legs syndrome or periodic leg movement disorder. On the basis of the two small studies, gabapentin has potential utility in treating insomnia occurring in alcohol-dependent individuals and those undergoing alcohol withdrawal. Given this potential and the absence of abuse liability, gabapentin may also be useful in the treatment of the insomnia that frequently occurs in abstinent alcoholics.

Because both of these agents are renally eliminated, they should be used with caution in individuals with renal impairment or when there is any factor that impedes renal function. Caution should also be used in prescribing pregabalin to abuse-prone individuals.

Tiagabine

Overview. Tiagabine is a GABA reuptake inhibitor that is FDA approved for the treatment of partial seizures. It has been evaluated for the treatment of insomnia at dosages from 2 to 16 mg.[98,99] Tiagabine has a T_{max} of 1 to 1.5 hours and a $T_{1/2}$ of approximately 8 hours. It is primarily metabolized by the liver CYP3A4 isoenzyme, and as a result, factors that affect that enzyme can affect the elimination of tiagabine.

Evidence of Efficacy. Tiagabine has been evaluated in four placebo-controlled trials of the treatment of primary insomnia in adults and the elderly.[98-101] Whereas this agent reliably increases slow wave sleep in a dose-dependent manner, it does not improve sleep onset and has not consistently improved measures of sleep maintenance compared with placebo.

Profile of Adverse Effects. The most frequent side effects of tiagabine in placebo-controlled trials were sedation, dizziness, and nausea. There is also a small risk of precipitating seizures.

Potential Clinical Utility. The use of tiagabine could be considered in those patients requiring treatment for partial seizures who have accompanying insomnia. Tiagabine should be used with caution in those with significant hepatic disease.

ANTIHISTAMINES

The designation of an agent as an antihistamine generally indicates that the agent was developed or marketed with the treatment of allergies as its primary intended use. However, many other agents that are potent antagonists of the H_1 histamine receptor but were developed for the treatment of other conditions are not generally referred to as antihistamines. This includes the antidepressants doxepin and mirtazapine, which have more selective H_1 antagonist effects than any agents designated as antihistamines, as well as antipsychotics including olanzapine and quetiapine.[3] Agents designated as antihistamines that cross the blood-brain barrier include diphenhydramine, doxylamine, chlorpheniramine, and hydroxyzine. Of these, the agents most commonly used to treat insomnia are diphenhydramine and doxylamine, constituents of nearly all over-the-counter insomnia medications.

Diphenhydramine

Overview. Diphenhydramine is generally dosed at 25 to 50 mg for the treatment of insomnia. H_1 antagonism is its greatest effect; muscarinic cholinergic antagonism is the next most important effect of diphenhydramine.[102] This agent has a T_{max} of 2 to 2.5 hours and a $T_{1/2}$ of 5 to 11 hours.[3] The elimination of diphenhydramine primarily involves the CYP2D6, CYP1A2, CYP2C9, and CYP2C19 liver P450 isoenzymes.[3]

Evidence of Efficacy. Sleep-enhancing effects of diphenhydramine have been observed in placebo-controlled trials of sleep disturbance in a mixed group of psychiatric patients, outpatients in a primary care practice, nursing home residents, and primary insomnia patients.[102-107] Of note, the effects appear to be more consistent on sleep maintenance than on sleep onset. Whereas this may, at least in part, reflect the T_{max} of 2 hours, a greater effect on the maintenance of sleep than on sleep initiation was also noted with doxepin in low dose, at which it is a highly selective H_1 antagonist. In a study examining the sedating effects associated with daytime dosing of diphenhydramine, a sedating effect on the first day of treatment was no longer present on the fourth day of treatment.[108]

Profile of Adverse Effects. The most important adverse effects of diphenhydramine include sedation, dizziness, psychomotor impairment, cognitive impairment, dry mouth, blurred vision, constipation, urinary retention, and weight gain. Less commonly, diphenhydramine is associated with agitation and insomnia. Diphenhydramine does not have substantial abuse potential.

Potential Clinical Utility. The most promising use of diphenhydramine would be expected to be in the treatment of individuals with sleep difficulty occurring in conjunction with allergy symptoms or upper respiratory infections. The only data relevant to this potential use are from a study of 100 children with nocturnal cough in which diphenhydramine did not improve sleep compared with placebo according to parental reports.[109]

The primary relative contraindications to diphenhydramine use are closed-angle glaucoma, decreased gastrointestinal motility, urinary retention, asthma, chronic obstructive pulmonary disease, and severe liver disease.

Doxylamine

Overview. The characteristics of doxylamine are similar to those of diphenhydramine. It is generally used for insomnia in the same dosage range (25 to 50 mg) and also has a $T_{1/2}$ of 10 to 12 hours. Doxylamine has a slightly faster absorption with a T_{max} of 1.5 to 2.5 hours.[3] The primary elimination of doxylamine occurs through the CYP2D6, CYP1A2, and CYP2C9 liver isoenzymes.

Evidence of Efficacy. No clinical trials of insomnia treatment with doxylamine have been reported. However, one double-blind placebo-controlled trial demonstrated a significant therapeutic effect on self-reported sleep with doxylamine 25 mg in 2931 postoperative patients.[110]

Profile of Adverse Effects. The adverse effects profile of doxylamine is essentially the same as that of diphenhydramine, with the exception of case reports that coma and rhabdomyolysis can occasionally occur with the use of doxylamine.[111]

Potential Clinical Utility. The most promising clinical use of doxylamine is as a treatment of insomnia in those with associated allergy symptoms or upper respiratory infections. Relative contraindications to the use of doxylamine are the same as with diphenhydramine: closed-angle glaucoma, decreased gastrointestinal motility, urinary retention, asthma, chronic obstructive pulmonary disease, and severe liver disease.

CHLORAL HYDRATE

Overview. Chloral hydrate is thought to exert its effect through its metabolite trichloroethanol, which modulates $GABA_A$ receptors in a manner similar to the barbiturates.[112] The T_{max} of the active metabolite of chloral hydrate is approximately 30 minutes, and the $T_{1/2}$ is 7 to 10 hours. The dosage range for the treatment of insomnia is 250 mg to 2 g. Trichloroethanol undergoes hepatic conjugation with subsequent renal elimination.

Evidence of Efficacy. Chloral hydrate has been studied in placebo-controlled trials in elderly insomnia patients and mildly demented older adults.[113,114] A significant advantage over placebo was found only in the trial carried out in the mildly demented patients.[114]

Profile of Adverse Effects. The primary adverse effects of chloral hydrate include sedation, nausea, diarrhea, psychomotor impairment, and parasomnias.[112] There is significant abuse and dependence potential with this agent.[112] Hepatic and renal damage can also occur with chloral hydrate treatment. There is also a significant risk of death with overdoses of chloral hydrate.[112]

Potential Clinical Utility. Given the potential for serious adverse effects with this agent and that there are no conditions for which it has unique efficacy, it has been reasonably argued that this agent should not be used for the treatment of insomnia.[101] Whereas chloral hydrate is not recommended for insomnia therapy, there are some conditions for which its use is of special concern. This includes those at risk for suicide, neonates, substance abuse–prone individuals, those with

gastritis, patients with intermittent porphyria, and those with hepatic or renal failure.

SODIUM OXYBATE

Overview. Sodium oxybate is indicated by the FDA for the treatment of cataplexy and daytime sleepiness in those with narcolepsy. The mechanisms of action of this agent are unknown. It is dosed in the range of 2.5 to 4 g. The T_{max} is 0.5 to 0.8 hours and the $T_{1/2}$ is 0.3 to 1.2 hours.[51] Because the half-life is so short, it is dosed both at bedtime and again in the middle of the night. The metabolism of sodium oxybate is incompletely understood; however, enzymatic breakdown appears to be the most important element, involving a cytosolic enzyme gamma-hydroxybutyrate dehydrogenase and a mitochondrial transhydrogenase.

Evidence of Efficacy. Sleep-promoting effects of sodium oxybate have been demonstrated in terms of both onset and maintenance of sleep in placebo-controlled trials in patients with narcolepsy and fibromyalgia.[114,115] No trials of this agent have been carried out in any other groups of insomnia patients.

Profile of Adverse Effects. The primary adverse effects of sodium oxybate are headache, nausea and vomiting, excess salivation, parasomnias, and amnesia. Sodium oxybate has significant abuse and dependence potential. It can lead to coma and delirium in overdose.[51]

Potential Clinical Utility. Sodium oxybate should be considered a potential therapy for the treatment of sleep disturbance in narcolepsy patients and in those with fibromyalgia. Because of the side effect profile of sodium oxybate, it is relatively contraindicated in individuals vulnerable to respiratory depression, those with sleep disordered breathing, patients with chronic obstructive pulmonary disease, and those at risk for substance abuse.[51]

HERBALS

The use of plants for medicinal purposes dates back to antiquity. In the United States, herbals that are not claimed to treat a disorder are regulated differently from drugs. Such herbals are regulated by the Dietary Supplement Health and Education Act of 1994 and not according to the procedures the U.S. FDA employs for approving and overseeing prescription medications. Whereas herbal dietary supplements undergo regulation in terms of safety, it is critical to note that manufacturers of herbal remedies are not required to provide data on the safety, efficacy, or purity of their products. As a result, the constituents, effectiveness, and side effect profile of any given product remain uncertain.[116] Further complicating the situation for practitioners and consumers is that for many herbals, the active constituents and their interactions are not fully established. At the same time, as plant derivatives, the constituents of herbal preparations will vary according to the particular species of plant, where and when it was cultivated, and how it was processed.[116] Despite these limitations, herbal medications are frequently used around the world for the treatment of insomnia. Valerian, one of the most common herbal insomnia therapies, is reviewed here. This review is limited by uncertainty as to the identity of the active constituents and the composition of any of the valerian preparations studied.

Valerian

Overview. Valerian is an herbal substance derived from one of the more than 150 plants of the *Valeriana* genus that grow in temperate regions in North America, Europe, and Asia.[116] Of these plants, the most commonly used for medicinal purposes is *Valeriana officinalis L.* It is believed that this substance has been used for therapeutic purposes for more than a thousand years. Because the active ingredients of valerian are unknown, it is not possible to establish its pharmacokinetics or route of metabolism. The dosages of valerian most commonly employed in studies of insomnia treatment are 400 to 900 mg/day.

Evidence of Efficacy. Because of the uncertainty as to the dosing of active ingredients in the preparations of valerian studied, the evidence for efficacy in placebo-controlled studies of insomnia treatment with valerian can be reviewed only in general terms. Whereas a few studies have reported a therapeutic effect on sleep, the majority of placebo-controlled trials of the treatment of insomnia with valerian preparations have failed to find a therapeutic benefit over placebo.[117-120] At this point, it must be concluded that the efficacy of valerian in the treatment of insomnia is unestablished.

Profile of Adverse Effects. Valerian preparations have been associated with few side effects. Most studies have noted comparable rates of adverse effects in valerian- and placebo-treated patients. There is no evidence for significant abuse potential, and daytime sedation appears to be relatively uncommon.

Potential Clinical Utility. Given the uncertainty as to the efficacy of valerian preparations, it makes sense to limit the clinical use of this herbal therapy to individuals who are specifically interested in taking an herbal therapy for their insomnia. Four cases of hepatitis have been reported in individuals taking valerian in combination with skullcap, another herbal therapy.[116] The role of valerian in causing the hepatitis is uncertain; however, it is recommended that caution be used in combining valerian with skullcap and in treating those with liver disease.

SUMMARY OF CLINICAL CIRCUMSTANCES IN WHICH AGENTS OTHER THAN BZRAs MIGHT BE PREFERRED

As discussed, there are a number of circumstances in which the agents reviewed in this chapter might be preferred to BZRAs for the treatment of insomnia. One of these is when patients appear to be nonresponsive to or intolerant of several BZRAs. The other circumstances include the treatment of insomnia in individuals with substance abuse, allergies, psychosis, mania, or hypomania and the treatment of insomnia in individuals with major depression or chronic pain for whom single-agent therapy is desired.

Substance Use Disorders

Unlike the BZRAs, nearly all of the agents discussed in this chapter have minimal abuse potential. As a result, they might

be preferred in the treatment of those with a predisposition to substance abuse or those who are experiencing insomnia in the setting of substance abuse. The exceptions to this are sodium oxybate, chloral hydrate, pregabalin, and suvorexant, which have abuse potential. Among the non-BZRAs without abuse potential, only gabapentin and trazodone have been studied for the treatment of insomnia in substance abuse populations. Gabapentin was found to improve sleep and to have an acceptable safety profile in small studies of alcohol-dependent individuals and those going through alcohol withdrawal. Trazodone was found to be an effective treatment for insomnia in abstinent alcoholics in a placebo-controlled study. The utility of other non-BZRAs in the treatment of insomnia in those with substance use disorders remains to be evaluated.

Allergy-Associated Sleep Disturbance

A number of the non-BZRA agents appear to have sleep-enhancing effects as well as significant antihistaminergic effects (H_1 antagonism). As a result, they might be expected to be particularly useful in the treatment of insomnia occurring in association with allergies. These agents include doxepin, mirtazapine, diphenhydramine, doxylamine, amitriptyline, trimipramine, trazodone, olanzapine, and quetiapine. No placebo-controlled trials of the treatment of insomnia occurring with allergic rhinitis have been carried out.

Insomnia Occurring with Psychosis

Quetiapine and olanzapine are agents with antipsychotic efficacy that have been used to treat insomnia. The sleep-enhancing effects of these agents have been demonstrated in trials in normal sleepers or small open-label studies in those with schizophrenia.[83] Olanzapine and quetiapine should be considered for the treatment of sleep disturbance in individuals with psychosis. No placebo-controlled trials of the treatment of insomnia in patients with psychosis have been carried out.

Insomnia Occurring with Mania or Hypomania

Quetiapine and olanzapine also are indicated for the treatment of mania. These agents would be natural to consider for sleep difficulties in individuals with mania or hypomania, although no systematic trials of their effects specifically on sleep in this setting have been carried out.

Single-Agent Therapy for Insomnia Occurring with Major Depression

Several BZRAs may be effective and safe for the treatment of insomnia occurring with major depression. These studies have evaluated eszopiclone, zolpidem, zolpidem CR, and clonazepam as adjunctive therapy to antidepressant medications.[3] However, in some instances, single-agent therapy that addresses both the depression and the sleep disturbance may be preferred. Such instances include those in which cost or the complexity of the medication regimen is the primary concern. The importance of the complexity of the medication regimen is that it may contribute to nonadherence in some patients, particularly those in whom cognitive function is impaired. In such instances, a single agent that has established antidepressant effects as well as sleep-enhancing effects may be preferred. Of the non-BZRAs discussed in this chapter, a number of agents have demonstrated antidepressant effects and are therefore promising for use in this setting. These include amitriptyline, trimipramine, doxepin, mirtazapine, trazodone, and quetiapine. Of these, only trimipramine and doxepin have demonstrated efficacy in the treatment of insomnia in placebo-controlled trials. Another consideration is that for some agents, the antidepressant dosage is substantially higher than what is typically used to treat insomnia. Such a difference in dosage range precludes their utility for single-agent therapy for insomnia and depression or anxiety. Agents for which the dosing range for depression and insomnia is not substantially different and, as a result, could be considered for this purpose include trimipramine, doxepin, amitriptyline, mirtazapine, and quetiapine.

Single-Agent Therapy for Insomnia Occurring with Chronic Pain

The BZRAs eszopiclone and triazolam have been found to be safe and effective for the treatment of insomnia occurring with chronic pain.[3] However, as with single-agent therapy for insomnia occurring with depression, there may be circumstances in which single-agent therapy for insomnia occurring with chronic pain is desirable. Agents with sleep-enhancing effects that appear to have therapeutic effects on chronic pain include amitriptyline, doxepin, trimipramine, gabapentin, and pregabalin. Whereas none of these medications has been evaluated in controlled trials for the treatment of insomnia occurring with chronic pain, these agents could be considered for use when single-agent therapy for these conditions is indicated.

CLINICAL PEARL

Whereas the empiric support for the efficacy and safety of BZRAs in the treatment of insomnia is generally stronger than for non-BZRAs, there are some clinical circumstances in which non-BZRAs would be preferred to BZRAs, including intolerance of BZRAs, nonresponse to BZRAs, treatment of insomnia occurring with substance use disorders, allergy-associated sleep disturbance, insomnia occurring with psychosis, insomnia occurring with mania or hypomania, insomnia occurring with depression for which single-agent therapy is desired, and insomnia occurring with chronic pain for which single-agent therapy is desired.

SUMMARY

A number of agents other than BZRAs are used in the treatment of insomnia. This includes agents classified as antidepressants, antipsychotics, antihistamines, anticonvulsants, melatonin agonists, and herbals as well as chloral hydrate and sodium oxybate. These agents vary in the extent to which studies have been carried out on their efficacy and safety in the treatment of insomnia. In general, the empiric support for the efficacy and safety of the BZRAs is far greater than for the non-BZRA agents. Exceptions to this include doxepin, trimipramine, melatonin, and ramelteon. Each of the non-BZRA agents has unique properties that should be considered in determining whether they might be of utility for the treatment of a given patient with insomnia. In addition, there are some clinical circumstances in which non-BZRAs would be preferred to BZRAs. Awareness of the characteristics of the non-BZRA agents and of the clinical circumstances in which they might be the preferred pharmacotherapy is important for

optimizing the pharmacotherapy of insomnia in clinical practice.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grant NR010539.

Selected Readings

Krystal A. A compendium of placebo-controlled trials of the risks/benefits of pharmacologic treatments for insomnia: the empirical basis for clinical practice. *Sleep Med Rev* 2009;**13**:265–74.

Krystal AD. Current, emerging, and newly available insomnia medications. *J Clin Psychiatry* 2015;**76**(8):e1045.

Saper C, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* 2005;**437**:1257–63.

Spadoni G, Bedini A, Lucarini S, et al. Pharmacokinetic and pharmacodynamic evaluation of ramelteon: an insomnia therapy. *Expert Opin Drug Metab Toxicol* 2015;**11**(7):1145–56.

Winrow CJ, Renger JJ. Discovery and development of orexin receptor antagonists as therapeutics for insomnia. *Br J Pharmacol* 2014;**171**:283–93.

Zisapel N. Current phase II investigational therapies for insomnia. *Expert Opin Investig Drugs* 2015;**24**(3):401–11.

A complete reference list can be found online at ExpertConsult.com.

Section 12

Neurologic Disorders

Chapter 89

Narcolepsy: Genetics, Immunology, and Pathophysiology

Emmanuel Mignot

Chapter Highlights

- Narcolepsy is a sleep disorder characterized by excessive daytime sleepiness and abnormal transitions to REM sleep.
- When cataplexy is present, the cause is almost always an immune-mediated destruction of the hypocretin-producing (also called orexin) neurons in the hypothalamus, an abnormality that can be documented by measuring cerebrospinal fluid hypocretin-1. This disorder is now called *type 1 narcolepsy* in the new classification of sleep disorders.
- The treatment of narcolepsy typically uses two or three of the following drug classes: antidepressants (for cataplexy, acting on adrenergic and serotoninergic uptake), modafinil or amphetamine-like stimulants (for sleepiness, most likely acting on dopamine), and sodium oxybate (a sedative acting on all symptoms, most likely acting on gamma-aminobutyric acid B receptors).

In the original consensus description, narcolepsy was characterized by "excessive daytime sleepiness associated with cataplexy and other REM sleep phenomena such as sleep paralysis and hypnagogic hallucinations."[1] Cataplexy, the sudden occurrence of muscle weakness in association with laughing, joking, or anger, has long been considered a pathognomonic symptom for the disorder.[2-4] Because cataplexy is generally easy to recognize, a large number of studies have shown that the prevalence of narcolepsy with cataplexy is approximately 0.02% to 0.05%.[5-7] Slowly, a broader definition of narcolepsy has emerged that includes patients with sleepiness and abnormal rapid eye movement (REM) sleep, such as sleep-onset REM periods (SOREMPs) during the Multiple Sleep Latency Test (MSLT), sleep paralysis, or hypnagogic hallucinations ("narcolepsy without cataplexy"). Disturbed nocturnal sleep is less frequently mentioned but is also frequent and disabling.[8]

The definition of narcolepsy has recently been revised owing to rapidly evolving scientific discoveries.[9] In most narcolepsy patients with cataplexy and in a few without cataplexy, a deficiency in the hypocretin neuropeptide system is causal,[10-13] and this entity has been renamed *type 1 narcolepsy*.[9] Further, an increasing number of patients have been recognized in early childhood or close to disease onset, and in these individuals, cataplexy can be atypical, with spontaneous grimaces or jaw opening episodes with tongue thrusting or global hypotonia without any obvious emotional triggers.[14] A tight genetic

Table 89-1 International Classification of Sleep Disorders: Definitions and Pathophysiology

Condition	Diagnostic Criteria*	Pathophysiology
Type 1 narcolepsy	Presence of ≥2 of the following: cataplexy, positive MSLT, and low CSF hypocretin-1	Hypocretin deficiency 98% HLA-DQB1*06:02
Type 2 narcolepsy	Positive MSLT; most often with no or unclear cataplexy	Unknown, heterogenous ~16% Hypocretin deficiency ~40% HLA-DQB1*0602
Secondary narcolepsy	As above, but due to neurologic conditions	With or without hypocretin deficiency; various disorders (see Table 89-3)
Idiopathic hypersomnia	No cataplexy, no SOREMPs during the MSLT	Unknown, likely heterogeneous

*Abnormal Multiple Sleep Latency Test (MSLT): sleep latency ≤8 min, ≥2 sleep-onset REM periods (SOREMPs), including a nocturnal SOREMP. For details, see *International Classification of Sleep Disorders*, third edition.[9]
CSF, Cerebrospinal fluid.

association with the human leukocyte antigen (HLA)-DQB1*06:02 is also found in patients with cataplexy,[15] suggesting an autoimmune mediation of the hypocretin cell loss. In contrast, narcolepsy without cataplexy (type 2) is generally associated with normal cerebrospinal fluid (CSF) hypocretin and may not be HLA positive, suggesting a different etiology. The prevalence of type 2 narcolepsy is unknown[16] because systematic MSLT testing in large portion of the general population would need to be performed, with exclusion of confounding etiologies. Studies have shown several percent of the population have a positive MSLT,[17,18] but only approximately 0.2% may have real narcolepsy,[18] and only 0.02% are diagnosed.[6] As a result, in the most recent revision of the *International Classification of Sleep Disorders*, narcolepsy with and without involvement of hypocretin has been separated (Table 89-1).[9]

Narcolepsy treatments have also rapidly evolved, yet all remain symptomatically based.[19,20] Excessive daytime sleepiness is typically treated by amphetamine-like stimulants or modafinil. These compounds are effective in reducing daytime sleepiness but have little effect on cataplexy and abnormal REM sleep. Conversely, the most commonly used treatments for cataplexy, antidepressant drugs, alleviate cataplexy and other REM sleep abnormalities but have little effect on daytime sleepiness. Gamma-hydroxybutyrate (GHB or sodium oxybate) is effective in alleviating disturbed nocturnal sleep, cataplexy, and to a lesser extent daytime sleepiness. In this chapter, information regarding the pathophysiology and pharmacology of narcolepsy will be reviewed. The clinical, diagnostic, and therapeutic aspects are discussed in Chapter 90.

ANIMAL MODELS OF TYPE 1 NARCOLEPSY

In the past 20 years, narcolepsy research has been facilitated by the existence of a unique animal model, canine narcolepsy. Knecht and colleagues[21] and Mitler and Dement[22] first reported the existence of canine narcolepsy in 1973. Early attempts to establish genetic transmission were unsuccessful, suggesting a nongenetic etiology in most cases of canine narcolepsy. In 1975, two narcoleptic Dobermans were reported in a single litter.[23] Subsequent breeding experiments found autosomal recessive transmission with full penetrance, allowing the establishment of a Doberman Pinscher colony at Stanford University. Familial canine narcolepsy was also

Figure 89-1 Narcoleptic Doberman pinschers during attacks of cataplexy. Note that the eyes are open. Autosomal recessive forms of canine narcolepsy are due to mutations in the hypocretin receptor-2 gene.[28]

reported in Labrador Retrievers and Dachshunds.[23,24] Interestingly, experiments indicate that animals heterozygous for the canine narcolepsy gene have subclinical abnormalities; for example, the administration of drugs that increase cholinergic and reduce monoaminergic transmission (manipulations known to promote REM sleep) has been shown to induce cataplexy at specific developmental times.[25]

The parallel between human type 1 narcolepsy and canine narcolepsy is striking. In MSLT-like procedures, narcoleptic canines have been shown to have short sleep and REM sleep latencies.[19] Twenty-four-hour recording studies show increased sleep fragmentation and proportionally more daytime sleep than in control animals.[26] Finally, as in human narcolepsy, sudden episodes of muscle weakness akin to cataplexy can be observed in association with strong positive emotions, most typically during the presentation of appetizing food or while at play (Figure 89-1). These episodes usually last a few seconds and preferentially affect the hind legs, neck, or face. Cataplexy may also escalate into complete muscle paralysis with abolition of tendon reflexes. During these episodes, the animal is conscious and most often able to visually track nearby movement (see Video 89-1). Polygraphic recording indicates a desynchronized, wakelike electroencephalogram (EEG) pattern at the onset of cataplexy, followed by

increased theta activity and genuine REM sleep in long-lasting episodes.[27]

The cause of autosomal recessive canine narcolepsy was identified in 1999 through positional cloning.[24,28] The pathology was found to be due to mutations in a receptor for the newly identified neuropeptide system hypocretin (hypocretin receptor-2, also known as orexin receptor-2). Three different mutations causing a complete dysfunction of the receptor were identified in Doberman, Labrador, and Dachshund pedigrees.[24,28] Sporadic cases of canine narcolepsy were later shown to be associated with low CSF hypocretin and almost absent brain hypocretin peptide content,[29] as found in human narcolepsy.[11,12]

Several rodent models of narcolepsy are now available. Chemelli and colleagues[30] developed preprohypocretin knockout mice, reporting fragmented sleep, rapid transitions from wakefulness into REM sleep, and a reversible state of paralysis akin to cataplexy.[30] In another model, toxic transgenes are driven by the hypocretin promoter, producing various level of hypocretin-containing cell loss and narcolepsy.[31,32] In this model, cell loss higher than 90% produces a syndrome similar to that of the preprohypocretin knockout mice. A rat model with partial hypocretin cell loss[33] and mice lacking either of the two hypocretin receptors, hypocretin receptor-1 and -2, are now also available.[34-37] In these models, only hypocretin receptor-2 knockout animals experience behavioral arrest episodes similar to cataplexy, although the episodes are few and not as clear as in hypocretin knockout mice.[34] Interestingly, hypocretin receptor-1 knockout animals are almost normal, having only mildly fragmented sleep and REM sleep abnormalities but no behavioral arrests.[35] Dual hypocretin receptor-1 and -2 knockout mice, however, recapitulate the phenotype of hypocretin knockout. Overall, data in mice suggest a role for hypocretin receptor-1 in increasing the severity of the phenotype.[34]

The use of these new rodent models, together with recent developments that make it possible to increase firing rate selectively in hypocretin cells through optogenic stimulation,[38] is revolutionizing research in this area. It is important to keep in mind, however, that humans are more genetically diverse than inbred rodent lines and have distinct ecologic niches, most likely explaining interspecies differences in the regulation and functions of hypocretin.[39] This may explain why hypocretin receptor-2 mutants have more clear cataplexy in dogs versus mice, or why hypocretin receptor-2 mutations causing narcolepsy have not yet been identified in humans (e.g., the phenotype may be too mild to raise concern). Similarly, food deprivation (up to 31 hours) in mice increases wakefulness, most likely because of the acute need to search for food, but this response is severely blunted in mice lacking hypocretin.[40] Yet, metabolically, mice have high metabolic demands and low energy reserves, unlike humans. Further, in contrast to these results, patients with narcolepsy/hypocretin frequently use food restriction to stay awake and have binge eating abnormalities, especially at night.[41-43] Similarly, we found that the small decrease in food intake found in rodents with hypocretin deficiency could largely be explained by wake fragmentation in these models.[33] Finally, as discussed later, pharmacologic responses may vary across species. As an example, the α_2 antagonist yohimbine is a strong wake-promoting compound and anticataplectic agent in canines, yet it has little if any effect in humans.[19]

PHARMACOLOGY OF NARCOLEPSY

Adrenergic Uptake Inhibition Mediates the Anticataplectic Effects of Antidepressants

In the past, the most commonly prescribed anticataplectic agents were tricyclic antidepressants, and clomipramine is still widely used as a treatment in some countries. More recently, more selective monoaminergic reuptake inhibitors have become available. These compounds have a complex pharmacologic profile that includes monoamine (serotonin, norepinephrine, epinephrine, and dopamine) uptake inhibition and, for older tricyclic antidepressants, cholinergic, histaminic, and α-adrenergic blocking effects.[19,,45] Whereas cataplexy is difficult to quantify in humans (because it is unpredictable) or mouse models (because it is difficult to differentiate from REM sleep or sleep paralysis), this symptom can be easily measured in narcoleptic canines using a simple behavioral tool, the Food Elicited Cataplexy Test. In this test, pieces of dog food are lined up on the floor, and the animal is released in the room.[19] A normal animal will complete the test in less than 10 seconds, whereas narcoleptic dogs, excited by the food, exhibit multiple partial or complete cataplexy attacks that can be recorded in number and duration. This test has been used to pharmacologically dissect the mode of action of currently prescribed anticataplectic agents.[19]

In narcoleptic canines, the therapeutic efficacy of antidepressants is mainly due to inhibition of adrenergic but not dopaminergic or serotonergic uptake.[44,45] This observation fits well with available human pharmacologic data (Table 89-2).[19] Indeed, protriptyline, desipramine, viloxazine, and atomoxetine, four adrenergic-specific uptake blockers with no effect on serotonin transmission, are effective and potent anticataplectic agents (see Table 89-2). In contrast, escitalopram, zimelidine, fluoxetine, and other serotonin-specific reuptake inhibitors suppress cataplexy only at relatively high doses, an effect probably mediated by the weak adrenergic uptake effects of these compounds and their metabolites.[44] Alternatively, it may be that, in humans, blocking the serotonin reuptake site has minor therapeutic effects on cataplexy in contrast to canines because of increased genetic diversity in humans or species differences.

The observation that adrenergic uptake blockers are anticataplectic agents correlates well with the potent inhibitory effects of these compounds on REM sleep. It is well established that adrenergic transmission is reduced during REM sleep, and firing in the locus coeruleus nearly ceases during cataplexy in narcoleptic canines.[46] Adrenergic uptake blockers might thus increase adrenergic tone in projection sites involved in REM sleep regulation. This effect would lessen the effect of decreased locus coeruleus impulse flow normally occurring during natural REM sleep. The fact that serotonergic uptake blockers, also known to have inhibitory effects on REM sleep, have less or no effect on cataplexy is more surprising. Like adrenergic cells of the locus coeruleus, serotonergic cells of the raphe nuclei dramatically decrease their activity during REM sleep. This discrepancy could be explained by a preferential effect of serotonergic projections on REM sleep features other than atonia (e.g., in the control of eye movements). In this model, adrenergic projections may be more important than serotonergic transmission in the regulation of REM sleep atonia and thus cataplexy.[44] In favor of this hypothesis, serotonergic activity decreases moderately during cataplexy

Table 89-2 Commonly Prescribed Treatments and Their Pharmacologic Properties

Compound	Pharmacologic Properties
Stimulants	
Amphetamine	Increases monoamine release (DA > NE ≫ 5-HT); blocks monoamine reuptake and MAO at high doses. The D-isomer is more specific for DA transmission and is a more effective stimulant
Methamphetamine	More lipophilic and potent than amphetamine; increased CNS penetration; proportionally less peripheral side effects
Methylphenidate	Blocks monoamine uptake at lower dose than amphetamine; slightly less effect on monoamine release; short half-life
Pemoline	DA uptake inhibition; low potency; hepatotoxicity
Selegiline (L-deprenyl)	Monoamine oxidase B inhibitor; in vivo conversion into amphetamine
Modafinil	Fewer peripheral side effects. Mode of action is likely DA uptake inhibition. R-modafinil is the major active, long-lasting enantiomer
Anticataplexy Compounds	
Venlafaxine	Specific serotonin and adrenergic reuptake blocker (5-HT = NE); very effective but some nausea
Atomoxetine	Specific NE reuptake blocker; also improves daytime sleepiness
Fluoxetine	Specific 5-HT uptake blocker (5-HT ≫ NE = DA). Active metabolite norfluoxetine has more NE effects. High therapeutic doses are often needed
Protriptyline	Tricyclic antidepressant; monoamine uptake blocker (NE > 5-HT > DA); anticholinergic effects
Imipramine	Tricyclic antidepressant; monoamine uptake blocker (NE = 5-HT > DA); anticholinergic effects Desipramine is an active metabolite
Desipramine	Tricyclic antidepressant; monoamine uptake blocker (NE ≫ 5-HT > DA); anticholinergic effects
Clomipramine	Tricyclic antidepressant; monoamine uptake blocker (5-HT > NE ≫ DA); anticholinergic effects Desmethyl-clomipramine (NE ≫ 5-HT > DA) is an active metabolite. No specificity in vivo

5-HT, Serotonin; CNS, central nervous system; DA, dopamine; MAO, monoamine oxidase; NE, norepinephrine.

in narcoleptic canines,[47] in contrast with locus coeruleus activity.[46]

Increased Dopaminergic Transmission Mediates the Wake-Promoting Effects of Stimulants

Commonly prescribed stimulant compounds include amphetamine-like drugs, such as dextroamphetamine, methamphetamine, methylphenidate, pemoline, and modafinil (see Table 89-2). Like tricyclic antidepressants, amphetamine-like drugs are nonspecific pharmacologically. Their main effect is to globally increase monoaminergic transmission by stimulating monoamine release and blocking monoamine reuptake. Abuse and dose escalation can occur with amphetamines, especially in patients without cataplexy. Less abuse is reported with methylphenidate, and modafinil is not believed to be addictive. Recent studies have shown that the wake-promoting effect of these compounds is secondary to dopamine release and reuptake inhibition.[48,49] The mode of action of modafinil is debated, but this compound also selectively inhibits dopamine uptake.[50,51] All these compounds are ineffective in dopamine transporter knockout mice, suggesting a primary promotion of wake through dopaminergic systems.[49] Interestingly, compounds selective for dopaminergic transmission have no effect on cataplexy, whereas amphetamine-like compounds with combined dopaminergic and adrenergic effects have some anticataplectic properties at high doses.[44,52] Adrenergic effects of amphetamine-like stimulants also correlate with the respective effects of these compounds on normal REM sleep.[19,52] Dopaminergic-specific uptake blockers have little effect on REM sleep compared with adrenergic or serotonergic compounds.[19] The most important effects of dopaminergic uptake blockers are to reduce total sleep time and slow wave sleep.[53] This preferential effect of dopaminergic uptake blockers on non–rapid eye movement (NREM) sleep correlates with electrophysiologic data; as opposed to adrenergic or serotonergic neurons, the firing rate of dopaminergic neurons remains relatively constant during REM sleep.[54,55]

Interestingly, studies in humans and narcoleptic canines have shown that large doses of stimulants are needed to polygraphically "normalize" narcolepsy symptoms. In our narcoleptic Doberman population, doses equivalent to 60 mg/day were needed to reduce daytime sleep to control levels.[26] In both control and narcoleptic animals, however, dose-response curves for modafinil or amphetamine were parallel. This result suggests that there is no difference in sensitivity to stimulants in narcoleptic animals but rather that higher doses are needed in narcoleptic animals because of their extreme baseline sleepiness.[26] However, a recent study has suggested increased wake-promoting effects of modafinil in rodents with hypocretin deficiency.[56]

Sodium Oxybate (Gamma-Hydroxybutyrate)

Sodium oxybate, also called GHB, is a sedative known to increase slow wave sleep and, to a lesser extent, REM sleep.[19] Abuse in the context of rave parties has been reported, and the prescription of the compound is highly regulated, with

distribution centralized by one pharmacy in the United States. Because slow wave sleep is associated with growth hormone release, GHB also induces growth hormone release and has been abused by athletes. When administered at night, it consolidates sleep and improves daytime functioning. Because of its short half-life (approximately 30 minutes),[57] it must be administered twice a night. Interestingly, cataplexy and daytime alertness improve gradually, and sometimes the full therapeutic effect is achieved only after several months of treatment and dose adjustments.[20]

The mode of action of GHB on sleep and narcolepsy is unclear. GHB has a major effect on dopamine transmission, reducing firing rate and raising brain content of dopamine.[19,58,59] Other effects on opioid, glutamatergic, and cholinergic transmission have been reported.[58] Specific GHB receptors have been identified, but the compound is also a gamma-aminobutyric acid B ($GABA_B$) agonist.[58,59] Most studies to date suggest that the sedative hypnotic effect is mediated through $GABA_B$ agonist activity.[58-61] Whether this effect also mediates the anticataplectic effects after long-term administration is unknown.

Other Known Modulators of Narcolepsy Symptoms

The effects of more than 200 compounds with various modes of action have been examined in human patients and narcoleptic canines.[19] In almost all cases except a few, similar effects were found in humans and canines.[19] Almost all the significant effects have been reported for monoaminergic and cholinergic compounds. Because cataplexy is easier to study than sleep in canines, most research in narcoleptic dogs has focused on cataplexy. For cataplexy, the findings were generally consistent with pharmacologic studies of REM sleep. As is the case for REM sleep, the regulation of cataplexy is modulated positively by cholinergic systems and negatively by monoaminergic tone.[19] Muscarinic M_2 or M_3 receptors mediate the cholinergic effects, whereas monoaminergic effects are mostly modulated by postsynaptic α_1-adrenergic receptors and presynaptic D_2/D_3 autoreceptors.[19]

Although the primary cause of narcolepsy is hypocretin deficiency, studies have shown abnormal cholinergic and monoaminergic receptor density and neurotransmitter levels in human or canine narcolepsy brain and CSF samples.[19,62-66] Local injection studies in selected brain areas of narcoleptic canines have also shown functional relevance for some of these abnormalities.[67-69] As a result, cholinergic hypersensitivity, dopaminergic abnormalities, and abnormal histaminergic tone are likely to be critical downstream mediators of hypocretin deficiency in the expression of the narcolepsy symptoms.[63,66-69] The cholinergic-monoaminergic imbalances observed in narcolepsy are best illustrated by the finding that in asymptomatic animals heterozygous for the hypocretin receptor-2 mutation, a combination of cholinergic agonists with an α_1 blocker or a D_2/D_3 agonist can trigger cataplexy.[25] Further, recent results suggest an increased number of histaminergic cells in the brains of narcolepsy patients[64,65] and rescue of sleepiness in hypocretin receptor-2 mutant animals by expression of the receptor in the posterior hypothalamus.[37] A possible application of these findings is illustrated by the recent development of histamine-3 (H_3) antagonists, drugs that are known to stimulate histamine release through the H_3 receptor, as novel wake-promoting stimulants for the treatment of narcolepsy.[70,71]

HYPOCRETIN INVOLVEMENT IN NARCOLEPSY

Anatomy and Physiology of the Hypocretin (Orexin) Neuropeptide System

The hypocretin-orexin system was discovered almost simultaneously by two groups of scientists, hence the two conflicting names. De Lecea and colleagues first isolated the preprohypocretin transcript and suggested the existence of two resulting processed neuropeptides sharing extensive homology with each other and weak homology with secretin.[72] These neuropeptides were called *hypocretin 1* and *hypocretin 2*, to indicate *hypo*thalamic peptides of the in*cretin* family. Meanwhile, using cell lines expressing various orphan G-protein-coupled receptors (GPCRs), Sakurai and colleagues screened tissue extracts for GPCR agonist activity.[73] Two mature peptides were isolated and called orexin-A and orexin-B.[74] The name was chosen from Greek *orexis* (for "appetite") based on associated studies suggesting stimulation of appetite after central administration. Sakurai and colleagues[74] also identified a second receptor with high homology. The two G-protein-coupled receptors were initially called orexin receptor-1 and orexin receptor-2,[74] but their official names are hypocretin receptor-1 and hypocretin receptor-2 in genetic databases. Hypocretin receptor-1 binds hypocretin-1 (OxA) selectively, whereas hypocretin receptor-2 (OxR2) binds both hypocretin-1 (OxA) and hypocretin-2 (OxB) with similar affinity.[73] The two receptors mostly couple with Gq and stimulate cellular activity in most cell types.[24,75,76]

Only a few thousand cell bodies containing these two peptides are found in the entire brain, all within the perifornical area of the posterior hypothalamus. Contrasting with this discrete perikaria localization, the neurons project widely in the central nervous system.[24,75,76] Limbic system areas (including amygdala and nucleus accumbens), monoaminergic cell groups (adrenergic locus coeruleus), serotoninergic raphe nuclei, dopaminergic ventral tegmental area and substantia nigra, histaminergic tuberomammillary nuclei, other hypothalmic nuclei, and various periventricular organs are densely innervated.[24,75,76] Hypocretin immunoreactive varicose terminals are also seen in the cerebral cortex, spinal cord, and thalamus. The strongest extrahypothalamic hypocretin projection is found in the locus coeruleus. Hypocretin has also been suggested to be present in selected peripheral tissue (testis, gut) at very low levels of expression.

Hypocretin (Orexin) Deficiency and Type 1 Narcolepsy

As expected from the observation that most cases of human narcolepsy are sporadic and not fully genetic as in dogs or mice, extensive genetic screening studies did not identify preprohypocretin or hypocretin receptor mutations in human narcolepsy.[11,24,77] Surprisingly, even familial cases of narcolepsy (some of which were HLA-DQB1*06:02 negative) did not have any hypocretin mutations, suggesting further heterogeneity in genetic cases.[11] Rather, only one person with a signal peptide mutation of the preprohypocretin gene has been identified. This individual had an extremely early onset (6 months), severe narcolepsy-cataplexy, DQB1*06:02 negativity, and undetectable hypocretin-1 CSF levels.[78] This important observation indicates that hypocretin system gene mutations can cause narcolepsy like in animal models.

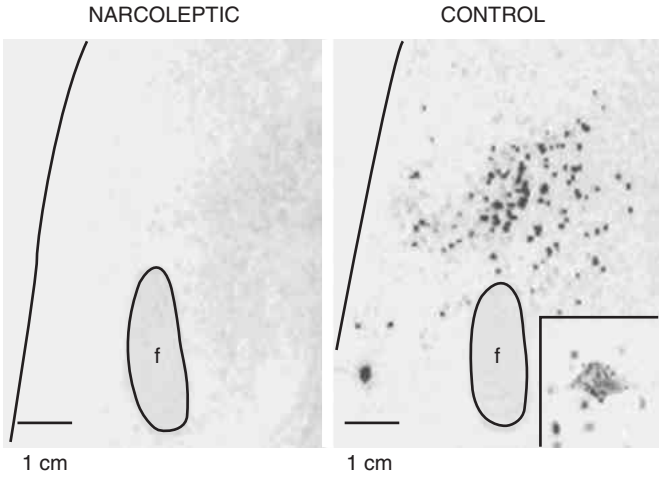


Figure 89-2 Hypocretin in the hypothalamus of control and narcoleptic subjects. Preprohypocretin messenger RNA molecules are detected in the hypothalamus of a control (*right*) but not a patient with narcolepsy (*left*). *Insert* shows exemplar high magnification of a preprohypocretin-positive neuron. f, Fornix, 3rd V, third ventricle. (Modified from Peyron C, Faraco J, Rogers W, et al. A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat Med* 2000;6[9]:991–7.)

Following on the cloning of the canine narcolepsy gene, we have also found that most sporadic, HLA-DQB1*06:02-positive narcolepsy patients with cataplexy have undetectable or low (≤110 pg/m) hypocretin-1 immunoreactivity in CSF.[10,79-83] Subsequent neuropathologic studies in 10 narcolepsy patients also indicated dramatic loss of hypocretin-1, hypocretin-2, and preprohypocretin messenger RNA (mRNA) in the brain and hypothalami of narcolepsy patients (Figure 89-2).[11,12] As mentioned earlier, these subjects have no hypocretin gene mutations and a peripubertal or postpubertal disease onset[84] as opposed to a 6-month onset in the subject with a preprohypocretin mutation.[11] Together with the tight HLA association,[85-87] a likely pathophysiologic mechanism in most narcolepsy patients could thus involve an autoimmune injury to hypocretin containing neurons in the central nervous system.

Hypocretin (Orexin) Transmission in Sleep Regulation

Since the discovery of the hypocretins, much has been learned about how they regulate sleep. Central (intracerebroventricular or local injections) but not peripheral administration of hypocretin-1 stimulates wakefulness and reduces REM sleep. Hypocretin antagonists are NREM and REM sleep promoting, including in humans.[88,89] In rats and monkeys, cisternal CSF hypocretin-1 fluctuates, with maximal levels at the end of the active period (night in rodents) and minimal levels at the end of the inactive period (amplitude is about 40% of the maximum).[90-92] Using in vivo dialysis, a similar profile is observed in rat brain tissue extracellular fluid.[90] In diurnal, wake-consolidated squirrel monkeys, cisternal hypocretin-1 levels also peak in the late evening, around bedtime (amplitude about 40%).[92] These results suggest that hypocretin may be important to promote wakefulness in the evening (in humans). In this model, hypocretin would oppose the sleep pressure that has accumulated since early morning, allowing a constant level of wakefulness through the day.[92] Additional studies suggest that diurnal fluctuations in hypocretin release are driven both directly by the circadian clock and indirectly by the increased sleep pressure.[39] It is also still unclear how hypocretin release and activity fluctuate across the various sleep stages (REM vs. NREM sleep). Of note, lumbar CSF hypocretin-1 only has limited diurnal fluctuation (10%), suggesting a dampening and delay of changes when reaching the lumbar sac (minimal levels in morning).[93] This finding is of practical importance because time of CSF collection has no significant effect for narcolepsy diagnostic purposes.[13,94]

The most dense reported hypocretin projections are to the monoaminergic cell groups of the locus coeruleus (norepinephrine), substantia nigra and ventral tegmental area (dopamine), raphe magnus (serotonin), and tuberomammillary (histamine) neurons. Dopamine and histamine cell groups have a very high hypocretin receptor-2 density[95] and may be especially important.[24,75,76,96] Increased dopamine level in the amygdala is one of the most consistent neurochemical abnormalities reported in canine narcolepsy.[62,66] Decreased histamine levels are also observed in the brain of narcoleptic canines,[97] although in humans, increased numbers of histidine decarboxylase–positive cells have been reported,[64,65] with decreased levels in the CSF.[98-100] In vivo dialysis studies have indicated a critical role for the dopamine mesolimbic and mesocortical system in the regulation of alertness and the triggering of cataplexy by emotions. Histaminergic transmission has long been recognized as a critical wake-promoting neurotransmitter.[101] Dopamine and histaminergic projections may thus be centrally involved in controlling both cataplexy and alertness.[37] Similar to REM sleep, cataplexy is controlled by pontine cholinergic REM-on cells and aminergic locus coeruleus REM-off cells.[19] Loss of excitatory hypocretin projections to monoamine cell groups could decrease monoaminergic tone and produce a cholinergic-aminergic imbalance, resulting in the sleepiness and abnormal REM sleep of narcolepsy. Hypocretin projections to the basal forebrain area, an area with cholinergic hypersensitivity in narcoleptic animals, are also likely to be involved.[19] It is also likely that hypocretin helps integrate sleep regulation with metabolic status, although the importance of this effect could be species dependent.

GENETIC EPIDEMIOLOGY OF HUMAN NARCOLEPSY

Familial Aspects of Human Narcolepsy

Westphal[102] was the first to report familial occurrence of narcolepsy-cataplexy in 1877. In recent studies, the risk for a first-degree relative to develop narcolepsy-cataplexy has been shown to be 1% to 2%.[5] A larger portion of relatives (4% to 5%) may have isolated daytime sleepiness, when other causes of daytime sleepiness have been excluded.[5] These figures are important to keep in mind because they are helpful in reassuring patients regarding the risk to their children and relatives. A 1% to 2% risk is 10- to 40-fold higher than in the general population but remains manageable. A 4% to 5% risk for daytime sleepiness is not negligible, but similar values have been reported for excessive daytime sleepiness in the general population independent of narcolepsy.[103-105]

Twin Studies and Environmental Factors in Narcolepsy

Approximately 25 reports of monozygotic twins with narcolepsy are available in the literature.[5] Seven to nine are

concordant for narcolepsy, depending how strictly concordance is determined clinically.[5,106-108] This low concordance thus suggests that narcolepsy requires the influence of stochastic and environmental factors for the pathology to develop. This is also substantiated by the fact that onset is not at birth but rather in adolescence,[84,109] suggesting the existence of triggering factors.

Upper Airway Infections such as Pandemic H1N1 2009 as Triggers of Narcolepsy

Upper airway infections such as influenza and *Streptococcus pyogenes* are now known to be triggers of narcolepsy onset, at least in children. In the past, head trauma,[110-112] sudden change in sleep-wake habits,[113] or various infections[114,115] were often cited. Starting in the 2000s, increased recognition of younger children seen closer to onset led to the realization that narcolepsy often followed strep throat and that recent-onset subjects were often positive for antistreptolysin-*O*, a marker of streptococcus pyogenes.[116] Epidemiologic studies also indicated a link.[7] Han and associates studying onset in more than 1000 Chinese patients found strong annual fluctuations of onset of narcolepsy in children, with onsets six times more frequent in spring and summer than in winter (Figure 89-3), suggesting that winter infections may trigger a process that over a few months results in hypocretin cell loss.[117]

In the spring of 2010, a number of events converged to indicate that the 2009 H1N1 pandemic influenza significantly triggered the onset of narcolepsy in young children (see Figure 89-3). To review the background: In the spring of 2009, a new strain of influenza A (H1N1), likely of swine origin, appeared in Mexico, spreading rapidly in humans and affecting young adults with a high reported case fatality rate of 0.4%.[118] This caused alarm, as with such a high mortality rate, millions of death were likely worldwide when the new virus would hit the world population the following winter. Faced with such a threat, the World Health Organization (WHO) and other organizations encouraged vaccine makers to initiate large-scale production of vaccines targeting the new strain, which had not been included in the 2009–2010 regular trivalent seasonal flu vaccine.[119] To generate these vaccines, almost all manufacturers used a A/California/7/2009 (H1N1)-pdm09-like reassortant virus containing hemagglutinin HA type 1, neuraminidase NA type 1 (thus H1N1), and polymerase basic 1 (PB1) proteins from A/California/7/2009, on a backbone H1N1 virus PR8, derived from an older, A/Puerto Rico/8/1934, H1N1 virus.[120,121]

As predicted, pandemic 2009 H1N1 spread rapidly and became the dominant influenza strain the following winter. Fortunately, however, mortality was not as high as anticipated, ranging closer to that of a regular seasonal flu.[122] Soon after, in the spring of 2010, we noted that a much higher number of children with recent-onset narcolepsy were referred to Stanford compared with prior years.[123] Further, in China the number of children with new-onset narcolepsy increased threefold to fivefold over prior years in the spring and summer of 2010 (see Figure 89-3, *B*), peaking 4 to 6 months after the peak in H1N1 infections.[117,124]

In parallel with this and perhaps most strikingly, in both Finland[125,126] and Sweden[127] cases of childhood-onset narcolepsy were reported a few months after vaccination with a particular pH1N1 vaccine formulation called Pandemrix, documenting about a 10-fold increased risk for developing narcolepsy after vaccination.[123] Other studies confirmed that this particular vaccine had similar effects in Norway,[128] England,[129] France,[130] and Ireland,[131] although it is important to realize that only approximately 1 in 15,000 children vaccinated with Pandemrix developed narcolepsy (including DQ0602 siblings and in at least one case a discordant twin). Interestingly, however, it is unclear whether other pH1N1 vaccines increase the predisposition to narcolepsy.

Pandemrix is a unique vaccine, manufactured by GlaxoSmithKline using a manufacturing process to isolate surface antigens (typically purifying mostly the HA protein, which is dosed at 3.75 mcg in this vaccine).[120,121,132] In addition, a specific adjuvant, AS03A, a mix of squalene, DL-α-tocopherol, and polysorbate 80, was added. The AS03A adjuvant is potent at stimulating $CD4^+$ T-cell responses,[134] and it is clear that vaccine efficacy was high; one injection was sufficient to obtain high coverage as measured by the hemagglutinin inhibition assay (an assay measuring antibodies targeting the HA protein).[120,121]

Other H1N1 vaccines were manufactured using different protocols to isolate surface antigens or different adjuvants. Arepanrix, a vaccine also produced by GlaxoSmithKline is identical to Pandemrix, except that it was made at a different site using a slightly different process for isolating surface antigens (the Flulaval process).[121,132,133] Focetria, a Novartis vaccine, is another vaccine relatively similar to Pandemrix. It uses a MF59 adjuvant, containing squalene and 7.5 μg of H1 and of polysorbate and contains a more pure H1 preparation.[120,132] Arepanrix, which was used in Canada, also increased the risk for narcolepsy, but much more weakly, 1.5- to 3-fold.[135] Although no study has been formally done, Focetria has not been reported to trigger narcolepsy.

In the United States, only nonadjuvanted or live attenuated H1N1 vaccines were used. Of interest is the fact all seasonal trivalent split or subunit vaccines that have been used since 2009 still contain A/California/7/2009 (H1N1)-pdm09-like reassortant as one of the three strains covered. Although this has not been well studied and sporadic cases have been reported, the narcolepsy risk of these vaccines appears small or nonexistent.[136]

In summary, it appears that in the spring and summer of 2010, a larger than usual number of children with narcolepsy were observed in China and probably in other countries independent of any vaccination. In addition, narcolepsy in children also occurred in reaction to Pandemrix, although overall risk was small. The effects of other pH1N1 vaccines was either much milder or nonexistent. Antigens common to the wild-type virus and Pandemrix may have been involved.[121,133]

HUMAN LEUKOCYTE ANTIGEN IN NARCOLEPSY

HLA genes, also called major histocompatibility complex genes, are a specialized set of immune-regulatory genes located on human chromosome 6. These genes and resulting proteins are unusual because they are extremely polymorphic, thus the need for matching HLA in transplantation. Importantly, however, all the polymorphisms are located in a small region of the protein that functions as a peptide-binding area. Within this area, HLA molecules present foreign antigens (e.g., peptide fragments derived from a virus or bacteria) to T cells, the major effector cells of the immune system. Because each

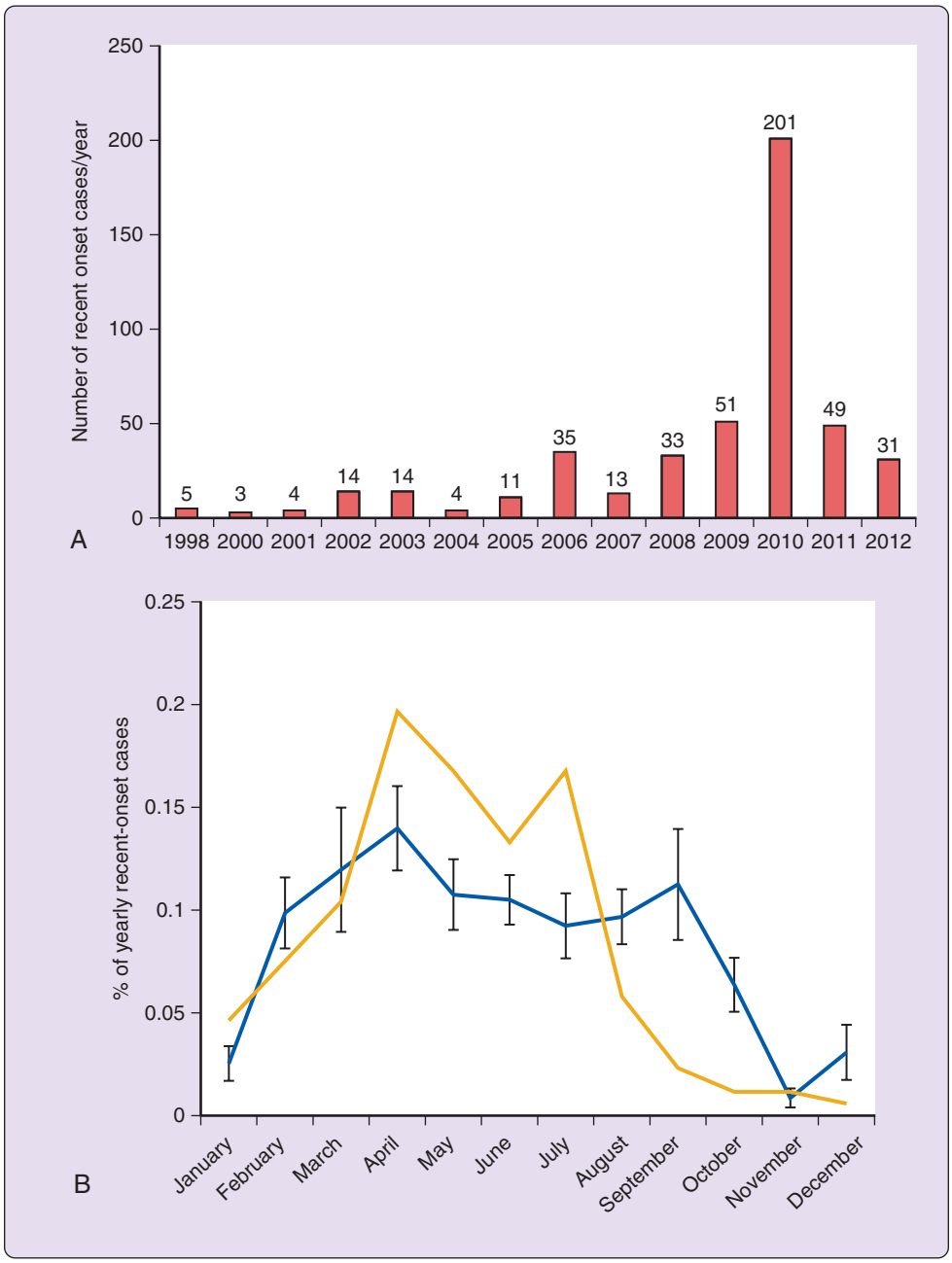


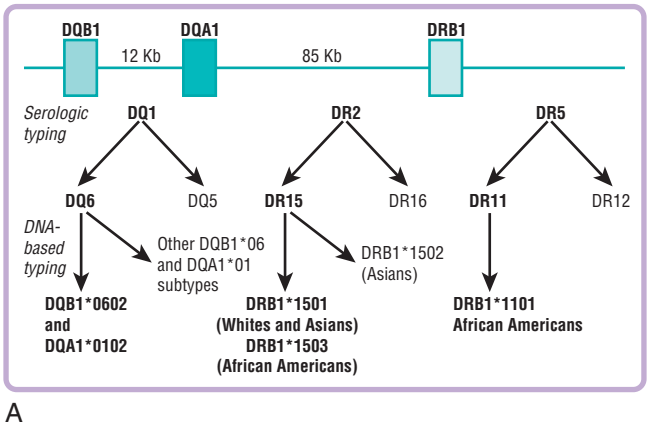
Figure 89-3 **A,** Yearly occurrence of recent onset (diagnosis within 1 year of onset) showing a dramatic increase in 2010, following the H1N1 pandemic of 2009, with return to baseline condition the following years.[117,124] **B,** Seasonal pattern of onset of narcolepsy in Chinese patients showing highly increased risk in spring and summer versus early winter. Data are represented as a yearly fraction of 12 months, that 0.083 (8.3%) would be the expected value of each month if onsets were randomly distributed across the year. The *blue line* represents the mean ±SEM of this yearly fraction for the years 2002 to 2009, with very low levels of new onset in late winter. The *orange line* represents the numbers for 2010, following the 2009 pH1N1 pandemic, and shows a more pronounced circannual pattern peaking in spring and summer.

person has a specific set of HLA genes, these bind slightly different sets of peptide antigens (peptide repertoire), and thus T cells of each individual can only "see" a specific set of antigenic peptides. HLA polymorphisms are therefore at the origin of much of the interindividual genetic variation in immune responses. As such, immune responses to infection vary with HLA subtype (and these genes are submitted to strong evolutionary pressure), although because most infections involve thousands of epitopes and multiple genes, the effect at the level of individual genes is modest, and HLA polymorphisms modulate infectious disease severity rather than occurrence. Rather, discussed later, most of the diseases that are strongly associated with HLA are autoimmune, likely involving the presentation of a few specific autoantigens by specific HLA subtypes and therefore explaining the strong associations.

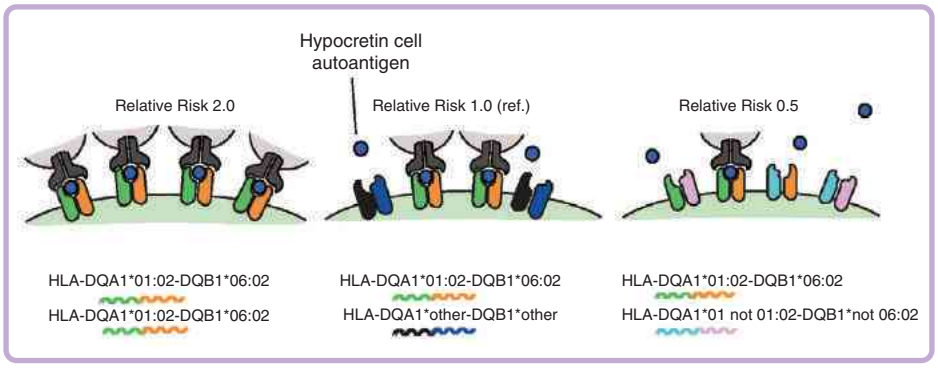
HLA-DQA1*01:02-DQB1*06:02 (DQ0602) Is Associated with Narcolepsy

The observation that narcolepsy was strongly associated with HLA was first reported in Japan[85,137] with HLA-DR2 and -DQ1. It was quickly confirmed in Europe[138-141] and North America,[142,143] with 90% to 100% of all patients with cataplexy carrying the HLA-DR2 subtype. Since this discovery, HLA-DR and -DQ typing has changed from serologic to molecularly based, and the broad DR2 and DQ1 subtypes were further divided into DR15, DQ6, and then DRB1*15:01, DQA1*01:02, and DQB1*06:02 (Figure 89-4, *A*). The key gene involved was then found to be DQB1*06:02, a subtype of DQ1. This is especially important in African American patients, many of whom are DQB1*06:02 positive but DR2 negative.[86,144-146] Subjects were also found to be DQA1*01:02 positive,[146-148] a less specific marker than DQB1*0602. Novel DNA markers developed in the HLA-DQ region have been tested to further map the narcolepsy susceptibility region within the DQA1-DQB1 interval.[149] This segment was entirely sequenced and the gene segment (haplotype) shown to contain no new genes.[149]

In all narcolepsy susceptibility DR-DQ haplotypes identified, two polymorphic genes, DQA1*01:02 and DQB1*06:02, are always present together as a single block within a very small segment of 20 kb.[147] The functional HLA-DQ protein is a heterodimer composed of a DQα (encoded by DQA) and a DQβ (encoded by DQB) chain, and because these genes are located close together, there is high linkage disequilibrium between these two genes so that specific combinations of DQA1 and DQB1 alleles are found as haplotypes. The presence of these nonrandom combinations of DQA1-DQB1 haplotype has been shaped by evolutionary constraints because some DQα and DQβ chains can bind together. Indeed, functional studies have shown that the DQα chains of the broad DQ1 subtype (subseparated into DQ5 and DQ6) can heterodimerize together, but not with other subtypes that are very distinct in term of their sequence homology. The sequence homology is reflected by the nomenclature, with gene products of the DQ1 family being associated with DQA1*01 and DQB1*05 or DQB1*06 subtypes. As a consequence, only very rarely would a DQB1*06 subtype be found adjacent to a non-DQA1*01 DQα chain because in these case a nonfunctional heterodimer would result.



A



B

Figure 89-4 Human leukocyte antigen (HLA)-DR and HLA-DQ alleles typically observed in narcolepsy. **A,** The DR and DQ genes are located very close to each other onto chromosome 6p21 and are part of the HLA class II family of HLA genes. These genes encode heterodimeric HLA proteins composed of an α and a β chain that interact with T-cell receptors (TCRs) located on $CD4^+$ T cells. In the DQ locus, both the DQα and DQβ chains have numerous polymorphic residues and are encoded by two polymorphic genes, DQA1 and DQB1, respectively. Polymorphism at the DR(αβ) level is mostly encoded by the *DRB1* gene, so only this locus is depicted in this figure. DQB1*06:02, a molecular subtype of the serologically defined DQ1 antigen (later split into DQ5 and DQ6) is the most specific marker for narcolepsy across all ethnic groups. It is always associated with the DQA1 subtype, DQA1*01:02, forming the DQ(αβ) heterodimer DQ0602. **(B).** Studies suggest that the allelic dose of DQ0602 influence narcolepsy risk. Because of this, homozygotes are at approximately twice the risk for developing narcolepsy, whereas subjects heterozygous with other DQ1 subtypes are at reduced risk. (Modified from Ollila HM, Fernandez-Vina M, Mignot E. HLA-DQ allele competition in narcolepsy: a comment on Tafti et al. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep* 2014;38:147–51.)

A number of other DR-DQ DQ1 haplotypes in the population carry DQA1*01:02 without DQB1*06:02, and those do not predispose to narcolepsy.[146] Conversely, although DQB1*06:02 subjects are almost always DQA1*01:02 positive, rare haplotypes with DQB1*06:02 but without DQA1*01:02 are observed in the control population but not in narcolepsy subjects.[146] Thus both the DQA1*01:02 and the DQB1*06:02 alleles (the DQ0602 heterodimer) are likely needed for narcolepsy predisposition,[146] which is logical because polymorphism in both the DQα and DQβ regions contributes to peptide binding.

Dosage of DQ0602 Influences Type 1 Narcolepsy Risk

Across multiple ethnic groups, DQ0602 (the combination of DQA1*01:02 and DQB1*06:02) is a near prerequisite for developing narcolepsy, and individuals homozygous for DQ0602 are at approximately two times greater risk for developing narcolepsy,[86,87,150,151] suggesting that the amount of DQ0602 heterodimer increases risk.[152] In addition, DQB1*05:01, DQB1*06:01, DQB1*06:03, and other DQ1 alleles that are non-DQ0602 appear protective.[86,87,151,153-155] As mentioned earlier, DQ1 is a broad DQ subtype that includes the DQα alleles encoded by DQA1*01 and DQβ alleles encoded by DQB1*05 and 06 subtypes. These DQ1 alleles, unlike those of the other broad DQ groups (DQ2, 3, and 4), are "compatible" with each other, meaning that they have sequence similarity and proper folding as selected by invariant chain binding (in contrast, non-DQ1 subtypes such as DQ2 and DQ3 are generally compatible with each other). Estimating relative risk, we noted that risk for DQ0602/other DQ1 is about one half of DQ0602/other, suggesting that there is competition of *trans*-encoded DQ1 alleles that are non-DQ0602, reducing the amount of DQ0602 and thus risk, a phenomenon we call *allele competition*[87,151] (see Figure 89-4, *B*).

HLA-DQB1*03:01 Influences Age of Onset

Intriguingly, DQ0602/DQB1*03:01 increases risk versus other combinations,[86,87,151,153] an effect difficult to explain because it occurs in the context of multiple DQα-associated alleles (DQA1*03:01, DQA1*03:02, DQA1*05:05, and DQA1*06:01), suggesting it is not mediated through DQα/β heterodimers. This effect also occurs in trios using transmission dysequilibrium tests, a design in which power is enhanced by the removal of alleles that are located together with DQ0602 in DQ0602-positive parents and thus never transmitted.[156] The reason for this additional effect is unclear, but unlike DQ0602 dose, the presence of DQB1*03:01 reduces age of onset, so the association is stronger in patients with early-onset narcolepsy.[157] The mechanism mediating this effect is unknown.

DP and HLA Class I Also Modulate Narcolepsy Risk

Although HLA-DQ is the primary genetic factor associated with narcolepsy (DQ0602 dose and DQB1*03:01 for early onset), additional effects within the HLA region add further modulation. In a recent study, Ollila and colleagues[15] compared controls and narcolepsy subjects fully matched for HLA-DQ and found protective effects of DPB1*04:02 (odds ratio [OR] = approximately 0.5) and susceptibility effects of the Asian subtype DPB1*0501 (OR = approximately 2). Additional matching for HLA-DP (i.e., matching for DR, DQ, and DP) revealed weak but significant residual effects in the HLA class I region (Figure 89-5).

These results are reminiscent of findings obtained in other HLA class II–associated autoimmune diseases such as type 1 diabetes, in which the primary association is with HLA-DR or -DQ, even though additional HLA-DP and class I effects can be detected. Because HLA class II interacts with $CD4^+$ T cells (see later), whereas HLA class I interacts with $CD8^+$ T and NK cells, these data suggest that multiple immune populations are likely involved in the pathophysiology of narcolepsy.

Role of HLA Typing in Clinical Practice

The usefulness of HLA typing in clinical practice is limited by many factors. First, the HLA association is very high (98%) only in patients with hypocretin deficiency/type 1 narcolepsy, and most of these patients have clear cataplexy, so the diagnosis can be made on clinical grounds alone. Second, a large

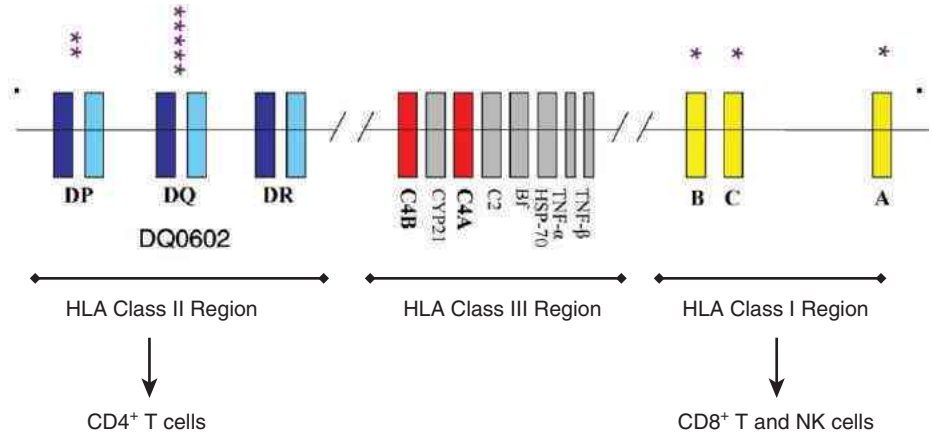


Figure 89-5 Effects of HLA loci other than HLA-DQ. HLA loci other than DQ also influence narcolepsy, notably HLA class II genes DPB1*04:02 (protective) and DPB1*05:01 (predisposing). In addition, effects of specific HLA class I gene alleles are also evident. HLA class I "A, B, C" genes present peptides to T-cell receptors located on $CD8^+$ cytotoxic T cells, or they may interact with natural killer cells, suggesting involvement of these cells in the pathophysiology of narcolepsy.[15] C4A, C4B, complement C4 genes; TNF, tumor necrosis factor.

number of control individuals (approximately 12% in Japan, 25% in whites and Chinese, 38% in African Americans) have the HLA-DQB1*06:02 marker without having narcolepsy. Third, low CSF hypocretin cannot be interpreted in the context of coma or serious neurologic defects (e.g., head trauma) because it can be reversibly low.[158] It is also interesting to note that in a recent study, 40% of nine non-DQB1*06:02 subjects with documented low CSF hypocretin-1 were DPB1*09:01 positive, a rare subtype in most ethnic groups (approximately 3%).[159] Practically, determining whether a patient carries DQB1*06:02 is only helpful if the clinician wishes to exclude hypocretin deficiency as the cause of the clinical complaint, especially before a lumbar puncture for CSF hypocretin-1 determination. Whether having an ultimate biologic basis for symptoms is important is a matter of individual clinician preference. In my opinion, clarification of hypocretin status helps clinicians decide whether to be more aggressive with selected therapies, whereas symptoms without a clear cause must lead to a constant reevaluation of the situation.

GENETIC FACTORS OTHER THAN HLA

As mentioned previously, genetic factors other than HLA are likely to be involved in narcolepsy predisposition. The increased familial risk in first-degree relatives (10-fold in Japanese, 20- to 40-fold in whites) cannot be solely explained by the sharing of HLA subtypes, estimated to explain a 2- to 3-fold increased risk.[5] Other results indicate that a polymorphism in the catechol-*O*-methyltransferase gene, a key enzyme in the degradation of catecholamines, also modulate disease severity.[160,161]

Additionally, the existence of rare HLA-negative families suggests disease heterogeneity and the possible involvement of other genes in HLA-negative narcolepsy. Many of these individuals have normal CSF hypocretin but lack mutations in hypocretin receptor genes.[159] Linkage analysis in HLA-DQB1*0602-positive Japanese families has suggested the existence of a susceptibility gene on 4q13-23.[162] Finally, in a rare multiplex family with DQB1*06:02 negativity and low CSF hypocretin levels, a mutation in myelin oligodendrocyte glycoprotein was found to be associated with narcolepsy.[163] Myelin oligodendrocyte glycoprotein mutations have not yet been found in any other HLA-negative patients with low CSF hypocretin.[159,159] Interestingly, however, a single case of prepro-hypocretin mutation was found in a very early-onset (6 months of age) individual with cataplexy and HLA negativity.[11]

More recently, studies have moved toward systematic genome coverage using genome-wide association (GWA) designs.[155,157,164-167] In a study of 222 patients and 389 controls and replication in 159 patients versus 190 controls, Miyagawa and colleagues[164] found involvement of rs5770917, a polymorphism located between *CPT1B* and *CHKB*, two genes that regulate cholinergic metabolism and β-chain fatty acid oxidation, respectively. A similar effect of rs5770917 (and a significant HLA association) was also observed in patients with "essential hypersomnia syndrome," a milder form of narcolepsy without cataplexy defined in Japan by sleepiness, refreshing naps, and no cataplexy,[164] suggesting a disease continuum. Interestingly, however, although the effect replicated weakly in Koreans,[164] the same polymorphism had no effect in a large sample of white[165] and Chinese[168] narcolepsy patients

Using a larger sample of whites and Asians, Hallmayer and colleagues[165] found that narcolepsy is strongly associated not only with HLA but also with a specific polymorphism in the T-cell receptor-α (TCR-α) gene.[165] Although genetic risk was not high (OR = ~2) when compared with effects found with HLA polymorphism, the finding was nonetheless remarkable because it further demonstrated a role of the immune system in narcolepsy. It was also unusual because none of the other autoimmune disorders subjected to GWA analysis have TCR loci as a susceptibility factor.

Further studies in larger samples that also included other ethnic groups, notably Chinese, Japanese, and African Americans, have led to the identification of other associated genes, most known to be involved in other autoimmune diseases.[157,166,167] Other associated loci include the TCR-α and -β genes; TNFSF4 (also called OX40L), a costimulatory receptor for T-cell activation and associated with multiple autoimmune diseases; cathepsin H, an enzyme likely involved in antigen processing and associated with type 1 diabetes; ZNF365, a transcription factor associated with inflammatory bowel disease (IBD); and IL10RB-IFNAR1, a region associated with IBD and other autoimmune diseases.

Of additional interest is the association with the PPAN-P2RY11-EIF3G gene region, 10 kb from the DNA methylase gene 1 (*DNMT1*).[166] This finding is notable because this gene region is not known to be associated with other autoimmune diseases, although P2RY11, an adenosine triphosphate receptor, regulates cell death, notably in immune cell subsets. Interestingly, in a parallel exome sequencing project of rare dominant phenotypes with narcolepsy, we found that a rare disease consisting of late-onset narcolepsy with deafness, cerebellar ataxia, and dementia (ADCA-DN) was secondary to a mutation in exon 21 of the *DNMT1* gene, resulting in late-onset neurodegeneration, with a likely effect on hypocretin cells[169] (see Narcolepsy in Genetic Syndromes). Further mapping of the GWA signal confirmed location within P2RY11-EIF3G and not extending to *DNMT1*, although regulatory elements for the latter could still lie within the nearby region. In favor of this hypothesis, although P2RY11 is a pseudogene in rodents, the syntenic block containing PPAN-P2RY11-EIF3G-DNMT1 is conserved from zebrafish through mammals.

ELLUSIVE AUTOIMMUNE MECHANISM IN TYPE 1 NARCOLEPSY

The finding of hypocretin cell loss in narcolepsy, together with the demonstration that HLA-DQ0602 is responsible for most of the association signal within the HLA region, has rekindled the hypothesis of autoimmunity, with hypocretin cells as the logical target. Surprisingly, however, autoantibodies targeting hypocretin peptides have not been found,[170-172] and immunostaining of hypothalamic tissue with human narcolepsy sera has not revealed autoantibodies targeting colocalized antigens on these neurons.[173-175] Several nonreproducible findings were made. Passive transfer experiments of human sera in mice have been published,[176,177] suggesting the presence of functional autoantibodies with modulating effects on spontaneous colonic migrating motor complex contractions or reactions of rodent bladder strips to muscarinic stimulation,[178,179] but we could not replicate the finding (data not shown). These investigators also suggested that peripheral

injections of human narcolepsy sera in mice caused narcolepsy symptoms, but when we attempted replication, all mice had seizures causing behavioral arrests that may have been confused with narcolepsy events. Two of the five treated animals died, and hypocretin neurons were intact postmortem in all animals, including the three animals after recovery. Further, we sent blinded sera samples for the bladder strip assays, but opposite results to the initial findings were returned to us. Using a BAC-based transgenic animal model expressing a Flag-tagged poly(A)-binding protein (*Pabpc1*) cyclic DNA sequence in hypocretin neurons, Cvetkovic-Lopes and colleagues[180] isolated transcripts believed to be increased in hypocretin cells, including the protein Tribbles homologue 2 (Trib2). The authors went on to demonstrate increased Trib2 autoantibodies in individuals with recent-onset narcolepsy and some cross-reactivity of sera with hypocretin neurons.[173-175]

Unfortunately, however, these authors may have been on the right track for the wrong reasons. Shortly after the TRIB2 publication, the TRIB2 antibody finding was replicated[181,182] using sera samples from subjects collected between 1990 and 2005, but staining of hypocretin neurons with narcolepsy sera was not observed.[173-175,183] Further studies using a similar approach but an mRNA binding protein other than Pabpc1, the protein P10, also found that few of the genes expressed in hypocretin neurons as reported by Cvetkovic-Lopes and colleagues,[180] including *TRIB2*, were enriched in hypocretin neurons.[184] These latter results were also confirmed by our own multiple expression array studies.[185,186] Pursuing this line of investigation, TRIB2 autoantibodies were later found to be generally absent in more recent narcolepsy samples.[123] It is our hypothesis that TRIB2 autoantibodies may have marked a coinfection present together with a narcolepsy trigger in some patients with disease onset in the 1990s and 2000s, a result substantiated by the finding of a correlation between A/H1N1 and TRIB2 autoantibody levels in a recent study.[89] Interestingly, a recent study, reminiscent of the older Australian studies mentioned previously,[176] reported that local injections of purified immunoglobulins of narcolepsy-TRIB2-positive individuals, but not controls, produced hypocretin cell loss and narcolepsy symptoms.[177] Careful reading of this manuscript, however, does not support the conclusion of the study because no hypocretin cell count statistics were provided, only an exemplar hypothalamic section showing widespread local cell loss that is more extensive than just hypocretin loss. Further, the authors reported "narcolepsy-like immobilization attacks" without associated EEG studies in six animals, which may have been seizures considering their mean duration (66 to 464 seconds), much longer than typically reported in murine cataplexy (2 to 60 seconds).[30]

Finally, a recent report from our group suggested that specific portions of hypocretin that bind to DQ0602 were the likely autoantigen and that T-cell cross-reactivity and molecular mimicry with pH1N1 was involved.[187,188] We felt a blood test for narcolepsy might be close at hand, but on finding these data were invalid, we withdrew the paper.[187,188]

Unfortunately, although rare cases of type 1 narcolepsy coexisting with multiple sclerosis, lupus, or other common autoimmune conditions have been reported, there are no systematic studies of co-occurrence in cases or family members. Rather, the possibility that narcolepsy is an autoimmune disease is supported by the rare existence of autoimmune narcolepsy with documented hypocretin deficiency as part of two other autoimmune diseases. The first is a rare paraneoplastic syndrome associated with anti-ma2 antibodies in the context of seminomas.[189,190] Interestingly, in one case, a $CD8^+$ T-cell infiltration largely restricted to the hypothalamus and an almost complete loss of hypocretin neurons were observed.[191] Similarly, in one case of late-onset hypoventilation syndrome, a disorder with reported hypothalamic abnormalities,[192] we found very low CSF hypocretin-1 levels in an individual with otherwise unexplained sleepiness and cataplexy-like episodes.[13] More recently, this disorder has been renamed rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation (ROHHAD) and suggested to be autoimmune based on the finding of extensive infiltrates of lymphocytes and histiocytes in the hypothalamus of some patients.[193,194] An excellent response to anticataplectic therapy was observed in this patient.

This brief discussion exemplifies the difficulties of narcolepsy research. Currently, the circumstantial evidence strongly suggests that narcolepsy is an autoimmune disease, but we believe like others[195] that, as yet, there is no direct evidence for this hypothesis. Figure 89-6 summarizes the likely autoimmune process involved in narcolepsy, with the hypothesis that $CD8^+$ T cells are involved.

CEREBROSPINAL FLUID HYPOCRETIN-1 AS A DIAGNOSTIC TOOL FOR TYPE 1 NARCOLEPSY

The observation that CSF hypocretin-1 levels are decreased in patients with narcolepsy provides a new test to diagnose this disorder (Figure 89-7; see Table 89-1). Using a large sample of patients and controls, we recently conducted a quality receiver operating curve (QROC) analysis to determine the CSF hypocretin-1 values most specific and sensitive to diagnose narcolepsy.[13] A cutoff value of 110 pg/mL was the most predictive in patients with cataplexy (30% of mean control values). In patients without cataplexy, QROC analysis indicated that a cutoff value of 200 pg/mL is most predictive, suggesting that in some of these patients, partial hypocretin deficiency could be present (some patients may evolve to more complete symptomatology with cataplexy and even lower levels[196]). Most positive samples had undetectable levels of hypocretin-1 (<40 pg/mL in most assays), whereas a few samples had detectable but very diminished levels. Hypocretin levels appear normal in patients with idiopathic hypersomnia, sleep apnea, restless legs syndrome, or insomnia.

Using the 110 pg/mL cutoff, the measurement of CSF hypocretin-1 is especially predictive in patients with definite cataplexy (100% specificity, 83% sensitivity). Sensitivity and specificity are higher for this test than for the MSLT. In most case series, approximately 12% of narcolepsy patients with cataplexy or hypocretin deficiency do not have a positive MSLT. Measuring CSF hypocretin may therefore be helpful in patients with cataplexy when the MSLT is negative, to exclude a possible conversion disorder. CSF hypocretin-1 measurements have a more limited predictive power in cases with atypical or absent cataplexy. Specificity of the CSF hypocretin-1 measurement is still extremely high (99%), but sensitivity is low (16%) for CSF levels below 110 pg/mL (see Figure 89-6),[13] with most patients having normal levels.[13,81,83] Raising the cutoff of CSF hypocretin-1 to 200 pg/mL after demonstrating HLA positivity increases sensitivity to 41%,

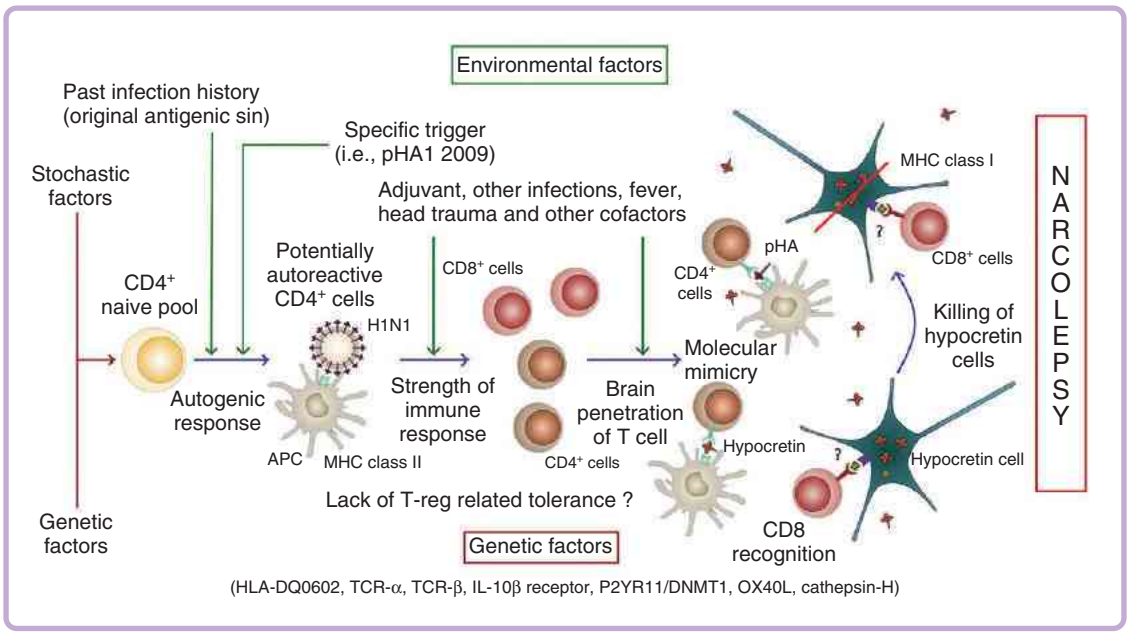


Figure 89-6 Hypothetical pathophysiologic model for autoimmune narcolepsy. Narcolepsy is likely an autoimmune disease involving specific subpopulations of $CD4^+$ and $CD8^+$ T cells and may have little if any B-cell, autoantibody involvement. The development of narcolepsy probably results from many unlikely events. First, DQ0602 must almost always be present, and genetic background at other loci also influence predisposition. $CD4^+$ T cells bearing specific receptors are generated stochastically, although with influence of T-cell receptor (TCR) loci polymorphisms, and autoreactive cells are removed by the thymus, whereas others are released as naïve T cells. We hypothesize that only a subset of the population may have naïve T cells with the potential to become pathogenic and narcolepsy inducing. In many cases, these cells may become anergic or will be moved into a regulatory T-cell compartment that rather inhibits autoimmunity. In some unlucky individuals, an immune reaction to an external infection such as 2009 H1N1 may engage these pathogenic, cross-reactive naïve T cells. These cross-reactive T cells recognize a yet unknown autoantigen on hypocretin cells. This may be more likely to occur in younger individuals with a particular history of infections (or lack of). Indeed, in many subjects, such infection would first stimulate memory $CD4^+$ T cells that recognize shared epitopes and are not pathogenic. Other factors that may participate in tilting the balance toward developing narcolepsy could be the strength of the immune response (coinfection, adjuvant), a leaky blood–brain barrier, etc. Following engagement of $CD4^+$ T cells, $CD8^+$ T cells are also likely to be involved and may be important in mediating the actual hypocretin cell killing.

suggesting that 200 pg/mL may be more appropriate in these cases after HLA typing, although specificity may be more questionable if other neurologic disorders are present.[158] For the patient, the diagnostic value of CSF hypocretin-1 measurement needs to be weighed against the trauma associated with obtaining CSF. Because HLA is positive in 98% of cases when CSF hypocretin is below 110 pg/mL, we advise measuring CSF hypocretin-1 only after verifying that the subject is DQB1*06:02 positive. If the subject is negative, low hypocretin is very unlikely, about 2% probability if the subject has cataplexy and probably less without cataplexy.

Another potential application for CSF hypocretin-1 testing lies in the complex field of narcolepsy and hypersomnia related to neurologic disorders associated with trauma, tumor, infection, degenerative diseases, and genetic disorders (Table 89-3). Von Economo was the first to suggest that narcolepsy may have its origins in the posterior hypothalamus. In his classic study of encephalitis lethargica, von Economo recognized three categories of patients: a group with hypersomnia and eye movement abnormalities (somnolent-ophthalmoplegic), a group with insomnia and hyperkinetic movement disorder (sometimes with reversal of the sleep cycle), and a group with parkinsonism ("amyostatic-akinetic," often as a residual form).[197] Neuropathologic studies revealed involvement of the midbrain periaqueductal gray matter and posterior hypothalamus in the hypersomnolent variant (with extension to the oculomotor nuclei, explaining the oculomotor symptoms). Involvement of the anterior hypothalamus with extension into basal ganglia was observed in the insomnia variant (explaining the frequently co-occurring chorea). This led von Economo to speculate that the anterior hypothalamus contained a sleep-promoting area, whereas an area spanning from the posterior wall of the third ventricle to the third nerve (encompassing part of the posterior hypothalamus and the periaqueductal midbrain region) was involved in promoting wakefulness. The cause of idiopathic narcolepsy that had been described some 50 years earlier[102] was also speculated to involve this general area.[198] This hypothesis was further refined by others, who noted that tumors or lesions located close to the third ventricle were also associated with secondary narcolepsy.[110,199] A postulated hypothalamic cause of narcolepsy was widespread until the 1940s but was subsequently ignored during the psychoanalytic boom, to be replaced by a brainstem hypothesis.[200] Interestingly, only two cases of postencephalitis genuine narcolepsy with cataplexy had even been clearly documented, and in one case reversal of symptoms was reported with recovery.[201,202]

Reports of lesions near the third ventricle (hypothalamus and upper midbrain) in association with narcolepsy (such as tumors) have been described for more than 80

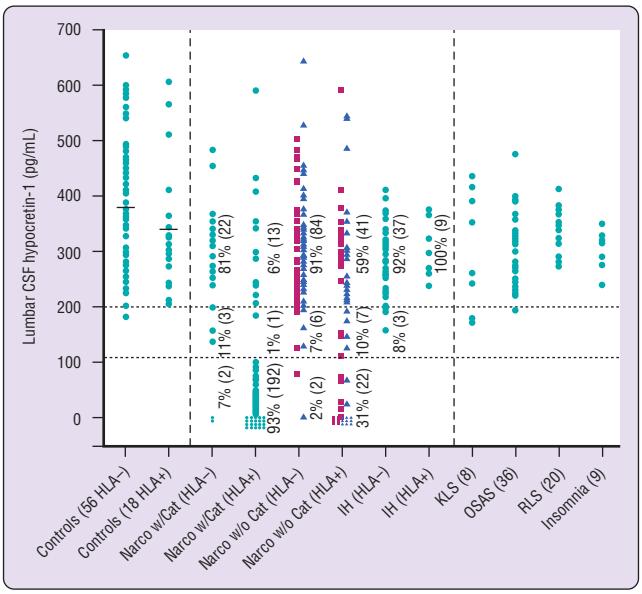


Figure 89-7 Lumbar cerebrospinal fluid (CSF) hypocretin-1 concentrations in controls versus subjects with narcolepsy and other sleep disorders (from the Stanford Center for Narcolepsy Research database). Each point represents the concentration of hypocretin-1 as measured in unextracted lumbar CSF of a single individual. Subjects are differentiated according to HLA-DQB1*06:02 status and include controls (samples taken during both night and day). Patients are classified as having narcolepsy with or without cataplexy, and individuals without cataplexy are subdivided into those with atypical (*triangle*) and no (*squares*) cataplexy. Clinical subgroups include narcolepsy patients (Narco) with (w) and without (w/o) cataplexy (Cat), idiopathic hypersomnia (IH), Kleine-Levin syndrome (KLS), obstructive sleep apnea syndrome (OSAS), restless legs syndrome (RLS), and insomnia. Individuals with secondary narcolepsy/hypersomnia are not included. The *dashed lines* indicate hypocretin levels that are low (<110 pg/mL), intermediate (111–200 pg/mL), or normal (>200 pg/ml). Note that these pg/mL values are largely artificial and are meant to represent approximately 30% of mean control value as tested in a given center using direct radioimmunoassay and a set of healthy controls.[13] Mean CSF hypocretin-1 concentration was not significantly different between HLA-DQB1*06:02 positive and negative controls. The percentage and number of subjects are specified for each group of subjects according to the two CSF hypocretin thresholds.

years.[198,199,201-203] Narcolepsy-like symptoms have also been reported after traumatic brain injury, acute disseminated encephalomyelitis, hypothalamic sarcoidosis or histiocytosis X, and in association with multiple sclerosis and Parkinson disease.[158,204,205] In some patients, lesions of the hypothalamic hypocretin neuron region have been clearly identified using magnetic resonance imaging, as in bilateral multiple sclerosis or anti-aquaporin-4 lesions in the hypothalamus, and tumors of the third ventricle.[158,204,205] Cataplexy may be present in these individuals, and CSF hypocretin-1 levels may be either in the narcolepsy range (<110 pg/mL) or in the intermediate range[13,158] (see Table 89-3). Similarly, postmortem studies have shown a 30% and 50% decrease in hypocretin cell counts in Huntington and Parkinson diseases,[158] respectively, with maintenance of normal CSF hypocretin-1 levels.[13] Such intermediate or normal levels may reflect damage to nearby hypocretin projection sites, with sufficient preservation of cell bodies to maintain detectable levels of hypocretin-1. Alternatively (or additionally), other regions in the upper midbrain may also contribute symptoms, especially sleepiness, as initially proposed by von Economo.[198]

Therefore, as illustrated previously, CSF hypocretin-1 levels can be helpful in complex clinical situations in which the history, polysomnography, or MSLT data may be difficult to interpret (Table 89-4). Hypocretin values, however, should be cautiously interpreted. In a large series of individuals with various neurologic disorders, we found that up to 15% had CSF hypocretin-1 values within the intermediate range; most of these patients had severe brain pathology, most notably head trauma, encephalitis, and subarachnoid hemorrhage.[79,158] Decreased hypocretin-1 levels in these patients may reflect damage to hypocretin-1 transmission or may be related to changes in CSF flow. Other authors have shown that CSF hypocretin-1 increases with locomotor activity and decreases with treatment with serotonin reuptake inhibitors (but never to near-undetectable, narcolepsy-like levels).[93] Therefore the finding of hypocretin-1 levels in this intermediate range should alert the clinician to the possibility of underlying brain pathology, which may require additional clinical evaluation, laboratory testing, or imaging. Whether genuine hypocretin deficiency explains abnormal sleep in these neurologic disorders is in need of further investigation.[158]

NARCOLEPSY IN GENETIC SYNDROMES

Genetic disorders such as ADCA-DN (MIM 604121, *DNMT1* mutations), Coffin-Lowry Syndrome (MIM 303600, *RSK2* mutations),[206] Moebius syndrome (MIM 157900, heterogeneous),[207-209] myotonic dystrophy (MIM 160900 and 602668, *MD1* and *MD2* mutations),[210] Niemann-Pick disease type C1 (MIM 257219, *NPC1*),[13,211-213] Norrie disease (MIM 310600, Xp11.4-p11.23 deletions encompassing the *NDP* and in cases with cataplexy, the *MAO* genes),[207,214,215] and Prader-Willi syndrome (MIM 176270, most often due to 15q11.2 deletions encompassing the paternal copies of the imprinted *SNRPN* and *NDN* genes)[13,207] can result in daytime sleepiness, SOREMPs, and cataplexy-like symptoms. HLA may be negative or positive by chance in these patients. We have explored CSF hypocretin-1 levels in such pathologies and have found that most have normal or intermediate CSF hypocretin-1 levels (<110 pg/mL).[13] These special etiologies are described subsequently.

ADCA-DN, a syndrome first described by Melberg and associates,[216] is a late-onset neurodegenerative disorder (ages 30 to 40 years) with ataxia, deafness, and narcolepsy-cataplexy and intermediate to low CSF hypocretin-1. Narcolepsy is an early manifestation, and at this stage CSF hypocretin-1 may be normal, dropping only in the late stage of the disorder.[169,216,217] Deafness is an early symptom, followed by cerebellar ataxia, ocular nerve atrophy, and neurodegeneration, leading to death in 5 to 10 years. The disorder was recently found to be caused by mutations in exon 21 of the DNA methylase 1 (*DNMT1*) gene, in a regulatory region of the gene.[169] It is closely related to *DMNT1* mutations in exon 20 and hereditary sensory and autonomic neuropathy type 2 (MIM 614116), in which the syndrome is similar but manifests first with a peripheral neuropathy, a symptom that is rarely the primary, early manifestation of ADCA-DN.[217]

In Coffin-Lowry syndrome, mental retardation is present, and cataplexy is atypical and more likely to represent atonic seizures because other types of events; for example, tonic-clonic or absence seizures can be induced by emotions such

Table 89-3 Cerebrospinal Fluid Hypocretin-1 Measurements in Secondary Narcolepsy and Hypersomnia*

Secondary hypersomnia with hypocretin-1 <110 pg/mL
Acute disseminated encephalomyelitis (ADEM) with sleepiness
Autosomal dominant cerebellar ataxia, deafness, and narcolepsy (*DNMT1* mutations)
Large pituitary adenoma with probable hypothalamic involvement
Late-onset hypoventilation syndrome, probably rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation (ROHHAD)
Multiple sclerosis cases with bilateral hypothalamic plaques
Paraneoplastic syndrome (seminoma) with anti-Ma2 and sleepiness, cataplexy
Hypothalamic tumors and surgery after removal of such tumors
Prader-Willi syndrome (15q11-q13)
Steroid-responsive encephalopathy associated with autoimmune thyroiditis
Whipple disease with central nervous system involvement (with hypersomnia)†
Secondary hypersomnia with intermediate hypocretin-1 levels (110–200 pg/mL)
Autosomal dominant cerebellar ataxia, deafness, and narcolepsy (*DNMT1* mutations)
Moebius syndrome
Niemann-Pick disease type C with cataplexy (*NPC1* mutations)
Norrie syndrome (deletion of *MAO* genes on the X chromosome)
After diencephalic stroke with hypothalamic and midbrain lesions
After head trauma (but unclear if mediating symptoms)
Prader-Willi syndrome (15q11-q13)
Secondary hypersomnia with normal hypocretin-1 levels(>200 pg/mL)
Acute disseminated encephalomyelitis (ADEM) with sleepiness
Human immunodeficiency virus encephalopathy with sleepiness
Hypersomnia with depression
Hypersomnia in association with Parkinson disease
Hypersomnia in association with myotonic dystrophy type 1 (*DM1* mutations)
Myotonic dystrophy type 1 (*DM1* mutations)
Neurocysticercosis cysts in the hypothalamus and other locations
Niemann-Pick disease type C without cataplexy (*NPC1* mutations)
Narcolepsy-cataplexy after hypothalamic irradiation
Pontine lesions
Thalamocortical strokes
Other pathologies with intermediate or low levels (no reported but possible sleep symptoms)‡
Traumatic brain injury (probably transient)
Encephalitis of infectious origin
Guillain-Barré syndrome and other inflammatory neuropathies
Coma caused by infection or trauma

*Cases tested at Stanford University or in centers that used the same standard cerebrospinal fluid (CSF) samples for comparisons are included; measurements in other laboratories cannot be compared.
†In another patient with Whipple disease with long-lasting resistant insomnia without hypersomnia, intermediate CSF hypocretin levels were found.
‡These may reflect nonspecific effects or be indirectly affecting the hypothalamus, with potentially reversible changes in CSF hypocretin-1.

as surprise at a sudden noise.[206] In our opinion, the disorder is not genuinely associated with true narcolepsy.

Moebius syndrome is a heterogeneous set of brainstem anomalies involving minimally the sixth and seventh cranial nerves with resulting congenital facial palsy and impairment of ocular abduction; it is often associated with a 13q12.2 deletion encompassing the *MBS1* gene and occasionally other skeletal abnormalities. Cataplexy with or without other sleep abnormality has been reported in several cases (all without skeletal abnormalities),[207,209] and in one case in which CSF hypocretin was measured, it was intermediate.[208] One case was responsive to antidepressant. In these cases, brainstem impairment affecting REM regulating centers of the pons is likely causal. The condition can be complicated by the existence of sleep-disordered breathing.

In its typical manifestation, Norrie disease is a X-linked recessive disorder characterized by very early childhood blindness due to degenerative and proliferative changes of retinal neurons, and frequently hearing deficits and neuropsychiatric symptoms.[218] Many cases are secondary to a Xp11.3-p11.4 deletion that encompass the *NRD* gene. Because mutations of *NRD* also cause Norrie disease with its ocular manifestations, the syndrome is causal to this gene. Cases with cataplexy have been described in various microdeletion syndromes but with isolated *NDP* mutations.[215,207,214] Further, families or cases with deletion encompassing both *MAO* genes and that did not affect the *NDP* gene show severe developmental delay, intermittent hypotonia, and stereotypical hand movements.[219,220] The author saw one patient with cataplexy, mental retardation without ocular symptoms, and a deletion limited to the *MAO* genes, suggesting these monoaminergic genes are indeed crucial to this manifestation. This observation may be of interest considering that optogenetic stimulation of the adrenergic locus coeruleus has recently been shown to produce arousal followed by behavioral arrests.[221]

Niemann-Pick disease type C is a lysosomal storage disease associated with mutations in *NPC1* and *NPC2* genes and abnormal cholesterol metabolism. Onset is usually before 10 years of age (with death before 20 years), but the disease can manifest at much later age. Niemann-Pick disease type C has

Table 89-4 Cerebrospinal Fluid Hypocretin-1 and HLA Results: Selected Examples in Secondary Narcolepsy and Hypersomnias

Clinical Case	CSF Hypocretin-1	Notes
An 8-year-old boy without cataplexy (onset within 6 mo)	88 pg/mL HLA+, MSLT+	Type 1 narcolepsy, treatment with modafinil, later developed cataplexy and treated with venlafaxine
A 17-year-old boy with rape hallucinations, suspicious and difficult to interview, possible cataplexy	Undetectable (<40 pg/mL) HLA+, refused MSLT testing	Type 1 narcolepsy, associated with psychosis, positive effect of venlafaxine
A 16-year-old woman with a 5-year history of depression and drug-resistant insomnia; cataplexy on interview	Undetectable (<40 pg/mL) HLA+, MSLT not interpretable	Type 1 narcolepsy, now successfully treated with sodium oxybate, modafinil, and atomoxetine
A 32-year-old man, after resection of hypothalamic craniopharyngioma, very impaired, possible cataplexy	152 pg/mL HLA–, MSLT impossible	Secondary narcolepsy, possibly lesions of other areas, partial effect of stimulants
A 33-year-old woman successfully treated with D-amphetamine and fluoxetine; no cataplexy; no MSLT	Undetectable (<40 pg/mL) HLA+	Type 1 narcolepsy, no change in treatment but considering modafinil
A 15-year-old girl with sleepiness, no cataplexy	310 pg/mL HLA+, MSLT+	Type 2 narcolepsy, treatment with modafinil
A 67-year-man diagnosed with narcolepsy without cataplexy at age 50 years, currently with falls not typically triggered, AHI = 25 events/hr, CPAP noncompliant	Undetectable (<40 pg/mL) HLA+	Type 1 narcolepsy, first tried on venlafaxine without effect, then on sodium oxybate with very positive response

AHI, Apnea-hypopnea index; CPAP, continuous positive airway pressure; CSF, cerebrospinal fluid; HLA+, HLA-DQB1*06:02 positive; MSLT+, Mean Sleep Latency Test positive (mean sleep latency <8 min, ≥2 sleep-onset REM periods [SOREMPs] including a nocturnal SOREMP).

a wide clinical spectrum that may include hepatosplenomegaly and a wide range of neurologic abnormalities (cerebellar ataxia, tremor, seizures, dysphagia, dysarthria, hypotonia, dystonia, psychosis, dementia, and other psychiatric symptoms). Vertical gaze palsy with involvement of the third cranial nerve is often an early and typical feature This condition is remarkable because cataplexy can be clear, triggered by typical emotions (laughing),[211] and partially responsive to anticataplectic treatment. To our knowledge, all cases with cataplexy have been reported in young children with *NPC1* mutations. CSF hypocretin-1 has been measured in multiple patients, and intermediate levels have been found in several cases with cataplexy.[13,213]

Some diseases are associated with the development of both narcolepsy and sleep-disordered breathing, such as myotonic dystrophy[210] and Prader-Willi syndrome[13]; in such patients, primary hypersomnia should only be diagnosed if excessive daytime sleepiness does not improve after adequate treatment of sleep-disordered breathing. Mytotonic dystrophy due to *DM1* mutations, the most common, is X-linked and is pleiotropic in its clinical manifestations. It is due to a germline and subsequently somatic expansion of a CTG trinucleotide repeat in the noncoding region of *DMPK* that can vary across tissue and causes a misregulation of alternative splicing events. It is characterized by progressive muscle wasting or weakness, cataracts, cardiac conduction defects, gastrointestinal motility defects, baldness, endocrinopathies, and infertility. Age of onset can be extremely variable. Because of the muscle weakness, hypoventilation and sleep-disordered breathing are common. Cataplexy has never been reported in myotonic dystrophy, but positive narcolepsy-like MSLTs and cognitive abnormalities are well demonstrated to be independent of sleep-disordered breathing. CSF hypocretin is generally normal.

Prader-Willi syndrome is caused by a loss of function of genes in the 15q11.2 region, with likely involvement of the small nuclear ribonucleoprotein polypeptide N *(SNRPN)* and nectin (*NDN*) genes leading to a complex cascade of genetic dysregulations. Most cases (70%) are due to a deletion of paternal origin (in this case a loss of function occurs when the maternal copy is imprinted), and most of the other cases are due to maternal uniparental disomy (inheritance of two inactivated maternal copies). Prader-Willi syndrome is characterized by a typical facial appearance, mental retardation, hypotonia, hyperphagia, and many other symptoms. Because of obesity and hypotonia, hypoventilation plus sleep-disordered breathing is a complication that can contribute to sleepiness, although MSLTs with SOREMPs are typically observed and central nervous system effects are almost surely involved.[222] Prader-Willi syndrome is interesting because cataplexy, although rare, can be typical, triggered by laughing. In one patient with a 15q deletion, obesity was not a striking feature thanks to food restriction, and cataplexy was improved by the adrenergic reuptake inhibitor atomoxetine, although at higher dose it exacerbated absence seizures also observed in this patient. CSF hypocretin is generally intermediate.[223]

In conclusion, a few genetic disorders are associated with REM sleep abnormalities and cataplexy-like features. It is likely that the study of the downstream mechanisms involved could shed light on REM sleep regulation and type 2 narcolepsy.

HYPOCRETIN COMPOUNDS AS POTENTIAL THERAPEUTIC TARGETS

Researchers have studied the effects of hypocretin on sleep after nasal, systemic, and central administration (e.g., intracerebroventricular injection or local perfusion in selected brain areas). Central administration of hypocretin-1, for example, into the lateral ventricle of wild-type rodents or normal canines, is strongly wake promoting[224] and reverses cataplexy and sleep abnormalities in narcoleptic mice.[225] The effect is likely to be partially mediated by the hypocretin receptor-2 as intracerebroventricular hypocretin-1 at the same dose (10 to 30 nmol) has no effect in hypocretin receptor-2 mutated narcoleptic canines.[224] Interestingly, hypocretin-2 administration has few if any central effects even in normal animals, perhaps because it is biologically unstable and rapidly degraded. The instability may also explain why hypocretin-1 but not hypocretin-2 is detectable in native CSF.[226]

Experiments conducted after intravenous administration of hypocretin-1 have been performed in hypocretin receptor-2 mutated canines and in two hypocretin-deficient narcoleptic dogs. Despite a previous report,[227] we were unable to detect any significant effect even at extremely high doses in hypocretin receptor-2 mutated animals. This result was not surprising considering the lack of effects after central administration of the same dose in these animals lacking hypocretin receptor-2 (see earlier).[224] More interestingly, a possible very slight and short-lasting suppression of cataplexy was observed in a single hypocretin-deficient narcoleptic animal at extremely high doses.[224]

The effect of hypocretin-1 intranasal administration in rodents and humans has also been studied,[228-231] with the hope that hypocretin-1 would penetrate into the brain through the cribriform region. Deadwyler and colleagues found a significant reversal of magnetic resonance imaging abnormalities induced by sleep deprivation after intranasal administration of hypocretin-1 (1 μg/kg) in rhesus monkeys. Weinhold and colleagues[231] found improved attention, decreased wake-REM sleep transitions, and less REM sleep in narcolepsy patients after intranasal hypocretin-1 (435 nmol), but no effects on daytime tests of maintenance of wakefulness, suggesting limited clinical significance. We also examined the possibility of intrathecal administration (up to the very large dose of 96 μg/kg) by implanting a Medtronic pump with catheterization of the cisterna magna in a single hypocretin-deficient narcoleptic canine.[232] Our hope was that at high dose, some reverse flow would occur back into deeper brain structure, providing therapeutic relief. A positive result would have had therapeutic application because these pumps are frequently used in humans for the treatment of pain or spasticity using intrathecal administration. Disappointingly, however, we did not observe any significant effect on cataplexy,[232] probably because the hypocretin did not diffuse into upper ventricular compartments. Additional studies using intraventricular rather than intracisternal injections will be needed to verify that hypocretin-deficient narcoleptic canines are responsive to supplementation.[228]

The discovery of hypocretin and its involvement in narcolepsy has led some drug companies to develop hypocretin receptor antagonists for insomnia. One of these compounds, suvorexant, a dual hypocretin receptor-1 and -2 antagonist (called DORA, or dual orexin receptor antagonist), has been approved for the treatment of insomnia.[89] More relevant to narcolepsy, hypocretin agonists are being sought by companies and academic laboratories. Although the discovery of agonists with brain penetration has been difficult, it is likely that such compounds will become available in the future, to the benefit of patients with type 1 narcolepsy.

CLINICAL PEARLS

- Cataplexy is the most predictive symptom of hypocretin deficiency and therefore should be carefully defined. In children close to onset, cataplexy can be atypical and severe, presenting as episodes of tongue protrusion, jaw openings, or generalized weakness not necessarily triggered by emotions. Later in the life of the patient or in adults, cataplexy becomes typical, triggered by emotions such as laughing, joking, and anger. Cataplexy can improve with age and with the introduction of sodium oxybate; some patients do not have any residual cataplexy when well treated.
- Type 1 narcolepsy is a lifelong disease, whereas little is known about the evaluation of type 2 narcolepsy. For this reason, treatment of type 2 narcolepsy should be more conservative.
- HLA-DQB1*06:02 typing is useful to request before measuring CSF hypocretin. If HLA is negative the probability that the patient has hypocretin deficiency is extremely low, especially if the patient does not have clear cataplexy.
- Rapid weight gain is a troubling symptom that often occurs in children when onset is abrupt. Aggressive treatment close to onset is important in these cases.
- Hypocretin receptor antagonists, also called DORAs, have recently been introduced as hypnotic compounds. Based on animal data, these compounds are likely to exacerbate narcolepsy because patients still have a few hypocretin cells left. Hypocretin agonists, compounds that should be beneficial in narcolepsy, are under development.
- Pandemrix, a flu vaccine directed against the 2009 H1N1 swine flu, was shown to have significantly increased the risk for developing narcolepsy in Europe. No other vaccine, including U.S. preparations, has been implicated.

SUMMARY

Type 1 narcolepsy is both a common neurologic disorder and a model disorder to further our understanding of REM sleep and sleepiness regulation. During the past 50 years, research has been greatly facilitated by the existence of a canine narcolepsy model, now rapidly replaced by rodent models. Narcolepsy-cataplexy, now reclassified as type 1 narcolepsy, is most commonly caused by a loss of hypocretin-producing neurons in the hypothalamus. Low CSF hypocretin-1 can be used to diagnose type 1 narcolepsy. The disorder is associated with HLA-DQB1*06:02, other HLA genes, and polymorphisms in immune system genes, indicating that the cause in most patients is an autoimmune destruction of these neurons. Studies in narcoleptic canines have substantiated a tight parallel between the pharmacologic control of cataplexy and that of REM sleep. Using this model, it was found that the mode of action of stimulant medications is presynaptic stimulation of dopaminergic transmission, whereas anticataplectic compounds exert their therapeutic effects primarily through adrenergic uptake inhibition. The hypocretin system

sends strong excitatory projections onto monoaminergic cells. The loss of hypocretin is likely to create a cholinergic-monoaminergic imbalance in narcolepsy. Abnormally sensitive cholinergic transmission and depressed dopaminergic and histaminergic transmission are believed to underlie abnormal REM sleep and daytime sleepiness in canine narcolepsy (see Pharmacology of Narcolepsy).

Although most cases of narcolepsy-cataplexy are caused by hypocretin neuron loss, some patients with cataplexy and most without cataplexy have normal CSF hypocretin-1 levels. This may either reflect disease heterogeneity or a partial loss of hypocretin neurons without significant CSF hypocretin-1 decrements. A critical area in need of further inquiry is the role of CSF hypocretin-1 testing in predicting therapeutic response to medications already in use to treat the symptoms of narcolepsy.[94] Developing an assay that could reliably measure hypocretin-1 in plasma may be possible and would also be extremely useful if low levels were observed in narcolepsy.[94] Measuring hypocretin-1 levels may some day facilitate development of therapies that interrupt or delay the development of disease. Additional work comparing drug responses in patients with type 1 and 2 narcolepsy is needed.

There is increasing evidence that the autoimmune destruction of hypocretin cells is triggered by upper airway infections. In children, when narcolepsy onset is often more abrupt, a history of strep throat is often reported. The onset of narcolepsy in children is seasonal and peaks in spring and summer, strongly suggesting that winter infections may precipitate narcolepsy a few months later. Most strikingly, increased incidence of childhood narcolepsy was found in early 2010 in China, following the 2009 H1N1 swine flu pandemic, and possibly in the United States as well. Strongly suggesting that H1N1 itself is involved, H1N1 vaccinations in Europe using Pandemrix increased the number of new narcolepsy cases many fold. More mysteriously, however, risk was not increased with other vaccines, maybe because the content of the various vaccines varied and AS03 is a particularly strong adjuvant for stimulating $CD4^+$ T cells.

Explaining the link between the HLA-TCR association and hypocretin deficiency must be a high priority. It may be possible to use CSF hypocretin-1 testing to evaluate the extent of hypocretin cell loss in early stages of the disease (e.g., in children), thus facilitating the development of treatments that may arrest or at least delay disease progression. Similar strategies using immunosuppression have been used in other autoimmune diseases, such as type 1 diabetes mellitus. In one case, 2 months after an abrupt onset, we tried high-dose prednisone but did not observe significant effects on symptoms and CSF hypocretin-1 levels[233]; however, in this case, very low hypocretin-1 levels were already observed, suggesting the possibility that irreversible damage to cells had already occurred. In other patients with recent onset, intravenous immunoglobulin administration was reported to have positive effects in some open-label studies,[234,235,234-239] suggesting the need for placebo-controlled studies.[240] We anticipate that in time, patients at risk for narcolepsy will be identified through genetic typing monitored with biomarkers (e.g., measuring T-cell populations bearing specific TCR idiotypes, occurrence of specific infections), therefore allowing early intervention to stop hypocretin cell immune destruction.

Selected Readings

Daniels LE. Narcolepsy. *Medicine* 1934;**13**(1):1–122.

Goldbart A, Peppard P, Finn L, et al. Narcolepsy and predictors of positive MSLTs in the Wisconsin Sleep Cohort. *Sleep* 2014;**37**(6):1043–51.

Han F, Lin L, Warby SC, et al. Narcolepsy onset is seasonal and increased following the 2009 H1N1 pandemic in China. *Ann Neurol* 2011;**70**(3):410–17.

Jacob L, Leib R, Ollila HM, et al. Comparison of Pandemrix and Arepanrix, two pH1N1 AS03-adjuvanted vaccines differentially associated with narcolepsy development. *Brain Behav Immun* 2015;**47**:44–57.

Mignot E, Lammers GJ, Ripley B, et al. The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. *Arch Neurol* 2002;**59**(10):1553–62.

Nuvigil (armodafinil): tablets prescribing information. <http://www.cephalon.com/newsroom/assets/nuvigil_prescribing_information.pdf>; [Accessed September 24, 2015].

Olafsdottir BR, Rye DB, Scammell TE, et al. Polymorphisms in hypocretin/orexin pathway genes and narcolepsy. *Neurology* 2001;**57**(10):1896–9.

Ollila H, Ravel J, Han F, et al. Novel genetic loci in HLA-DPB1 and HLA-class I region confer risk and protection for narcolepsy. *Am J Hum Genet* 2015;**96**(1):136–46.

Partinen M, Saarenpaa-Heikkila O, Ilveskoski I, et al. Increased incidence and clinical picture of childhood narcolepsy following the 2009 H1N1 pandemic vaccination campaign in Finland. *PLoS ONE* 2012;**7**(3):e33723.

A complete reference list can be found online at ExpertConsult.com.

Narcolepsy: Diagnosis and Management

Michelle T. Cao; Christian Guilleminault

Chapter 90

Chapter Highlights

- Narcolepsy disrupts the maintenance and orderly occurrence of wake and sleep stages. Cataplexy is a highly specific symptom of narcolepsy, but many other symptoms can be seen in a variety of sleep disorders.
- The diagnosis of narcolepsy with cataplexy (type 1) is straightforward; however, the diagnosis of narcolepsy without cataplexy (type 2) is challenging and requires a good understanding of available diagnostic tests and their limitations.
- Narcolepsy with cataplexy is caused by loss of the hypocretin-producing neurons. As hypocretins influence many neurologic functions, narcolepsy is associated with multiple comorbidities, and it is important to recognize and to treat these comorbid conditions in addition to treating narcolepsy.
- Although there is no cure for narcolepsy, treatments are often effective and include both behavioral and pharmacologic approaches.

Gélineau[1] first coined the term narcolepsy in 1880 to designate a pathologic condition characterized by irresistible and brief episodes of sleep recurring at close intervals. Although Westphal[2] and Fisher[3] previously published reports of patients with excessive daytime sleepiness (EDS) and episodic muscle weakness, Gélineau was the first to characterize narcolepsy as a distinct syndrome. He wrote that falls, or "astasias," sometimes accompanied attacks. Henneberg[4] later referred to these attacks as cataplexy. In the 1930s, Daniels[5] emphasized the association of daytime sleepiness, cataplexy, sleep paralysis, and hypnagogic hallucinations with the syndrome. Referring to these symptoms as the clinical tetrad, Yoss and Daly[6] and Vogel[7] reported nocturnal sleep-onset rapid eye movement (REM) periods in narcolepsy patients, a finding confirmed in the following years.[8-11] However, it was von Economo who first hypothesized that narcolepsy is caused by injury to neurons in the lateral hypothalamus. He also indicated an association between narcolepsy and the influenza pandemic that occurred in 1918–1919.[9,10]

The discovery of the hypocretin peptides (often called orexins by basic researchers),[11,12] the mapping of hypocretin projections from the lateral hypothalamus to many parts of the brain (including the sleep-wake and autonomic nervous regulatory system),[13] and the discovery that the hypocretin-producing neurons are destroyed in the brains of people with narcolepsy[14,15] are key elements in our current understanding of narcolepsy. Honda and Juji[16] were the first to discover that narcolepsy is strongly associated with human leukocyte antigen (HLA) DQB1*06:02. Further findings of other immune function genes strongly support the concept that narcolepsy with cataplexy in its most common form is likely an autoimmune disorder of the brain (see Chapter 89).[17] Although this is the most common cause of narcolepsy, destruction of the hypocretin neurons or their projections by other pathologic processes (e.g., trauma, tumors, infections)[18] can cause secondary cataplexy.[19]

Hypocretins control many functions, and narcolepsy is more than just sleepiness or dysfunctional REM sleep. Narcolepsy causes instability of wake and sleep stages; patients have the capacity to achieve wakefulness and non-REM (NREM) and REM sleep but are unable to maintain these states. In addition, patients with narcolepsy can have intermixed fragments of sleep states, such as cataplexy, which likely represents a combination of the waking state and the paralysis of REM sleep.[20]

CLINICAL FEATURES

The third edition of the *International Classification of Sleep Disorders* reclassified this syndrome into narcolepsy type 1 (narcolepsy with cataplexy, Na-1) and narcolepsy type 2 (narcolepsy without cataplexy, Na-2).[21] The classic tetrad of symptoms includes EDS with variable amounts of cataplexy, sleep paralysis, and hypnagogic hallucinations. EDS occurs in all patients with narcolepsy, but only about one third of patients have all four of these symptoms. Automatic behaviors and disrupted nighttime sleep also commonly occur. Symptoms suggestive of narcolepsy can occur in any person who is severely sleep deprived, but *only* cataplexy is unique to narcolepsy. Narcolepsy has a population prevalence of 0.02% to 0.06%[22-25] and affects both sexes equally.

Sleepiness

Unwanted episodes of sleep recur several times a day, not only under favorable circumstances, such as during monotonous sedentary activity or after a heavy meal, but also in situations in which the patient is fully involved in a task. The duration of sleep may vary from a few seconds to several minutes if the patient is in an uncomfortable position to longer than 1 hour if the patient is reclining. Narcolepsy patients characteristically wake up feeling refreshed, and there is a refractory period of relative alertness for 1 hour to several hours before the next

episode occurs. These relatively short but refreshing naps can help differentiate patients with narcolepsy from patients with idiopathic hypersomnia, who often take long and unrefreshing naps. Despite feeling sleepy during the day, narcolepsy patients generally do not sleep more in 24 hours compared with those without narcolepsy. Apart from sleep attacks, patients complain of feeling abnormally drowsy, resulting in poor performance at work, memory lapses, and even gestural, ambulatory, or speech automatisms. These lapses are related to unconscious microsleep episodes that can become more frequent as the patient fights against falling asleep.

Cataplexy

Cataplexy occurs in 60% to 70% of narcolepsy patients.[21,26] It is an abrupt and reversible decrease or loss of muscle tone, most frequently elicited by strong emotions such as laughter, anger, and surprise. It may involve certain muscles or the entire voluntary musculature. Most typically, the jaw sags, the head falls forward, the arms drop to the side, and the knees buckle. Awareness is preserved throughout the attack.

The severity and extent of cataplexy attacks can vary from a state of absolute powerlessness, which seems to involve the entire voluntary musculature, to limited involvement of certain muscle groups or to no more than a fleeting sensation of weakness. Some patients may report blurred vision, possibly due to eyelid or extraocular muscle weakness. Speech may be impaired, and respiration may become irregular during an attack. Apneas have never been recorded, but short pauses similar to those seen during nocturnal REM sleep in healthy subjects may occur.

Rarely in a cataplexy attack, there is complete loss of muscle tone that can lead to total body collapse and serious injuries, such as skull or other bone fractures. However, in most cases, cataplexy is not that extreme and may even go unnoticed by individuals nearby. An attack may consist of only slight buckling of the knees. Patients may perceive this abrupt and short-lasting weakness and simply stop or stand against a wall. Speech may be broken or slurred because of intermittent weakness affecting the voice. If the weakness involves only the jaw or speech, the patient may present with wide masticatory movement or an unusual attack of stuttering speech. If it involves the upper limbs, the patient may complain of "clumsiness," reporting activity such as dropping cups or plates or spilling liquids when surprised, laughing, and the like. Many patients have both complete and partial attacks. At the onset of cataplexy, the abrupt muscle inhibition can be interrupted by recurrent bursts of normal muscle tone that can mimic a tremor.

Short attacks are the most common presentation of cataplexy. Because they do not resemble the "classic" full-blown attack of cataplexy, they are often missed even by skilled physicians without the aid of electromyographic recording.[27] By the same token, the skilled physician must be cautious not to overdiagnose normal phenomena, such as the "rubber knees" that precede the anxiety of public speaking or "rolling on the ground laughing." The duration of each cataplexy attack—whether partial or complete—is variable, lasting from a few seconds to 30 minutes, but in most cases it is 30 seconds to 2 minutes. Status cataplecticus is a rare manifestation of cataplexy, characterized by prolonged cataplexy lasting hours, and it can be triggered by sudden withdrawal of cataplexy-suppressing medications or insufficient sleep.

Attacks can be elicited by emotion or stress. Laughter and anger seem to be the most common triggers, although "laughing excitedly" is a more potent trigger than simple laughter.[28] These attacks can also be induced by a feeling of elation while listening to music, reading a book, or watching a movie, but a certain intensity of the associated emotion triggered by the abrupt new situation is an important element. Cataplexy can be induced merely by remembering a happy or funny situation, and it may rarely occur without clear precipitating acts or emotions, particularly if the patient is sleepy. It often occurs when the patient is telling a joke and frequently occurs when the patient simply anticipates saying something humorous. A large phenotypical diversity has been reported for cataplexy that may render it difficult to recognize.[28]

The following features are seen in narcolepsy but can also occur in normal individuals.

Sleep Paralysis

This is a terrifying experience that occurs on falling asleep or awakening from sleep. Patients find themselves paralyzed, suddenly unable to move the limbs, to speak, or even to breathe deeply. The patient is fully aware of the condition and can recall it completely later. This state is frequently accompanied by hallucinations. In many episodes of sleep paralysis, but especially the first occurrence, the patient may experience extreme anxiety associated with fear of dying. This anxiety is often greatly intensified by the terrifying hallucinations that may accompany the sleep paralysis. Patients often interpret the experience as being "scared stiff." With more experience with the phenomenon, however, the patient usually learns that episodes are brief and benign, rarely lasting longer than a few minutes and always ending spontaneously. Sleep paralysis may occur as an independent and isolated phenomenon, and 3% to 5% of the general population may be affected by it.[29]

Hallucinations

Patients may have vivid and often unpleasant auditory or visual hypnagogic hallucinations at sleep onset, either during daytime naps or at night.[30] Hallucinations may occur on awakening; these hypnopompic hallucinations may be more characteristic of narcolepsy than the hypnagogic ones at sleep onset. The visual hallucinations usually consist of simple forms (colored circles, parts of objects) that are either constant or changing in size. The image of an animal or a person may present abruptly and more often in color. Auditory hallucinations are also common, although other senses are seldom involved. The auditory hallucinations can range from a collection of sounds to an elaborate melody. Hypnopompic hallucinations are often perceived as so vividly realistic that the patient acts on them on awakening. For example, patients with narcolepsy have been known to call the police because of intruders being in the home, only to discover after authorities have searched the house that it was all a hallucination. The exact boundary between hypnagogic/hypnopompic hallucinations and dreams is not a clear one. In some cases of unrecognized narcolepsy with daytime hypnagogic/hypnopompic hallucinations, the patient may be mistakenly diagnosed as having a psychosis.[31]

Another common and interesting type of hallucination reported at sleep onset involves elementary feelings (i.e., experiencing picking, rubbing, or light touching), changes in location of body parts (e.g., arm or leg), or feelings of levitation

or extracorporeal experiences (e.g., the body in space or floating above the bed) that may be elaborate. For example, the patient may report, "I am above my bed and I can also see my body below" or "I am a few feet up and people are jumping over my body." The abrupt motor inhibition that involves the spinal cord motor neurons may alter the feedback of information that is normally used by the central nervous system to gauge the position of the body and limbs.

Sleep Disruption

Narcolepsy patients commonly experience fragmented nocturnal sleep. Night sleep is often interrupted by repeated awakenings and sometimes accompanied by terrifying dreams. Ironically, patients complain of difficulty in staying asleep at night, although they may fall asleep repeatedly during daytime. Rarely, insomnia and secondary daytime tiredness are the initial complaints. The sleep disruption may be enhanced by the presence of periodic limb movements, REM sleep behavior disorder, or obstructive sleep apnea (OSA), which appear to be common in narcolepsy patients.[32,33]

COMORBID ASSOCIATIONS

Narcolepsy is caused by loss of the hypocretin neurons, and this can result in additional symptoms as hypocretins regulate other functions in addition to sleep-wake regulation. Weight gain is common, and it seems to be worse in younger children at the onset of narcolepsy. Particularly in Chinese children, body mass index tends to increase and fatty infiltration of the tongue may occur, decreasing the size of the upper airway and consequently leading to OSA. The association with OSA was noted years ago,[32] but its high frequency was appreciated only in later years.[33]

Hypocretin deficiency affecting the autonomic nervous system can affect blood pressure and heart rate.[34,35] It also affects core body temperature, which can affect sleep tendencies.[36] Sexual dysfunction (difficulty with erection and ejaculation)[37] was also documented long ago, a problem that may be worsened by some anticataplexy drugs. Increased frequency of migraines has also been reported.[38] Increase in pain threshold and chronic pain is also frequent in narcolepsy.[39]

Psychiatric comorbidities are frequent, especially major depressive disorder, with one study reporting an increased risk of suicide.[40] In Chinese children who developed narcolepsy with cataplexy during prepubertal age, severe schizophrenia was noted in about 10% of these children. The schizophrenia poorly responded to treatment and was also associated with abnormalities particularly in the prefrontal lobe documented by positron emission tomography studies compared with other patients with narcolepsy or schizophrenia.[41]

These comorbidities may impair quality of life, not only for the patient but also for the caretaker. Poor quality of life in this group has been documented in studies from multiple countries.[42-45] A large study in the United States involving 55,871 subjects, including 9312 in the narcolepsy group and 46,559 matched controls, characterized health care utilization, costs, and productivity in a large population of patients diagnosed with narcolepsy; it showed that narcolepsy with its comorbidities was associated with substantial personal and economic burdens, as indicated by significantly higher rates of health care utilization and medical costs. Narcolepsy is also associated with higher rates of accidents, employee absence due to short-term disability, job dismissal, and early retirement.[46]

ONSET OF CLINICAL SYMPTOMS

In white people, the first symptom often develops near the age of puberty; the peak age at which reported symptoms occur is 15 to 25 years,[47] but narcolepsy and other symptoms have been noted as early as 2 years and at 6 months of age in one child with a hypocretin gene mutation.[48] In China, the onset of narcolepsy often occurs during the prepubertal years.[49] A few years ago, many children and adolescents developed narcolepsy with cataplexy soon after receiving Pandemrix, a specific brand of H1N1 influenza vaccine no longer in use. In these post-vaccination patients, narcolepsy with cataplexy began at an earlier age than previously reported in white children.[50-52]

A survey of 157 narcolepsy patients found that about 80% experienced symptoms before the age of 30 years,[53] but a second, small peak of onset has been noted between 35 and 45 years and near menopause in women. In those who developed narcolepsy at age 60 years or later, cataplexy was the most common initial symptom.[53] Diagnosis of narcolepsy is often delayed more than 10 years, especially if cataplexy is initially absent.[54]

EDS and irresistible sleep attacks usually occur as the first symptoms, either independently or associated with one or more symptoms. Sleep attacks are enhanced by high environmental temperature, indoor activity, and idleness. Symptoms may lessen with age but never resolve completely. When cataplexy occurs years after onset, the diagnosis of narcolepsy type 1 versus type 2 can be challenging. Cataplexy occasionally occurs before the abnormal sleep episodes or EDS, in which case it is a major source of difficulty in diagnosis. Cataplexy can vary in frequency from a few episodes during the patient's entire lifetime to one or several episodes per day. Cataplexy attacks have an overall tendency to decrease in frequency with aging. Disturbed nocturnal sleep usually worsens with age.[54]

DIAGNOSIS OF NARCOLEPSY

The diagnosis of narcolepsy requires a clinical history of EDS and a positive Multiple Sleep Latency Test (MSLT) result, with a mean sleep-onset latency of 8 minutes or less and two or more sleep-onset REM periods (SOREMPs).[21] It is important to categorize narcolepsy as type 1 (Na-1, narcolepsy with cataplexy) or type 2 (Na-2, narcolepsy without cataplexy). Na-1 is a distinct phenomenon with consensus on presentation and diagnostic criteria. Na-2, however, remains an "evolving diagnosis," and there is much debate among researchers on whether it is a distinct phenomenon or part of the spectrum of narcolepsy and idiopathic hypersomnia.

Na-1 requires a history of EDS and at least one of the following: (1) cataplexy and a positive MSLT result or (2) cerebrospinal fluid (CSF) hypocretin-1 deficiency. Hypocretin-1 deficiency is defined as CSF hypocretin-1 level less than one third of normal (or ≤110 pg/mL if analyzed using the Stanford reference). Low CSF hypocretin-1 levels are highly specific for Na-1 and can confirm the diagnosis.[55] Nocturnal SOREMPs (within 15 minutes of sleep onset) can be included in the tally of SOREMPs.[21]

The diagnosis of Na-2 is more of a problem and challenging even to the most experienced clinician.[56] Na-2 requires a history of EDS and a positive MSLT result. Na-2 is a clinical diagnosis or rather a diagnosis of exclusion. A positive MSLT result is not specific for narcolepsy; many symptoms of narcolepsy are nonspecific, and biomarkers are lacking in Na-2. CSF hypocretin-1 levels are almost always normal, and cataplexy is absent. Moscovitch et al[57] also cautioned against the diagnosis of narcolepsy if EDS and two or more SOREMPs on MSLT are the only positive findings. The authors also found that only 84% of individuals with complaints of EDS and cataplexy presented with two or more SOREMPs on MSLT, a replication of findings by van den Hoed et al[58] 10 years earlier. When the MSLT was repeated daily for 4 days, all subjects with EDS and cataplexy had two or more SOREMPs in at least one MSLT. A recent consensus paper addresses the challenges of diagnosing narcolepsy without cataplexy and in standardizing its definition and diagnostic criteria.[56]

EVALUATION OF SLEEPINESS

The Stanford Sleepiness Scale,[59] a 7-point scale, was developed to quantify the subjective sleepiness of patients throughout the day, but it is often difficult for patients to accurately rate themselves every 15 to 20 minutes. The Epworth Sleepiness Scale is more commonly used as an index of subjective sleepiness in adults (see Chapter 169). The Pediatric Daytime Sleepiness Scale is validated for use in children and teenagers.

Multiple Sleep Latency Test

The MSLT (see Chapter 169) was designed to measure physiologic sleep tendencies in the absence of alerting factors.[60] This test consists of five scheduled naps, usually at 10 AM, noon, and 2, 4, and 6 PM, during which the subject is polygraphically monitored in a comfortable, soundproof, dark bedroom while wearing street clothes. The MSLT records the latency for each nap (time between lights out and sleep onset), the mean sleep latency, and the presence or absence of REM sleep during any of the naps.[61] On the basis of polygraphic recording, REM sleep that occurs within 15 minutes of sleep onset is considered a SOREMP.[62] After each 20-minute monitoring period, the patient stays awake until the next scheduled nap. A positive MSLT result requires a mean sleep-onset latency of 8 minutes or less plus two SOREMPs.

There is controversy about the reliance on a positive MSLT result in diagnosing narcolepsy. With increasing age, there is a progressive decrease in the number of SOREMPs and an increase in the mean sleep latency on the MSLT, suggesting that the standard criteria may be too stringent in older patients.[63] General population investigations have shown that normal subjects can present with two or more SOREMPs, and there seems to be a greater chance of having narcolepsy with the presence of two or more SOREMPs, but the likelihood of having narcolepsy in relation to an increasing number of SOREMPs is unknown.[64] In Japan, there is less reliance on the MSLT, and a positive history of cataplexy associated with EDS is systematically required for the diagnosis of narcolepsy.[65] Cataplexy is pathognomonic of narcolepsy, but it may be difficult to rely on this symptom alone, particularly when cataplexy is partial (i.e., limited to the head and neck or neck and upper arms). Anic-Labat et al[66] developed a self-administered questionnaire that validated cataplexy in 1000 subjects.

Although the MSLT is an objective measurement of sleepiness, it comes with limitations. A positive MSLT result is not specific for narcolepsy. Studies have found a high prevalence of SOREMPs and a positive MSLT result in the general population.[64,67] In the normal population, MSLT scores vary with age, sex, and puberty. Prepubertal children between the ages of 6 and 11 years appear to be hyperalert. In postpubertal subjects, mean MSLT scores below 8 minutes are generally considered to be in the pathologic range; those above 10 minutes are considered to be normal. Mean latencies of 8 to 10 minutes represent a gray area.[57] There is a progressive decrease in the number of SOREMPs and an increase in the mean sleep latency on the MSLT as a function of age.[63] Recent data from the Wisconsin Sleep Cohort reported that in older adults, the diagnostic value of the MSLT is strongly affected by shift work and chronic sleep deprivation.[68] The MSLT can produce false-positive results in these groups. Therefore, it is especially important to document prior sleep history by use of actigraphy or sleep log. Data from the Wisconsin cohort also found that positive MSLT results were much more common in men.[68]

An MSLT performed alone has drawbacks and should not be valid as it measures sleepiness regardless of its cause, which may simply be due to sleep deprivation or another sleep disorder. The MSLT also ignores repetitive microsleeps that can lead, in borderline cases, to daytime impairment not scored by conventional analysis. It is important to exclude factors or other sleep disorders that may cause false-positive results with the MSLT, such as sleep apnea, sleep deprivation, shift work, or medications.[69-73] Another potential drawback of the MSLT is related to methodology. Despite published guidelines from the American Academy of Sleep Medicine on how to conduct the test,[74] sleep laboratories vary widely in methodology, including interscorer reliability in identifying SOREMPs. On the night preceding the MSLT, the subject should undergo standard nocturnal polysomnography. Throughout the total nocturnal sleep period, any sleep-related biologic abnormalities responsible for sleep fragmentation and sleep deprivation are recorded. Unfortunately, some sleep laboratories do not uniformly perform polysomnography on the preceding night or use actigraphy or sleep log before conducting the MSLT. Guidelines on when to discontinue medications that may interfere with SOREMPs are also lacking, and protocols are not uniform across sleep laboratories.

Finally, actigraphy for 2 weeks (or sleep logs if actigraphy is not available) should be performed before the sleep study to document sleep-wake schedules and to rule out circadian disorders or insufficient sleep. Once the nocturnal sleep recording has eliminated specific diseases, the MSLT aids in the diagnosis of narcolepsy with a mean sleep latency of 8 minutes or less and the presence of two or more SOREMPs. To date, the MSLT is the best test available in diagnosing Na-2.

Browman et al[75] proposed adding to the MSLT a test for the maintenance of wakefulness (MWT). The MWT tests the patient's ability to remain awake in a comfortable sitting position in a dark room for different trials given at 2-hour intervals during the daytime. Traditionally, the test was composed of five sessions at 2-hour intervals (10 AM, noon, and 2, 4, and 6 PM), similar to the MSLT. Each opportunity for remaining

awake lasts 20 minutes. More recently, the test consists of four 40-minute sessions administered at 2-hour intervals (9 and 11 AM, 1 and 3 PM), with a cutoff point around 30 minutes before falling asleep. The MWT has been validated with different performance tests, including driving simulation.[76] The MWT may be helpful in drug trials, in evaluating a patient's response to treatment, and in evaluating the risk of falling asleep associated with specific jobs or activities.[75]

Polysomnography

Overnight polysomnography is used to rule out other sleep disorders and to evaluate sleep quality and quantity. The occurrence of REM sleep within 15 minutes of sleep onset on the preceding polysomnogram before the MSLT can now be included in the requirement of two or more SOREMPs on the MSLT.[21] This new recommendation is based on the finding that the occurrence of REM sleep 15 minutes or less after sleep onset is highly specific for Na-1 and associated with low CSF hypocretin-1 levels.[73]

Genetic Testing

Genetic testing is sometimes used in the clinical diagnosis of narcolepsy. HLA DQB1*06:02 is the most specific genetic marker for narcolepsy across all ethnic groups, and it is found in 95% of patients with Na-1 (with cataplexy).[17,77] In Na-2 (no cataplexy), only 40% of subjects have DQB1*06:02,[17] so HLA testing is generally unhelpful. Furthermore, genetic testing alone is insufficient for the diagnosis of narcolepsy as DQB1*06:02 is found in 18% to 35% of the general population. However, previous studies indicated that homozygosity for DQB1*06:02 doubles to quadruples the risk for narcolepsy. The investigation of children in East Asia led to the recognition of an HLA association between HLA DQB1*06:02 and DQB1*03:01 that also increased the risk for narcolepsy. Conversely, the presence of DQB1*05:01 and DQB1*06:01 in heterozygote individuals decreased the risk for development of narcolepsy.[66,78,79]

Hypocretin-1 Measurement in Cerebrospinal Fluid

Hypocretin neurons are selectively damaged in patients with narcolepsy with cataplexy.[80] By a lumbar puncture, a very low or nonexistent level of hypocretin can confirm the diagnosis of narcolepsy with cataplexy.[81-85] CSF hypocretin-1 levels below 110 ng/L (measured using the Stanford University technique) have a high positive predictive value (94%) for narcolepsy type 1.[55,80] In Na-2, however, hypocretin-1 levels are usually normal. A study by Andlauer et al[86] found that only 24% of Na-2 patients had low CSF hypocretin-1 levels (defined as <110 pg/mL). A low CSF hypocretin-1 level without cataplexy is also more common in African Americans.[86] Currently, CSF hypocretin-1 measurement is the most accurate diagnostic technique available. CSF hypocretin-1 can help confirm narcolepsy and distinguish between type 1 and type 2. In rare cases, low CSF hypocretin-1 levels may be due to a neurologic condition that injures the hypocretin neurons, such as brain tumors, encephalitis, vascular diseases, and brain trauma.[80]

CHILDREN AND NARCOLEPSY

Prepubertal children developing narcolepsy often present with severe symptoms, with cataplexy and partial cataplexy predominantly involving the face, and a variable amount of movement abnormalities, such as "tics" or other neurologic problems, may be simultaneously noted.[87] Frequent irritability and aggressive behavior have been mentioned in large series. Rapid and tremendous weight gain is common in these early appearance cases.[88] In one series, rapid development of obesity in children with narcolepsy was associated with later development of schizophrenia.[41] Children with early narcolepsy onset may also present with precocious puberty.[89]

TREATMENT

The goal of all therapeutic approaches is to optimize control of narcolepsy symptoms and to allow the patient to have a full personal and professional life. Treatment goals should focus on improving EDS, cataplexy attacks, hypnagogic/hypnopompic hallucinations, sleep paralysis, nocturnal sleep, and psychosocial difficulties. When selecting medications, clinicians must take into account possible side effects because narcolepsy is a lifelong illness and patients will have to receive medication for years. Tolerance or addiction may occur with some compounds. In addition, hypertension and psychosis are the most commonly reported complications associated with the long-term use of stimulant medications. The treatment of narcolepsy must balance the maintenance of an active life with the avoidance of side effects and tolerance to medications.[90]

Behavioral Approaches

Part of the difficulty in treating patients with narcolepsy involves the patient's frustration over the delay in diagnosis. The consequence of this, particularly in a young patient, is the development of reactive depressive symptoms. One of the most important initial treatments is a referral to patient support groups organized by sleep disorders centers, such as those sponsored by the U.S. National Sleep Foundation and the Narcolepsy Network. Other support groups exist in most western European countries and in North America. Patients can also find helpful information on the websites of the American Academy of Sleep Medicine, Stanford University, and Harvard University.

Career counseling is also important because patients and their employers must be educated about jobs that patients with narcolepsy should avoid, including shift work, on-call schedules, driving and the transportation industry, and any job necessitating continuous attention for long hours without breaks, particularly under monotonous conditions. Some of these difficulties can be overcome if the employer recognizes the importance of short 15- to 20-minute naps every 4 hours during the daytime. In addition to scheduled naps, other important behavioral treatment targets include a regular sleep-wake schedule, avoidance of frequent time zone changes, and overall good sleep hygiene (Tables 90-1 and 90-2).

Pharmacologic Treatments

Pharmacologic treatments are listed in Tables 90-1 to 90-3.

All Major Symptoms of Narcolepsy (Excessive Daytime Sleepiness, Cataplexy, Disturbed Nocturnal Sleep)

Sodium Oxybate. Sodium oxybate is the sodium salt of gamma-hydroxybutyrate (GHB) and is approved by the Food and Drug Administration for the treatment of cataplexy and sleepiness in narcolepsy. In 1979, Broughton and Mamelak[91]

Table 90-1 Examples of Initial Treatment Packages for Children

Prepubertal Children	Adolescents
General Measures	
Contact school to alert teachers	Contact school to alert teachers
Nap at lunchtime	Emphasize need for regular nocturnal sleep schedule
Nap at 4 or 5 PM	Try to obtain 9 hours of nocturnal sleep
	Nap at lunchtime and 4 or 5 PM
Medications for Sleepiness	
Modafinil 100–200 mg*	Modafinil 100–400 mg*
Sodium oxybate 6–8 g	Sodium oxybate 6–9 g
Methylphenidate 5 mg (2–4 tablets†)	Methylphenidate 5 mg (2–6 tablets†) or 20 mg sustained-release tablet in AM (on empty stomach)
Atomoxetine 10–25 mg	Atomoxetine 10–25 mg
Medications for Cataplexy‡	
Sodium oxybate 6–8 g	Sodium oxybate 6–9 g
Venlafaxine XR 75–150 mg in AM	Venlafaxine XR 75–150 mg in AM
Fluoxetine 10–20 mg in AM	Fluoxetine 10–40 mg in AM

*Modafinil is started at 100 mg in the morning for 5 days, and a second dose of 100 mg is then added at lunchtime, if needed. This is usually sufficient in prepubertal children. Pubertal children may require a further increase (after 5 days) to an additional 100 mg in the morning and, if still needed later, another 100 mg at noon.
†Usually 10 mg when waking up on an empty stomach, 5 mg around lunchtime, and 5 mg at 3 PM.
‡No medications have specifically received approval by the Food and Drug Administration for use in patients with narcolepsy younger than 16 years. The use of antidepressants for cataplexy has not been approved by the Food and Drug Administration.

first suggested that GHB improved cataplexy, daytime sleepiness, and disturbed nocturnal sleep in patients with narcolepsy. GHB, a naturally occurring central nervous system metabolite found in highest concentrations in the hypothalamus and basal ganglia, acts as a sedative to consolidate sleep. Sodium oxybate is now a treatment of choice for narcolepsy with cataplexy. In addition to its use as an anticataplexy medication, it also improves EDS and nocturnal sleep disturbances. Sodium oxybate treats cataplexy through an unknown mechanism that is thought to be related either to its consolidation of REM sleep or to a secondary interaction with dopamine secretion.[91-96]

GHB is a neurotransmitter and neuromodulator that affects dopamine, serotonin, gamma-aminobutyric acid (GABA), and endogenous opioids. It is postulated to act mainly through $GABA_B$ receptors as mice lacking $GABA_B$ receptors show little response to GHB and similar drugs. In large, multicenter studies, sodium oxybate has shown efficacy in improving the EDS and cataplexy of narcolepsy.[95,97] The drug is normally taken at bedtime with the patient already in bed to avoid falls. A second dose is taken approximately 2.5 to 4 hours after the first one while the patient is in bed. Its half-life is 90 to 120 minutes. Many patients start with a total nightly dose of 4.5 to 6 g, and the dose is gradually titrated up during 2 to 3 months to 6 to 9 g, based on improvements in EDS and cataplexy. These doses produce a small increase in nocturnal total sleep time and a decrease in sleep paralysis, hypnagogic hallucinations, and nightmares. Deeper nighttime sleep is apparent early in treatment, but it may take more than 3 months to see the full benefits of the medication on EDS and cataplexy. Overall, a positive effect on cataplexy is reported. The response is gradual but is seen much more rapidly than the effect on alertness, with a significant decrease in the frequency of cataplexy attacks. If patients wake 1 to 2 hours after dosing, they may be confused and disoriented, and they may have episodes of enuresis, particularly at a high dosage of 9 g and also during initial use of the drug. Nausea as well as sluggishness in the early mornings may be seen with high dosage. In clinical trials, sodium oxybate produced dose-related improvements in cataplexy and EDS, with no acute rebound effect with withdrawal after 4 weeks of therapy. There is a sustained increase in efficacy on cataplexy during 12 months, with maintenance of efficacy on daytime sleepiness. Sodium oxybate shows a dose-related increase in slow wave sleep and reduction in nocturnal awakenings at doses of 7.5 to 9 g. Studies have not shown negative effects on respiratory parameters.[95-104] It is not recommended in pregnancy.

Atomoxetine. Atomoxetine, a selective noradrenergic reuptake inhibitor also used in the treatment of cataplexy, has been effective in improving daytime sleepiness, particularly in children. It may allow prescription of only one medication to deal with both main symptoms (daytime sleepiness and cataplexy), but it is less effective than modafinil and sodium oxybate in older teenagers and adults.[105-107]

Excessive Daytime Sleepiness

Modafinil. Modafinil is a first-line pharmacologic treatment for EDS in narcolepsy. Modafinil is a wake-promoting drug whose mechanism is unknown, but it has been hypothesized to selectively activate wake-generating sites in the hypothalamus. The mechanism of action of modafinil (2-[(diphenylmethyl)sulfinyl]acetamide) is different from that of the amphetamines.[108-111] In narcoleptic dogs, the drug increases extracellular dopamine in a hypocretin receptor 2–independent manner.[111] In dopamine transporter knockout mice, the lack of response to the wake-promoting action of modafinil indicates that dopamine transporters do play a role and are necessary for wake-promoting actions of this drug.[111]

Modafinil improves EDS in narcolepsy patients with relatively few side effects. Headache is the most common complaint, followed by nervousness, nausea, and dry mouth. These symptoms can be reduced by a slow and progressive increase in dosage. Blood pressure should be monitored as the drug may cause elevated pressures. Modafinil is not addictive and has a low potential for abuse. There is no evidence that tolerance develops to the effects of modafinil on EDS. Most European studies suggest twice-daily administration in the morning and at lunchtime. Its elimination half-life is 10 to 12 hours. Modafinil can be administered concurrently with anticataplexy medications. Modafinil can reduce the efficacy of oral contraceptives, and therefore patients should use alternative methods of birth control.

Table 90-2 Examples of Initial Treatment Packages for Adults

General Measures

Avoid shifts in sleep schedule.
Avoid heavy meals and alcohol intake.
Regular timing of nocturnal sleep: 10:30 PM to 7 AM
Naps: Strategically timed naps if possible (e.g., 15 minutes at lunchtime, 15 minutes at 5:30 PM)

Medications for Sleepiness

The effects of stimulant medications vary widely among patients. The dosing and timing of medications should be individualized to optimize performance. Additional doses, as needed, may be suggested for periods of anticipated sleepiness.
Modafinil* 100–200 mg (taken when waking up in the morning) and 100–200 mg at lunchtime *or*
Sodium oxybate† at bedtime: dosage must start low at 2.25 g taken twice while in bed (at bedtime and 2.5–4 hours after bedtime); increase to total dosage of 5 to 6 g within 2–4 weeks. This initial dose is usually ineffective, so increase to 3 g at bedtime and 3 g approximately 2.5 to 4 hours after bedtime if tolerated. Depending on response, dosage can be increased to as high as 9 g total nightly dose. Do not increase above 9 g because of risk of serious side effects during sleep. It may take more than 2 months for daytime symptoms to improve, and cataplexy may improve faster than excessive daytime sleepiness. If the patient is already taking a daytime stimulant, it may be possible to reduce the stimulant dose or to discontinue it once a therapeutic dosage of sodium oxybate has been reached.
Methylphenidate 5 mg (3 or 4 tablets; 10 mg when waking up; 5 mg 30 minutes before lunch; 5 mg near 3 pm; better action is always obtained if the drug is taken on an empty stomach) or 20 mg SR in the morning (on an empty stomach)

If Persistent Difficulties

Modafinil 200 mg in the morning and 200 mg at lunch (total daily dosage, 400 mg) *or*
Add sodium oxybate (GHB) at bedtime: dosage must start low as indicated above
Methylphenidate (SR): 20 mg in the morning; 5 mg after noon nap; 5 mg at 4 pm *or*
Possibly (more in teenagers) atomoxetine: start at 0.5 mg/kg within 1 week to appropriate dosage of 1 to 1.2 mg/kg taken in the morning

If No Response

Dextroamphetamine sulfate: 15 mg on awakening; 5 mg after noon nap; 5 mg at 3:30 or 4 PM (or 15 mg at awakening and 15 mg after noon nap)

Medications for Cataplexy‡

Sodium oxybate (see above)
Venlafaxine 150–300 mg
Fluoxetine 20–60 mg
Duloxetine 60 mg

If No Response

Clomipramine 75–125 mg, *or*
Viloxazine 150–200 mg, *or*
Imipramine 75–125 mg

*Modafinil works best in naive subjects. It should be the drug of first choice in children and adults.
†Response to sodium oxybate is slow.
‡Medications may be taken in the evening near bedtime (sodium oxybate, clomipramine, imipramine), only in the morning (fluoxetine), or in the morning and at lunchtime (viloxazine, venlafaxine). The only medications specifically approved for use in narcolepsy by the Food and Drug Administration are modafinil and sodium oxybate.
SR, Sustained-release tablet.

As modafinil mainly blocks dopamine reuptake, it can be less potent and have fewer side effects than amphetamine, which can block reuptake of dopamine, norepinephrine, and serotonin. As a consequence, when patients are switched from an amphetamine to modafinil, they may report "less control over sleepiness" and worse cataplexy. The combination of modafinil and venlafaxine can provide better control of cataplexy.

Armodafinil. Armodafinil is the active *R*-enantiomer of modafinil and is approved by the Food and Drug Administration for the treatment of EDS in narcolepsy. In a multicenter, randomized, double-blind, placebo-controlled trial of 196 subjects with narcolepsy, armodafinil significantly improved EDS throughout the day.[112,113] It may produce slightly longer improvements in EDS than regular modafinil. The usual dosing is 150 to 250 mg/day. Side effects are similar to those of modafinil, which include headache, nausea, dizziness, and insomnia.

Amphetamines and Amphetamine-like Central Nervous System Stimulants. Stimulants include amphetamines, such as dextroamphetamine and methamphetamine, and amphetamine-like drugs, including methylphenidate. Like modafinil, these drugs block the reuptake of dopamine, but they also block the reuptake of norepinephrine and serotonin and can cause efflux of these monoamine neurotransmitters from nerve terminals. These neurotransmitters all promote

Table 90-3 Narcolepsy Drugs Currently Available

Drug	Usual Dosage* (All Drugs Administered Orally)
Treatment of EDS	
Stimulants†	
Modafinil	100–400 mg/day
Sodium oxybate	6–9 g/day (divided in two doses)
Methylphenidate	10–60 mg/day
Atomoxetine	10–25 mg/day
Dextroamphetamine	5–60 mg/day
Methamphetamine	20–25 mg/day
Treatment of Auxiliary Effects (e.g., Cataplexy)	
Sodium oxybate (gamma-hydroxybutyrate)	6–9 g/day (divided in two doses)
Antidepressants‡	
Without Atropinic Side Effects	
Venlafaxine XR	75–300 mg/day
Fluoxetine	20–60 mg/day
Viloxazine	50–200 mg/day
Duloxetine	60 mg/day
With Atropinic Side Effects	
Protriptyline	2.5–20 mg/day
Imipramine	25–200 mg/day
Clomipramine	25–200 mg/day
Desipramine	25–200 mg/day

*On occasion, depending on clinical response, the dose may be outside the usual dosage range.
†Most stimulants should be administered in divided doses, commonly in the morning and at lunchtime. This is recommended for amphetamines and modafinil. Methylphenidate has a fast elimination rate, so the slow-release (SR) formula may be helpful in the morning (e.g., 20 mg SR). If it is administered by 5-mg increments, the usual timing of methylphenidate administration is every 3 to 4 hours until 3 PM.
‡The only medications approved for use in narcolepsy by the Food and Drug Administration are modafinil and sodium oxybate.
EDS, Excessive daytime sleepiness.

wakefulness and suppress REM sleep, and consequently, stimulants improve EDS, increase the latency to NREM and REM sleep, and reduce the percentage of REM sleep. At standard doses, amphetamines can enhance performance on simple motor and cognitive tasks; improve coordination; and increase strength, endurance, and mental and physical alertness, especially in monotonous situations.[114]

In adults, methylphenidate and amphetamines at dosages of more than 60 mg/day tend to produce side effects, including frequent worsening of the nocturnal sleep disruption and higher frequency of psychosis, paranoia, and psychiatric hospitalizations. Rebound hypersomnia is more frequent with higher dosages of amphetamines. These stimulants have a high potential for abuse and development of tolerance and therefore should be administered at the lowest effective doses. The drug is usually administered in three divided doses with a maximum of 20 mg in the morning, 20 mg at lunch, and 20 mg no later than 3 PM as these medications can disrupt sleep. The slow-release form may provide gradual and delayed responses during the daytime.

Our patients subjectively rate the effectiveness of methylphenidate on a level similar to that of modafinil. Black and Houghton[115] performed a double-blind placebo controlled comparison trial between modafinil, sodium oxybate, or the combination for the treatment of EDS in 270 adult narcolepsy patients and found that both drugs are effective in treating EDS and are additive when used together. Both the stimulant medications and the anticataplexy drugs are hepatically metabolized, so any liver dysfunction will affect both classes of drug. Thus, checking baseline liver function is recommended.

Cataplexy and REM Sleep–Related Symptoms

Cataplexy is thought to be an intrusion of REM sleep atonia into wakefulness. REM sleep is strongly suppressed by norepinephrine and serotonin, and drugs that increase levels of these neurotransmitters are often effective in suppressing cataplexy.

Monoamine Nonspecific Reuptake Inhibitors. Tricyclic antidepressants were the first medications used to treat cataplexy. The older tricyclics include imipramine, clomipramine, and protryptiline.[116-118] Tricyclic antidepressants inhibit monoamine (serotonin, norepinephrine, dopamine) reuptake and block cholinergic, histaminic, and alpha-adrenergic transmission and were the first drugs of choice, particularly protriptyline. However, they have significant anticholinergic side effects, including dry mouth, sweating, constipation, tachycardia, difficulty in urinating, and particularly sexual dysfunction that led to impotence in more than 40% of male narcolepsy patients. Tricyclic antidepressants are now rarely used to treat cataplexy long term as better options are available.

Selective Serotonin Reuptake Inhibitors. Several selective serotonin reuptake inhibitors have an active noradrenergic reuptake blocker metabolite; the prototype is fluoxetine and its active metabolite norfluoxetine.[119] We recommend starting with 20 mg of fluoxetine in the morning and increasing to 60 to 80 mg/day in two divided doses if needed. Fluvoxamine at 25 to 200 mg/day has been shown to be mildly effective for cataplexy. Compared with the classic tricyclic antidepressants, selective serotonin reuptake inhibitors can be less efficacious but have fewer side effects. Adverse effects include insomnia, nausea, and sexual difficulties. Tolerance to this class does not develop.

Serotonin-Norepinephrine Reuptake Inhibitors. These newer antidepressants have been found to be effective for cataplexy, sleep paralysis, and hypnagogic/hypnopompic hallucinations and are the recommended drugs because of fewer side effects and greater efficacy. The most commonly used drug of this class is venlafaxine XR, a potent inhibitor of serotonin and noradrenergic reuptake and a weak inhibitor of dopamine reuptake. At a dose of 75 to 150 mg, venlafaxine has been the most widely used of these compounds both in adults and in children; it has good efficacy and is better tolerated than the tricyclics. Duloxetine (60 mg each morning) is another serotonin-norepinephrine reuptake inhibitor that reduces cataplexy.[120] Atomoxetine (18 to 100 mg once a day or in two divided doses) is a highly specific noradrenergic reuptake inhibitor and has been tried in cases of resistant cataplexy after failure of fluoxetine, venlafaxine, and other serotonin reuptake inhibitors.

Serotonin-norepinephrine reuptake inhibitors and selective serotonin reuptake inhibitors are category C drugs when they are used during the last trimester of pregnancy and may lead to serious respiratory and feeding side effects in the neonate immediately after birth. Pregnant women may need to be progressively withdrawn from these medications before the third trimester. The withdrawal must be slow to avoid a marked cataplexy rebound that usually occurs on day 3 or 4 and peaks near day 10 after completion of withdrawal.

Medication Side Effects

Common side effects have been well described in previous reports, but some side effects that are rarely mentioned may be more important in narcolepsy patients, such as the appearance of periodic limb movements during sleep[121,122] and the development of REM sleep behavior disorder, particularly in older subjects.[63] Rebound cataplexy and other REM sleep–related symptoms, such as sleep paralysis and hypnagogic/hypnopompic hallucinations, can be seen with abrupt drug withdrawal. Patients should be withdrawn from these medications slowly after chronic drug use; the recommended withdrawal schedule is one dose every 4 days.[123]

Treatment of Children

On the basis of our clinical experience,[124] we recommend modafinil for the treatment of EDS in children with narcolepsy. A study of 13 children (mean age, 11.0 years) receiving modafinil (mean dose, 346 mg/day) showed reduction in sleep attacks in 90% of subjects and prolongation of sleep latency on MSLT (from 6.6 minutes to 10.2 minutes), and it appeared to be safe and well tolerated for more than a year.[125] Modafinil seems to work best when it is administered in the morning and at noon, but this may be a problem with children, so a single dose in the morning can been prescribed. The drawback of single-dose administration is that the therapeutic effect may not last late into the day. A 5-mg methylphenidate tablet may be added when the child returns from school, if needed. The drug must not be given too late to avoid inducing sleep-onset insomnia.

If modafinil cannot be prescribed, methylphenidate is the second best option; it has a short half-life and is available in a slow-release form.[126] In children, the recommended dose is based on weight, and we try to maintain a maximum of 30 mg/day administered as the slow-release form. The drug is given in the morning (15 mg) and at lunchtime (15 mg maximum) in the best case or as a 20- or 30-mg slow-release preparation in the morning.

In our experience, sodium oxybate also improves EDS and cataplexy in children with narcolepsy. The therapeutic response was sustained over time and without development of tolerance. Nausea, constipation, and mood changes were reported as initial side effects, and in a small series, side effects occurred in 40% of the children but rarely led to discontinuation of the drug. NREM parasomnias (e.g., sleep walking, night terrors) and enuretic episodes have also been reported, particularly at higher dosages.[127] Worsening of untreated OSA has been reported, and risk is probably higher in obese children.

Emerging Therapies and Pharmacologic Agents Under Investigation

Emerging treatments for narcolepsy can be classified into five broad categories, which include hypocretin-based treatments, immunotherapy, thyrotropin (TRH) analogues and promoters, histamine (H_3) antagonists, and combinations of or variations of currently used therapies and other therapies.[128,129]

A major limitation to hypocretin-1 replacement therapy is that it has to cross the blood-brain barrier to reach the central nervous system.[130-134] Because of the connections between the central nervous system to the outside environment through olfactory and trigeminal nerves, intranasal administration may deliver hypocretin into the nervous system.[129] A double-blind, randomized, crossover, placebo-controlled, within-subject design study administered human recombinant hypocretin-1 intranasally to eight Na-1 patients at bedtime and showed a reduction in REM sleep time and fewer wake to REM sleep transitions. A study of the same design conducted in 14 Na-1 subjects with MWT, attention testing, and second night of polysomnography[131] showed less REM sleep and less wake-REM transitions with treatment. Asahi et al[134] and Lang et al[135] found that the entire hypocretin protein is not necessary for biologic activity and selectivity for the corresponding receptors.[134,135] They determined the minimal sequences needed for receptor activation by synthesizing different combinations of C-terminal and N-terminal truncated peptides, in addition to fragments of central sequences of hypocretin-1 and hypocretin-2. They reported several analogues that selectively activated hypocretin-2 receptors.

Transplantation techniques, such as those used in Parkinson disease for dopaminergic neurons, could be applied, although survival of grafted hypocretin-producing neurons and immune reactions are limiting factors.[134] There is some evidence that TRH and TRH agonists have stimulant, antidepressant, and neurotrophic effects, and TRH has been shown to increase wakefulness and to decrease cataplexy in canine narcolepsy.[128]

Drugs that increase histamine signaling may be useful for improving EDS in narcolepsy.[129,136] Histamine signaling promotes wakefulness, whereas decreased activity leads to sleepiness. Histamine H_3 receptors are inhibitory autoreceptors on histamine and other monoamine neurons, and thus drugs that block these receptors (H_3 receptor antagonists or inverse agonists) should improve EDS. Thioperamide, a potent H_3 antagonist, significantly enhances wakefulness in hypocretin/orexin-deficient narcoleptic mice.[137] In a phase II study of people with narcolepsy, the H_3 inverse agonist pitolisant (40 mg each day) improved EDS compared with placebo. The most frequent side effects were headache, nausea, insomnia, and a fainting sensation. This study did not examine the effect of pitolisant on other symptoms of narcolepsy, such as cataplexy.[138] In 2012, a case series of four teenagers with narcolepsy with cataplexy was published demonstrating some positive data for off-label use of pitolisant for 13 months.[139]

The hypothesis of an autoimmune etiology has led some investigators to explore immunotherapy as a treatment for narcolepsy, including corticosteroids, plasmapheresis, and intravenous immune globulin.[140-143] The cases studies reported limited success and reduction of symptoms, suggesting that if a therapeutic benefit is obtained, it may be limited by timing of treatment (i.e., at onset of disease). Case reports present opposite findings.[142,143] Currently, intravenous immune globulin and corticosteroids are not recommended, and further trials are needed to determine the possible role of immunomodulators in the treatment of narcolepsy.[143]

ADX-N05, a unique phenylalanine derivative with dopamine and noradrenergic activities, is being investigated for the treatment of daytime sleepiness in narcolepsy patients.[144]

Preliminary findings show that at doses of 150 to 300 mg/day, the agent is well tolerated and significantly improved objective and subjective symptoms of sleepiness in 100 adults with narcolepsy. At 12 weeks, the average sleep-onset latency on the MWT was 12.8 minutes with the active drug compared with 2.1 minutes with placebo. Epworth Sleepiness Scale scores also significantly improved with active drug.

CLINICAL PEARLS

- Narcolepsy is characterized by the clinical tetrad of EDS, cataplexy, sleep paralysis, and hypnagogic/hypnopompic hallucinations. All narcolepsy patients have EDS, but other symptoms are present only in some individuals. Cataplexy is unique to narcolepsy, but other symptoms are nonspecific.
- Narcolepsy is associated with multiple comorbid conditions, including childhood obesity, sleep apnea, and psychiatric disorders including depression.
- The diagnosis of narcolepsy type 1 requires a history of EDS and one of the following: (1) low CSF hypocretin-1 level or (2) cataplexy and a positive MSLT result. The diagnosis of narcolepsy type 2 requires a history of EDS and a positive MSLT result.
- Genetic testing for HLA DQB1*06:02 is positive in 95% of patients with narcolepsy type 1, in 40% of patients with narcolepsy type 2, and in 18% to 35% of the general population without narcolepsy. Therefore, HLA testing is not recommended as a diagnostic tool.
- The MSLT must be preceded by nighttime polysomnography to rule out other sleep disorders and to document adequate sleep. The MSLT result can be falsely positive in other sleep disorders, such as shift work, sleep apnea, or sleep deprivation, and it is influenced by age, sex, and puberty.
- Sodium oxybate (GHB) can improve cataplexy, EDS, and fragmented sleep.
- Modafinil and armodafinil can reduce EDS without many of the side effects associated with older stimulants.

SUMMARY

Narcolepsy is a chronic neurologic sleep disorder due to hypocretin neuron loss resulting in EDS, disturbed nocturnal sleep, and intrusions of aspects of REM sleep in wakefulness, such as cataplexy, sleep paralysis, and hypnopompic/hypnagogic hallucinations. The syndrome is associated with multiple comorbidities. Multiple sleep latency testing after overnight polysomnography shows short sleep latencies (mean sleep-onset latency of 8 minutes or less) and two or more SOREMPs. In its most classic presentation, narcolepsy is associated with destruction of hypocretin neurons located in the lateral hypothalamus and absence of or significant reduction of hypocretin-1 measured in CSF. The destruction of hypocretin neurons is thought to be related to an autoimmune process. The treatment of narcolepsy has improved over time, with the newest compounds providing fewer side effects compared with classic stimulants for EDS and tricyclic antidepressants for cataplexy. Modafinil or armodafinil improves daytime sleepiness, and sodium oxybate improves both cataplexy and daytime sleepiness, and these are the drugs of choice. Narcolepsy is a complex syndrome that involves dysfunction in the timing of changes in wake and sleep stages. Medications with an alerting effect can help but will not completely control the multifaceted problems associated with this syndrome.

Selected Readings

Andlauer O, Moore H 4th, Hong SC, et al. Predictors of hypocretin (orexin) deficiency in narcolepsy without cataplexy. *Sleep* 2012;**35**:1247–55.

Andlauer O, Moore H, Jouhier L, et al. Nocturnal REM sleep latency for identifying patients with narcolepsy/hypocretin deficiency. *JAMA Neurol* 2013;**70**:891–902.

Baumann CR, Mignot E, Lammers GJ, et al. Challenges in diagnosing narcolepsy without cataplexy: a consensus statement. *Sleep* 2014;**37**:1035–42.

Black J, Reaven NL, Funk SE, et al. The Burden of Narcolepsy Disease (BOND) study: health-care utilization and cost findings. *Sleep Med* 2014;**15**:522–9.

Goldbart A, Peppard P, Finn L, et al. Narcolepsy and predictors of positive MSLTs in the Wisconsin Sleep Cohort. *Sleep* 2014;**37**:1043–51.

Kotagal S, Krahn LE, Slocumb N. A putative link between childhood narcolepsy and obesity. *Sleep Med* 2004;**5**:147–50.

Mattarozzi K, Bellucci C, Campi C, et al. Clinical, behavioral and polysomnographic correlates of cataplexy in patients with narcoleptic/cataplexy. *Sleep Med* 2008;**9**:425–33.

Mignot E, Hayduk R, Black J, et al. HLA-DQB1*0602 is associated with narcolepsy in 509 narcoleptic patients. *Sleep* 1997;**20**:1012–20.

Mignot E, Lammers GJ, Ripley B, et al. The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. *Arch Neurol* 2002;**59**:1553–62.

Sansa G, Iranzo A, Santamaria J. Obstructive sleep apnea in narcolepsy. *Sleep Med* 2010;**11**:93–5.

Sasai-Sakuma T, Kinoshita A, Inoue Y. Polysomnographic assessment of sleep comorbidities in drug-naïve narcolepsy-spectrum disorders—a Japanese cross-sectional study. *PLoS One* 2015;**10**(8):e0136988.

Singh M, Drake CL, Roth T. The prevalence of multiple sleep-onset REM periods in a population-based sample. *Sleep* 2006;**29**:890–5.

U.S. Xyrem Multicenter Study Group. Sodium oxybate demonstrates long-term efficacy for the treatment of cataplexy in patients with narcolepsy. *Sleep Med* 2004;**5**:119–23.

Wozniak DR, Quinnell TG. Unmet needs of patients with narcolepsy: perspectives on emerging treatment options. *Nat Sci Sleep* 2015;**7**:51–61.

A complete reference list can be found online at ExpertConsult.com.

Chapter 91

Idiopathic Hypersomnia

Yves Dauvilliers; Claudio L. Bassetti

Chapter Highlights

- Idiopathic hypersomnia (IH) is a rare etiology of central nervous system hypersomnolence characterized clinically by excessive daytime sleepiness with long and unrefreshing naps, prolonged and undisturbed nocturnal sleep, and great difficulty waking up and "getting going" (sleep drunkenness) after sleep.
- Polysomnography often shows normal nighttime sleep with high sleep efficiency and slow wave sleep percentages, without sleep apneas or periodic limb movements. A short mean sleep latency (<8 minutes) on the Multiple Sleep Latency Test is required with no, or a maximum of one, sleep-onset REM periods, or a prolonged sleep time on prolonged continuous polysomnography or actigraphy (total sleep time ≥11 hours on 24-hour monitoring).
- The pathophysiology of IH remains unclear. In the absence of a specific biologic marker, IH is a diagnosis of exclusion with a broad differential diagnosis, including atypical forms of depression, narcolepsy without cataplexy, sleep apnea syndrome, and behaviorally induced insufficient sleep syndrome.
- Pharmacologic options for IH are similar to the treatments for narcolepsy, including modafinil, methylphenidate, pitolisant, mazindol, and dextroamphetamine.
- Symptoms of IH may be severe and long lasting, but spontaneous improvement in hypersomnolence can occur in one fourth of patients.

HISTORY

The term idiopathic hypersomnia (IH) was used as early as 1829 (*"die idiopathische chronische Schlafsucht"*) for excessive daytime sleepiness (EDS) of undetermined origin.[1] Bedrich Roth was the first in the late 1950s to describe a syndrome characterized by EDS, prolonged sleep, and sleep drunkenness and by the absence of "sleep attacks," cataplexy, sleep paralysis, and hallucinations. The terms "independent sleep drunkenness" and "hypersomnia with sleep drunkenness" were initially suggested.[2-5] Overlapping features with narcolepsy were identified from the beginning and led to the use of such labels as essential narcolepsy, independent narcolepsy, and non–rapid eye movement (NREM) sleep narcolepsy.[6,7] Other terms, including idiopathic central nervous system hypersomnolence, functional hypersomnia, hypersomnia with automatic behavior, harmonious hypersomnia, and IH, were also put forward in the 1990 version of the *International Classification of Sleep Disorders* (ICSD). The 2005 version of the ICSD differentiated between IH with long sleep time (>10 hours; (polysymptomatic, classic form of IH) and IH without long sleep time (monosymptomatic form). The recent ICSD3 now pools both conditions (with and without long sleep time) into one heterogeneous condition because researchers were unable to objectively separate both forms of the disease based on the length of nocturnal sleep; patients above the cut-off of 10 hours of sleep showed no significant differences in daytime sleepiness assessed by the Epworth Sleepiness Scale (ESS), Multiple Sleep Latency Test (MSLT), and percentage of subjects with sleep drunkenness or unrefreshing naps.[8,9]

Although recent research provides some intriguing findings, the pathophysiology of IH remains unclear. The absence of a specific biomarker for IH, together with the presence of subtle forms of sleep-disordered breathing and chronic sleep insufficiency that may mimic IH, raises questions about the "true" frequency and clinical picture of IH.

EPIDEMIOLOGY

In the absence of systematic studies, the exact prevalence and incidence of IH remain unknown. A few reports suggest that IH patients represent only about 1% of patients seen in neurologic sleep centers and are 5 to 10 times less common than patients with narcolepsy.[8,10] Hence, the prevalence of IH in the general population may be estimated at approximately $50/10^6$, a figure much lower than the 300 to $600/10^6$ suggested until the early 1980s.[4,5,11] In a large series recently published, patients with IH represented 1% of 6000 patients seen at a single respiratory sleep center. Because IH was 60% as prevalent as narcolepsy, questions about the diagnostic accuracy arise.[12]

The age of onset of symptoms varies, but it is frequently between 10 and 30 years. In contrast to narcolepsy with cataplexy, the age of onset is sometimes difficult to pinpoint

because of the insidious onset of the condition over several weeks or months. When established, symptoms are generally stable and long lasting. Spontaneous improvement in EDS may be observed, however, in up to one fourth of patients.[10,12,13]

A female preponderance was found in some but not all series.[9,10,12,13] In one to two thirds of cases (see later), IH appears to be familial.

PATHOGENESIS

Genetic and Environmental Factors

IH can present in families, and these individuals are more likely to have long sleep time. On rare occasions, IH and narcolepsy may occur in the same family.[10,8,14] An autosomal dominant mode of inheritance has been discussed, and females may be affected more frequently[9]; however, no well-done studies in the field have been performed. In a few series, an association with diabetes or obesity was observed.[10,15]

Given the existence of overlapping features between IH and narcolepsy (see later), there has been an interest in potential HLA markers for IH. Despite reports of an increase in HLA-DQ1,[10] -DR5, -Cw2,[16] and -DQ3,[17] and of a decrease of HLA-Cw3,[18] no consistent findings have emerged. HLA typing currently does not play a role in the diagnosis of IH.

A recent study investigated the dynamics of the expression of circadian clock genes in dermal fibroblasts of 10 patients with IH compared with healthy controls. The amplitude of the rhythmically expressed *BMAL1, PER1,* and *PER2* was dampened in cells from IH patients over two circadian periods, and the overall expression of *BMAL1* was significantly reduced.[19]

Hypersomnia usually starts insidiously. Occasionally, EDS is first experienced after transient insomnia, abrupt changes in sleep-wake habits, overexertion, mood change, general anesthesia, viral illness, or mild head trauma.[10]

Neurochemistry

Montplaisir and colleagues found a decrease in dopamine and indoleacetic acid in both patients with IH and those with narcolepsy.[20] Faull and associates found dysregulation of the dopamine system in narcolepsy and of the norepinephrine system in IH.[21-23] These metabolic data suggest the possibility of a dysfunction of aminergic arousal systems in IH. Additional experimental and human data give some support to this hypothesis. In the cat, both hypersomnia and disturbances of monoamines can be induced by a lesion of ascending noradrenergic pathways.[24] Cerebrospinal fluid (CSF) hypocretin-1 levels are normal in IH.[15,25-27] Several studies have examined CSF levels of histamine in patients with IH and other central hypersomnias with discordant results.[28,29] Based on a highly sensitive and selective ultraperformance liquid chromatography tandem mass spectrometry assay to quantify simultaneously histamine and its major metabolite tele-methylhistamine, Dauvilliers and colleagues found no differences for these two amines between patients with narcolepsy with or without cataplexy or IH and patients with a complaint of EDS without any objective underlying etiologies.[30] Moreover, no association was found between CSF histamine and tele-methylhistamine, subjective or objective daytime sleepiness, or use of psychostimulants.

Rye and associates examined modulators of gamma-aminobutyric acid (GABA) signaling in patients with non-hypocretin-deficient central nervous system hypersomnia, some of whom met criteria for IH. They found that in the presence of GABA, CSF from these patients enhanced $GABA_A$ receptor function in an in vitro electrophysiologic assay (i.e., a whole-cell patch-clamp recording to measure the potentiation of $GABA_A$ receptor currents).[31] In this assay, inhibitory chloride currents were increased when cells were exposed to GABA combined with CSF from patients with hypersomnia compared with controls. Moreover, flumazenil, a drug that antagonizes the sedative-hypnotic actions of benzodiazepines, reversed this enhancement of $GABA_A$ signaling and may have improved vigilance in some hypersomnolent patients. This recent discovery is exciting but requires replication because CSF alone from IH patients had no effect on $GABA_A$ signaling. Also, no differences were observed between different diagnostic categories, and no correlation was found between GABA potentiation and measures of vigilance. Finally the bioactive CSF component remains unknown.

Neurophysiology

Researchers have hypothesized that the sleepiness of IH could be caused by homeostatic and circadian disturbances of sleep regulation as well as deficient activity in arousal systems.

Some investigators have reported an abnormally high level of slow wave activity (SWA) due either to an abnormal slow decay of SWA or to a normal decay of an enhanced level of SWA.[32,33] A recent study showed more sleep instability (i.e., more sleep stage shifts and more time spent in stage 1 sleep) in lighter sleep in IH than in narcolepsy with cataplexy (NC); more stable slow wave sleep (SWS) in IH than in normal subjects; and reduced cyclic alternating pattern (CAP) rate in IH, with higher values than NC in light NREM sleep and lower than normal subjects in SWS. These findings suggested an alteration of sleep microstructure in IH, with potentially less restorative sleep and daytime sleepiness in IH.[34]

Reports of increased sleep spindles activity at the beginning and at the end of sleep and of a delayed start (and decline) of melatonin and cortisol secretion suggest a primary circadian deficit in IH.[35,36] Unfortunately, core temperature recordings have not yet been reported in IH.

The detection of delayed and smaller P300 potentials after awakening in IH documents a cortical activation problem in these patients, but the origin of this phenomena is unknown.[33,37]

CLINICAL FEATURES

Excessive Daytime Sleepiness

Patients describe a constant, daily excessive daytime sleepiness (EDS) that only rarely leads to involuntary naps ("sleep attacks"). The ESS score is typically high (>11/24). As with other forms of EDS, alcohol, exercise, heavy meals, and warm environments can accentuate EDS. Daytime drowsiness leads to naps that are usually prolonged (typically >1 hour) and frequently unrefreshing. The unrefreshing quality of napping and the sleep drunkenness associated with awakenings lead patients to fight sleepiness as long as they can.

Patients who do not nap are particularly prone to episodes of drowsiness and automatic behavior, which are usually signaled by blank stares. During these episodes, patients act in

an unplanned and often inappropriate way. Patients have reported finding themselves miles from their homes while driving or performing inappropriate actions such as sprinkling salt on coffee, putting dirty plates in a clothes dryer and turning on the machine, writing incoherent sentences during classes, having loud and irrelevant bursts of speech, and so on. Amnesia of such occurrences is common, although patients are usually aware that they have had one of their "drowsy" episodes when they are later confronted with the results of their automatic behavior.

Some IH patients may present with symptoms that overlap with narcolepsy. A few patients with IH may report occasional episodes of irresistible sleep as well as short and refreshing naps.[10] Conversely, patients with narcolepsy with cataplexy may have nonimperative EDS, prolonged naps, and prolonged nocturnal sleep, and both conditions can occur in the same family (see earlier).

Nocturnal Sleep

Nocturnal sleep is typically subjectively long and undisturbed, with more than 10 hours of sleep. A few patients may report sleep times of 12 to 19 hours per day on weekends and holidays. In most IH patients, EDS does not improve with prolonged sleep, so extreme sleep times (sleep >12 hours per day) during weekdays is uncommon.[10]

Awakening after nocturnal sleep is typically difficult. Sleep drunkenness, confusional arousal (also called *syndrome d'Elpénor* after the youngest of Ulysses' comrades who killed himself during an episode of incomplete awakening) is reported by 40% to 60% of patients[5,10,12] but can be seen also in other forms of EDS. Patients are hard to awaken; they can be aggressive and verbally and physically abusive during that twilight state if they are awakened, even at their own request. The patient may be confused and unable to react adequately to external stimuli on awakening. The time it takes to "get going" in the morning may be as long as 2 to 3 hours (sleep inertia); when severe, it is sometimes called sleep drunkenness. Sleep inertia is also common when patients are awakened from naps.

Associated Features

Sleep paralysis and hallucinations are common in narcolepsy but less frequent in IH.[10,27,38] Cataplexy-like episodes and nightmares are also possible.

Depressive symptoms were noted in 15% to 25% of patients in Roth's series and confirmed in later studies.[5,10,14,39,40] Mood changes not qualifying for the diagnosis of affective disorder may precede or follow the onset of EDS and evolve independently. The presence of major depression is incompatible with the diagnosis of IH (see later).

Headache is a frequent symptom in patients with EDS. Migraine- and tension-type headaches are reported in about 30% of patients with IH. Pain complaints of other localizations are occasionally observed.

Patients with IH were found to have lower scores in most domains of the quality-of-life inventory, excluding "bodily pain," "social functioning," and "mental health," whereas treatment seems to have a positive influence only on "role limitations due to emotional problems" and "mental health."[41]

Neurovegetative symptoms such as cold hands or feet, lightheadedness on standing up, orthostatic hypotension, or syncope have been observed.[3,10,13,42,43] The frequency of these symptoms is, however, similar in narcolepsy and IH.[13]

An increased body mass index has been noted in a few series.[10,15]

DIAGNOSIS

Definition Criteria

The ICSD3 has redefined the criteria of IH, resulting in a more heterogeneous diagnosis than in the previous classification (ICSD2), which distinguished two forms with and without long sleep time. ICSD3 diagnostic criteria include the following:

1. Daily periods of irresistible need to sleep or daytime lapses into sleep for more than 3 months
2. Absence of cataplexy
3. MSLT showing <2 sleep-onset REM periods (SOREMPs) or no SOREMPs if the REM sleep latency on the preceding polysomnogram is 15 minutes or less
4. The presence of at least one of the following:
   - - MSLT showing a mean sleep latency of 8 minutes or less
   - - Total 24-hour sleep time of 660 minutes or longer (typically 12 to 14 hours) on 24-hour polysomnography (PSG) monitoring (performed after the correction of chronic sleep deprivation) *or* by wrist actigraphy in association with a sleep log (averaged over at least 7 days with unrestricted sleep)
5. Insufficient sleep syndrome ruled out (if deemed necessary, by lack of improvement of sleepiness after an adequate trial of increased nocturnal time in bed, preferably confirmed by at least 1 week of wrist actigraphy)
6. Hypersomnolence or MSLT findings that are not better explained by another sleep disorder, medical or neurologic disorder, mental disorder, medication use, or substance abuse. Drugs known to affect sleep, sleep latency, and daytime alertness must be carefully evaluated and should be withdrawn for a minimum of 2 weeks before objective tests.

Polysomnography

PSG findings of IH include a short sleep latency, a high sleep efficiency (usually >90%), and increased amounts of deep (slow wave) NREM sleep (Figures 91-1 and 91-2).[3,10,12,43] These findings are nonspecific and can be seen also in patients with behaviorally induced insufficient sleep syndrome (BIISS; see later). Amounts of sleep spindles (throughout the sleep period or at the beginning and end of the night) have been occasionally reported to be elevated in IH[8,35] (Figure 91-3).

In IH, sleep-onset REM sleep episodes are rare, and typically the arousal, apnea-hypopnea, and periodic limb movements indexes are less than 5 to 10 per hour. Some authors believe, however, that periodic limb movements in patients with EDS but without restless legs symptoms may not necessarily preclude the diagnosis of IH. In a few patients otherwise satisfying all other criteria for IH, higher indexes have been observed.[10,12,44,45]

Multiple Sleep Latency Test

MSLT typically shows a mean sleep latency score of less than 8 minutes, and a few IH patients have mean sleep latencies of less than 5 minutes.[9,10,12] SOREMPs occur in 3% to 4% of naps, but never more than once.[10,12] Subjective

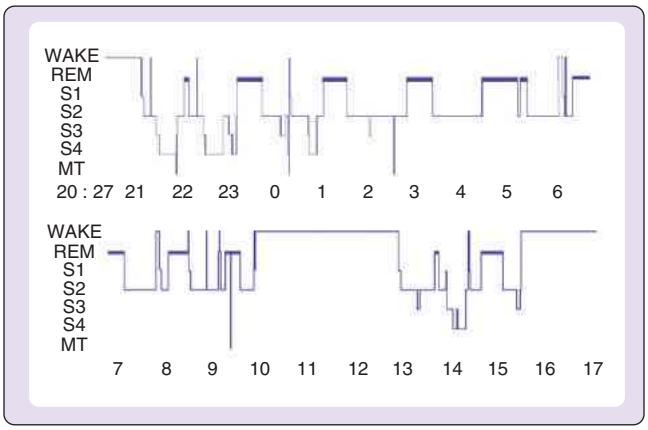


Figure 91-1 A 24-Hour Continuous Hypnogram in Idiopathic Hypersomnia. A 24-year-old woman presented with excessive daytime sleepiness (EDS), prolonged unrefreshing sleep, and sleep drunkenness. Her Epworth Sleepiness Scale score was 18/24. She had frequent hallucinations while falling asleep or on awakening, but no sleep paralysis or cataplexy. Her body mass index was 24, and she had a positive family history for EDS. The polysomnogram was notable for a sleep efficiency of 98%, sleep latency to NREM 2 sleep of 10 minutes, and slow wave sleep 26% of total sleep time. She had no snoring, apneas, or periodic limb movements. Her Multiple Sleep Latency Test showed a mean sleep latency of 4.8 minutes but no sleep-onset REM period. A 24-hour continuous polysomnogram revealed a total nighttime sleep time of 12 hours, 44 minutes and daytime sleep of 2 hours, 36 minutes (a single long nap). Hypocretin-1 in cerebrospinal fluid was normal. HLA-DQB1*0602 positive. Psychiatric assessment was normal. Her EDS improved slightly with modafinil up to 600 mg/day, but her sleep inertia was unchanged.

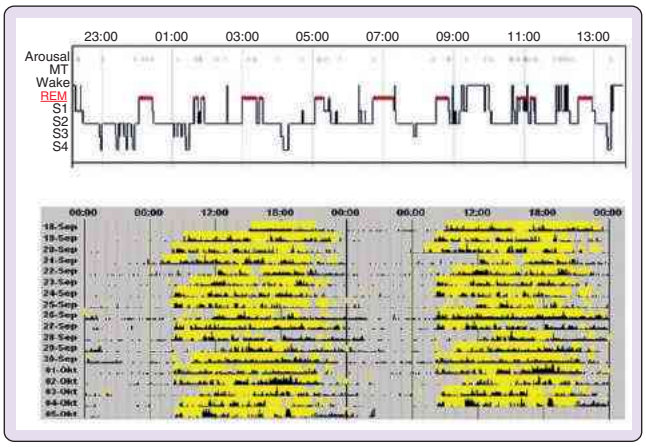


Figure 91-2 Hypnogram and Actigraphy in Idiopathic Hypersomnia. A 32-year-old man presented with excessive daytime sleepiness (EDS), prolonged unrefreshing sleep up to 13 to 18 hours, and sleep drunkenness. His Epworth Sleepiness Scale score was 19/24. He had occasional sleep paralysis and hallucinations. He had migraine headaches, a body mass index of 27, and no family history of EDS. Polysomnography ad libitum over 16 hours showed a sleep efficiency of 92% with a total sleep time of 14.7 hours, latency to NREM 2 sleep of 22 minutes, and slow wave sleep 6% (of sleep period time). He had no snoring, apneas, or periodic limb movements. His Multiple Sleep Latency Test showed a mean sleep latency of 4.3 minutes and no sleep-onset REM period (SOREMP). His Maintenance of Wakefulness Test showed a mean sleep latency of 4.3 minutes and no SOREMP. Two weeks of actigraphy revealed a mean sleep time (rest/sleep) over 48% of the recording time. His cerebrospinal fluid hypocretin-1 level was normal, and HLA-DQB1*0602 was negative. Psychiatric assessment was normal. He did not improve with modafinil, methylphenidate, or melatonin.

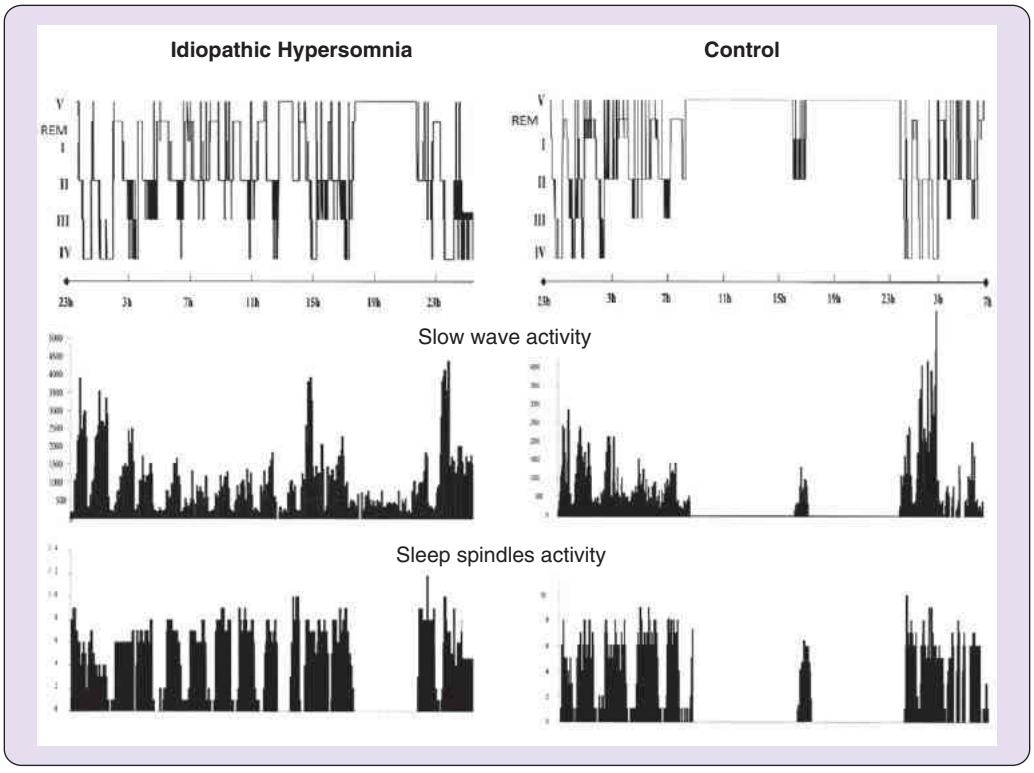


Figure 91-3 Increased Sleep, Slow Waves, and Spindles in Idiopathic Hypersomnia (IH). Prolonged (32 hours) polysomnograms in a 26-year-old woman with IH and long sleep time (*left*) and a 28-year-old control (*right*). The patient with IH has increased total sleep time, persistently high slow wave activity (SWA), and persistent sleep spindles. This patient also has an increase in daytime sleep, including high SWA at about 3 PM.

awareness of sleep during naps is often higher than in patients with narcolepsy.

MSLT is of limited diagnostic value in IH patients with long sleep time.[8,9] The first reason is the usual difficulty keeping the patient awake before the test and between sessions of the test. The second reason is the obligation to wake the patient in the morning to perform the MSLT, thus precluding the recording of prolonged nighttime sleep, which is a typical symptom of IH with long sleep time. Regarding these limits, some patients may have mean sleep latencies longer than 8 or even 10 minutes.[8,9,10,12] In addition, a recent study of 36 patients demonstrated poor test-retest reliability of the MSLT in a clinical population of patients with non-hypocretin-deficient central nervous system hypersomnias.[46] No correlation was found between the mean sleep latencies on the first and second tests, with a change in diagnosis occurring in 42% of patients because of differences in the mean sleep latencies.

Other Polysomnography Protocols

As a result of the limits mentioned earlier of the PSG-MSLT procedure, a prolonged (up to 24 to 32 hours) continuous PSG on an ad libitum sleep-wake protocol was proposed as a diagnostic tool for IH.[8] This protocol allows the documentation of a major sleep episode (>10 hours) and of daytime sleep episodes of more than 1 hour's duration (see Figures 91-2 and 91-3). Spontaneous sleep periods of up to 19 hours were previously reported in IH, with normal MSLT latency.[47] In a series of 11 patients, a total sleep time of 699 ± 130 minutes at night was observed.[48] Several 24-hour recordings have even been suggested to ensure that the patient is saturated with sleep.[43,49]

These protocols, however, have several drawbacks. First, standardization and validation, especially regarding the level of physical and social activity allowed during the recording, is still lacking. (Does the patient remain in bed during the full recording or perform some physical activity, and to what degree? Do these prolonged polysomnograms need to be performed in an ambulatory or only in laboratory setting? Do age and gender modify both night and daytime quantity of sleep obtained in normal controls and in IH patients?). Second, prolonged sleep times are not specific for IH and are also seen in patients with hypersomnia related to neurologic disorders and rarely in depressed patients with EDS.[50-53] Third, a standard MSLT for the diagnosis of narcolepsy without cataplexy would still need to be performed separately. Fourth, a prolonged PSG recording adds considerable cost to the diagnostic workup.

Other Findings

Instead of additional 24-hour PSG, ambulatory actigraphy monitoring over several days can demonstrate the prolonged rest episodes characteristic of IH (see Figure 91-2).[15] In addition, actigraphy is helpful to rule out BIISS (see later) and other circadian disorders that may lead to EDS. However, actigraphy protocols have not been standardized or validated in IH and the differentiation between sleep and rest while awake (clinophilia) may be difficult, especially in the context of mild depression.

Brain magnetic resonance imaging, HLA typing, and assessment of CSF hypocretin-1 levels are not useful in the diagnosis of IH (see earlier) but may be considered to rule out other causes of EDS. Testing for abnormal $GABA_A$ receptor functioning in patients with IH should be considered experimental until it is further validated (see earlier).

DIFFERENTIAL DIAGNOSIS

Narcolepsy with and without Cataplexy

Narcolepsy is a common differential diagnostic consideration, but it is not the most difficult, as demonstrated also by large comparative studies.[8,10,12,13] The presence of EDS with clear-cut (definite) cataplexy, two or more SOREMPs on the MSLT, and low or undetectable CSF hypocretin-1 levels are diagnostic for type 1 narcolepsy (ICSD3). Conversely, long sleep times, sleep drunkenness, long unrefreshing naps, high sleep efficiency, high amounts of SWS on PSG, and absence of (or one at maximum) SOREMPs during MSLT suggests the diagnosis of IH. As pointed out before, some patients fulfilling the current international criteria for IH report a variable EDS associated with irresistible sleep episodes, as well as short and refreshing naps. On the other hand, a few patients with narcolepsy with cataplexy present with nonimperative EDS, prolonged naps, and nocturnal sleep.[10,38,54] Patients with narcolepsy may initially present as EDS without or with mild or rare cataplexy, and the positive diagnosis may be in doubt for months or even years. Finally, an overlap may exist between IH and narcolepsy without cataplexy because similar clinical symptoms may be shared in both conditions, and changes in the number of SOREMPs over two consecutive MSLTs occur in patients with central hypersomnia that may modify the diagnosis.

Unfortunately, CSF hypocretin-1 level has been found to be normal in most patients with either narcolepsy without cataplexy (80% of patients) or IH (100% of patients).[15,25,26]

Sleep-Disordered Breathing Syndromes

IH should be distinguished from sleep-disordered breathing syndromes, including upper airway resistance syndrome (UARS).[55] Patients with UARS complain of isolated EDS and snoring. Examinations reveal a triangular face or a steep mandibular plane, a highly arched palate, a class II malocclusion, and, at times, retroposition of the mandible, but patients need not be obese. Cephalometric radiographs have indicated the presence of a small space behind the base of the tongue (posterior airway space), often near the location of the hyoid bone. These subjects presented repetitive short ("transient") alpha electroencephalographic arousals lasting 3 to 14 seconds that regularly interrupted the abnormally high inspiratory efforts.[55] Standard PSG recordings of these subjects evoked the diagnosis of UARS from the presence of these repetitive transient arousals, increases in snoring just before the arousal, and an increase in inspiratory time and a decrease in expiratory time, which were determined with the use of well-calibrated sensors. No significant change in Sao_2 was seen, and the respiratory disturbance index was low (<5).

A nasal pressure cannula and, in rare, doubtful cases, esophageal pressure monitoring must be included in all sleep studies to confirm the diagnosis of UARS. In patients with isolated snoring and IH, a continuous positive airway pressure trial may be warranted. Lack of improvement would support the latter diagnosis.[10,45]

Behaviorally Induced Insufficient Sleep Syndrome, Chronic Sleep Insufficiency, and Long Sleepers

A careful history is needed to differentiate patients with IH from those with BIISS (chronic insufficient nocturnal sleep) who present with EDS.[56,57] The patient should establish a regular sleep-wake schedule and complete sleep diaries (or actigraphy) for at least 2 weeks before PSG. Typically, BIISS patients sleep 2 to 3 hours longer on weekends than weekdays. PSG and MSLT findings in BIISS may be similar to those in IH (see earlier).

Actigraphy is often necessary to rule out BIIS because the history may be misleading or inconclusive. Particularly difficult is recognizing relative sleep insufficiency in long sleepers. It is possible that besides long sleepers, other individuals may also be more prone to develop hypersomnia in association with sleep insufficiency (Figure 91-4).

Some authors have suggested that IH may represent an extreme phenotype of long sleepers.[8,33] However, patients with insufficient sleep, but rarely those with IH, exhibit a subjective and objective improvement of EDS with prolongation of sleep times or when allowed to sleep ad libitum.

Hypersomnia Associated with Psychiatric Disorders

Hypersomnia associated with psychiatric disorders (atypical depression, bipolar depression, dysthymia or neurotic depression, neurotic hypersomnia) can be difficult to differentiate from IH. Both conditions can include nonimperative EDS, long unrefreshing naps, long sleep times, sleep inertia, and depressed mood. The term *atypical* or *vegetative depression* has been used for the association of major depression with these symptoms. Particularly difficult is the differential diagnosis between IH and mild depression or dysthymia. PSG findings may be very similar, although patients with hypersomnia associated with psychiatric disorders generally have higher amounts of NREM stage 1, less SWS, and lower sleep efficiency.[58] In patients with hypersomnia associated with psychiatric disorders, the MSLT typically shows normal mean sleep latencies. In addition, these patients may spend a large amount of time in bed and acknowledge resting more than sleeping (clinophilia). Patients with hypersomnia associated with psychiatric disorders may exhibit mean sleep times of up to 50% to 60% of the entire actigraphy recording times.[15] Finally, worse EDS in winter months, obesity, and improved EDS with antidepressants are other typical features of atypical depression.

Hypersomnia associated with psychiatric disorders may also, however, be accompanied by abnormal MSLT findings (in 36% of cases in one series[59]), and conversely, patients with IH may exhibit normal MSLT findings.[11,15,60] In unclear cases, formal psychiatric assessment is needed. Treatment with activating antidepressants (selective serotonin reuptake inhibitors [SSRIs], monoamine oxidase inhibitors [MAOIs], norepinephrine reuptake inhibitors) rather than stimulants may be considered in patients in whom a psychiatric disorder is likely to cause the EDS.

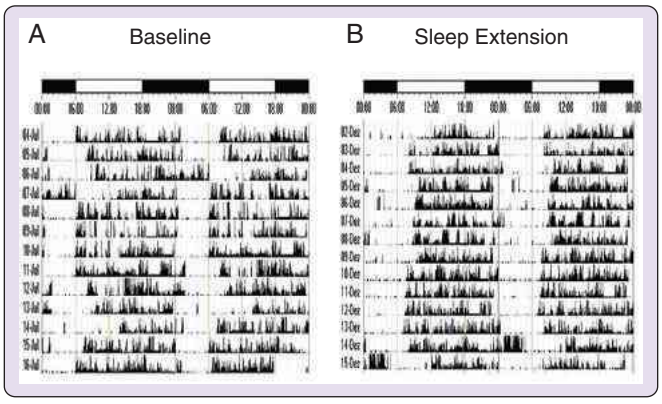


Figure 91-4 Actigraphy in Hypersomnia Associated with Chronic Sleep Insufficiency. A 22-year old man presented with excessive daytime sleepiness (EDS) and sleep drunkenness. His initial Epworth Sleepiness Scale (ESS) score was 20/24. He had occasional sleep paralysis on awakening. He had no cataplexy or hallucinations. His body mass index was 24, and there was no family history of EDS. Initially, he denied sleep deficiency. His Multiple Sleep Latency Test showed a mean sleep latency of 6 minutes and 3 sleep-onset REM periods. **A,** Two weeks of actigraphy (during working days) showed irregular sleep-wake rhythms with a mean sleep time (rest/sleep) more than 35% of the recording time. Hypocretin-1 in cerebrospinal fluid was normal. HLA-DQB1*0602 was positive. Psychiatric assessment was normal. **B,** After sleep extension of more than 1 hour/day, with a mean sleep time (rest/sleep) more than 41% of the recording time, he had complete resolution of his subjective sleepiness and normalization of the ESS score (4/24).

Chronic Fatigue Syndrome

Chronic fatigue syndrome (CFS) is characterized by persistent or relapsing fatigue that does not resolve with sleep or rest. One of the clinical difficulties is that CFS is a poorly defined diagnosis, and patients and clinicians may have difficulty differentiating fatigue from a desire for sleep, EDS, and need for sleep. The clinical presentation of CFS is similar to that seen in patients with hypersomnia associated with psychiatric disorders. In addition to fatigue, patients complain of cognitive difficulties, poor mood, anxiety, fever, and myalgias. PSG may show decreased sleep efficiency and recurrent alpha intrusions unlike IH. The MSLT is typically normal. A few patients with CFS may have a specific sleep diagnosis (e.g., sleep apnea, restless legs syndrome [RLS], periodic limb movement disorder [PLMD]).[61]

Restless Legs Syndrome and Sleep-Related Movement Disorders

Sleep-related movement disorders include several conditions such as RLS and PLMD. Patients or bed partners may complain of movements before or during sleep. Nocturnal sleep disturbances and complaints of fatigue and EDS are common.[62]

A clinical interview should address symptoms such as nocturnal restlessness and discomfort of the extremities during periods of inactivity that are relieved with movement. PSG shows periodic and highly stereotyped limb movements during sleep, exceeding more than 10 to 15 per hour in adults. PLMD must be interpreted in the context of a patient's related complaint, with an important overlap between symptomatic and asymptomatic patients according to the leg movement index.

Circadian Disorders

Rarely, EDS during the morning hours in patients with a delayed sleep phase syndrome and EDS during the afternoon hours in patients with advanced sleep phase syndrome may lead to questions regarding the diagnosis, but an appropriate sleep history, sleep logs covering 15 days, and, if needed,

actigraphy indicate a normal total sleep time but abnormal sleep-onset and wake-up times.

Hypersomnia and Neurologic and Medical Disorders

A medical condition may produce hypersomnia and mimic IH with EDS, automatic behaviors, prolonged sleep episodes, and sleep drunkenness. EDS is usually associated with other manifestations of the underlying condition. Rarely, hypersomnia, EDS, or fatigue may be the only or main symptom.

Several neurologic disorders may cause hypersomnia and EDS, with large variation in severity and clinical presentation.[63-65] Brain tumors, encephalitis, stroke, and other lesions in thalamus, hypothalamus, or brainstem can cause hypersomnia that may mimic clinical symptoms of IH but typically also including alteration in sleep continuity.[66] Neurodegenerative conditions such as Alzheimer disease, Parkinson disease, or multiple-system atrophy are also associated with EDS and hypersomnia.[67] Although intrinsic hypersomnia exists in those neurologic disorders, other possible causes of EDS, such as sleep-disordered breathing, drugs, and periodic limb movements, need to be excluded.

Several genetic disorders (Norrie disease, Niemann-Pick type C, Prader-Willi syndrome, myotonic dystrophy) may be associated with central hypersomnia but also with sleep-related breathing disorders or periodic limb movements that may lead to EDS.

Hypersomnia and EDS are occasionally observed in diabetes, metabolic encephalopathy (e.g., hepatic, uremic, hypercarbic), hypothyroidism, and acromegaly, and these disorders can also cause sleep-related breathing and PLMD. Hypophyseal insufficiency and obesity without sleep-disordered breathing can be associated with apathy, EDS, or hypersomnia.[68,69]

Posttraumatic hypersomnia is another etiology of neurologic hypersomnia and is discussed in detail in Chapter 99. Abnormal sleepiness may be observed within 6 to 18 months after head trauma.[70] Clinical symptoms may be similar to IH, but with poor sleep efficiency and frequent association with headaches, memory loss, and lack of concentration. Posttraumatic complaints of EDS have been associated with variable degrees of impaired daytime functioning.

After an acute viral infection (e.g., mononucleosis, pneumonia), patients may develop a (postviral) syndrome of EDS with features similar to those of chronic fatigue and IH.[71] In some of these patients, an encephalitic process or elevated levels of inflammatory cytokines may play a role.[72,73]

Human African trypanosomiasis, which is due to the transmission of trypanosomes by tsetse flies, is a frequent cause of severe hypersomnia in western Africa (*Trypanosoma brucei ganbiense*) and eastern Africa (*Trypanosoma brucei rhodesiense*). After an extensive immune reaction during the initial stage, severe sleep and wakefulness impairment follows, and the disorder at this point is referred to as "sleeping sickness." The possibility of human African trypanosomiasis needs to be assessed in travelers and individuals migrating from Africa with EDS.

Periodic Hypersomnias

Periodic and recurrent hypersomnias, including the Kleine-Levin syndrome and the recurrent hypersomnia associated with the menstrual cycle, occasionally alternating with episodes of hypersomnia, are usually easy to differentiate from the chronic sleepiness of IH by history alone.[74-78]

Drugs and Substance Use and Abuse

Many medications can cause fatigue, EDS, and hypersomnia, including beta blockers, other antihypertensive agents, dopaminergic agents, antidepressants, and opioids.[79] In doubtful cases, it may be helpful to screen for drugs in urine at the time of MSLT.

TREATMENT

The underlying causes of IH are unknown, and therefore treatment is symptomatic. Prolongation of sleep times has been suggested by Roth but has generally proved to be of no help.[10] Behavioral approaches and sleep hygiene are always recommended to prevent insufficient sleep. Restriction of time in bed and short duration of planned naps may be advantageous to decrease sleep inertia on awakening but may have little positive impact alone.

The pharmacologic options for IH are similar to the treatments of EDS in narcolepsy. Treatment response, however, is often less frequent and robust than in narcolepsy and especially in patients with prolonged nighttime sleep.[10] The list of medications that have been prescribed for IH patients is extensive and includes mainly the stimulant drugs (i.e., modafinil, methylphenidate, mazindol, and dextroamphetamine) but also tricyclic antidepressants, MAOIs, SSRIs, clonidine, levodopa (isolated or in combination), bromocriptine, selegiline, and amantadine. Overall, only about 50% to 70% of patients report significant improvement, most commonly with modafinil or amphetamines.[10,12]

Modafinil is a first-line treatment for IH. The dose of is typically started at 100 mg and gradually increased to effectiveness to a maximum of 600 mg. The most common side effect is headache, and this negative effect may be reduced if the dose is increased gradually. A positive effect has also been observed in children.[80] In some series, a good treatment response was found in a more than half of IH patients.[10,12,81,82] Modafinil seems to have good benefit-to-risk ratio in IH, similar to its effect in narcolepsy.[82] Differences between studies are probably due to the more or less strict criteria used by different authors to exclude other causes of hypersomnia (some of which respond to stimulants). A recent randomized, placebo-controlled study including 33 patients with IH reported efficacy of modafinil compared with placebo on ESS but not on the Maintenance of Wakefulness Test.[83] Another randomized, crossover, double-blind placebo-controlled trial showed improvement on driving performance with modafinil in patients with IH in the same way as in narcolepsy.[84]

Other medications that increase monoamine signaling may be effective in IH. Recently, two retrospective studies reported a favorable benefit-to-risk ratio of mazindol (a tricyclic, anorectic, nonamphetamine stimulant) and pitolisant (a wake-enhancing drug that increases the histamine release in the brain by blocking presynaptic histamine-3 reuptake)[85] in drug-resistant patients with hypersomnia including IH.[86,87] However, postawakening confusion (sleep drunkenness) often persists in IH despite medication.[8] Treatment with activating antidepressants (SSRIs and MAOIs) rather than stimulants may be first considered in patients in whom a psychiatric or

depressive disorder cannot be ruled out with certainty. Fantini and Montplaisir reported improvement in half of their 10 patients, in whom melatonin (2 mg of slow-release form at bedtime) was attempted.[45] One of the authors could not confirm this observation (CB, unpublished observation). Sodium oxybate may be also of interest in certain patients, but it has not been studied systematically.

Based on recent evidence for abnormal $GABA_A$ receptor activity in IH, single-blind interventions using oral flumazenil were proposed in a few patients with hypersomnia, with improvement of vigilance as assessed on psychomotor vigilance and subjective alertness tests.[31] Because flumazenil is formulated and approved only for intravenous use, clarithromycin, an oral agent with potential antagonism of the $GABA_A$, receptor has been used in a single sleep center. Clinical experience with clarithromycin has been reported in 53 patients with hypersomnia refractory to conventional psychostimulants, whose spinal fluid potentiated $GABA_A$ receptor function in vitro.[88] Two thirds of the patients reported subjective improvement in sleepiness during short-term treatment and 38% with long-term treatment, but potential side effects include gastrointestinal problems, antibiotic resistance, and infection. This clinical intervention was single-blinded, and further studies are needed to confirm these preliminary data.

To date, no drug had been approved by the U.S. Food and Drug Administration or other agencies specifically for the treatment of IH. These recommendations for treating the EDS of IH with stimulants is based on expert opinion only.

CLINICAL COURSE AND PREVENTION

The overall psychosocial burden of IH is similar to that of narcolepsy.[89-91] At times, the severity of the impairment can place lives in jeopardy; for example, patients with IH have sustained third-degree burns during automatic behavior episodes or have turned on gas furnaces or stoves without lighting them, leading in one case to a severe explosion.[91] In IH, symptoms may be stable and long lasting, but spontaneous improvement in EDS may occur in up to one fourth of patients.[10,12,13,46] No prevention is possible in this disorder.

PITFALLS

Because IH is rare and essentially a diagnosis of exclusion, the main pitfall is not making an accurate diagnosis. The terms IH and hypersomnia of unknown origin are not synonymous. Different research groups have historically used different diagnostic criteria, making comparisons across studies difficult. This is particularly true of case series that did not exclude mild forms of sleep-disordered breathing, BIISS, and hypersomnia associated with psychiatric disorders.

Careful history, sleep questionnaire or actigraphy, full physical examination, overnight PSG, and MSLT are essential for diagnosis and, more important, to rule out other causes of EDS. Demonstration of increased sleep times using prolonged PSG or actigraphy, formal psychiatric testing, CSF hypocretin-1 measurements, and brain magnetic resonance imaging may be necessary in unclear cases to confirm the diagnosis and rule out other causes of EDS. Treatment with stimulants is not always satisfactory.

The main controversies relate to (1) the clinical and neurophysiologic overlap between IH and hypersomnia associated with psychiatric disorders, mild sleep apnea, narcolepsy without cataplexy, and BIISS; (2) the potential for spontaneous improvement or change in diagnostic category; and (3) the currently unknown pathophysiology of IH. Further studies are required to understand the pathophysiology of IH, to determine whether there are different clinical subtypes of IH (forms with and without long sleep time), and to validate the specificity and sensitivity of biomarkers involved with diagnostic and therapeutic significance. Finally, prospective studies are needed to obtain objective evidence for the efficacy of medications in treating IH and to clarify whether mood changes in IH are consequent to difficulty adapting to the disease or indicate a primary brain dysfunction.[92]

CLINICAL PEARLS

- Clinicians should consider IH in the differential diagnosis of young patients with EDS (mean age at onset, 15 to 25 years), prolonged unrefreshing sleep, and difficulty waking up in the morning or after a nap.
- Before diagnosing IH, clinicians should rule out insufficient sleep syndrome, atypical forms of depression, sleep apnea syndrome, and narcolepsy.
- The symptoms of IH are often severe and long lasting, and treatment requires stimulant drugs as in narcolepsy.
- IH can remit in about 25% of patients, and clinicians should periodically reevaluate the diagnosis and treatment plan.

SUMMARY

IH is a rare disorder characterized clinically by EDS with long and unrefreshing naps, prolonged and undisturbed nocturnal sleep, and great difficulty waking up and "getting going" (sleep inertia) after sleep. Physiologically, the key features of IH are high sleep efficiency and SWS percentages on PSG, decreased mean sleep latencies, one or no SOREMs on MSLT, and long sleep times on prolonged PSG or actigraphy. The pathophysiology of IH is unknown but may include genetic factors and deficient monoaminergic and GABAergic signaling. In the absence of a specific biologic marker, IH is a diagnosis of exclusion with a broad differential diagnosis, including narcolepsy without cataplexy, atypical forms of depression, and BIISS. The response to stimulants is variable, and remission is possible.

Selected Readings

Ali M, Auger RR, Slocumb NL, et al. Idiopathic hypersomnia: clinical features and response to treatment. *J Clin Sleep Med* 2009;**15**:562–8.

Anderson KN, Pilsworth S, Sharples LD, et al. Idiopathic hypersomnia: a study of 77 cases. *Sleep* 2007;**30**:1274–81.

Billiard M, Dauvilliers Y. Idiopathic hypersomnia. *Sleep Med Rev* 2001;**5**: 351–60.

Dauvilliers Y, Dellallée N, Jausssent I, et al. Normal cerebrospinal fluid histamine and tele-methylhistamine levels in hypersomnia conditions. *Sleep* 2012;**35**:1359–66.

Dauvilliers Y, Lopez R, Ohayon M, Bayard S. Hypersomnia and depressive symptoms: methodological and clinical aspects. *BMC Med* 2013;**11**:78.

Lavault S, Dauvilliers Y, Drouot X, et al. Benefit and risk of modafinil in idiopathic hypersomnia vs. narcolepsy with cataplexy. *Sleep Med* 2011; **12**:550–6.

Mayer G, Benes H, Young P, et al. Modafinil in the treatment of idiopathic hypersomnia without long sleep time-a randomized, double-blind, placebo-controlled study. *J Sleep Res* 2015;**24**(1):74–81.

Pizza F, Jaussent I, Lopez R, et al. Car crashes and central disorders of hypersomnolence: a French study. *PLoS One* 2015;**10**:e0129386.
Rye DB, Bliwise DL, Parker K. Modulation of vigilance in the primary hypersomnias by endogenous enhancement of GABAA receptors. *Sci Transl Med* 2012;**4**:161ra151.
Šonka K, Šusta M, Billiard M. Narcolepsy with and without cataplexy, idiopathic hypersomnia with and without long sleep time: a cluster analysis. *Sleep Med* 2015;**16**:225–31.
Trotti LM, Saini P, Bliwise DL, et al. Clarithromycin in γ-aminobutyric acid-Related hypersomnolence: a randomized, crossover trial. *Ann Neurol* 2015;**78**:454–65.
Trotti LM, Staab BA, Rye DB. Test-retest reliability of the multiple sleep latency test in narcolepsy without cataplexy and idiopathic hypersomnia. *J Clin Sleep Med* 2013;**9**:789–95.
Vernet C, Arnulf I. Idiopathic hypersomnia with and without long sleep time: a controlled series of 75 patients. *Sleep* 2009;**32**:753–9.
Yaman M, Karakaya F, Aydin T, et al. Evaluation of the effect of Modafinil on cognitive functions in patients with idiopathic hypersomnia with P300. *Med Sci Monit* 2015;**21**:1850–5.

A complete reference list can be found online at ExpertConsult.com.

Chapter
92

Parkinsonism

Claudia Trenkwalder; Isabelle Arnulf; Ronald Postuma

Chapter Highlights

- Sleep problems are extremely common in Parkinson disease (PD) and are a major cause of disability. They are primarily caused by neurodegeneration in sleep regulatory structures and motor impairment.
- Although sleep apnea is not clearly increased in PD, treatment of severe apnea may be useful for severe somnolence in some patients. Inspiratory nighttime stridor can be associated with sudden respiratory death and mandates continuous positive airway pressure treatment in most cases.
- Insomnia is generally characterized by sleep maintenance problems and can be treated with optimization of nighttime motor treatment, eszopiclone, doxepin, cognitive behavioral therapy, and possibly atypical neuroleptics.
- Somnolence is more common in advanced PD and is usually caused by degeneration of arousal systems and medications for PD. Treatment options include medication reduction, modafinil, caffeine, or methylphenidate.
- REM sleep behavior disorder occurs in up to 50% of PD patients and notably often starts years before motor manifestations. Primary treatments include bed safety measures, clonazepam, and melatonin.

Parkinsonism is a frequent and disabling condition that affects 2% to 3% of adults older than 65 years. During the past decades, there have been major advances in optimizing the treatment of motor symptoms, at least in idiopathic Parkinson disease (PD), and treatment-resistant symptoms (e.g., falls, dementia) are now major determinants of prognosis. The contribution of sleep disorders to this already complex picture has become increasingly recognized within the movement disorders community. In the past, abnormal sleep in parkinsonian patients was mostly considered "collateral damage" until key observations were made during the past decade. This increasing interest in nonmotor symptoms of PD has led to the first treatment trials for sleep and other nonmotor symptoms of PD.

DEFINITIONS

Parkinsonism

Parkinsonism describes a syndrome characterized by rigidity, a 4- to 6-Hz "pill-rolling" tremor, and bradykinesia, of which PD is the most common cause. Motor PD is characterized by asymmetrical parkinsonism; a slow, progressive worsening; and a good, sustained response to levodopa.[1,2] In recent years, symptoms at onset and disease progression have been fully characterized and correlated to specific areas of neurodegeneration, with a final common clinical spectrum of motor and cognitive sympoms.[3] PD is caused by diffuse synuclein-mediated degeneration of neurons, often first involving lower brainstem areas (e.g., nucleus of the vagus nerve) before spreading to the substantia nigra and eventually the cortex.[4] PD is primarily a disease of aging. Its prevalence increases with age from about 0.9% among persons aged 65 to 69 years to 5% among persons older than 80 years, with a 1.5-fold male preponderance. Only a small percentage develop parkinsonism before age 45 years. Young-onset patients are more likely to have genetic causes. The most important genetic causes include mutations in *Parkin*, PINK-1, LRRK2, glucocerebrosidase, and the α-synuclein gene.[5]

Since the introduction of L-dopa and dopamine agonists, relatively few major therapeutic advances have been accomplished in the treatment of motor PD, except functional neurosurgery. Often, after several years of "honeymoon" on dopaminergic treatment, patients develop motor complications, including dyskinesias (abnormal involuntary movements) at the peak of effect, and generalized slowing and painful dystonia (often confused with cramps) when the effect of L-dopa wears off. Patients then use dopaminergic treatment on daily schedules that become stricter and more frequent as the disease progresses. It is not uncommon for patients with a 10-year duration of PD to take 20 tablets per day, including frequent doses of levodopa, dopamine agonists, monoamine oxidase inhibitors, and various medications to avoid the side effects (nausea, orthostatic hypotension, hallucinations) of motor treatments or to treat nonmotor symptoms.

Recently, there has been increasing recognition that PD is not simply a motor disease. Pathologic staging systems of PD have documented that a variety of structures, including those involved in control of autonomic regulation, olfaction, mood, and sleep, degenerate early in PD, whereas cognition is impaired mostly later in the disease.[6] In fact, prodromal

symptoms of PD can include diverse nonmotor symptoms such as constipation, urinary dysfunction, orthostatic hypotension, depression, anxiety, sleep disorders such as REM sleep behavior disorder (RBD), olfactory loss, and cognitive and behavioral changes. These symptoms are generally not levodopa responsive and may occur years before the first motor symptoms.[7] After PD has commenced, further levodopa-resistant symptoms, such as falls, freezing, dysarthria, drooling, and swallowing dysfunction, often occur.

Especially prominent in PD pathology is widespread degeneration of brainstem nuclei that control sleep and wakefulness. Sleep disturbances include fragmented sleep with reduced sleep maintenance insomnia, abnormal movements during sleep (periodic limb movements during sleep [PLMS], RBD, and rarely fragmented myoclonus), excessive daytime sleepiness, and even sleep apnea. Many of these difficulties worsen with increasing duration of disease and are exacerbated by medications used in PD. A detailed interview, sometimes followed by a video polysomnogram, is an essential tool to understand and adequately treat these patients.

Atypical Parkinsonism

It is important to differentiate PD from *atypical parkinsonism* because treatment and prognosis are different. Indeed, more than 10% of patients have atypical parkinsonism characterized by different neuropathologies, primarily other synucleinopathies such as multiple system atrophy (MSA)[8] or tauopathies. The tauopathies comprise progressive supranuclear palsy (either Parkinson subtype or the Steele-Richardson-Olszewski type with parkinsonism, saccade slowing, vertical gaze impairment, swallowing problems, dysarthria, frontal cognitive impairment, and early falls)[9] and the corticobasal syndrome.[10] All these diseases show reduced or absent dopaminergic response; high doses usually improve the condition only partially. In general, the prognosis of atypical parkinsonism is worse. Compared with PD, sleep problems can be even more frequent and severe in atypical parkinsonism, especially in MSA. A specific sleep-associated respiratory symptom, stridor, occurs in some patients with MSA and must be looked for because it can be associated with severe nocturnal oxygen desaturation and sudden death. Stridor has also been recently described in a rare antibody-mediated tauopathy.[11] Vascular parkinsonism results from multiple infarcts in the basal ganglia and occurs either as a separate disease or as a comorbidity of PD, especially in patients with hypertension and hypercholesterolemia. It is characterized by prominent gait dysfunction and symmetrical akinetic-rigid symptoms without most nonmotor features. Further rare diseases are Guadeloupean parkinsonism,[12] Parkinson-dementia complex of Guam, and Parkinson-hypoventilation syndrome.[13]

Finally, dementia with Lewy bodies (DLB) is the second most frequent cause of dementia after Alzheimer disease. These patients develop cognitive impairment mostly affecting visual-spatial performance, early hallucinations, and parkinsonism (frequently levodopa sensitive), with fluctuations of alertness and vigilance over days or weeks.[14] Like PD, DLB is a synucleinopathy that overlaps profoundly with PD; many consider these to be on the same disease spectrum.[3,15] A recent pathologic study suggested that RBD was rare in Alzheimer disease and other non–Lewy body dementias and that the existence of RBD in a dementia syndrome is a strong diagnostic marker for DLB.[16,17] Mild parkinsonian symptoms can also be observed in patients with Huntington disease and those with spinocerebellar ataxia. All these neurodegenerative disorders are classified according to the mechanism of neuronal loss (Table 92-1), whether it is associated with a deposit of α-synuclein in Lewy bodies, of *tau* protein, or secondary to a genetic polyglutamine disease. Although the major emphasis in this chapter is on PD, much of the discussion also applies to the other degenerative disorders associated with parkinsonism.[18]

Table 92-1 Prevalence of REM Sleep Behavior Disorder in Neurodegenerative Diseases

Disease	Prevalence (%)
Synucleinopathies	
Parkinson disease[32,36,132,133]	15–60
Multiple system atrophy[34,134]	88–90
Dementia with Lewy bodies[16,17]	76–86
Tauopathies	
Progressive supranuclear palsy[55,56]	10–11
Alzheimer disease[135]	4.5–7
Corticobasal degeneration[17]	Case reports
Frontotemporal dementia[17]	0
Pallidopontonigral degeneration[136]	0
Guadeloupean parkinsonism[137]	78
Genetic Diseases	
Huntington disease[138]	8
Spinocerebellar ataxia type 3[139]	56
Parkin mutation[140,141]	9–60

CLINICAL FEATURES

In clinical practice, parkinsonian patients are usually referred to sleep centers for three main complaints: insomnia, abnormal movements when asleep, and daytime sleepiness. These symptoms may arise alone or in association.

Insomnia and Causes of Sleep Fragmentation

The sleep history in PD should include all the features the physician would obtain from any patient with a sleep complaint. Additionally, the physician should query disease-specific questions on nocturnal akinesia, daytime fatigue in relation to medication intake, and psychiatric symptoms. A careful description from the bed partner is essential to determine the presence and frequency of movements during sleep (and their timing), arousals and awakenings, and periods of daytime sleepiness. Two scales, the PD Sleep Scale and the SCOPA-sleep can be helpful for research purposes on sleep quality in PD.[19,20] A checklist of useful questions when investigating a sleep complaint in a PD patient is displayed in Table 92-2.

In PD, many patients report two to five long awakenings during the night (twice more than controls), lasting 30% to 40% of the night.[21] Generally, PD patients have more problems with sleep maintenance—especially in the second half of the night—than with sleep onset.[22] In patients with moderate to severe PD, special attention should be paid to anxiety, depression, hallucinations, and sensory-motor symptoms

Table 92-2 Sleep and Night Problems in Parkinson Disease

Problem	Potential Diagnosis	Proposed Management
Frequent Nocturia (≥2 episodes/night)		
Normal volumes	Sleep apnea syndrome	Check for sleep apnea and treat appropriately
Small volumes, poor stream	Prostatism	Refer to urologist
Small volumes, good stream	Parkinsonism-associated nocturia	Intranasal desmopressin, oral amitriptyline, or transdermal rotigotine patch; if detrusor instability: oxybutynin, tolterodine
		Decrease evening fluid intake; empty bladder prior to bed; avoid evening dosing with diuretics, antihypertensives, and vasodilators; have a urinal at the bedside table
Difficulty Initiating Sleep		
Early in the evening	Too early lights-off	Switch off lights later
	Anxiety, or behavioral insomnia	Sleep hygiene; treat anxiety; evening melatonin, eszopiclone, doxepin
With restlessness	Restless legs syndrome	Check for low ferritin; remove antidepressant drugs; if the diagnosis is uncertain, consider polysomnography with legs monitoring; try gabapentin or opiates such as tramadol if not confused
Late in the night	Altered circadian cycle	Sleep hygiene; decrease levodopa/dopamine agonists in the evening; melatonin 1–2 hours before the desired bedtime
Late in night, hypomanic	Assess for impulse control disorder	Decrease dopamine agonists; keep on levodopa monotherapy; close neuropsychological follow-up
Difficulty Resuming Sleep		
With cramps, muscle pain, slowness	Nocturnal bradykinesia	Immediate-release levodopa with a glass of water during awakenings; continuous drug delivery (ropinirole transdermal patch, pramipexole or ropinirole extended-release, apomorphine infusion; intrajejunal levodopa-carbidopa infusion)
		Satin bed sheets to aid movement in bed
With restlessness	Restless legs syndrome	See above
With anxiety	Anxious disorder	Evening antidepressants (mirtazapine, amitriptyline, doxepin, paroxetine)
With low mood	Depressive disorder	Treat the depression
Nightmares, Agitation		
Confused at night when awake	Hallucinations, psychosis, confusion	Remove or reduce the evening dose of dopamine agonist or antidepressant; assess for sleep apnea; antipsychotics (quetiapine, clozapine)
Kicks, shouts, slaps	REM sleep behavior disorders	Secure the bed environment; discontinue antidepressant; assess likelihood of sleep apnea (video-polysomnogram before treating); melatonin 3–9 mg in the evening; clonazepam 0.5–2 mg in the evening
Daytime Sleepiness		
Falls asleep unexpectedly	Sleep attack	Check for possible sedating drugs (e.g., dopamine agonists) and remove or change them; warn patient not to drive
Falls asleep more often than before		Consider the Epworth sleepiness score; ask about associated hallucinations; consider polysomnogram and Multiple Sleep Latency Test
		Treat sleep apnea if severe
		Decrease/stop the dopamine agonist during daytime, and other sedative drugs
		Caffeine; modafinil; methylphenidate

during the night. Sensory discomfort during the night is common in PD. It includes a variety of symptoms such as pain, motor restlessness, and painful dystonia. Restless legs syndrome occurs in 12% to 21% of PD patients (it can be difficult to distinguish from leg pain because the urge to move is common in PD).[23] Early-morning dystonia is common and often severe; it consists of long-lasting, painful contractions of the toes in flexion or extension (sometimes with internal rotation of the ankle) that occur both during the end of the night and on awakening and that make walking difficult. Severe dystonia or rigidity of the leg, neck, and back muscles may occur only at night, when dopamine levels are low. Patients also commonly experience nighttime bradykinesia, with difficulty turning in bed, readjusting the blankets or pillow, sitting up, and walking to the toilet.[24] Difficulty moving during the night can further increase anxiety levels. All these symptoms may curtail sleep by one third if not more, with mostly long intranight awakenings. If nocturnal bradykinesia is present, nighttime dopaminergic therapy should be considered. On the other hand, a deleterious effect on sleep of dopaminergic agonists (which can be considered stimulants) when given at bedtime can be observed.[25] At the extreme of this spectrum, some patients may use supraoptimal doses of levodopa and dopamine agonists day and night, and stay awake and hyperactive at night, using computers, gambling, shopping compulsively on the Internet, or experiencing hypersexuality. This impulse control disorder affects up to 11% patients with PD and is mostly seen with high doses of dopamine agonists.[26]

Abnormal Movements During Sleep and REM Sleep Behavior Disorder

Abnormal movements during sleep are usually either a parasomnia or PLMS. If movements are stereotyped and periodic, then PLMS is likely. Nonstereotyped movements suggest a parasomnia, usually RBD. Other potential sleep movements include myoclonus, rest tremor, dystonia, and bruxism (which are absent in sleep but can emerge during arousals).

RBD consists of dream-enactment behavior such as laughing, talking, crying, kicking, fighting invisible enemies during sleep, and so forth.[27] RBD is covered in detail in Chapter 103. RBD can be violent enough to disrupt sleep and induce self-injury or bed-partner injury, although this is relatively uncommon in PD (Video 92-1). Compared with patients with idiopathic RBD, who were usually selectively referred for violent behaviors, RBD in PD is less frequent and violent.[28] Indeed, quiet behaviors, including drinking soup, singing a song, or giving a lecture, are often mentioned incidentally by the bed-partner.[29] In addition to fully expressed complex behaviors, patients can also have abrupt movements and jerks, such as simple, aborted, proximal or distal movements of the limbs. In general, RBD is the most frequent parasomnia in parkinsonism. Unlike non-REM parasomnia, RBD patients generally do not walk, do not interact with the environment (e.g., they may grab something in their immediate vicinity, but will not reach for objects), and will not talk back to bed partners during an episode.[30] The eyes are usually closed. If woken from an episode, RBD patients usually have a normal level of consciousness, whereas non-REM parasomnia patients appear only half-awake. Some of the patients remember a dream, but not always. These features can usually distinguish a non-REM parasomnia from RBD with reasonable reliability. One should note, however, that PD patients with both sleepwalking and RBD (i.e., parasomnia overlap) have been occasionally reported.[31] Patients with RBD generally have normal sleep patterns except for higher numbers of PLMS than those without. RBD does not generally cause daytime sleepiness or insomnia. Some PD patients are disturbed by nightmares and awakenings because of vocalizations or by violent movements, including falling from bed.

A new area of research pertains to the disappearance of parkinsonism during an RBD episode. In a large series, spouses of PD patients reported that the patients had unusually strong and rapid movements during RBD, as if they were transiently cured of PD.[32] This clinical improvement has been confirmed on sleep and video monitoring, performed after a 12-hour withdrawal of dopaminergic drugs (Video 92-2). Similar findings have been reported in MSA and progressive supranuclear palsy.[33,34] Understanding the cause of this spontaneous improvement could lead to development of novel treatments (e.g., novel surgical targets) and shed light on motor control during sleep.

Whereas RBD is experienced by 30% to 59% of PD patients, there is increasing evidence that RBD marks a subtype of PD. Patients with RBD have less tremor, more falls and freezing, more cardiovascular autonomic dysfunction (especially orthostatic hypotension), and more cognitive dysfunction on detailed neuropsychological testing in an early population.[35] These associations have not always been detected.[36] Of special significance is a link to PD dementia; in prospective studies, the presence of baseline RBD was strongly linked to the risk for dementia (with relative risks ranging from 5 to 50).[35,37]

Arguably, the most important implication of RBD is that it can predict development of PD and other synucleinopathies. In a 5-year follow-up of their original idiopathic RBD cohort, Schenck and colleagues found that 38% eventually developed PD.[38] On continued follow-up, 81% developed a neurodegenerative disease.[39] This finding has now been confirmed in three independent cohort studies. The Barcelona cohort reported a 45% risk for neurodegenerative disease at 5 years follow-up, which rose to 76% at 10 years and 91% at 14 years.[40,41] The Montreal cohort recently reported 60% conversion to defined neurodegenerative disease at 6 years. Slightly lower risks (38% at 9 years) were reported from a Hong Kong cohort.[42] Finally, a clinical history suggestive of dream enactment behavior (i.e., possible RBD) was linked to risk for mild cognitive impairment and dementia in a population-based study.[43] Most cohorts find that patients are at approximately equal risk for both primary parkinsonism (PD and MSA) and DLB. This very high risk has critical implications for neuroprotective therapy against PD because idiopathic RBD patients are ideal candidates for neuroprotective trials and future candidates for definitive neuroprotective therapy when it becomes available.

Daytime Sleepiness

Some patients with PD experience sleep attacks (or sudden onset of sleep) when treated with dopamine agonists.[44] Examples of sleep attacks include patients falling asleep during stimulating life conditions, such as while eating a meal (the head drooping in the plate), walking, attending work, carrying a child in an escalator, and the most dangerous situation, driving a car. Because of the high risk for accidents in sleepy

drivers, the level of daytime sleepiness must be regularly checked in PD patients, especially when dopaminergic treatment is changed. This is particularly true in de novo, young PD patients who often receive high doses of dopamine agonists. Daytime sleepiness in PD can be self-reported but is usually more reliably reported by the caregiver because some patients are unaware of being abnormally sleepy.[45] Napping after lunch is frequent in PD patients and often perceived as beneficial. The Epworth sleepiness score has been well validated in PD,[46] with an abnormal threshold above 10, but it is poorly predictive of sleep attacks.[47] Some authors have added specific questions on the ability to fall asleep when driving, eating, working, or having a home routine activity,[47] which better predicts the risk for driving accidents. Questioning patients with daytime sleepiness about hallucinations, and vice versa, is useful because these two symptoms are often associated and are underreported (because many patients are afraid of being considered mentally ill).

Stridor

Stridor is caused by partial obstruction of the larynx, resulting in a harsh, high-pitched inspiratory noise (Video 92-3). Unlike obstructive sleep apnea, which does not expose patients to immediate risk, stridor is a life-threatening condition. The larynx obstruction usually begins during the night and is observed in 42% of unselected patients with MSA.[48] It can be recognized quite easily by mimicking it to the caregiver, or by an audio recording during the night, but is not detected on usual apnea monitoring devices. Of interest, nighttime stridor is alleviated by the application of nasal continuous positive airway pressure (CPAP), which can avoid tracheotomy and provides a long-term benefit on quality of sleep and median survival time.[48,49] Recently, a new syndrome associating rapid eye movement (REM) and non–rapid eye movement (NREM) parasomnia, severe stridor, hypoventilation, and early death has been described, in association with a brainstem tauopathy and antibodies against a neuronal cell adhesion molecule.[11]

Sleep Benefit

In PD, a specific, positive restoration of fluent mobility may happen on awakening from sleep, before drug intake, and is named *sleep benefit*. Based on questionnaires, the frequency of this phenomenon varies from 10% to 55% of patients.[50,51] The restored mobility lasts a mean of 84 minutes, and patients may be able to skip their first dose of levodopa.[50] When examining 10 patients with sleep benefit on awakening after a polysomnogram, however, the motor benefit was small and unrelated to the sleep structure, levodopa serum levels, or sleep chronotype.[52]

EPIDEMIOLOGY OF SLEEP DISORDERS IN PARKINSONISM

Insomnia

Sleep problems, and especially insomnia, are common in all forms of parkinsonism. A community-based survey determined that 60% of PD patients had sleep problems, significantly more than in patients with diabetes mellitus (45%) or in aged controls (33%).[22] These percentages increase up to 76% of PD patients who claim having "broken sleep" in hospital samples.[24] As many as 53% of patients with MSA complain of sleep fragmentation, compared with 39% of PD patients.[53] The most severe and specific insomnia is observed in patients with progressive supranuclear palsy.[54-56]

REM Sleep Behavior Disorder

REM sleep behavior disorder affects 30% to 90% of patients with synucleinopathies but is rare in other neurodegenerative disorders.[17] The prevalence of RBD in the various diseases with parkinsonism is shown in Table 92-1. Cross-sectional studies generally find that one third of PD patients report at least one RBD episode per week.[32,57] Studies using polysomnograms (which can detect asymptomatic RBD or asymptomatic loss of REM atonia), generally find RBD in 40% to 60%.[32,36] Note that RBD often waxes and wanes throughout the course of disease, so the lifetime prevalence of RBD in PD is likely higher. Most,[28,57] but not all,[32,36] series of RBD patients with parkinsonism contain more men than women, a ratio that is less consistently observed than in idiopathic RBD.[58] Of note, RBD is very common in other synucleinopathies, occurring in approximately 75% of patients with MSA and DLB.

Excessive Daytime Sleepiness

Excessive daytime sleepiness affects on average one third of patients with PD. Case-control studies performed in various countries consistently find higher sleepiness scores and higher percentage of subjects with abnormal somnolence in PD patients (16% to 74%) than in age- and sex-matched controls.[21,59,60] Sleepiness is infrequent at PD onset in the absence of medications[61] but develops with time, with an incidence of 6% per year in a prospective series.[62] Sleepiness may, however, precede the onset of PD because sleepy adults in a large Asian longitudinal study developed PD later in life 3.3 times more often than nonsleepy adults.[63] The percentage of PD patients having experienced sleep attacks or "sudden onset of sleep without a prodrome" varies from 1% to 14%, whereas 1% to 4% of PD patients report having experienced sleep attacks while driving.[47,64] Fortunately, most patients have some warning beforehand.

PATHOGENESIS

The disruption of the normal sleep and wakefulness in patients with parkinsonism may be caused by neurodegenerative damage in brain areas responsible for sleep and arousal regulation; by behavioral, respiratory, and motor system phenomena accompanying the disease; and by deleterious effects of medications that may induce nightmares, nocturnal movements, insomnia, or sleepiness. These three elements contribute in various degrees to the clinical symptoms. There is stronger evidence for neuronal loss as the primary cause of RBD than there is for insomnia.

Insomnia and Sleep Fragmentation

Insomnia is a nonspecific symptom that does not necessarily result from a selective lesion in any sleep system, except in progressive supranuclear palsy, in which brainstem cholinergic neuron loss may disrupt the REM sleep executive systems.[65] General factors, such as aging, anxiety, and depression, may account for some nocturnal sleep disruption in PD. Poor sleep quality correlates with depression and anxiety scores,[66] but motor phenomena and disability at night are perhaps the major causes of trouble maintaining sleep.[67] The frequency of

insomnia increases with advanced motor stages of PD and a need for higher daily dose of dopaminergic therapy, an indicator of dopamine denervation.[68] Slow movements, nocturnal akinesia with difficulties turning in bed and adjusting blankets, pain, cramps, nocturnal and early morning dystonia (a clawlike contraction of the toes), and frequent need to urinate are reported by patients with advanced PD as main causes of their insomnia.[24] The difficulty moving during the night can further increase anxiety, whereas depression, present in one third of patients with PD, also affects sleep. This may lead to early awakenings as in patients with major depression. Older patients with on-off phenomena and those with hallucinations are particularly likely to have severe sleep disruption.[69] In patients with advanced PD, the alleviation of nocturnal bradykinesia and painful dystonia using subthalamic nucleus stimulation at night can severely decrease time awake at night[70] and improve subjective sleep quality.[71]

Restless legs syndrome (RLS), a frequent cause of insomnia in the general population, has a 15% to 20.8% prevalence in PD.[23] Except in patients with a family history of RLS, they seem to reflect a secondary phenomenon. There is no evidence that RLS symptoms early in life predispose to the subsequent development of PD.[23,72] In clinical practice, RLS may be difficult to distinguish from leg pains, which are frequent in patients with PD,[73] and from the urge to move, which is common in patients with motor fluctuations. PD patients with RLS have lower ferritin levels than those without.[23]

There is still controversy about whether the circadian system is affected in PD, especially in patients displaying nighttime insomnia and daytime sleepiness. Some patients may have circadian phase advance because even patients with mild PD (i.e., without any motor problem) consistently report an early-to-bed and early-to-rise pattern. The nocturnal profile of body core temperature is similar in patients with PD and in controls, whereas the nocturnal temperature fall is blunted in patients with MSA.[74] However, recent studies have suggested a blunting of the circadian rhythm of melatonin secretion in PD, more pronounced in patients with somnolence.[75]

In addition, deleterious effects on sleep of medications can be observed because dopaminergic agents are considered as stimulants, and selegiline is metabolized to amphetamines. Therefore, attempts to improve motor state can also paradoxically worsen sleep in patients prone to hallucinations or psychosis.

REM Sleep Behavior Disorder

The exact cause of RBD in parkinsonism is unknown, but a nondopaminergic lesion of the system controlling atonia during REM sleep is highly suspected.[76,77] The substantia nigra does not seem to be implicated in generating RBD. There is no consistent evidence that levodopa or dopamine agonists can trigger RBD or elicit REM sleep without atonia and no strong evidence that RBD can be treated by dopaminergics.[78] Of interest, animal models of RBD have been created by lesioning the locus coeruleus peri-alpha (pons) in the cat, an area adjacent to the noradrenergic locus coeruleus. These animals displayed complex behaviors during REM sleep, including grooming, leaping, being on watch, and hunting invisible prey.[79] Muscle atonia is also reduced in rat models after lesion of an analogous region, the sublaterodorsal nucleus.[80,81]

The equivalent of these nuclei in humans is the subcoeruleus nucleus, a pontine region known to degenerate in both PD and MSA.[40,82] Notably, the MRI-neuromelanin signal in the coeruleus-subcoeruleus complex is decreased in PD patients with RBD (and not in PD patients without RBD and controls), in proportion to the loss of muscle atonia in REM sleep.[83] Cholinergic neurons located in the pedunculopontine tegmental nucleus have also been hypothesized to contribute to muscle atonia during REM sleep, but they are similarly damaged in DLB and MSA patients with and without RBD.[84]

In murine models, phasic and tonic REM sleep muscle activities are controlled by different mechanisms, and the role of glutamate and glycine and their altered pathways to spinal motor activity may be the substrate for the disinhibition of motor phenomena during RBD.[85] This could explain the clinical observation that REM sleep without atonia may occur without dream enactment behaviors in some patients, whereas others have retained chin muscle atonia yet have prominent phasic limb activity.

In addition, RBD can be triggered or exacerbated pharmacologically, especially by the use of antidepressants like serotonin reuptake inhibitors. This may be due to direct serotonergic activation of motor neurons.[77] Because RBD symptoms in PD are associated with relative neocortical, limbic cortical, and thalamic cholinergic denervation but not with differential serotoninergic or nigrostriatal dopaminergic denervation,[86] the presence of RBD symptoms might signal cholinergic system degeneration.

In atypical parkinsonism, the lesions are extensive, complicating clinicopathologic correlations. As shown in Table 92-1, synucleinopathies are more often associated with RBD than the tauopathies and polyglutamine diseases, suggesting that the neurons responsible for muscle atonia in REM sleep are more vulnerable to α-synuclein-related neurodegeneration than to other mechanisms of damage.

The key regions that cause RBD degenerate early in PD. Lewy bodies containing α-synuclein deposits are detected in the brain of 10% to 15% of clinically normal people older than 60 years; when not associated with PD, this is termed *incidental Lewy body disease*. The subcoeruleus-coeruleus complex is frequently affected in incidental Lewy body disease.[87] Moreover, recent staging systems of PD have proposed that α-synucleinopathy initially affects the olfactory areas and medulla oblongata and progresses to more rostral brain areas in a hierarchical sequence.[87] In this model, the subcoeruleus nucleus, which is located in the pons, should be affected earlier than the substantia nigra, located in the mesencephalon. This provides a pathologic basis for the fact that RBD anticipates PD. It should be noted, however, that only one third of patients with PD develop RBD before parkinsonism onset, and RBD prevalence is not 100% even in patients with advanced PD, suggesting that this staging is not perfectly followed (at least clinically).

Excessive Daytime Sleepiness

The mechanisms of sleepiness and sleep attacks in PD may include complex drug-disease interactions. Most arousal systems (Table 92-3) are affected by neuronal loss and Lewy bodies in PD brains, including the noradrenaline neurons in the locus coeruleus,[88] the serotonin neurons in the raphe,[88] the hypocretin neurons (also affected in primary narcolepsy) in

Table 92-3 Neuronal Loss in Arousal Systems in Patients with Parkinson Disease

Nucleus	Main Neurotransmitter	Neuronal Loss in PD Brain, %
Locus coeruleus[88]	Noradrenaline	40–50
Median raphe [88]	Serotonin	20–40
Ventral periaqueductal gray matter[142]	Dopamine	9
Pedunculopontine tegmental nucleus[143]	Acetylcholine	50
Tuberomammillary nucleus[94]	Histamine	Unchanged enzymatic activity
Lateral hypothalamus[89,90]	Hypocretin	23–62
Basal forebrain[88]	Acetylcholine	32–93

the hypothalamus,[89,90] and the cholinergic neurons in the pedunculopontine tegmental nucleus and basal forebrain.[88] Of note, deep brain stimulation of the pedunculopontine nucleus area in patients with PD can cause immediate sleepiness when the nucleus is blocked by high-frequency current and alertness when it is stimulated by low-frequency current.[91,92] In contrast, the wake-active dopamine neurons in the ventral periaqueductal gray matter and the histamine neurons in the hypothalamus are intact in PD brains.[93,94] In MSA, there is also a marked decrease of hypocretin neurons.[95] As many as 28% of patients with MSA complain of daytime sleepiness,[96] whereas 36% of MSA patients show a narcolepsy-like phenotype.[60]

In PD, sleepiness is more frequent in patients with advanced disease.[59] Sleep deprivation due to insomnia is often considered the cause of somnolence, but in PD, daytime sleepiness is actually associated with longer sleep time at night, suggesting a central hypersomnia.[97-99] Similarly, sleep apnea (observed in 20% to 30% of PD patients), PLMS, and sleep fragmentation do not correlate with daytime sleepiness,[97-99] suggesting they do not contribute much to the mechanisms of sleepiness (although this assertion could be wrong in individual cases). As mentioned previously, dopamine medications increase the risk for sleep attacks.[64,100] In contrast, levodopa is less often sedative, although some patients may sleep half an hour after levodopa intake. Why drugs that are supposed to stimulate the alerting brain system (and do so when given at bedtime) can also be sedating is yet unknown. This effect cannot be explained, at these high doses, by the biphasic effect (presynaptic sedative effect at low dose, postsynaptic alerting effect at high dose) of dopamine agonist that has been described in animals.[101] Rather, a different selectivity for D1 or D2 receptors may be important; D1 agonists and small doses of dopamine increase the firing of hypocretin neurons in rat hypothalamus, whereas high concentrations of dopamine and D2 agonists decrease or even block this firing.[102] If one applies this concept to PD, patients with a partial hypocretin deficiency would be sedated by D2 and D3 agonists or high doses of levodopa.

POLYSOMNOGRAPHY OF PARKINSON DISEASE

As with other diseases, the clinical interview is key for diagnosing the causes of sleep complaints in patients with parkinsonism (see Table 92-2). However, in parkinsonism, the information brought by videopolysomnography is important and useful, provided that all aspects of the disease (motor aspects, during both sleep and wake) and sleep are carefully analyzed. The recording of a patient with parkinsonism in a sleep unit can be difficult when the team is not used to monitoring disabled, anxious patients, who may get up several times to urinate and who may need to be given drugs every 3 hours, to be helped regularly during the night, to receive massage when they have violent cramps or dystonia, and to be reassured when they are confused or subject to hallucinations.

The minimal montage should include the usual sleep montage (electroencephalogram [EEG], electrooculogram [EOG], chin electromyogram [EMG]), electrocardiogram [ECG], nasal pressure, thorax and abdomen efforts, oxygen saturation, and leg EMG, but also audio monitoring because stridor may be mistaken for snoring. If RBD is suspected, a synchronized infrared video and upper-limb EMG electrode (hand rather than shoulder or arm muscles) increases the sensitivity to detect RBD.[103] The video monitoring also allows recognition of other frequent night-related motor problems in PD, including cramps, dystonia, tremor, and restless legs behavior, and even dyskinesias and their role in prolonging awakenings. Sleep scoring may be particularly difficult and time-consuming in patients with parkinsonism. Video monitoring is also helpful for scoring sleep in these patients because EEG features of sleep may altered, RBD may be confused with wakeful behaviors, and stridor needs to be identified. For example, the number of sleep spindles during NREM is reduced in PD,[104] and EEG alpha activity may occur in all sleep stages in PD, giving the false picture of complete wakefulness throughout the night. By contrast, in patients suffering DLB or progressive supranuclear palsy, the alpha background rhythm during wakefulness may turn to a slow, regular rhythm closer to 5 to 7.5 Hz, with a further shift of all NREM sleep frequencies toward slow wave activity, so that an awake patient with open eyes may have a concomitant EEG suggesting full-blown N3 sleep. In addition, sequences of slow or rapid eye movements may be observed during NREM sleep.[105]

In contrast to the quiescence of sleep in normal persons, increased muscle tone and abnormal simple and complex movements are common and also complicate the scoring of polysomnograms in PD patients. Tremor may produce a 4- to 6-Hz regular artifact at the level of the chin or the legs during wakefulness (Figure 92-1), and it disappears with the onset of N1 sleep.[106] It may persist as a polygraphic finding rather than a clinical movement in N1 with awakenings, arousals, and body movements.[107] Patterns of simple motor activity during sleep include repeated blinking at sleep onset, rapid eye

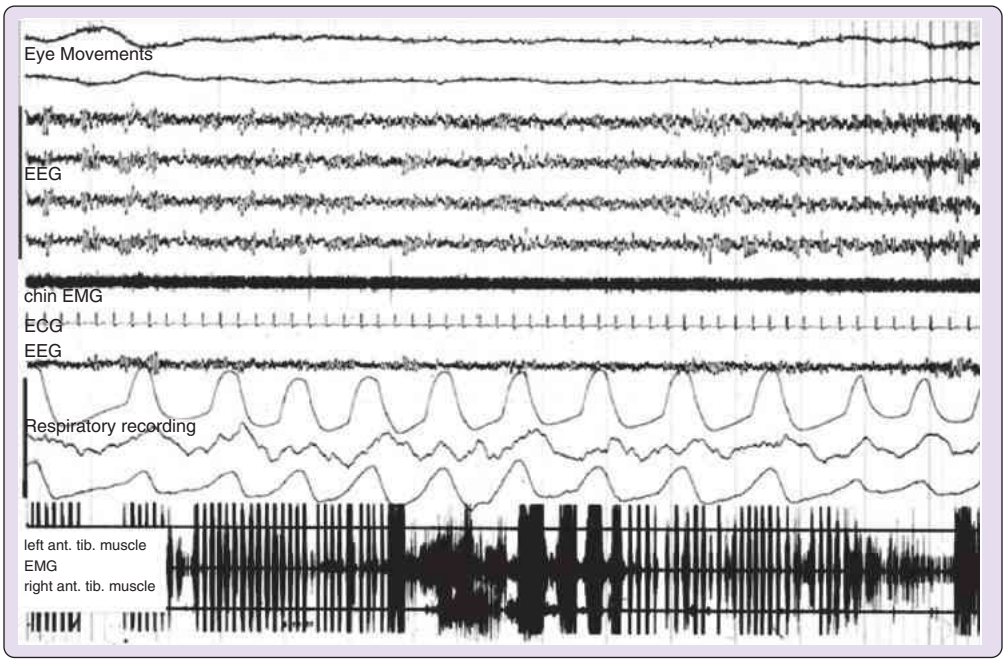


Figure 92-1 Polysomnogram in a patient with Parkinson disease and nocturnal rest tremor of the left leg occurring between wakefulness and stage 1 sleep. The regular unilateral rest tremor is interrupted by bilateral motor activity. Channels from top to bottom: 1–2: eye movements; 3–6: electroencephalogram (EEG); 7: chin electromyogram (EMG); 8: electrocardiogram (ECG); 9: additional EEG; 10–12: respiratory recording with thoracic and abdominal belts; 13: no recording; 14: EMG of right anterior tibialis muscle; 15: EMG of left anterior tibialis muscle.

movements during NREM sleep, blepharospasm at REM sleep onset, and prolonged tonic muscle activity of limb extensor or flexor muscles during NREM sleep.[108] Chin muscle tone may be enhanced during REM sleep, a feature frequently associated with clinical RBD. An example of sleep recordings during an RBD episode is shown in Figure 92-2.

After the EEG and the video are analyzed, and nighttime sleep and Multiple Sleep Latency Test (MSLT) are analyzed, the hypnogram generally shows severely fragmented sleep. The first part of the sleep report should include a description of the EEG aspect, and especially the frequency of the alpha background rhythm, because slow (e.g., 6.5 to 7.8 Hz) rhythms have been associated with RBD and possibly cortex degeneration. The presence of abnormal sleep stages, including eye movements during NREM sleep, and REM sleep without atonia should also be specified. As many as 51% patients with de novo PD (vs. 15% of controls) exhibit minor behavioral events during REM sleep that do not qualify for RBD but may precede its onset.[109]

Sleep architecture has been studied in two large, case-control series. At PD onset, there is no change in sleep structure compared with controls, except for longer REM sleep latency.[109] Patients with treated or more advanced PD have shorter total sleep time, lower sleep efficiency, longer REM latency, higher N1 percentage, and lower REM sleep percentage than controls.[110] Examples of hypnograms obtained in PD patients are shown in Figure 92-3. As for the MSLT, in a series of 54 PD patients, more than half of sleepy patients fell asleep within 5 minutes, a sign of pathologic sleepiness.[97] Moreover, 41% of sleepy patients had REM sleep in two or more MSLT naps.[97,111] This "narcolepsy-like" pattern of sleepiness with multiple sleep-onset REM periods on MSLT was observed in 15% of unselected PD patients,[98] in some cases of MSA, and in dementia with Lewy bodies, but not in progressive supranuclear palsy.[60] In PD patients with severe hallucinations, the hallucinations are temporally associated with REM sleep during the night and with sleep onset in REM periods during the daytime, as in narcolepsy.[105,112] In other PD patients, hallucinations occur after REM sleep during the night and during wakefulness or N1 sleep during the daytime.[113] Sleep attacks may also consist of rapid transitions into N2 sleep.[114]

TREATMENT

The management of sleep disorders in parkinsonism depends on the causes identified by the history and sleep laboratory tests. A list of common problems and corresponding, experience-based management is given in Table 92-3.

Insomnia

The comfort of the PD patient during the night can be improved by using sheets that enable easy movement, silk pajamas without buttons, levodopa tablets and a bottle of water on the bed table, and transitory blockade of nighttime polyuria with an evening intranasal dose of desmopressin. In advanced stages of the disease, the patient's spouse might be inclined to sleep in a different bed or room, but unfortunately these patients are in most need of a nocturnal caregiver for getting out of bed, using the bathroom, and taking their nocturnal levodopa; inadequate rest for the spouse or other caregiver may make the patient's sleep disturbance intolerable and lead to institutionalization. Despite the fact that insomnia is frequent in PD, little research has focused on its treatment. The treatment of PD insomnia is therefore not codified and is more experience than evidence based.[115] Two

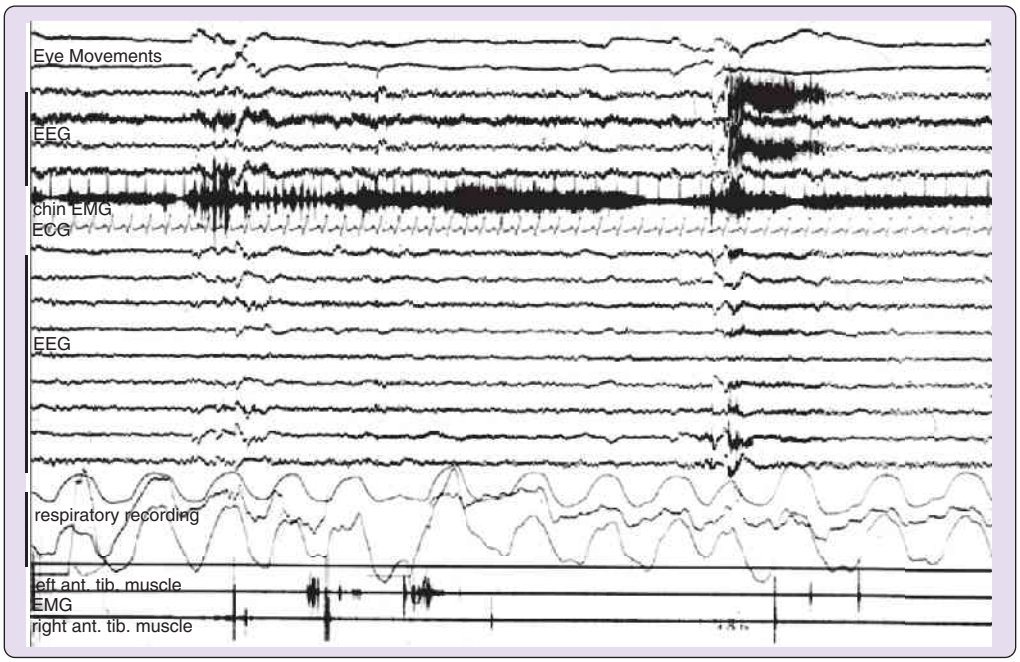


Figure 92-2 Polysomnogram in a patient with Parkinson disease with REM sleep behavior disorder. The chin electromyogram (EMG) shows a highly elevated muscle tone, and motor activity occurs in the recording, with typical muscle twitches in the EMG channels of both legs. Channels from top to bottom: 1–2: eye movements showing a typical REM sleep pattern; 3–6: electroencephalogram (EEG); 7: chin EMG with increased muscle tone; 8: electrocardiogram (ECG); 9–17: further EEG channels, showing muscle artifacts of eye movements; 18–20: respiratory recording with thoracic and abdominal belts; 21: no recording; 22–23: EMG of right and left anterior tibialis muscle.

directions can be drawn: (1) to improve sleep continuity (and in this case, the sleeping patient will not need any nocturnal dopamine treatment); and (2) to reduce movements and muscle tone during the night. Recent randomized control trials have shown equivocal benefit of eszopiclone.[116] A small three-arm study suggested benefits of doxepin and cognitive-behavioral therapy.[117] Pimavanserin, a serotonin 5-HT2A inverse agonist now under development, had beneficial effects on sleep disturbances in PD patients with psychosis.[118] Hypnotics and small doses of sedative antidepressants (provided they do not worsen preexisting RLS, RBD, hallucinations, or daytime sleepiness) are frequently used in clinical practice. When disturbing hallucinations are present, patients can benefit from the combined antihallucination and sedative effects of small, evening dose of atypical neuroleptics such as clozapine (with an appropriate monitoring of leukocyte count) or quetiapine.

Reestablishing continuous dopaminergic stimulation during day and night can be a strategy to improve nighttime motor disability. A careful record of the time of the symptoms, in correlation with time of drug intake, can help identify a nighttime gap in dopaminergic stimulation. The benefit of dopaminergic agents in the evening and night has, however, to be weighed against their alerting effect. As an example, pergolide in the evening worsened sleep efficiency and fragmentation in patients with repeated nocturnal awakenings caused by PD symptoms.[25] In contrast, the use of transdermal rotigotine during the day and night improved most aspects of sleep and night quality, including early-morning akinesia.[119] When comparing rotigotine to oral pramipexole three times a day for motor symptoms, similar results were obtained in advanced PD.[120] The use of levodopa 200 mg in the evening or at bedtime improves subjective sleep quality and decreases night-time movements.[121] Sustained-release forms of levodopa in the evening have not been compared systematically to normal-release forms of levodopa. Sustained-release preparations do not improve subjective aspects of sleep (general quality, sleep-onset latency, total sleep time, number of awakenings), but they have a mild benefit on nocturnal akinesia in small, unblinded trials. We have had good experience with rapid release (dispersible) forms of levodopa soluble in water during the night taken on demand. In contrast, subthalamic nucleus stimulation consistently improves sleep duration, reduces nocturnal awakenings, and reduces early morning dystonia in patients with advanced PD,[70,71,122] but can worsen restless legs symptoms (probably owing to decrease of dopamine agonists).

The treatment of nocturnal RLS in PD may be complex, apart from iron supplementation when it is deficient. Evening gabapentin and derivatives, and more rarely opiates (in non-demented patients without hallucinations), can be tried. No studies or guidelines are available for this purpose. The emergence of RLS and PLMS during subthalamic stimulation suggests that RLS is not controlled by the basal ganglia.[70,123] RLS might correspond to a deficit of dopamine stimulation at night (i.e., would benefit from an evening additional dose of dopamine agonist) or, on the contrary, to "augmentation" from chronic dopaminergic treatment of otherwise subclinical RLS (which would benefit from decreasing the daily dopamine dose which is usually impractical given motor disability). The occurrence of RLS years after (and not before) the onset of parkinsonism supports this last hypothesis.

Figure 92-3 Twenty-four-hour hypnograms in the following patients: a healthy 60-year-old woman with normal sleep **(A)**; a 54-year-old woman with Parkinson disease (PD) treated with levodopa 300 mg/day and bromocriptine 30 mg/day, reporting frequent night-time awakenings, daytime sleepiness and hallucinations (indicated by the *arrow*: saw a stranger in the room, here synchronous with abnormal daytime REM sleep onsets [narcolepsy-like phenotype]) **(B)**; an 80-year-old man with mild PD, complaining of falling asleep 2 to 3 hours after each levodopa intake (*arrow*), who displayed severe hypersomnia **(C)**; and a 72-year-old man with advanced PD and on-off motor fluctuations **(D)**. After a midnight awakening (*arrow*), he developed severe bradykinesia with an axial, painful dystonia that prevented him from resuming sleep. Later, at 4:30 AM, he had foot dystonia (early morning dystonia). All these motor phenomena lengthened his time awake. The *x* axis displays the time of night and day (clock hours), the *y* axis displays the stages of sleep and wakefulness, with awakening (A); non-REM sleep stage 1 (1), stage 2 (2), stage 3 (3), and stage 4 (4); and REM sleep (R).

REM Sleep Behavior Disorder

For patients with potentially dangerous manifestations of RBD, it is important to secure their sleeping environment and ensure the safety of their bed partner. This might include placing a barrier between them and their partner, using twin beds, placing the mattress directly on the floor, and removing objects close to the bed. If antidepressants triggered the RBD and are no longer necessary, they should be stopped (although paradoxically, in established RBD, antidepressants may actually reduce RBD episodes by reducing REM sleep).

Clonazepam (0.5 to 2 mg at bedtime) was the first medication described as useful for RBD.[58] The mechanism of action is unclear; it has no clear effect on tonic REM sleep but may reduce phasic EMG activity and reduce aggressiveness of dream content.[58] In a recent systematic observational study, clonazepam treatment provided moderate or greater improvement in 78% of patients.[124] Although generally well tolerated, in PD clonazepam can cause daytime somnolence, worsen sleep apnea, exacerbate cognitive impairment, and increase risk for falls. Therefore careful monitoring for these side effects is required.

Melatonin (3 to 9 mg at bedtime) is the other major therapy for RBD and may directly increase atonic REM sleep. The effectiveness of melatonin was confirmed in a small, preliminary, randomized trial.[125] In a systematic observational study, melatonin provided moderate or greater improvement in 48% of patients (lower than clonazepam), but with a mean reduction in a visual analogue scale similar to clonazepam.[124] However, melatonin may have less cognitive or gait side effects than clonazepam; 61% of patients on clonazepam reported side effects compared with 33% on melatonin.[124] Therefore

melatonin may be first-line therapy for RBD in PD, if symptoms are mild or moderate.

Otherwise there are some reports that dopaminergic therapy helps in RBD; one prospective study suggested benefit, but others have found none. We have personally experienced variable effects of dopaminergic therapy on RBD, including both reduction and exacerbation of dream enactment. Other potential treatments, which have been documented only in case studies or small series, include donepezil, levodopa, and zopiclone.[126]

Excessive Daytime Sleepiness

Finding and treating the cause of daytime sleepiness in PD requires a careful interview on nocturnal disturbances, hallucinations, recent changes in dopaminergic and psychotropic treatment, and sometimes a nighttime polysomnogram followed by MSLT. These tests may identify treatable severe sleep apnea or a central disorder of arousal. Although sleep apnea syndrome is no more frequent in PD than in aged controls, the benefit of CPAP on sleepiness (measured with MSLT but not with the Epworth Sleepiness Scale) was recently demonstrated in the short term in PD patients with sleepiness and a moderately increased apnea-hypopnea index.[127] If efforts to reduce any sedative drugs (clonazepam, other benzodiazepines, dopamine agonists, sedative antidepressants, opioids, atypical neuroleptics) are without effect or worsen the motor symptoms, the solution could be to add a psychostimulant during the daytime. Caffeine provided a borderline improvement in Epworth Sleepiness Scale score, with a positive clinical global impression of change.[128] Modafinil is well tolerated in PD patients[129] and is reported to have neuroprotective effects in animal models of dopamine depletion.[130] The alerting effect is limited, however, with less than one third responding.[129] Methylphenidate is another option because this drug given at 1 mg/kg decreased excessive daytime sleepiness (by 3 points on Epworth Sleepiness Scale) without worsening sleep quality in a recent controlled trial aiming at alleviating freezing of gait.[131] Caution is required in using stimulants in cognitively impaired patients because confusion and violent behaviors may occur.

CLINICAL PEARLS

- In patients with parkinsonism, clinicians should inquire about nighttime bradykinesia, off-dystonia, early morning akinesia, tremor, RLS, respiratory disturbances, insomnia, RBD, or psychosis from medication-associated factors and their contribution to the sleep problem.
- Specific interventions may then improve quality of sleep and reduce daytime sleepiness, thus enhancing quality of life in patients with PD.
- Nocturnal stridor is a life-threatening symptom in MSA that can benefit from continuous positive airway pressure.

SUMMARY

Altered sleep and vigilance are among the most frequent nonmotor symptoms in parkinsonism. As many as 60% patients with PD suffer from insomnia, 30% to 60% from REM sleep behavior disorder, and 30% from excessive daytime sleepiness. These frequencies can be even higher in atypical parkinsonism. The disruption of sleep and wakefulness in parkinsonism is caused by a combination of neurodegenerative damage in brain areas responsible for sleep, arousal, and circadian regulation; behavioral, respiratory, and motor system phenomena accompanying the disease; and side effects of medications. Sleep maintenance insomnia may be intrinsic to PD, but it is also promoted by motor disability, painful dystonia, RLS, dysuria, anxiety, and depression. Improving motor control during the night with levodopa or long-acting dopamine agonists, benzodiazepine receptor agonists, doxepin, pimavanserin, and quetiapine can help, as does the specific treatment of nocturia, anxiety, and depression. RBD is often violent, enacted dreaming that can cause nighttime injuries to patients and bed partners, generally caused by neurodegeneration in the REM sleep atonia system. Recent longitudinal studies indicate that RBD often precedes parkinsonism (or dementia with Lewy bodies), with up to 90% of idiopathic RBD patients eventually developing a synucleinopathy. Accordingly, idiopathic RBD patients often have other signs of neurodegeneration, including olfactory, cognitive, and autonomic disturbances; decreased dopaminergic transmission in functional imaging; and slowed EEG rhythms. Parkinsonism can be temporarily reversed during RBD episodes in some PD patients. Among PD patients, those with RBD have more gait dysfunction, cardiovascular autonomic loss, hallucinations, and cognitive impairment than patients without RBD. When bothersome or potentially dangerous, RBD is treated primarily with melatonin and clonazepam. Daytime sleepiness and a narcolepsy-like phenotype are also seen in PD. In addition, dopaminergic medications can increase risk for sleep attacks, notably when driving, suggesting a drug-disease interaction. Patients with multiple system atrophy also can develop a progressive, life-threatening laryngeal obstruction (stridor) during sleep that should be rapidly treated with positive airway pressure. Sleep problems in parkinsonism are common, disabling, and treatable; therefore prompt recognition and treatment can substantially improve the patient's quality of life.

Selected Readings

Arnulf I, Konofal E, Merino-Andreu M, et al. Parkinson's disease and sleepiness: an integral part of PD. *Neurology* 2002;**58**:1019–24.

Chaudhuri KR, Schapira AH. Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment. *Lancet Neurol* 2009;**8**: 464–74.

De Cock VC, Vidailhet M, Leu S, et al. Restoration of normal motor control in Parkinson's disease during REM sleep. *Brain* 2007;**130**:450–6.

Iranzo A, Tolosa E, Gelpi E, et al. Neurodegenerative disease status and post-mortem pathology in idiopathic rapid-eye-movement sleep behaviour disorder: an observational cohort study. *Lancet Neurol* 2013;**12**: 443–53.

Kuźniar TJ, Morgenthaler TI, Prakash UB, et al. Effects of continuous positive airway pressure on stridor in multiple system atrophy-sleep laryngoscopy. *J Clin Sleep Med* 2009;**5**(1):65–7.

Mollenhauer B, Trautmann E, Sixel-Doring F, et al. Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology* 2013;**81**:1226–34.

Postuma RB, Iranzo A, Hogl B, et al. Risk factors for neurodegeneration in idiopathic rapid eye movement sleep behavior disorder: a multicenter study. *Ann Neurol* 2015;**77**:830–9.

Tholfsen LK, Larsen JP, Schulz J, et al. Development of excessive daytime sleepiness in early Parkinson disease. *Neurology* 2015;**85**(2):162–8.

A complete reference list can be found online at ExpertConsult.com.

Chapter 93

Sleep and Stroke

Claudio L. Bassetti

Chapter Highlights

- Sleep-wake disorders and stroke are common and intertwined neurologic problems; each may cause the other, and they can arise from similar predisposing factors.
- Clinicians who treat patients with sleep-wake disorders or stroke should be aware of this potential comorbidity and its clinical implications.
- In patients with stroke, treatment of sleep-disordered breathing and other sleep-wake disorders can reduce the risk for subsequent stroke and improve short- and long-term outcomes.

Since the 1990s, the link between sleep and stroke, which was suggested already in the nineteenth century, has received increasing attention. This is mainly a result of better recognition of the strong link between sleep-disordered breathing (SDB) and cardiovascular and cerebrovascular diseases, the high incidence of sleep-wake disturbances in stroke victims, and the effect that SDB, sleep-wake disturbances, and their treatment have on the risk for and outcome of stroke.

STROKE

Stroke is a focal neurologic deficit of acute onset and vascular origin. Stroke has an incidence rate of 2 to 3 per 1000 per year and is the most common neurologic cause of hospitalization. Among patients with stroke, about 65% have ischemic strokes, 15% intracerebral hemorrhage, and 20% transient ischemic attacks (TIAs), in which neurologic deficits typically resolve within 1 hour. Risk factors for stroke include atrial fibrillation, arterial hypertension, dyslipidemia, disorders of glucose metabolism, overweight (abnormal waist-to hip ratio), excessive alcohol consumption, cigarette smoking, and physical inactivity.[1] Patients with heart disease, asymptomatic carotid stenosis, history of TIA, depression, psychosocial stress, and age older than 65 years are also at higher risk for stroke.[1]

Primary prevention of stroke includes treatment of risk factors, regular physical exercise, reduction of body mass index to less than 25, anticoagulation for atrial fibrillation, and endarterectomy in patients with at least 70% carotid stenosis.[2] Emergency treatment includes the systemic use of fibrinolytic agents within the first 4.5 hours and endovascular treatment (thrombectomy) within the first 12 hours after onset of symptoms.[3] Management of acute stroke includes placement of patients in a stroke unit, early recognition of medical complications, and prescription of agents that inhibit platelet aggregation. Surgery may be considered in patients with accessible (e.g., cerebellar) hemorrhages and malignant middle cerebral artery strokes. After stroke, therapies for preventing further events includes platelet antiaggregants, blood pressure–lowering medications, statins, treatment of risk factors, and in selected patients anticoagulation and endarterectomy.[2]

SLEEP-DISORDERED BREATHING AND STROKE

Epidemiology

Sleep-Disordered Breathing as a Risk Factor for Stroke

SDB is strongly associated with stroke. Habitual snoring is an independent risk factor for stroke with a pooled risk of about 1.5.[4] A study of 1022 patients showed that obstructive sleep apnea (OSA) is associated with an increased odds ratio (OR) of 2.0 for stroke and death, even after adjusting for multiple cardiovascular risk factors. The risk was even higher (OR = 3.3) in patients with severe OSA (apnea-hypoxia index [AHI] >36/hour).[5] In a single center study of 1387 male patients with OSA followed for up to 10 years, patients with severe OSA (AHI >30/hour) had a significantly higher incidence of fatal and nonfatal cardiovascular events including stroke compared with patients with mild to moderate OSA, OSA patients treated with continuous positive airway pressure (CPAP), 377 simple snorers, and 264 controls.[6] Other studies have confirmed that SDB is associated with an increased risk for stroke.[7,8] A meta-analysis of prospective studies showed that OSA was associated with incident stroke (OR = 2.24, 95% confidence interval [CI] = 1.57 to 3.19), and cardiovascular mortality (OR = 2.09, 95% CI = 1.20 to 3.65), especially in patients with a high AHI.[9]

SDB is also linked with white matter disease on magnetic resonance imaging and silent strokes.[10-12] In a recent study of 503 elderly individuals who were free of previously diagnosed cardiovascular and neurologic diseases, moderate to severe OSA (AHI ≥15) was independently associated with the presence of white matter changes (OR = 2.08, 95% CI = 1.05 to 4.13) compared with no OSA even after adjustment for hypertension.[12]

Sleep-Disordered Breathing in Acute and Postacute Stroke Patients

In the 1990s, three large, systematic studies demonstrated a very high frequency of SDB in patients with stroke and TIA.[13-15] In a meta-analysis of 29 studies with a total of 2343 ischemic or hemorrhagic stroke and TIA patients, the frequency of SDB with AHI greater than 5 was 72% and with AHI greater than 20 was 38%. Males had a higher percentage of SDB (AHI >10) than females (65% vs. 48%; $P = .001$). Patients with recurrent strokes had a higher prevalence of SDB (AHI >10) than those with first strokes (74% vs. 57%).[16] In a literature search until December 31, 2014, we found 54 publications in which SDB frequency was assessed on a total of 4293 stroke patients, solidly confirming these findings.

A few studies have assessed the prevalence of SDB in patients with TIA and have found it to be similar to that of patients with stroke.[16] A few studies have reported a higher frequency of SDB in specific stroke subgroups such as hemorrhagic, brainstem, nocturnal/wake-up, and recurrent strokes.[17-19]

SDB often improves across the acute to the subacute phase.[20-23] In general, central events improve more than obstructive events. One study suggested a greater improvement in SDB in hemorrhagic compared with ischemic strokes.[24]

Clinical Features

Breathing Disturbances During Sleep

The most common form of SDB in stroke patients is OSA (Figure 93-1). Moderate to severe OSA (AHI >30) is found in about 20% to 30% of patients with stroke.[16] Occasionally, patients may present with both OSA and Cheyne-Stokes breathing (CSB) (Figure 93-2). OSA is often worse in rapid eye movement (REM) sleep, whereas CSB is usually worse in light non–rapid eye movement (NREM) sleep.[13,25] In the first few days after stroke, CSB and other forms of central periodic breathing are present during at least 10% of the recording time in about one third of patients.[20,26-28]

Breathing Disturbances During Wakefulness

Different breathing abnormalities can occur during wakefulness after hemispheric strokes. Abnormalities include selective impairment of behavioral or volitional respiratory control.

Hemispheric strokes in the frontal cortex, basal ganglia, or internal capsule may cause respiratory apraxia, with impaired voluntary modulation of breathing amplitude and frequency, leaving patients unable to take a deep breath or hold the breath.[29,30]

Brainstem strokes can be associated with different forms of abnormal breathing patterns. Sustained respiratory rates above 25 to 30/min in the absence of hypoxemia (neurogenic hyperventilation) were originally described in six comatose patients with ventrotegmental pontine strokes but were subsequently attributed to pulmonary edema (and stimulation of lung and chest wall afferent reflexes).[31] Neurogenic hyperventilation after stroke usually indicates a poor prognosis.[32] Inspiratory breath holding (apneustic breathing), originally described in two patients with bilateral ventrotegmental mediocaudal (infratrigeminal) pontine stroke,[33] is rare and usually secondary to basilar artery occlusion. Erratic variations in breathing frequency and amplitude (ataxic or Biot breathing) and failure of automatic breathing (central sleep apnea or

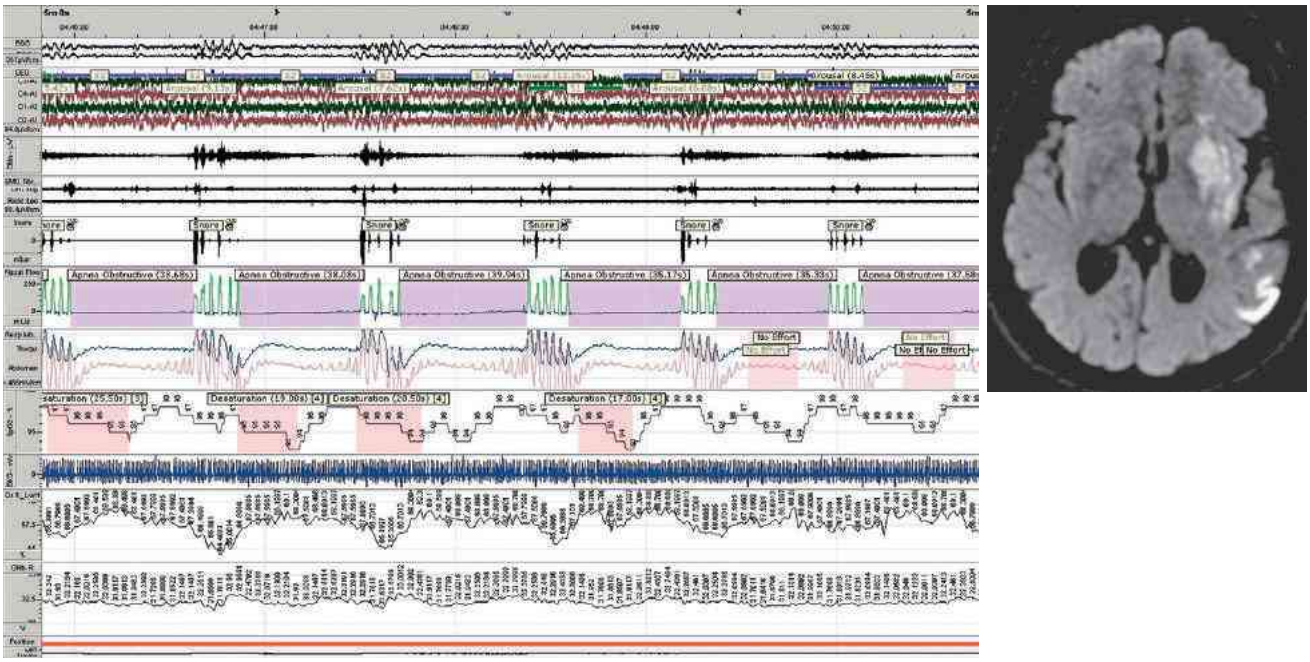


Figure 93-1 Obstructive sleep apnea in acute ischemic stroke. This 70-year-old man has left middle cerebral artery stroke, carotid artery occlusion, and atrial fibrillation. He has habitual snoring but no excessive daytime sleepiness. Aphasia and severe hemiparesis are clinically apparent. National Institutes of Health stroke score is 16, and there are no signs of heart failure. Polysomnography 2 days after stroke onset shows an apnea-hypoxia index (AHI) of 79 and minimum oxygen desaturation of 85%. The AHI normalized (<5/hr) with continuous positive airway pressure. (MRI pictures courtesy Professor A. Valavanis, Institute of Neuroradiology, University Hospital, Zürich, Switzerland.)

Figure 93-2 Central sleep apnea in acute ischemic stroke. This 63-year-old man had a left subcortical stroke of unknown origin, with arterial hypertension and habitual snoring but no excessive daytime sleepiness. He had a mild hemiparesis, with a National Institutes of Health stroke score of 8 and no signs of heart failure (cardiac ejection fraction, 55%). Polysomnography the first night after stroke onset showed an apnea-hypoxia index (AHI) of 53 (mainly central apneas). The patient spontaneously improved after 1 week (AHI 16). (MRI pictures courtesy Professor G. Schroth, Institute of Neuroradiology, University Hospital, Bern, Switzerland.)

Ondine's curse), usually imply a lateral medullary stroke, often bilateral.[34,35] Damage to the medullary reticular formation and nucleus ambiguous may cause a loss of automatic breathing, whereas a lesion that includes the nucleus of the solitary tract is necessary to cause failure of both automatic and voluntary respiration.[36] Volitional breathing can be impaired by brainstem strokes involving corticobulbar and corticospinal pathways at pontine and medullary levels.[30]

Spinal cord stroke can impair both automatic and voluntary breathing. Anterior spinal artery strokes can affect reticulospinal pathways, located anteriorly in the lateral columns of the first three cervical segments, which are crucial for automatic breathing.[37] In contrast, posterior spinal artery strokes can damage corticospinal pathways in the dorsolateral spinal cord and impair voluntary control of breathing.[38] Strokes that extend up to the C1 level usually cause severe respiratory insufficiency and necessitate ventilatory support.

Repetitive yawning can accompany hypersomnia (e.g., in patients with thalamic or posterior hypothalamus stroke) and can also occur as a release phenomenon in patients with brainstem and supratentorial lesions. Yawning can also occur with insular and caudate lesions.[39]

Pathophysiology

Several acute and chronic consequences of nocturnal respiratory events may explain the link between SDB and increased risk for stroke. Sympathetic hyperactivation, intermittent hypoxemia and oxidative stress, and inflammation likely increase the risk for atherosclerosis and cardiovascular morbidities. However, this link has been difficult to prove because sleep apnea and stroke share many of the same risk factors such as obesity and diabetes.

Sleep-Disordered Breathing as a Risk Factor for Stroke: Acute Effects

Acutely, apneas and hypopneas during sleep can be accompanied by decreased cardiac output, cardiac arrhythmias, systemic hypotension or hypertension, vasodilation due to hypoxia and hypercapnia, and increased intracranial pressure. These factors lead to a roughly 15% to 20% reduction in cerebral blood flow velocities during respiratory events.[40,41] Large fluctuations in cerebral blood velocities (and flow) may be particularly detrimental because patients with SDB have been shown to have diminished vasodilator reserve and impaired cerebral autoregulation.[41]

Type, duration, and timing of respiratory events affect hemodynamic consequences. Near infrared spectroscopy (NIRS) studies have shown that SDB can disrupt autoregulatory mechanisms and cause brain hypoxia, particularly with severe SDB (AHI > 30)[42] (Figure 93-3). These effects may be particularly detrimental to the ischemic region (penumbra) bordering the stroke.[43] Transcranial Doppler studies have shown that obstructive apneas of long duration and occurring during REM sleep may also be particularly detrimental.[44] CSB and central apneas also can alter cerebral blood flow.[44] Paradoxical embolization due to right-to-left shunting in patients with patent foramen ovale during long apneas is another potential mechanism of stroke.[17,45,46]

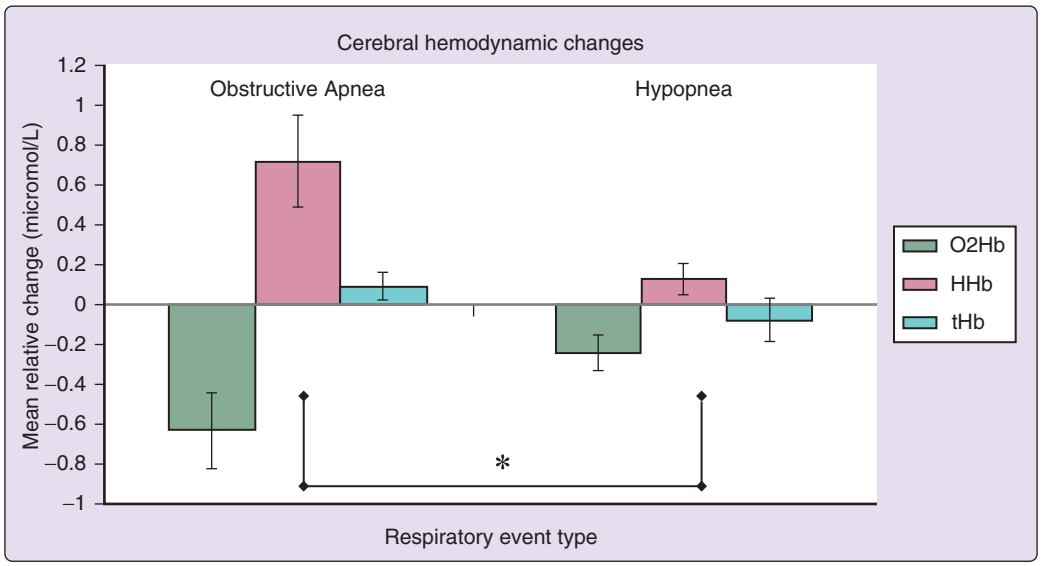


Figure 93-3 Cerebral hemodynamic alterations in patients with sleep-disordered breathing as estimated by near infrared spectroscopy (NIRS). Patients with snoring (n = 7, apnea-hypopnea index [AHI] = 2 ± 2/h); mild SDB (n = 7, AHI = 14 ± 8/h, range); and severe obstructive sleep apnea syndrome (n = 5, AHI = 79 ± 20/h) were studied. NIRS data associated with different respiratory events (obstructive apnea and hypopnea) were averaged for each patient. Subsequently, corresponding cerebral hemodynamic (and peripheral oxygen saturation, SpO_2) relative changes were assessed via integrals adjusted for duration. The relative changes in brain tissue parameters (concentrations of oxyhemoglobin [O_2Hb], deoxyhemoglobin [HHb], and total hemoglobin [tHb]) were significantly larger during obstructive apneas than during hypopneas.[42]

In light of these observations, it is not surprising that snoring and SDB have been associated with the onset of cerebrovascular events at night and strokes that are apparent on awakening (so-called wake-up strokes).[23,46-48]

Sleep-Disordered Breathing as a Risk Factor for Stroke: Chronic Effects

Chronically, SDB and even habitual snoring are associated with hypertension, which is a major risk factor for stroke. In the Wisconsin sleep cohort, an AHI of greater than 15 was independently associated with a threefold increased risk for developing new hypertension within a 4-year period.[49] In a prospective cohort study of 1889 participants without hypertension, Marin and colleagues found an increased risk for incident hypertension within a 12-year period in patients with an AHI of 5 or greater who were ineligible for CPAP therapy (OR = 1.33, 95% CI = 1.01 to 1.75), among those who declined CPAP therapy (1.96; 95% CI = 1.44 to 2.66), and among those nonadherent to CPAP therapy (1.78; 95% CI = 1.23 to 2.58), whereas the risk was lower in patients with AHI of 5 or greater who were treated with CPAP therapy (0.71, 95% CI = 0.53 to 0.94).[50] Importantly, these effects were independent from changes in body weight.

SDB is associated also with coronary heart disease, myocardial infarction, heart failure, and atrial fibrillation, all of which are also risk factors for stroke.[51,52]

Several observations support the hypothesis that SDB worsens atherogenesis. The intima media thickness of the common carotid artery is increased in SDB patients compared with controls matched for age and vascular risk factors.[53] Patients with SDB have increased arterial stiffness, a recognized marker of cardiovascular risk and long-term morbidity.[54]

A few studies have shown that CPAP produces reductions in mean arterial blood pressure. Pepperell and colleagues reported that therapeutic CPAP reduced mean arterial blood pressure by 2.5 mm Hg, whereas subtherapeutic CPAP levels increased blood pressure by 0.8 mm Hg. Such an effect could reduce stroke risk by about 20%.[55] Although the effects of CPAP on hypertension and cardiovascular events are debated, CPAP has been shown to reduce blood pressure in OSA patients with cardiovascular disease or multiple cardiovascular risk factors.[56,57] Treatment with CPAP also can improve other detrimental hemodynamic, neural, and molecular effects of SDB such as factor VII clotting activity, fibrinogen levels, and platelet activation or aggregation.[58]

Sleep-Disordered Breathing as a Consequence of Stroke

Although patients at risk for cerebrovascular disease frequently have SDB before they experience a stroke, some develop SDB as a consequence of stroke. The observation that recovery from stroke may be accompanied by improvement of nocturnal respiratory parameters indirectly supports the hypothesis that SDB can worsen or may even appear de novo after stroke (see earlier). Several factors may be involved.

First, CSB was traditionally attributed to CO_2 hypersensitivity secondary to bilateral and severe strokes with heart failure and a decreased level of consciousness,[59] but subsequent studies have challenged this view.[26,28,60] Changes in ventilatory sensitivity to inhaled CO_2 can also occur in patients with rostrolateral medullary lesions and with unilateral, small strokes of the central autonomic network (including the insula) and without heart failure.[28] Noteworthy, strokes in these areas can also cause acute cardioautonomic changes, including atrial fibrillation.[61] Nonneurogenic factors including older age, left ventricular failure, coronary heart disease,

acute caudorostral fluid shifts related to the nocturnal recumbent body position, and carotid stenosis may also contribute to the appearance of CSB.[26,28,62,63] Of note, CSB in the context of asymptomatic carotid stenosis (in 39% of patients in one study) was associated with a shift in sympathicovagal balance secondary to an increased baroreflex and chemoreflex sensitivities in the carotid body.[64]

Second, patients with stroke may have SDB from weak upper airway muscles or poor coordination between upper airway, intercostal, and diaphragmatic muscles due to brainstem or hemispheric lesions that impair cranial nerve function.[65] Accordingly, dysphagia and hypoglossal nerve dysfunction are associated with poststroke SDB.[66,67]

Third, acute brain damage may affect respiratory drive (see earlier).

Finally, other factors such as hypoxemia due to aspiration or respiratory infections, reduction of voluntary chest movements on the paralyzed side, supine position, and sleep fragmentation secondary to stroke or stroke complications can also lead to SDB.

Acute and Chronic Clinical Effects and Consequences

SDB in patients with cerebrovascular disorders can have a variety of detrimental effects. Acutely, SDB detected the first night after brain infarction is associated with early neurologic worsening, longer hospitalization, and detrimental physiologic changes such as higher blood pressure and reductions in cerebral oxygenation[43,47,68,69] Higher blood pressure in the acute phase of stroke is linked with a poorer clinical outcome.

Chronically, habitual snoring and SDB also worsen stroke outcome. A recent systematic review of the effects of SDB on recurrence and death of stroke and TIA patients generally supports a dose-response relationship between severity of SDB and risk for recurrent events and all-cause mortality.[70]

SDB can also worsen functional outcome. Stroke patients with nocturnal oxygen desaturations had a poorer functional outcome than those without desaturations.[71] In 61 patients in a stroke rehabilitation unit, OSA was associated with more functional impairment and longer hospitalization.[72] In a series of 120 stroke patients, those with an AHI of more than 10 within 24 hours of stroke onset had worse functional outcome (as assessed by the Barthel Index) and higher mortality at 6 months.[73] Similarly, in a study of 60 patients, patients with an AHI of more than 15 had a poorer functional outcome at 6 months.[74] More recent studies have shown that SDB may negatively affect clinical outcome within the first weeks after stroke onset.[75-77]

Diagnosis

Based on SDB's impact on stroke outcome and treatability, the American Heart Association and American Stroke Association now recommend that assessment of SDB should be considered in all patients with TIA and stroke.[2] On a practical level, screening and treatment of SDB in stroke patients are also cost-effective.[78]

The suspicion of SDB should be particularly high in obese male patients with a history of habitual snoring, witnessed apneas, hypertension, diabetes mellitus, and sleep-onset/wake-up stroke.[15,23,47] In clinical practice, asking the patients and relatives about sleep-related breathing symptoms (using, e.g., the Berlin Questionnaire) and excessive daytime sleepiness preceding the onset of stroke was shown to be helpful.[79,80]

Bedside assessment of sleep breathing can provide more accurate estimates of SDB. Different forms of unattended respirography or polysomnography (e.g., AutoSet, ApneaLink, and LifeShirt) may be sufficiently accurate to diagnose SDB and estimate its severity even in the acute setting.[19,23,69,81] Full polysomnography is eventually needed only in a minority of patients.

The optimal timing of sleep studies after stroke or TIA is unknown. Although studies within days of a stroke might be less representative of the patient's baseline, treatment of SDB soon after stroke could potentially minimize further brain injury and improve outcome.

Treatment of Sleep-Disordered Breathing

Treatment of SDB in stroke patients can be a clinical, technical, and logistical challenge. Treatment strategies should always include prevention and early treatment of secondary complications (e.g., aspiration, respiratory infections, pain) and cautious use or avoidance of alcohol and sedatives that may worsen breathing during sleep. Patient positioning in the acute phase can improve oxygen saturation and reduce severity of OSA by 20%.[82-84] In heart failure patients, a lateral sleeping position was shown to reduce the severity of central sleep apnea.[85]

The effect of supplemental oxygen in the acute phase after stroke is unclear.[86] Even the most recent guidelines of the European Stroke Organization and American Heart Association/Stroke Association, while pointing to the necessity to keep oxygen saturations above 92% to 94%, do not specify how to measure or correct nocturnal oxygen saturation.[87] In a recent trial of SDB patients, oxygen was found to be inferior to CPAP in reducing blood pressure levels in patients with high cardiovascular risk.[57]

Initial reports questioned the feasibility and utility of treatment with CPAP and other nocturnal ventilator devices in stroke patients. Based on a review of the 26 studies published until December 31, 2014 including a total of 901 treated patients, we recommend a more optimistic approach to treatment (see later).

Treatment of Obstructive Sleep-Disordered Breathing

CPAP is usually the treatment of choice for obstructive SDB, but CPAP compliance can be a challenge because most patients with stroke and SDB lack excessive daytime sleepiness and may not perceive much benefit from CPAP. In addition, stroke patients may have trouble using CPAP if they have dementia, delirium, aphasia, anosognosia, pseudobulbar or bulbar palsy, or severe motor impairment.

Initially, CPAP studies were mainly performed in the subacute phase of stroke (>1 month after onset). In a study of 105 stroke patients treated in a rehabilitation unit, CPAP was accepted by 70%, and poor acceptance was associated with aphasia and severe stroke. CPAP use over 10 days led to an improvement of subjective well-being and lower night-time blood pressure values.[88] CPAP compliance was poor (29%) over the first month of treatment in a series of 51 patients,[89] but over the next 18 months, CPAP use was associated with a five times reduction in vascular events. In a subsequent study, 5-year mortality was reduced in 28 stroke patients with mild to moderate SDB (AHI ≥20) who were treated with CPAP compared with 68 patients with SDB who did not tolerate treatment.[90]

More recently, several studies have examined the feasibility and effects of CPAP treatment started in the acute setting. Sandberg and colleagues reported a good effect on depressive symptoms even when CPAP was started within 2 to 4 weeks from stroke onset.[91] First series emphasized poor CPAP compliance in this clinical setting (ranging from 12% to 22% in patients followed for 2 to 60 months).[22,23,92,93] Recent studies on this topic reported better compliance when CPAP treatment was begun in the first one to three nights after hospitalization. These better results may have arisen from careful selection of patients, the use of new respiratory devices (including adaptive servoventilation [ASV] machines) and headgears, higher motivation and instruction of the treating teams, and frequent contact with patients.[75,94] One randomized trial showed a favorable effect on stroke severity (National Institutes of Health stroke scale) in patients treated with auto-CPAP within 48 hours of hospitalization.[75] In 45 patients with acute TIA, auto-CPAP had acceptable adherence and reduced the risk for recurrent stroke 82% over a treatment period of 90 days.[95] In stroke patients treated within 24 hours with CPAP, stroke severity was improved in those using CPAP more than 4 hours/night.[77] In the largest study thus far, Parra and colleagues reported an improved 1-month neurologic recovery and fewer cardiovascular events by 24 months in 71 patients with moderate to severe SDB (AHI ≥20) started on CPAP within the first 3 to 6 days after stroke onset compared with 69 untreated SDB patients.[81] At 5 years, the treated SDB patients also had better survival than the untreated SDB stroke group.[96]

Treatment of Central Sleep-Disordered Breathing

In patients with mainly central apneas and CSB, breathing disturbances can be improved with oxygen and possibly also lateral sleeping position (see earlier). Clinical experience and a few studies suggest that adaptive ASV may be effective in poststroke CSB and central sleep apnea patients, including those nonresponsive to conventional CPAP.[97]

Tracheostomy and mechanical ventilation may become necessary in patients with central hypoventilation, central apneas, and ataxic breathing.

Hiccups can be treated with neuroleptics or baclofen.

In sum, current data suggest the feasibility of CPAP and ASV treatment in patients with poststroke obstructive and central SDB with an acceptable compliance even when diagnosis and treatment take place in the acute setting. Further studies are needed to better define the best approach in different patient subpopulations according to type and severity of SDB and interval after the acute cerebrovascular event.

SLEEP-WAKE DISTURBANCES AND STROKE

Epidemiology

Sleep-Wake Disturbances and Short Sleep Duration as Risk Factors for Stroke

Several studies have linked insomnia with increased risk for cardiovascular events and death.[98,99] Two meta-analyses have confirmed that short sleep duration (≤6 hours) slightly increases the risk for ischemic stroke (OR = 1.2, 95% CI = 1.0 to 1.3).[100,101] A large study of 21,438 Asian insomniac patients and 64,314 matched noninsomniac patients found a 54% higher risk for stroke over a 4-year follow-up period.[102] Another large study of 30,934 U.S. subjects reported that short sleep duration (<5 hours) independently increased the risk for stroke 61%.[103] This increase in risk may be due to an increase in sympathetic activity secondary to sleep loss and fragmentation and recurrent arousals with consequent hypertension and impaired glucose metabolism.[104,105]

Conversely, a few studies have suggested an association between long sleep duration or excessive daytime sleepiness and an increased risk for stroke.[100,101,103,106] In a study of 2088 elderly community residents, more than 44% of the cohort reported no daytime dozing, 47% some dozing, and 9% significant daytime dozing. Compared with those reporting no daytime dozing, individuals reporting significant dozing had a 74% increased risk for ischemic stroke.[106]

Restless legs syndrome (RLS) and periodic limb movements during sleep (PLMS) may also increase the risk for cardiovascular diseases, including stroke.[105,107]

Whether shift work increases the risk for stroke remains controversial.[108]

Sleep-Wake Disturbances in Stroke Patients

Clinical experience and a few studies suggest that SWD are frequent after strokes, but few studies have assessed their prevalence and determinants systematically.

In a series of 100 consecutive stroke patients assessed in the acute phase, 22% had an Epworth Sleepiness Scale score of 10 or higher (excessive daytime sleepiness) or increased sleep need (2 or more hours sleep per 24 hours compared with the prestroke situation).[109]

In a series of 235 patients assessed 0.3 to 2 years after stroke, 46% reported abnormal fatigue (fatigue severity scale ≥4.0).[110] Other studies have suggested similarly high frequencies (up to 70% of patients) of poststroke fatigue, making it the most common SWD in stroke survivors.[110-112]

In a series of 277 consecutive patients evaluated 3 months after stroke, insomnia was reported by 38% of patients, and in 18% of these insomnia patients, their insomnia appeared de novo after stroke.[113] In a recent Brazilian study, insomnia was found in 38% of 40 stroke patients.[114]

In a series of 137 patients assessed 1 month after stroke, RLS symptoms were found de novo in 12% of patients.[115]

Other SWD after strokes include an abnormal transition from wakefulness to sleep and vice versa, with dream-reality confusion (oneiric state), dream changes, and an altered perception of time (*Zeitgefühl*).[116,117]

Clinical Features, Pathophysiology, and Treatment

Hypersomnia and Excessive Daytime Sleepiness

After a stroke, some patients may require more sleep (hypersomnia) and or have excessive daytime sleepiness (EDS). Reduced arousal because of lesions involving the ascending arousal pathways is the most common cause of poststroke hypersomnia.[109] The most severe and persistent hypersomnia occurs in patients with bilateral lesions of the thalamus (Figure 93-4), subthalamic and hypothalamic area, tegmental midbrain, and pons (Figure 93-5), where fibers of the ascending arousal pathways are bundled and can be severely injured even by single small lesions. Hypersomnia after hemispheric stroke (Figure 93-6) usually occurs with large lesions, on the left more than on the right, and anteriorly more than posteriorly.[118-121] In large hemispheric strokes, loss of arousal and coma can occur with injury to the upper brainstem secondary to brain edema and herniation.

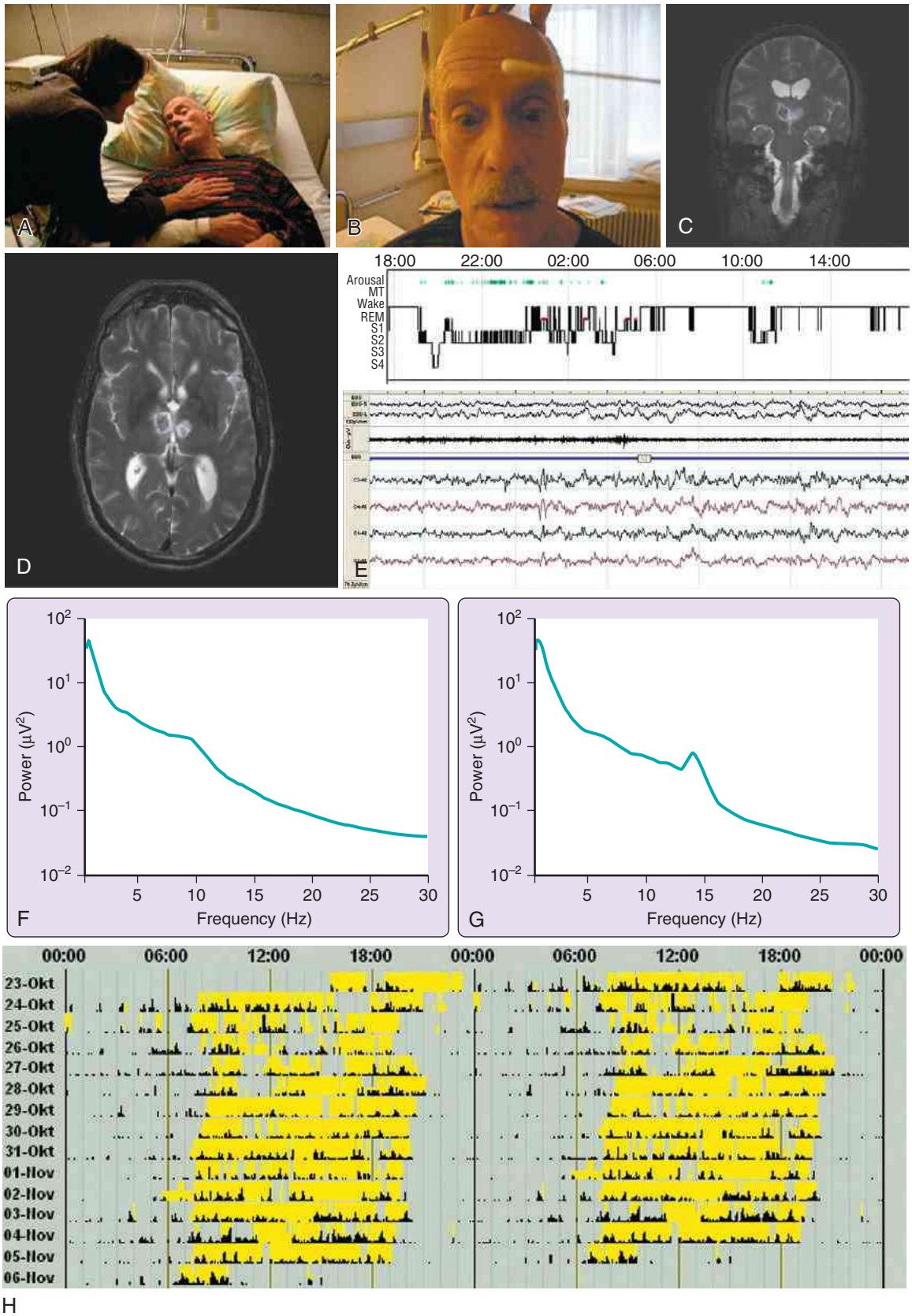


Figure 93-4 Hypersomnia after bilateral paramedian thalamic stroke. This 65-year-old man had initial coma, followed by severe hypersomnia **(A),** vertical gaze palsy **(B),** amnesia, and disturbed time perception (Zeitgefühl). Brain magnetic resonance imaging (MRI) showed bilateral paramedian thalamic strokes **(C, D).** Polysomnography 12 days after stroke onset demonstrated a drastic reduction of sleep spindles **(E)** and loss of spindle peak **(F)** (12 to 14 Hz activity) on spectral analysis (compared with a normal control **[G]**). Severe central apnea (apnea-hypopnea index, 54/hr, >90% central) was observed in the acute phase (in the absence of any signs of cardiac dysfunction) but not on follow-up a few months later. Actigraphy performed in the first month after stroke shows time "asleep" (rest or sleep) was 61% of the recording time (2 weeks) **(H).** One year after stroke, the patient still reported increased sleep need (15 hours per day), apathy (athymormia), and attentional and memory deficits. Modafinil at a dose of 200 mg per day improved his hypersomnia. (MRI pictures courtesy of Professor A. Valavanis, Institute of Neuroradiology, University Hospital, Zürich, Switzerland.) (From Bassetti CL, Hermann DM. Sleep and stroke. In: Vinken PJ, Bruyn GW. *Handbook of Clinical Neurology: Sleep Disorders.* New York: Elsevier; 2010.)

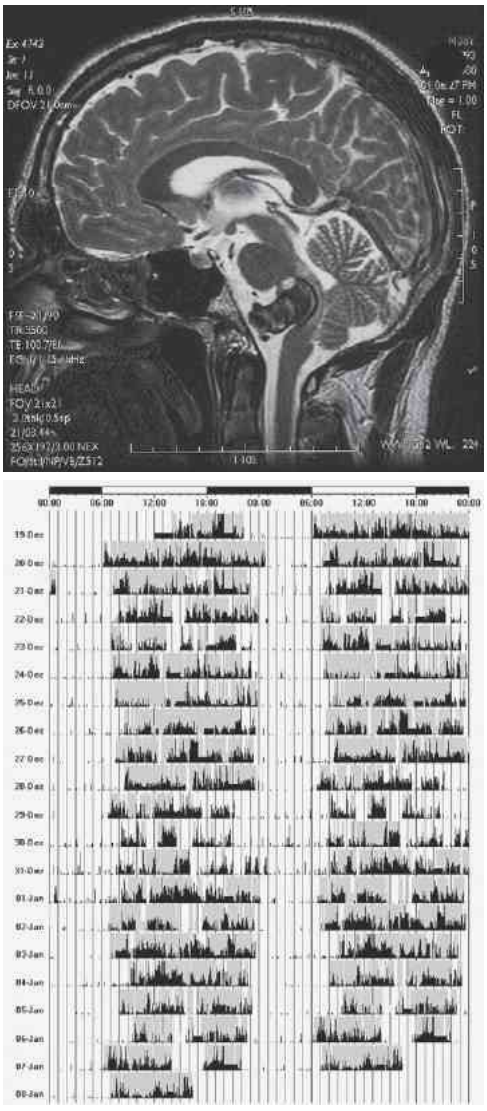


Figure 93-5 Hypersomnia and excessive daytime sleepiness after pontomedullary stroke. This 39-year-old man had pontomedullary ischemia following subarachnoid hemorrhage and embolization of a giant aneurysm of the basilar artery. This resulted in a brainstem syndrome with hiccups; left IX, X, XII palsies; dysarthria; gait ataxia; and mild left hemiparesis. Sleep symptoms postintervention were severe excessive daytime sleepiness (Epworth Sleepiness Scale score of 23/24) and increased sleep need (12 to 14 hours/day). Polysomnography showed sleep efficiency 97%, slow wave sleep 8% of total sleep time, no sleep apnea, and no periodic limb movements in sleep. The Multiple Sleep Latency Test showed mean sleep latency of 1 minute and no sleep-onset rapid eye movement periods. Actigraphy showed that time "asleep" (rest or sleep) was 43% of the recording time (2 weeks). Cerebrospinal fluid levels of hypocretin-1 were normal. (MRI pictures courtesy Professor A. Valavanis, Institute of Neuroradiology, University Hospital, Zürich, Switzerland.)

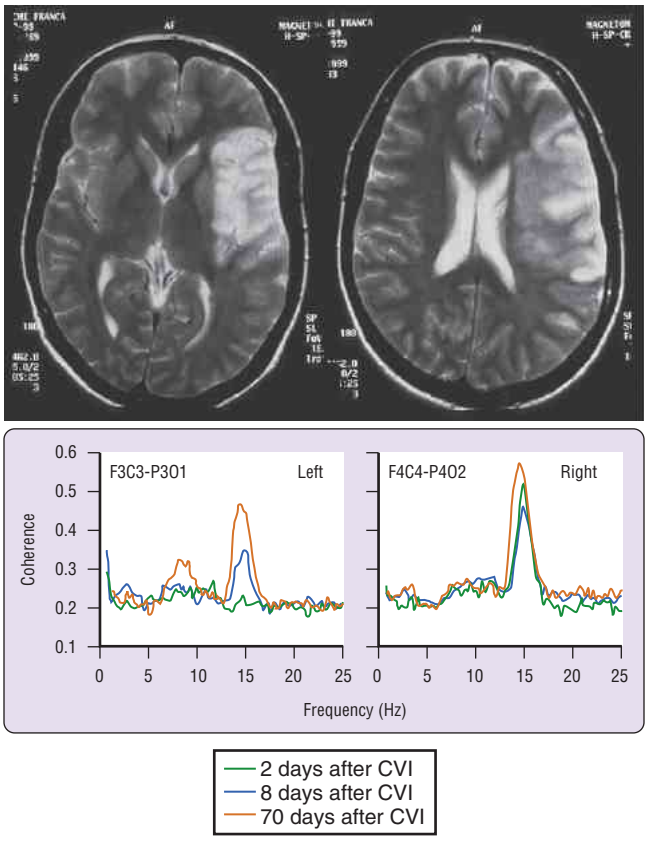


Figure 93-6 Hypersomnia and altered sleep architecture after middle cerebral artery stroke. This 39-year-old woman had aphasia, right hemiparesis, depressed mood, and crying spells. National Institutes of Health stroke score was 16. Before her stroke, the patient slept 7 hours/day, but in the first 1 to 2 weeks after her stroke, she slept 12 hours/day and had mild excessive daytime sleepiness (Epworth Sleepiness Scale score of 12). At 12 months, the patient reported an improvement in sleep need to 10 hours/day. Repeated polysomnograms on day 2, day 8, and day 70 after stroke demonstrated progressive recovery of spindling activity (coherent activity at about 12 Hz) over both the affected (*left*) and the unaffected (*right*) hemisphere. CVI, chronic venous insufficiency. (MRI pictures courtesy Professor G. Schroth, Institute of Neuroradiology, University Hospital, Bern, Switzerland.)

Hypersomnia has occasionally been documented polysomnographically in patients with thalamic, mesencephalic, and pontine strokes.[122-124] Such strokes may cause initial coma or, conversely, manic delirium, hyperalertness, and insomnia before hypersomnia evolves. Mental arousal seems to be affected more severely by medial lesions, whereas motor arousal (including spontaneous motor activities) is impaired more strongly by lateral lesions.[125,126]

In patients with strokes that injure arousal pathways or the paramedian thalamus, hypersomnia may alternate with insomnia (see earlier). For example, one 78-year-old patient with a tegmental mesencephalic infarct had severe, persistent hypersomnia accompanied by an inversion of the sleep-wake cycle with nocturnal agitation.[127]

With deep (subcortical) hemispheric and thalamic strokes, patients may exhibit so-called presleep behavior, during which they yawn, stretch, close their eyes, curl up, and assume a normal sleeping posture, while complaining of a constant sleep urge.[128] Some of these patients are able to control this behavior when stimulated or given explicit, active tasks to perform. This "presleep behavior" may be compulsive in that removal of the patient from bed can result in repeated attempts to lie down and adopt a sleeping posture. However, during what appear to be daytime sleep periods, relatively quick responses to questions or requests suggest wakefulness. For this peculiar dissociation between lack of autoactivation (spontaneous engagement in mental and motor activities) in the presence of preserved heteroactivation (mental and motor activities secondary to external stimulation), Laplane suggested the term athymormia, or "pure psychic akinesia".[129]

In some patients, hypersomnia evolves to extreme apathy with lack of spontaneity and initiative, slowness, poverty of movement, and catalepsy, a condition for which the term akinetic mutism was coined. Akinetic mutism, and its less severe form—usually referred to as abulia—may persist despite normalization of vigilance or even after appearance of insomnia. Some of these patients are eventually diagnosed to have poststroke fatigue (see later) or poststroke depression.

Narcolepsy-like phenotypes (with, however, atypical or questionable cataplexy) are rare with strokes but can occur in the absence of HLA positivity and cerebrospinal fluid hypocretin-1 deficiency. One 23-year-old patient had hypersomnia from bilateral diencephalic strokes; his hypocretin-1 levels were low, suggesting a link between poststroke hypersomnia and deficient hypocretin neurotransmission.[130]

Treatment of poststroke hypersomnia is often difficult, but improvements have been reported with amphetamines, modafinil, methylphenidate, and dopaminergic agents.[109,123] Bromocriptine may improve apathy and presleep behavior.[128] Treatment of an associated depression with stimulating antidepressants may also improve poststroke hypersomnia. It is noteworthy that a favorable influence on early poststroke rehabilitation was reported for both methylphenidate (5–30 mg/day, 3 weeks' trial) and levodopa (100 mg/d, 3 weeks-trial), an effect that may be partially related to improved arousal.[131,132]

Fatigue

A continuum exists among hypersomnia, depression, and fatigue, which is defined as a feeling of physical tiredness, exhaustion with lack of energy accompanied by a strong desire for sleep with usually normal or (paradoxically decreased) sleep propensity. Fatigue may be more common with brainstem strokes.[133] The Epworth sleepiness scale can sometimes help differentiate fatigue from EDS.[134]

An overlap exists between poststroke fatigue and poststroke depression. Psychological stress in coping with stroke probably plays an important role, as suggested by the absence of a correlation between poststroke fatigue and stroke size and site and by a similarly high frequency of fatigue after myocardial infarction (without brain damage).[135] Some have proposed that poststroke fatigue is caused by dysfunction of arousal and attentional circuits.[133] Activating antidepressants and amantadine can be tried for poststroke fatigue.[136] One study reported no benefit with fluoxetine.[137]

Insomnia

On rare occasions, strokes can cause insomnia directly, presumably through disruption of sleep mechanisms. One patient with a pontomesencephalic stroke had almost complete insomnia for more than 2 months.[138] Two patients with locked-in syndrome due to pontomesencephalic or bilateral basal pontine strokes had almost no sleep for more than 1 month.[139,140] Patients with caudate or subcortical (Figure 93-7), thalamic, thalamomesencephalic, and tegmental pontine stroke can have insomnia accompanied by an inversion of the sleep-wake cycle, with insomnia and agitation during the night and hypersomnia during the day.[138,140-142]

Aside from the brain damage, other factors may contribute to poststroke insomnia, including anxiety, dementia, medical disorders (e.g., heart failure, pulmonary disease), SDB, psychotropic medications, infections and fever, inactivity, environmental disturbances, stress, and depression.[113]

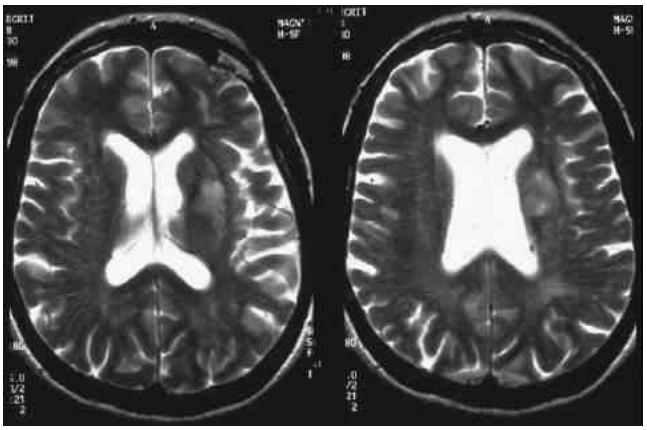


Figure 93-7 Insomnia after subcortical stroke. This 68-year-old woman had a left subcortical stroke in the corona radiata, with mild right motor hemiparesis. National Institutes of Health stroke score was 6. In the first poststroke week, she had almost complete insomnia and excessive daytime sleepiness (EDS). Two weeks later, her EDS was improved and she was sleeping 2 to 3 hours/night. Sleep-wake functions normalized after 4 weeks. (MRI pictures courtesy Professor G. Schroth, Institute of Neuroradiology, University Hospital, Bern, Switzerland.)

Treatment of poststroke insomnia should focus on behavioral measures such as placing patients in private rooms at night, protection from nocturnal noise and light, and mobilization with exposure to light during the day. If necessary, one could consider temporary use of hypnotics that are relatively free of cognitive side effects, such as zolpidem, zopiclone, and some benzodiazepines.[143,144] Still, these substances should be used cautiously because they can cause delirium and worsen neurologic deficits.[145]

Sleep-Related Movement Disorders and Parasomnias

REM sleep behavior disorder (RBD) has been reported to occur after strokes in the tegmentum of the pons.[146-148]

RLS has been observed de novo after stroke.[149-152] In a recent series of 137 patients with stroke, RLS was found mainly after pontine, thalamic, basal ganglia, and corona radiata strokes.[115] Most commonly, RLS was bilateral, appeared within 1 week after stroke, and was accompanied by PLMS in sleep.

After stroke, PLMS can worsen (and even appear de novo) and lead to insomnia (Figure 93-8). PLMS may also occur after unilateral hemispheric and spinal strokes.[105,153]

Hallucinations and Altered Dreams

Patients with strokes in the pons, midbrain, or paramedian thalamus may experience peduncular hallucinosis, characterized by complex, often colorful, dreamlike visual hallucinations, particularly in the evening and at sleep onset (Figure 93-9).[154-157] Peduncular hallucinosis may represent a release of REM sleep mentation. It can be associated with insomnia and usually resolves spontaneously within a few days.

The Charles Bonnet syndrome generally involves less complex visual hallucinations that also occur in the setting of diminished arousal. These hallucinations maybe limited to a hemianopic field and may be a "release phenomenon" in response to the lack of sensory input.[158,159]

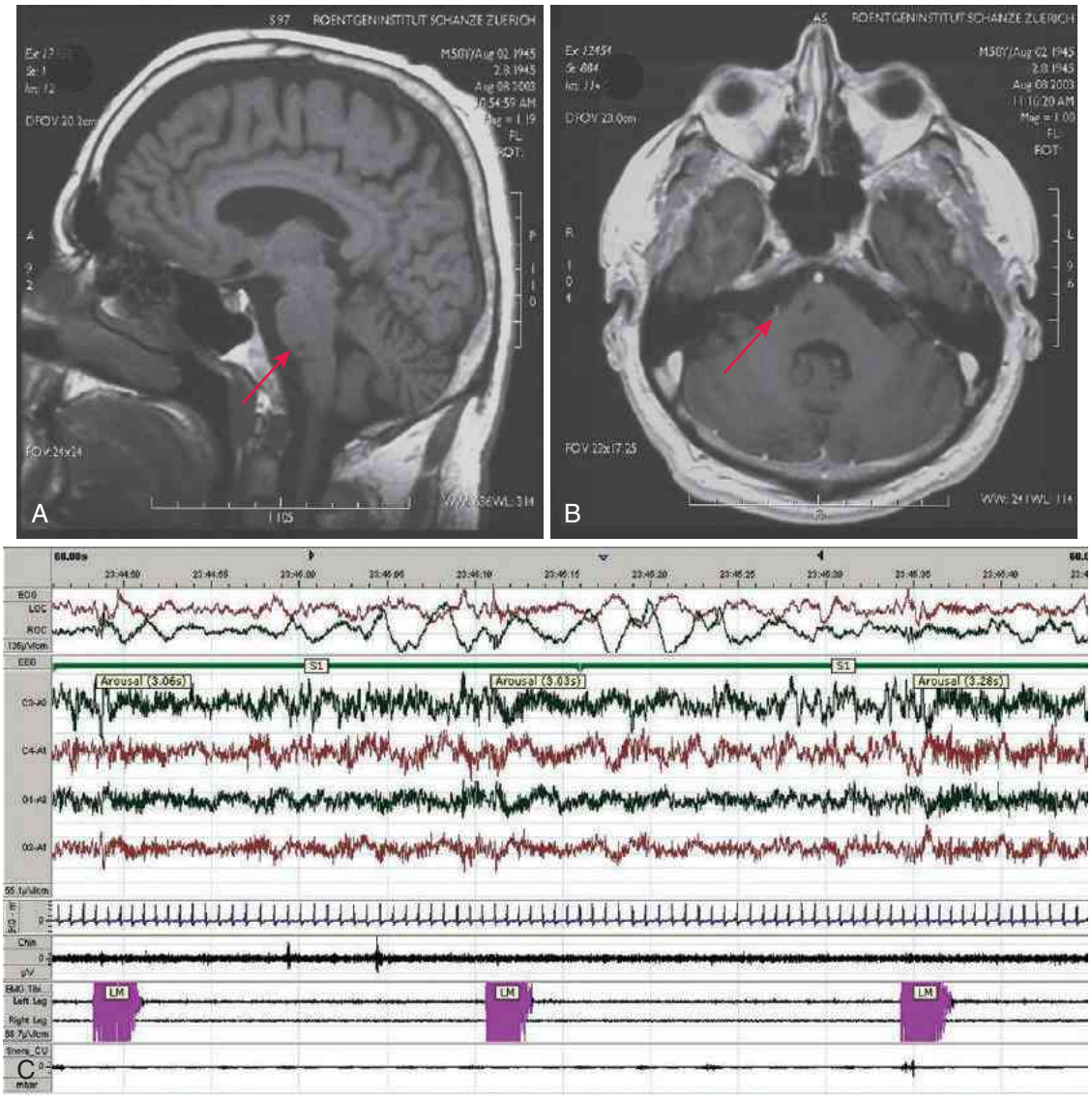


Figure 93-8 Insomnia and left-sided periodic limb movements after right paramedian pontine stroke. This 60-year-old man had a unilateral lacunar stroke in the right paramedian pons **(A** and **B).** He acutely developed severe insomnia, with involuntary, jerky, and tremorlike movements of the left leg and arm appearing intermittently at sleep onset and during sleep (periodic limb movements [LM]) **(C).** The patient denied restless legs symptoms. (MRI pictures courtesy Professor A. Valavanis, Institute of Neuroradiology, University Hospital, Zürich, Switzerland.)

Cessation or reduction of dreaming occurs in the Charcot-Wilbrand syndrome and is occasionally limited to alteration of the visual component of the dream.[160,161] This syndrome can occur in patients with parietooccipital, occipital, or deep frontal strokes, and the lesions are often bilateral.[162-165] Patients frequently show a deficient revisualization (i.e., an impairment to picture again something seen that was previously seen), topographic amnesia, and prosopagnosia. Conversely, REM sleep characteristics may be normal.[164] Severe insomnia and loss of dreaming have been reported with lateral medullary stroke.[142]

Focal (temporal) seizures secondary to stroke can lead to the syndrome of dream-reality confusion or to recurrent nightmares, which may be more frequent with right-sided lesions and can be controlled with antiepileptics.[166]

An increased frequency or vividness of dreaming may occur after stroke, particularly with thalamic, parietal, and occipital strokes.[162]

A few patients with severe motor deficits may report the persistence of seemingly normal motor function within their dreams for up to several years after stroke. Waking up in the morning is a source of great distress in these patients. In other patients, motor handicap may, conversely, be apparent in incorporated in dreams within a few days of stroke onset.

Clinical Significance of Sleep-Wake Disorder after Stroke

Considering the strong evidence that sleep promotes neuroplasticity and the importance of neuroplasticity in recovery after brain damage, it is possible that good quality sleep may

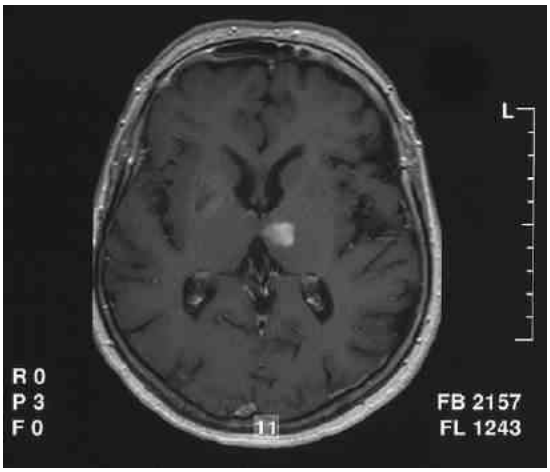


Figure 93-9 Dreamlike hallucinations after unilateral paramedian thalamic stroke. This 62-year-old woman had a left paramedian thalamic stroke and presented with confusion, abulia, anomia, and moderate to severe amnesia in the absence of major sleep-wake disturbances. In the first few days after stroke, the patient had recurrent episodes of visual and acoustic hallucinations in the form of human figures (mostly relatives, partial insight) seen on the right side of the visual field, which the patient described as dreamlike. At 7 months after stroke, the patient had persistent memory problems and reported almost daily episodes of psychic hallucinations ("sensed presence") and a disturbed time perception (Zeitgefühl). On polysomnography, she had no significant changes in rapid eye movement sleep. (MRI pictures courtesy Professor A. Valavanis, Institute of Neuroradiology, University Hospital, Zürich, Switzerland.)

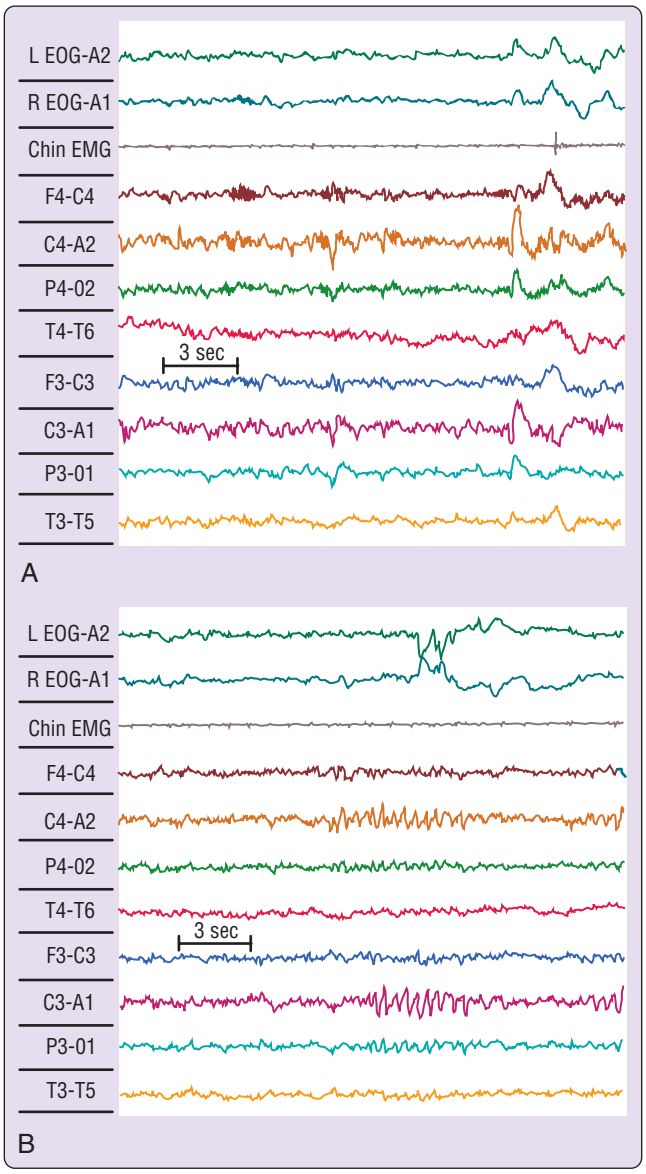


Figure 93-10 Sleep spindles and sawtooth waves after severe middle cerebral artery stroke. This 58-year-old man had a moderately severe left middle cerebral artery stroke (Scandinavian Stroke Scale score, 33/58). Polysomnography 9 days after stroke showed mild obstructive sleep apnea (apnea-hypoxia index = 16). **A,** In NREM sleep, spindling decreased ipsilaterally, with three spindles per hour recorded at C3 and 172 per hour at C4. **B,** In rapid eye movement sleep, sawtooth waves were symmetrical.

improve outcome after stroke. However, the evidence thus far is only indirect.

Several studies have shown that SWD disturbances after stroke are associated with cognitive and psychiatric (depression, anxiety) disturbances and worse outcome.[113,124,167-169] Fatigue at 2 years after stroke has been found to predict institutionalization and mortality.[170] Poststroke RLS also predicts a worse outcome at 3 and 12 months.[171] In humans, sleep between rehabilitation sessions was recently shown to positively influence motor recovery and learning.[172]

In animal models of stroke, sleep deprivation and fragmentation worsened stroke evolution and outcome, whereas sleep-promoting drugs improved functional outcome.[173-176] Sleep deprivation and disturbance in rats worsens functional recovery from stroke and underlying neuroplasticity processes.[177,178] Although the exact underlying molecular mechanisms remain unclear, axonal sprouting and synaptogenesis were significantly inhibited in rats that underwent sleep deprivation during the acute phase of stroke.[178] Conversely, sleep-promoting drugs such as γ-hydroxybutyrate and baclofen were found in both mice and rats to improve functional outcome and neuroplasticity after stroke.[173]

SLEEP ARCHITECTURE CHANGES

Abnormalities in sleep macrostructure and microstructure are common after acute stroke but result only in part from acute brain damage. Changes in sleep architecture depend on (1) health before the stroke (e.g., age, respiratory disturbances); (2) topography of the lesion; (3) complications of stroke (e.g., SDB, fever, infections, cardiovascular disturbances, depression, anxiety); (4) medications; and (5) time after stroke onset. Even acute myocardial infarction patients without brain damage who are admitted to an intensive care unit often have decreases in total sleep time, sleep efficiency, REM sleep, and slow wave sleep.[179]

Some changes in sleep architecture are more specifically related to brain damage, such as persistent alterations of spindling and slow wave sleep with supratentorial strokes and REM sleep abnormalities with infratentorial stroke.

Supratentorial Strokes

Reductions in NREM sleep, total sleep time, and sleep efficiency can follow acute supratentorial stroke.[180-186]

Reduced spindling can be observed with thalamic and cortical-subcortical strokes (see Figure 93-4; Figure 93-10).[168,187-190] With unilateral thalamic strokes, sleep

spindles may be preserved.[124,168,183,191] Rarely, spindling and slow wave sleep increase in the acute stage of large middle cerebral artery stroke.[182,189,192] In some cases, the increase in scored slow wave sleep may reflect an increase in delta activity during both sleep and wakefulness.[185,193]

Transient reductions in REM sleep can occur in the first days after supratentorial stroke.[183,189] Changes in REM sleep may persist after large hemispheric strokes with poor outcome.[182,186] Sawtooth waves can be decreased bilaterally in large hemispheric strokes, especially those that involve the right side.[183,194] Cortical blindness has been associated with a reduction of rapid eye movements.[195]

Paramedian thalamic strokes often reduce spindling and, to a lesser degree, slow wave activity and K-complexes.[124,168,191] These strokes can produce severe hypersomnia with an almost continuous state of light NREM stage 1 sleep, perhaps reflecting an inability to produce full wakefulness.[124] In these patients, REM sleep can occur at night and during the day.[124]

Like the electroencephalogram (EEG) of wakefulness, the sleep EEG reorganizes after acute damage, but data on this subject are scarce.[186,196,197] In patients with paramedian thalamic stroke, recovery from hypersomnia may occur despite the persistence of significant NREM sleep changes.[124,168,198] In hemispheric stroke, conversely, sleep EEG changes usually recover over time, even in patients with severe stroke (>50 mL in volume).[184] Some of these changes may reflect neuroplasticity during functional recovery and therapeutic interventions.[197,199,200] Low sleep efficiency, decreased spindles, K-complexes, slow wave sleep, and REM sleep predict poor outcome when found after hemispheric strokes.

Infratentorial Strokes

Bilateral, paramedian infarcts in the pontine tegmentum or large bilateral infarcts in the ventrotegmental pons can reduce NREM and, especially, REM sleep.[122,201-206] Normal sleep EEG features such as sleep spindles, K-complexes, and vertex waves may be completely lost.[202,207]

Patients with abnormal sleep architecture may complain of insomnia, but isolated REM sleep loss can persist for years without obvious cognitive or behavioral consequences.[204,208]

Bilateral infarction adjacent to the pontine tegmentum, or unilateral infarction of this area, usually does not alter sleep architecture.[122,207,209]

Occasionally, NREM or REM sleep may be altered selectively. Strokes that affect the pontomesencephalic junction tegmentum and the raphe nucleus can moderately decrease NREM sleep with no major changes in REM sleep.[124,139] Infarctions of the paramedian thalamus and of the lower pons have been associated with absence of slow wave sleep but preservation of REM sleep and appearance of REM sleep at sleep onset.[124,203] In contrast, infarction in the lower pons can cause an almost completely selective decrease in REM sleep.[207] Conversely, midbrain strokes can increase REM sleep.[205,210]

CIRCADIAN ASPECTS AND DISTURBANCES

Ischemic stroke, like myocardial infarction and sudden death, occurs most frequently in the morning hours, particularly after awakening, between 6 AM and noon. A meta-analysis of 31 publications reporting the circadian timing of 11,816 strokes found a 49% increase in stroke of all types (ischemic stroke, hemorrhagic stroke, TIA) between 6 AM and noon.[211] Another study reported that thrombotic, lacunar, and embolic strokes are 20% to 30% more frequent in the morning.[212] Possible explanations for this pattern have focused on circadian or postural changes in platelet aggregation, thrombolysis, blood pressure, heart rate, and catecholamine levels that occur with awakening and resumption of physical and mental activities.[213] In addition, the most prolonged REM sleep period, during which autonomic system instability is known to occur,[214-216] occurs close to awakening. The highest incidence in the early hours of the morning can be overestimated because of patients who awaken with stroke. Treatment with aspirin does not modify the circadian pattern of stroke onset.[217,218]

Whereas intracerebral and subarachnoid hemorrhages rarely occur at night, 20% to 40% of ischemic strokes present at night.[219] This suggests that sleep may represent a vulnerable phase for a subset of patients with cerebrovascular disease, and clinicians should consider evaluation for SDB.

Acute brain infarction—particularly when the right hemisphere and the insula are affected—can disturb normal circadian variation in autonomic functions (e.g., heart rate, blood pressure, temperature control) and breathing and contribute to increased poststroke cardiovascular morbidity.[28,220-226] Acute stroke also may alter other circadian functions such as sleep-related secretion of growth hormone and melatonin.[192,225] A loss of physiologic blood pressure dipping is found in about 50% of patients with acute stroke and predicts a less favorable outcome.[226,227] Hyperthermia can occur with diencephalic strokes and is often associated with a poor prognosis.[228]

CLINICAL PEARLS

- Clinicians should consider SDB and sleep-wake disorders as potential risk factors for stroke as well as modulators of its outcome.
- The study of SDB and sleep-wake functions in poststroke patients offers a unique opportunity to expand our knowledge about the brain mechanisms involved in sleep-wake regulation.

SUMMARY

SDB independently increases the risk for stroke. SDB occurs in 50% to 70% of acute stroke and TIA patients. Although obstructive SDB is more commonly observed, mixed SDB and even central SDB (the latter particularly in the first days after stroke) are also more common in stroke patients than in the general population. In the subacute phase of stroke, SDB (particularly central events) can improve. Treatment of SDB is feasible even in the acute phase and can improve short- and long-term stroke outcome. It remains unclear which treatment should be chosen for which patients, with which severity of SBD, and at which interval after stroke.

Sleep loss or insomnia, EDS, hypersomnia, and RLS are potential risk factors for stroke. These sleep-wake disturbances are also observed in 20% to 50% of stroke victims and negatively affect stroke outcome. Brain damage and stroke complications (e.g., pain, mood changes, immobilization, drugs) are pathophysiologically involved. More systematic data are needed on frequency, characteristics, and management of poststroke sleep-wake disturbances and their effects on stroke outcome.

Sleep EEG changes reflect the topography and severity of stroke but also are influenced by neuroplastic changes during recovery from stroke. High-density sleep EEG may become a sensitive tool not only to monitor brain reorganization after stroke but also the effects of treatment interventions.

Animal experiments support the hypothesis that sleep loss and disturbances negatively affect stroke evolution, and conversely, sleep enhancement (e.g., pharmacologic) may positively influence stroke outcome. The molecular mechanisms involved remain to be elucidated.

Selected Readings

Bassetti C, Aldrich M, Chervin R, Quint D. Sleep apnea in the acute phase of TIA and stroke. *Neurology* 1996;**47**:1167–73.

Bassetti C, Mathis J, Gugger M, et al. Hypersomnia following thalamic stroke. *Ann Neurol* 1996;**39**:471–80.

Birkbak J, Clark AJ, Rod NH. The effect of sleep disordered breathing on the outcome of stroke and transient ischemic attack: a systematic review. *J Clin Sleep Med* 2014;**10**:103–8.

Leng Y, Cappuccio FP, Wainwright NW, et al. Sleep duration and risk of fatal and nonfatal stroke. *Neurology* 2015;**84**:1072–9.

Parra O, Arboix A, Bechich S, et al. Time course of sleep-related breathing disorders in first-ever stroke or transient ischemic attack. *Am J Resp Crit Care Med* 2000;**161**:375–80.

Parra O, Sanchez-Armengol A, Capote F, et al. Efficacy of continuous positive airway pressure treatment on 5-year survival in patients with ischaemic stroke and obstructive sleep apnea: a randomized controlled trial. *J Sleep Res* 2015;**24**(1):47–53.

Pearce SC, Stolwyk RJ, New PW, Anderson C. Sleep disturbance and deficits of sustained attention following stroke. *J Clin Exp Neuropsychol* 2016; **38**(1):1–11.

Sarasso S, Santhanam P, Määtta S, et al. Non-fluent aphasia and neural reorganization after speech therapy: insights from human sleep electrophysiology and functional magnetic resonance imaging. *Arch Ital Biol* 2010;**148**:271–8.

Siegsukon CF, Boyd LA. Sleep enhances implicit motor skill learning in individuals poststroke. *Top Stroke Rehabil* 2008;**15**:1–12.

Young T, Finn L, Peppard OE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin Sleep Cohort. *Sleep* 2008;**31**:1071–8.

Zunzunegui C, Gao B, Cam E, et al. Sleep disturbance impairs stroke recovery in the rat. *Sleep* 2011;**34**:1261–9.

A complete reference list can be found online at ExpertConsult.com.

Chapter
94

Sleep and Neuromuscular Diseases

Michelle T. Cao; Christian Guilleminault

Chapter Highlights

- Patients with neuromuscular diseases are at risk for sleep-related problems, including sleep-disordered breathing and sleep hypoventilation, and when most severe, diurnal hypoventilation or respiratory insufficiency.
- Many factors can cause poor sleep quantity and quality in neuromuscular disease.
- The greatest advances in medical treatment of neuromuscular disease patients are in respiratory-related sleep disorders. The use of noninvasive ventilatory support devices has improved morbidity and mortality in this patient group.

Neuromuscular disorders are diseases of the motor unit comprising the lower motor neuron, nerve root, peripheral nerve, myoneural junction, and muscle. Any classification of neuromuscular disease may be somewhat arbitrary, and the astute clinician must keep in mind that the pathologic process may involve several segments of the nervous system and muscle. For example, neuropathies can lead to progressive, peripheral motor and sensory impairment along with autonomic dysfunction. Disorders such as amyotrophic lateral sclerosis (ALS) may progress rapidly toward death, whereas certain chronic polyneuropathies such as Charcot-Marie-Tooth disease, or autonomic syndromes such as familial dysautonomia, may have a slower evolution.

Patients with neuromuscular syndromes are at risk for sleep-related problems. Weakness, rigidity, and spasticity limit movement and posture changes during sleep, leading to discomfort, pain, and disrupted sleep. Difficulty maintaining comfortable positions may lead to cramping, abnormal uncontrolled movements, and weakness, which also contribute to poor sleep. Abnormal sphincter control may induce nocturia, incomplete emptying or incontinence, constipation, or painful defecation.

Sleep-related changes in respiration put the patient with a neuromuscular disorder at risk by impairing ventilation. Chronic respiratory muscle failure usually develops over years. It may initially present with sleep-disordered breathing (SDB), followed by progression to nocturnal hypoventilation, then diurnal hypoventilation, cor pulmonale, and eventual respiratory failure and end-stage disease. The slow progression of ventilatory failure in some disorders may go undetected for some time and contribute to increased mortality.

Limited attention is paid to the impact of sleep-related issues in this population, particularly because most clinics see a limited number of patients with neuromuscular disorders. Even in specialized neuromuscular clinics, a minority of patients are asked about their sleep problems or have been given a prior sleep evaluation.[1] Moreover, sleep specialists rarely manage the common problems in this population, such as spasticity, sphincter dysfunction, pain, abnormal movement, confusional arousal that can result in sleep fragmentation, insomnia, parasomnias, daytime fatigue, and hypersomnolence. Thus a multidisciplinary approach to treatment should be the standard of care in patients with neuromuscular disorders.

EPIDEMIOLOGY AND GENETICS

Each neuromuscular syndrome has distinct epidemiology and etiology. For example, ALS affects 0.005% of the U.S. population, and multiple sclerosis, a neurodegenerative disorder, affects 0.11%. There are no cumulative prevalence data that include all neuromuscular disorders. Many neurologic disorders, such as maltase deficiency, myopathy, myotonic dystrophy (MD), Rett syndrome, and familial dysautonomia, have a clear genetic origin. Other neuromuscular disorders can be secondary to traumatic, infectious, vascular, malignant, or degenerative diseases.

Although much research has addressed abnormal sleep and breathing in patients with neuromuscular diseases,[2-5] there are few large studies that examine the prevalence of SDB in these patients. One study from New Mexico[1] attempted to gather information from its entire neuromuscular clinic population. Although complete data were available for only 60 patients (20% of the clinic population), the investigators demonstrated that sleep and breathing abnormalities were present in more than 40% of patients.[1] Such a high prevalence is not surprising given the vulnerability of such patients to sleep-related reductions in muscle tone and overall ventilation. Patients with spinal cord injury (SCI) have been better studied. Compared with the normal population, individuals with SCI have significantly greater difficulties falling asleep and subjectively poor sleep; they also frequently require sleeping pill prescriptions, sleep more hours, take more and longer naps, and snore more.[6]

PATHOPHYSIOLOGY

The diaphragm is the major muscle of respiration during wakefulness and sleep. During non–rapid eye movement (NREM) sleep, there is an overall reduction in ventilation due

to altered chemosensation and increased impedance of the respiratory system. However, rib cage activity is maintained (albeit reduced), as is diaphragmatic activity. The importance of the diaphragm is particularly evident during rapid eye movement (REM) sleep. During REM sleep, there is postsynaptic inhibition of somatic motor neurons, causing further reduction or complete loss of tone in the intercostals and other muscles of respiration, but the diaphragm is relatively unaffected. Any process affecting the diaphragm, whether a myopathy or a process involving its innervation, can significantly reduce ventilation and oxygenation during REM sleep. In patients with bilateral diaphragmatic paralysis who are dependent on other respiratory muscles for breathing, marked oxygen desaturations can occur during REM sleep.[7,8] The REM sleep-related inhibition of intercostal and accessory muscles leads to profound hypoventilation during this sleep stage. As noted previously, the suppression of accessory respiratory muscle tone is a normal process of REM sleep and is seen in normal subjects.[9-11]

In some genetic neuromuscular disorders, muscle weakness can begin early in development, interfering with normal development of the skull and facial bones. For example, orofacial muscle weakness can affect growth of the maxilla and mandible, resulting in in the "long face" seen in congenital MD.[12] In rhesus monkeys, experimental reduction of nostril size by ligature with consequent impairment of nasal breathing leads to abnormal facial muscle contraction and secondary abnormal orofacial bone growth.[13,14] Similarly in humans, impaired nasal breathing leads to abnormal masseter contractions, consequently limiting orofacial growth.[15] Narrowing of dental arches, reduction in maxillary arch length, anterior crossbite, maxillary overjet, and overall narrowing of the facial skeleton are a result of changes in muscle contractions. These facial skeletal changes decrease upper airway diameter, consequently increasing its collapsibility during sleep.

Depending on the type of neuromuscular disorder, sleep-related breathing abnormalities may present as central apneas, obstructive apneas, nasal airflow limitation, or periods of prolonged hypoventilation, or there may be a combination that renders the neuromuscular patient at times difficult to treat. Sleep disruption with frequent cortical arousals may be due to discomfort with certain positions, muscle spasm, difficulty in clearing secretions, sphincter control, or increase in upper airway resistance from muscle weakness and secondary craniofacial changes. Periods of hypoventilation can contribute to arousals, reduced sleep time, and sleep deprivation from ventilatory and arousal responses to changes in oxygen and carbon dioxide (CO_2) levels. Although these changes may protect ventilation in the short term, over time ventilatory responses to changes in oxygen and CO_2 levels become blunted. This blunting leads to further worsening of hypoventilation, eventually occurring during both wakefulness and sleep.

CLINICAL FEATURES COMMON TO MOST NEUROMUSCULAR DISORDERS

Nonspecific complaints such as increased fatigue, daytime hypersomnolence, or disrupted sleep can be the initial manifestations of a slowly evolving neuromuscular disease of adult onset.[1] Such nonspecific complaints may also be the sole indication of a progressive neuromuscular disorder. Problems with clearance such as managing saliva or gastric contents can lead to significant drooling, esophageal reflux, or pulmonary infections from aspiration or retained secretions. Impairment of cough mechanisms may further impair the ability to clear secretions.

Autonomic dysfunction may be present in the form of abnormal sensitivity to temperature or pressure, with discomfort related to the use of sheets and blankets. The disease may psychologically affect individuals, leading to anxiety, depression, and insomnia, as can be seen with many other chronic illnesses. Pharmacologic agents that are prescribed in the evening may have alerting effects, whereas others used in the morning may lead to daytime sleepiness. In all, patients with chronic neuromuscular disorders can have many factors disrupting sleep, consequently worsening daytime function and quality of life. The addition of sleep-related problems complicates their already existing neurologic issues.

Neurodegenerative Diseases Involving the Motor Neuron

ALS is a degenerative motor neuron disease involving upper and lower motor neurons, leading to muscle weakness and atrophy throughout the body. Although ALS has not been shown to directly affect the sleep-regulating areas of the brain, it is likely that indirect effects of the disease contribute to sleep disruption.[16-18] Periodic limb movements associated with arousals and SDB may cause sleep disruption in ALS patients. SDB is reported in 17% to 76% of patients with ALS.[19] ALS patients with normal respiratory function, normal phrenic motor responses, and preserved motor units on electromyography may still have SDB with periodic oxygen desaturations independent of sleep stage (REM or NREM).[20] However, respiratory-related sleep disruption is generally not significant until phrenic motor neurons are involved and the diaphragm becomes weak. When there is involvement of the diaphragm, severe hypoventilation and hypoxemia occur during REM sleep, and nearly all these patients will need some form of ventilatory support. Some ALS patients without any respiratory disturbance or periodic limb movements still experience sleep fragmentation, independent of age. This suggests that other factors contribute to disturbed sleep, such as anxiety, depression, pain, choking, excessive secretions, fasciculation, cramps, and the inability to find a comfortable position or turn oneself freely in bed. Orthopnea, a frequent complaint in ALS, may also contribute to sleep disruption.[17,18]

Spinal Cord Disease

Poliovirus infection targets the nervous system in several ways by injuring cranial motor nuclei and spinal cord anterior horn cells, resulting in acute paresis. As a result, there are many effects on respiration. Abnormalities in central regulation of breathing in patients with acute and convalescent poliomyelitis were described in 1958 by Plum and Swanson.[21] Subsequently, central, mixed, and obstructive apneas have been noted.[22] Sleep and breathing abnormalities are seen not only in patients who are on respiratory assistance (e.g., rocking beds) during sleep but also before ventilatory assistance is initiated.[23] Sleep abnormalities include decreased sleep efficiency, increased arousal frequency, and varying degrees of apnea and hypopnea. After treatment of sleep and breathing abnormalities, many symptoms frequently attributed to the postpolio syndrome do improve. Although not all symptoms

can be explained,[24] excessive daytime sleepiness and fatigue can be explained by poor sleep quality related to abnormal respiration during sleep.[25]

Poliomyelitis can alter central and peripheral respiratory function decades after the acute infection, an important element of postpolio syndrome.[24] Muscle atrophy and immobility can lead to kyphoscoliosis and restricted ventilation. The anatomic deformities resulting from poliomyelitis may cause chronic pain and consequent sleep abnormalities. Also, bulbar involvement may affect upper airway muscles. SDB is reported in 31% of patients with postpolio syndrome.[24] Prolongation of REM sleep latency may result from prolonged recruitment time related to damaged neurons in the pontine tegmentum.[26]

Inherited metabolic diseases such as subacute necrotizing encephalomyelopathy (Leigh disease) typically appear in childhood and may be associated with respiratory disturbance. Rarely, this disease may appear in adulthood with respiratory failure during sleep.[27] Syringomyelia can be associated with central, mixed, and obstructive apneas when it involves the bulbar and high cervical neurons.[28] Malformations of the skull base or high cervical junction (platybasia, Chiari malformations) may cause both central and obstructive types of sleep apnea.[29]

SCI has dramatically increased in frequency over the past 30 years owing to an increase in traffic accidents and military conflicts. Incidence rates are highest in the second to fourth decades of life, and with improvements in long-term supportive care providing longer life expectancy,[30] the prevalence of SCI is destined to grow in the coming years. Overall, morbidity and mortality are higher with cervical and high thoracic spinal cord lesions, especially in ventilator-dependent individuals.[31-33] In a Stockholm, Sweden, epidemiologic survey,[6] muscle spasm, pain, paresthesias, and voiding problems were reported as the most important causes of sleep disturbance.

Gastrointestinal problems related to autonomic dysfunction are more often considered secondary problems. Gastrointestinal motility is also affected, and reflux is common.[32,33] In tetraplegic patients, the higher the spinal cord lesion, the more significant the impairment, not only with diaphragm, intercostal, and abdominal muscle weakness but also with impaired cough and other reflexes for laryngeal and lung clearance.[32,33] All these elements together can impair breathing, especially during sleep.[33,34] High cord lesions can interrupt pathways to the superior cervical sympathetic ganglion that regulate melatonin secretion.[35,36]

Pharmacologic agents used by patients with SCI (antispasmodic, analgesics, drugs for autonomic dysfunction, psychoactive substances) can also disrupt sleep and wakefulness. An important feature seen in cervical lesions is progressive ventilatory impairment during sleep noted between the 15th day and the 13th week after the injury, often after the patient has been released from an acute care setting. Such worsening may lead to a higher percentage of deaths during sleep, as reported in a cohort of patients with mid to lower cervical SCI during this time period.[37]

Polyneuropathies

The most common polyneuropathy associated with SDB is Charcot-Marie-Tooth disease.[38] It is characterized by chronic degeneration of peripheral nerves and roots, resulting in distal muscle atrophy that begins at the feet and legs and later involves the hands. SDB can occur in these patients as result of a pharyngeal neuropathy leading to upper airway obstruction (obstructive apnea, upper airway resistance syndrome)[39] or with diaphragmatic dysfunction.[40] Autonomic neuropathy, particularly when secondary to type 1 diabetes, may be associated with impaired chemosensitivity to CO_2, although the effects on sleep and breathing are inconsistent.[41]

Neuromuscular Junction Diseases

Myasthenia gravis (MG) is a disorder of the neuromuscular junction characterized by weakness and fatigability of skeletal muscles. Sleep breathing abnormalities can occur as a result of diaphragmatic or pharyngeal weakness. Risk factors for the development of sleep-related ventilatory problems in patients with MG include age, restrictive pulmonary syndrome, diaphragmatic weakness, and daytime hypoventilation.[42] Younger patients with a shorter duration of illness are less likely to experience sleep-related hypoventilation or hypoxemia,[43] whereas older patients with increased body mass index, abnormal total lung capacity, and abnormal daytime blood gases are more likely to develop hypopneas or apneas, particularly during REM sleep.[44] Sleep apnea was diagnosed in 60% of patients with MG even when the disease was in a clinically stable stage.[45,46] A prospective study by Nicolle and colleagues found that obstructive sleep apnea (OSA) was the predominant abnormality, occurring in 36% of MG patients, with significant associations with older age, male gender, elevated body mass index, and corticosteroid use.[47]

Other neuromuscular disorders that can disturb sleep include congenital myasthenic syndromes,[48] botulism, hypermagnesemia, and tick paralysis. Taking a careful history is extremely helpful in making the diagnosis in these circumstances. Dyspnea that worsens with activity, morning headache, paroxysmal nocturnal dyspnea, fragmented sleep, and daytime somnolence are among the symptoms that suggest SDB in these syndromes.

Muscle Diseases

Myotonic Dystrophy

MD is an autosomal dominant disorder that causes myotonia, muscle weakness, and daytime sleepiness. In this illness, there is consistent involvement of facial, masseter, levator palpebrae, sternocleidomastoid, forearm, hand, and pretibial muscles; MD is, in a sense, a distal myopathy. However, pharyngeal and laryngeal muscles may also be involved, as well as respiratory muscles, in particular the diaphragm.

Central nervous system abnormalities also occur in MD, causing excessive daytime sleepiness by different mechanisms.[12,48-51] For example, neurodegeneration in the dorsomedial nuclei of the thalamus can lead to a medial thalamic syndrome characterized by apathy, memory loss, and mental deterioration. Loss of 5-hydroxytryptamine (serotonin) neurons of the dorsal raphe nucleus and the superior central nucleus,[52] as well as dysfunction of the hypothalamic hypocretin-orexin system,[53] can result in short sleep latencies and sleep-onset REM periods on the Multiple Sleep Latency Test.[49,50] Excessive daytime sleepiness occurs in 33% to 77% of MD patients.[54]

Involvement of the respiratory muscles may result in SDB, including alveolar hypoventilation predominantly in REM sleep,[55-58] obstructive apneas,[59] and central apneas.[60] However, the development of SDB abnormalities in MD is not simply

caused by muscle weakness. When patients with MD were compared with patients with nonmyotonic respiratory muscle weakness, periods of hypoventilation and apneas (central and obstructive) occurred at higher frequencies in those with MD than in nonmyotonic patients who had the same degree of muscle weakness (measured by maximal inspiratory and expiratory pressures).[61] This finding suggests that changes in the central nervous system control of respiration contribute to abnormal breathing in patients with MD.

Similarly, decreased ventilatory responses to hypoxia and hypercapnia[62-64] and extreme sensitivity to sedative drugs suggest a central origin of the breathing impairments in MD. The differential diagnosis requires further testing. A standard technique for assessing control of respiration is to study the increase in ventilation as a response to increased arterial CO_2. However, when respiratory muscles are weak, as in MD, it may be difficult to interpret a reduced ventilatory response. That is, chemoreceptor activity and efferent signaling to respiratory muscles may be intact, but weak or inefficient respiratory muscles may not produce a normal ventilatory response to a hypercapnic or hypoxic stimulus. Another method of assessing impairment of respiratory center output is measurement of mouth pressure developed at the beginning of a transiently occluded breath (occlusion pressure, $P_{0.1}$).[65] In patients with MD, $P_{0.1}$ may be as high or higher than that of control subjects at rest and during stimulated breathing, although overall ventilation is lower.[63,66] The finding of a high transdiaphragmatic pressure (P_{di}) despite overall lower ventilation suggests that increased impedance of the respiratory system accounts for incomplete transformation into ventilation of normal or increased respiratory center output.

Magnetic stimulation of the cortex, in conjunction with phrenic nerve recordings, can also be used to test the corticospinal tract to phrenic motor neuron pathways and is a reliable method for diagnosing and monitoring patients with impaired central respiratory drive.[67] The use of transcortical and cervical magnetic stimulation demonstrates that greater than 20% of MD patients have impaired central respiratory drive.[68] The finding of neuronal loss in the dorsal central, ventral central, and subtrigeminal medullary nuclei in MD patients with alveolar hypoventilation[69] and the severe neuronal loss and gliosis in the tegmentum of the brainstem[70] also support a central abnormality.

Another problem in MD patients is orofacial growth impairment early in life. Craniofacial muscle weakness can negatively affect bone growth during development, particularly on orofacial muscles that are involved with stimulation of particular growth areas such as the intermaxillary synchondrosis that usually becomes inactive near 15 years of age. As a result of such muscular weakness, development of craniofacial structures in patients with MD is impaired. They experience more vertical facial growth than normal subjects and have more narrowed maxillary arches and narrow palate as measured between the palatal shelves, with deeper depths. These craniofacial changes may contribute to the development of obstructive sleep apnea owing to a smaller maxilla and mandible, consequently restricting the size of the upper airway and leading to upper airway collapse during sleep.

Other Myopathies

Abnormalities in sleep and breathing have been reported in isolated series of patients with various neuromuscular disorders, such as congenital myopathies (nemaline or congenital fiber-type disproportion myopathy)[71-73] or metabolic myopathies (mitochondrial myopathy such as Kearns-Sayre syndrome[74-77] and acid maltase deficiency[78,79]). In all these cases, there are various alterations in respiratory control and breathing pattern changes, including hypoventilation, obstructive apneas, and central apneas. All genetic myopathies with orofacial weakness have similar risks for impaired bone growth, particularly on the maxilla and mandible, as in MD.[13-15] Severe central sleep apnea and marked hypoxemia, particularly during REM sleep, resulting in hypoxia-induced pulmonary hypertension, excessive daytime sleepiness, heart failure, morning headaches, and rare nocturnal seizures, may be seen in patients with congenital muscular dystrophy.[80] OSA has also been described in Thomsen disease (myotonia congenita).[81]

Myopathies such as Duchenne muscular dystrophy (DMD) can cause restrictive lung disease and chest wall deformities.[82,83] These changes also contribute to ventilatory impairment, fragmented sleep, hypercapnia and hypoxemia (more profound during REM sleep),[84,85] development of deformities, chronic pain, and discomfort. There is a bimodal presentation of SDB in children with DMD, whereby OSA is more common in younger children in the first decade of life.[86,87] In younger children with DMD, OSA can improve with adenotonsillectomy, whereas in older children who have already developed hypoventilation, OSA is better managed with noninvasive ventilation.

Acid maltase deficiency myopathy can cause SDB, with rapid and significant diaphragmatic impairment noted long before the weakness of other skeletal muscles.[88] In fact, SDB and the secondary daytime fatigue may be presenting symptoms of the myopathy.[88]

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant disease and the third most frequent form of muscular dystrophy, after DMD and MD. Della Marca and colleagues evaluated FSHD patients and found that impaired sleep quality was directly correlated to the severity of the disease.[89] Among 46 FSHD patients, 27 had snoring and 12 reported respiratory pauses during sleep.[89]

DIAGNOSTIC EVALUATION

The evaluation should consider the type of neurologic disorder, the degree of sensory and motor impairment and the resulting disability, the associated autonomic defects, and the impact of the illness on the patient's mood. Understanding the patient's interaction with society and family is an important factor for subsequent treatment decisions. A detailed sleep history is required to outline the severity and type of sleep-related problems. General assessment should also determine the degree of pain and discomfort (particularly in the supine position and during sleep), the presence or absence of sphincter problems and urinary or digestive dysfunction during wake and sleep, and any evidence of autonomic dysfunction already present during wakefulness and suspected during sleep. A number of additional diagnostic tests may supplement the evaluation of sleep in the patient with neuromuscular disease. These include a disability index scale,[1] a sleep disorder questionnaire, and a sleep log or actigraphy (helpful for the investigation of daily rhythms and sleep-wake disturbances during the 24-hour period). The severe respiratory insufficiency questionnaire, a

multidimensional health-related quality-of-life instrument, may be used for patients with neuromuscular disorders on assisted ventilation.[90]

Clinicians should carefully look for craniofacial abnormalities, including high and narrow hard palate, teeth crowding, tongue indentations, and Mallampati or Friedman rating scales[90,91] evaluating the size of upper airway. Routine measures of pulmonary function (spirometry, lung volumes, diffusing capacity) and gas exchange (Pao_2 and $Paco_2$) should be performed in all patients at initial presentation. Static lung volume measurements, both upright and after 15 minutes in supine position, often demonstrate significant changes caused by respiratory muscle weakness, particularly diaphragmatic weakness. A forced expiratory volume in 1 second (FEV_1) or forced vital capacity (FVC) less than 40% of predicted, a $Paco_2$ greater than 45 mm Hg, and a base excess of 4 mmol/L or greater may indicate increased risk for sleep-related hypoventilation, and overnight polysomnography should be performed. Supine inspiratory vital capacity measurements of less than 40%, 25%, and 12% will likely result in hypoventilation during REM sleep, full night, and daytime, respectively.[83,92,93]

Overnight polysomnography is the key to a definitive evaluation of sleep and breathing in these patients. Although it can be done in many settings, including at home, in-laboratory evaluation allows additional measures such as video monitoring. More important, measurement of transcutaneous or end-tidal CO_2 allows continuous tracking of overall ventilation during sleep and can help guide the decision for nocturnal ventilatory assistance.

TREATMENT OF SLEEP ABNORMALITIES IN PATIENTS WITH NEUROMUSCULAR DISEASE

The greatest advances in the medical treatment of neuromuscular disorders have been for sleep-related abnormalities.[51] The goal is restoration of normal sleep architecture, with subsequent improvement of sleep, daytime function, and quality of life. Simple measures such as bedding are often overlooked. Specialized beds and mattresses are available with specifications allowing ease of positional changes, avoidance of skin lesions at pressure points, and segmental inflation or deflation (e.g., air mattresses), thus improving autonomic dysfunction, cramps, spastic contraction, and rigidity. Great efforts should be made to diminish pain and discomfort of any type. Treatment of abnormal behavior and confusional arousals may necessitate use of sedatives such as benzodiazepines, but such therapy should be considered only after careful evaluation of ventilatory function and risk for worsening the sleep-related abnormal breathing.

Judicious use of wake-promoting drugs, such as modafinil or armodafinil, can improve daytime alertness without nocturnal sleep disruption. Prior reports of patients with MD and ALS have shown beneficial effects of modafinil in improving daytime fatigue.[94-104] Baclofen can reduce muscle spasms and help nocturnal sleep but may worsen daytime sleepiness. Treatment of pain with opioids can complicate treatment of SDB by inducing sleep hypoventilation or central sleep apnea. Treatment of abnormal breathing during sleep should be based on polysomnographic findings and should be adjusted with regular follow-up considering clinical symptoms and polysomnographic studies. Various therapies may improve nocturnal hypoventilation or offset the attendant oxygen desaturation.

Supplemental oxygen has been used to alleviate the REM sleep–related oxygen desaturation in patients with DMD but does not clearly improve sleep.[105] Repeated nocturnal hypoxia may worsen muscle weakness, which begets further oxygen desaturation, and reversal of the hypoxemia may arrest the muscle weakness. In one patient with acid maltase deficiency treated with nocturnal oxygen, hypoxemia and muscle weakness did not progress over an 8-year period.[106] Because most of the hypoventilation occurs during REM sleep, pharmacologic suppression of REM sleep with a tricyclic antidepressant is a theoretical option. In a small study of patients with DMD, protriptyline markedly improved the nocturnal oxygen saturation profile.[107,108] However, anticholinergic side effects limit the widespread use of such therapy. Inspiratory muscle training has improved waking respiration in one patient with acid maltase deficiency,[88] with major improvement in the nocturnal oxygen saturation. In general, muscle training is often helpful in neuromuscular patients.[109]

In children, two syndromes have a high prevalence of sleep hypoventilation: DMD and spinal muscular atrophy. Both conditions can cause progressive hypoventilation during sleep and, as the disease progresses, hypoventilation during wakefulness. In this young age group, the appropriate time to begin airway clearance and to introduce noninvasive ventilatory support that can preserve or enhance lung growth and chest wall mobility must be carefully assessed. The presence of an imbalance between mechanical load and the capacity of the respiratory muscles must be evaluated because fatigue may occur, leading to respiratory failure. Inspiratory muscle training can significantly improve respiratory parameters after 1 month of training.[110-112] Children with impaired orofacial bone development as a consequence of abnormal muscle contractions secondary to the genetically induced generalized muscle dystrophy may benefit from myofunctional reeducation (i.e., orofacial muscle exercises aiming at improving suction, mastication, swallowing, and nasal breathing).[113-115]

Noninvasive Positive Airway Pressure

Mechanical ventilation has been a mainstay in supporting ventilation since the days of the poliomyelitis epidemic. Rocking beds, negative-pressure tank ventilators, positive-pressure ventilation through tracheostomy, and cuirass ventilation have been long-term options in the past.[116] However, all these options are cumbersome, severely limit the mobility of patients, and in the case of tracheostomy, may have unwanted complications. Therefore other forms of assisted ventilation have been developed, including phrenic nerve pacing,[117] and noninvasive positive-pressure ventilation devices.[118,119]

Positive airway pressure devices, including continuous positive airway pressure, bilevel positive airway pressure with or without backup respiratory rate (a noninvasive positive-pressure ventilation device), and the advanced devices targeting minute ventilation by way of pressure support have been used to treat hypoventilation of various causes. Treatment of hypoventilation requires adequate delivery of tidal volume and maintaining minute ventilation to effectively eliminate CO_2, therefore illustrating the limitations of continuous positive airway pressure in treating hypoventilation. Because bilevel positive airway pressure acts as a noninvasive ventilator and supports ventilation, it also treats CO_2 retention, which is

commonly seen in patients with advanced neuromuscular disorders. Low-flow oxygen can be bled into the nasal mask during nocturnal sleep to maintain adequate oxygenation if needed.

In the past two decades, noninvasive positive-pressure ventilation has improved the natural course of neuromuscular disorders, and it is often the treatment of choice. Noninvasive positive-pressure ventilation has been shown to improve quality of life and increase survival in neuromuscular disorders.[120-128] In postpolio syndrome, the median time for prolongation of life expectancy is more than 20 years. In patients with spinal muscular dystrophy types 2 and 3, DMD, and acid maltase deficiency, the median improvement in life expectancy is 10 years. In MD, the median improvement in life expectancy is 4 years, and in ALS it is 1 year.[120,121,124] Compared with ventilation through tracheostomy, noninvasive positive-pressure ventilation through nasal interfaces and portable ventilators is becoming the preferred means of assisting ventilation because it is much simpler to administer, is more comfortable, and reduces costs.

Nocturnal noninvasive positive-pressure ventilation had been used predominantly for patients with postpoliomyelitis and other neuromuscular disorders. With eradication of poliomyelitis in most of the world, ALS has become the most common neuromuscular disorder for which noninvasive positive-pressure ventilation is used. In ALS patients with orthopnea, maximum inspiratory pressure less than 60% of predicted, or symptomatic daytime hypercapnia, noninvasive positive-pressure ventilation significantly improved quality of life, sleep-related symptoms, and survival in those without severe bulbar dysfunction.[120] These improvements were greater than those achievable with any currently available pharmacotherapy. Therefore a trial of noninvasive positive-pressure ventilation in ALS patients is warranted even in those with severe bulbar dysfunction for palliative reasons.

Bilevel positive airway pressure therapy is an effective treatment for a number of neuromuscular diseases and in early stages of the disease may be as effective as an invasive conventional ventilator. However, patients may need better control of their hypoventilation during sleep and require more ventilatory support, particularly with progressive muscle weakness. Advanced noninvasive positive-pressure ventilation devices are available specifically for treatment of hypoventilation during sleep by targeting tidal volume or minute ventilation (e.g., average volume assured pressure support [AVAPS] and intelligent volume assured pressure support [iVAPS]).[129] Because the AVAPS and iVAPS devices adjust pressure support based on the patient's respiratory cycle breath by breath, they adapt to changes in severity of disease and therefore are ideal for sleep hypoventilation or respiratory insufficiency. Adaptive servoventilation, an anticyclic positive airway pressure device, has been considered when opioid intake complicates the clinical presentation because of a high number of central apneas, provided that hypoventilation is not significant.[130,131]

The choice of settings when choosing noninvasive positive-pressure ventilation can influence sleep architecture and quality in patients with various neuromuscular diseases. Tailoring the settings (options available depending on portable ventilator used) to the individual's respiratory effort rather than the usual or default parameters is associated with better nighttime gas exchange, percentage of REM sleep, and sleep quality.[132] The AVAPS mode, for example, requires predetermined tidal volume that is calculated from the patient's ideal body weight, with a recommended 6- to 8-mL/kg of ideal body weight used instead of actual body weight. However, if rib cage and abdominal muscles are weak, the recommended volume setting for the patient may be too high, and pain may develop during chest expansion; in these cases we recommend decreasing the predetermined tidal volume to 6 to 7 mL/kg.

Another parameter that is an important component in the patient's comfort and compliance is the "rise time" (i.e., the speed at which airflow is delivered from expiration to inspiration in 10ths of a second). It should be adjusted based on the severity of thoracic muscle weakness, lung inflation capabilities, and amount of secretion accumulated in the airway. This adjustment may be critical because the patient may not tolerate noninvasive positive-pressure ventilation if the rise time is inadequate for the condition.

In endotracheally intubated patients or those with a tracheostomy, the differential between inspiratory and expiratory positive airway pressure on the ventilator can be wide. However, unlike an endotracheal or tracheostomy tube, the upper airway is not rigid, and as a result there is variability in airway dilator contraction during inspiration due to multiple factors: the degree of local muscle impairment, sleep stage, sleep state, neck position, and narrowing because of anatomic factors (deviated septum, enlargement of nasal turbinates, presence of adenoids and tonsils). If the bilevel "pressure differential" becomes too wide, in the presence of nonlinear flow (turbulence) there will be a greater tendency for upper airway obstruction that translates into airflow limitation (visible when studying nasal flow during polysomnography).[133] We generally keep a differential of 6 cm H_2O to avoid induction of variable abnormal upper airway resistance. Besides determination of appropriate inspiratory and expiratory pressures during overnight polysomnography, the need for a backup respiratory rate may also be assessed if central sleep apnea complicates the picture. The backup rate is commonly set at about 8 to 10 breaths/min but will need adjustment over time based on the severity and evolution of the syndrome. In all, the use of noninvasive positive-pressure ventilation allows patients to return to work and even travel, something previously impossible when constrained by reliance on a rocking bed or the complications of tracheostomy for nocturnal ventilatory support.

Decision to Assist Nocturnal Ventilation

When patients present with disrupted sleep, snoring, excessive daytime sleepiness, and unexplained development of peripheral edema or polycythemia, sleep studies can help characterize a breathing disorder, and the decision to assist with ventilation is generally straightforward, with addition of low-flow oxygen bled into mask if needed. For a detailed review of noninvasive ventilation, see Chapter 115.

Nocturnal noninvasive positive-pressure ventilation should be started when nocturnal hypoventilation is present. Clinical symptoms and physiologic markers of hypoventilation assess disease severity and assist in the decision to initiate nocturnal noninvasive positive-pressure ventilation. Patients usually first develop nocturnal hypoventilation followed by diurnal hypoventilation with associated clinical symptoms, which can lead to acute respiratory failure. Continuous monitoring of arterial CO_2 by end-tidal CO_2 or transcutaneous CO_2 during an overnight sleep study is necessary to document

nocturnal hypoventilation, which may occur exclusively during REM sleep. Arterial blood gas and serum chemistry can document daytime hypoventilation with elevated arterial CO_2 ($Paco_2$), low arterial oxygen (Pao_2), relatively normal pH, and high serum bicarbonate. Many clinicians consider starting noninvasive positive-pressure ventilation with an arterial Pco_2 greater than 45 mm Hg and an arterial Po_2 less than 70 mm Hg. An isolated change in nocturnal oxygen saturation alone is insufficient for deciding whether the patient needs ventilatory assistance. However, sustained nocturnal oxygen desaturations may be an indicator of nocturnal hypoventilation.

In summary, there is long-standing consensus[134-136] on the management of severe progressive neuromuscular disorders in which respiratory failure plays a significant part of the natural history of the disease. The positive impact of noninvasive ventilatory support in neuromuscular disease patients has become clear in the past two decades. The most effective time to introduce noninvasive positive-pressure ventilation is when SDB develops, including nocturnal hypoventilation. Issues such as quality of life must be taken into account. Each patient must be assessed in detail, and the clinician must bear in mind that nocturnal (and later, 24-hour) ventilation will treat only one (albeit important) aspect of the disorder.

CLINICAL PEARLS

- Sleep is a vulnerable state for patients with neuromuscular disorders because normal REM sleep–related changes in ventilation are magnified as a result of muscle weakness, resulting in hypoventilation and oxygen desaturation.
- In addition to SDB, sleep may also be disturbed by spasticity, poor secretion clearance, sphincter dysfunction, inability to turn, pain, and autonomic dysfunction. All of these factors can impair sleep and worsen daytime disability.
- Noninvasive ventilation can aid neuromuscular disease patients by reducing morbidity and improving life expectancy.
- The most effective time to introduce noninvasive ventilatory support is when SDB develops, including nocturnal hypoventilation.
- New advanced noninvasive ventilatory equipment, such as AVAPS or iVAPS, is ideal for ventilatory insufficiency in neuromuscular patients because it targets tidal volume or minute ventilation and therefore treats hypoventilation and CO_2 retention.

SUMMARY

Neuromuscular disorders consist of central and peripheral neurologic disorders with impairment of the motor system. The disability of patients with a neuromuscular disorder worsens during sleep, and the abnormal sleep and secondary impairment of daytime function further degrade quality of life. Nocturnal sleep disruption can result from pain and discomfort related to weakness, rigidity, or spasticity that limits movement and posture. Sleep disruption may also be caused by autonomic dysfunction, poor sphincter control, problems with clearance of secretions, and abnormal movements and behaviors during sleep. Most important, sleep-related hypoventilation is common with neuromuscular disorders, and overlooking this may lead to death. Daytime evaluation will determine the severity of the disability but may not identify the presence and severity of an associated sleep-related disorder. Nonspecific symptoms of daytime fatigue and sleepiness can indicate poor sleep in these patients. Polysomnography with continuous monitoring of CO_2 is the only test that can objectively identify and evaluate the severity of sleep-related disorders as well as ventilatory impairment. By recognizing and treating sleep-related problems, these patients can enjoy improved survival and better quality of life.

Selected Readings

Aboussouan LS. Sleep-disordered breathing in neuromuscular disease. *Am J Respir Crit Care Med* 2015;**191**:979–89.

Alves RSC, Resende MBD, Skomro RP, et al. Sleep and neuromuscular disorders in children. *Respir Physiol Neurobiol* 2009;**169**:165–70.

Begin R, Bureau MA, Lupien L, et al. Pathogenesis of respiratory insufficiency in myotonic dystrophy. *Am Rev Respir Dis* 1982;**125**:312–18.

Bourke SC, Gibson GJ. Sleep and breathing in neuromuscular disease. *Eur Respir J* 2002;**19**:1194–201.

Camacho M, Certal V, Abdullatif J, et al. Myofunctional therapy to treat obstructive sleep apnea: a systematic review and meta-analysis. *Sleep* 2015;**38**:669–75.

Contal O, Janssens JP, Dury M, et al. Sleep in ventilator failure in restrictive thoracic disorders. Effects of treatment with non invasive ventilation. *Sleep Med* 2011;**12**:373–7.

Devivo MJ. Epidemiology of traumatic spinal cord injury: trends and future implications. *Spinal Cord* 2012;**50**:365–72.

Finder JD, Birnkrant D, Carl J, et al. Respiratory care of the patient with Duchenne muscular dystrophy: ATS consensus statement. *Am J Respir Crit Care Med* 2004;**170**:456–65.

Hukins CA, Hillman DR. Daytime predictors of sleep hypoventilation in Duchenne muscular dystrophy. *Am J Respir Crit Care Med* 2000;**161**:166–70.

Selim B, Junna M, Morgenthaler T. Therapy for sleep hypoventilation and central apnea syndromes. *Curr Treat Options Neurol* 2012;**14**:427–37.

Shneerson JM, Simonds AK. Noninvasive ventilation for chest wall and neuromuscular disorders. *Eur Respir J* 2002;**20**:480–7.

Skatrud J, Iber C, McHugh W, et al. Determinants of hypoventilation during wakefulness and sleep in diaphragmatic paralysis. *Am Rev Respir Dis* 1980;**121**:587–93.

Udd B, Krahe R. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurol* 2012;**11**:891–905.

A complete reference list can be found online at ExpertConsult.com.

Restless Legs Syndrome and Periodic Limb Movements During Sleep

Chapter 95

Richard P. Allen; Jacques Montplaisir; Arthur Scott Walters; Luigi Ferini-Strambi; Birgit Högl

Chapter Highlights

- Restless legs syndrome, also called Willis-Ekbom disease (RLS/WED), is defined by its clinical symptoms involving an urge to move the legs while resting in the evening and night. It has a motor sign of periodic limb movements during sleep that is present in almost all patients. This chapter describes how to make the diagnosis and also when to use medical and sleep tests for RLS/WED.
- RLS/WED varies both in severity from annoying to disabling and in natural course from intermittent (remission and reoccurrence) to gradually progressive. More severe disease affects work productivity, quality of life, and cardiovascular health.
- The close relation between iron status and RLS/WED is described in this chapter for epidemiology, biology, and clinical treatments. In some patients, iron treatment may improve disease course.
- The need for long-term pharmacologic treatment of RLS/WED alters treatment choices. Ropinirole and pramipexole, when given for 1 year or longer, may cause RLS/WED augmentation, altering the disease process to create worse RLS/WED symptoms. Newer drugs are as effective, with no risk for augmentation with α2δ ligands and possibly reduced risk with longer acting dopamine agonists.

DESCRIPTION AND CLINICAL CHARACTERISTICS

Sensory and Motor Manifestations

Restless legs syndrome, also called Willis-Ekbom disease (RLS/WED), has been described for centuries, with an early description by Willis in 1672,[1,2] but only in 1945 was it singled out as a distinct clinical entity by the Swedish neurologist Carl Ekbom.[3] Patients with RLS/WED report an urge to move (akathisia) focused on the legs usually, but not always associated with dysesthesia when at rest.[4] Patients with dysesthesias describe them very differently (e.g., uncomfortable, unpleasant, creepy-crawly, jittery, internal itch, or shocklike feelings); up to 50% of RLS/WED patients describe their sensations as painful. Some, however, describe only an urge to move, and they are unaware of a sensory component. Symptoms are usually felt over large areas of the thighs or calves (or both) and are usually experienced as coming from deep within the legs rather than as superficial.

The original term RLS fails to adequately describe the condition. First, the leg akathisia is not the fidgety, jittery overall restlessness that occurs when waiting or sitting too long; rather it comes as a focused strange feeling of an urge to move specific parts of the body that must involve at least part of one leg. Second, the disorder can also involve the arms and other body parts in addition to the leg, particularly when more severe. Almost 50% of patients with moderate to severe RLS/WED in one tertiary care population reported symptoms in the arms.[5] Given the problems of the term *restless legs*, the International Restless Legs Syndrome Study Group (IRLSSG) and several patient groups have adopted the new name *Willis–Ekbom disease*, recognizing those most responsible for defining the disease.[4]

Three critical features specify occurrence of the sensory symptoms defining RLS/WED (Box 95-1). First, the symptoms are engendered or worsened by rest or inactivity (Videos 95-1, 95-2, and 95-3).[4] Typically, patients describe exacerbation of symptoms in soporific situations, such as watching television, driving or flying long distances, and attending meetings. Symptoms worsen with a decrease in alertness. Second, activity relieves the symptoms.[4] Patients use different motor strategies to relieve the discomfort, producing motor features characteristic of the disease, such as moving their legs vigorously or flexing, stretching, or crossing the legs one over the other. In severe cases, they might walk around for hours in the evening or night to relieve the discomfort. The relief usually begins immediately or soon after the activity starts and persists as long as the activity continues. The symptoms, however, may return after movement stops, sometimes quickly in cases of more severe disease. Relief with walking is often complete, but some with more severe RLS/WED may find only little symptom relief even from strenuous movement. Third, the symptoms are worse in the evening or during the night.[4] The usual decreased activity and alertness in the evening do not account for this nocturnal increase in symptoms. Three studies using modified constant routine protocols investigated the pattern of RLS/WED symptom occurrence.[6-8] These demonstrated a circadian rhythm of the symptoms, maximal after midnight independent of effects of activity. The

Box 95-1 DIAGNOSTIC CRITERIA ESTABLISHED BY THE INTERNATIONAL RESTLESS LEGS SYNDROME STUDY GROUP

Essential Features*

1. An urge to move the legs usually accompanied by, or felt to be caused by, uncomfortable and unpleasant sensations in the legs
2. The urge to move the legs, with any accompanying unpleasant sensations beginning or worsening during periods of rest or inactivity, such as lying down or sitting
3. The urge to move the legs, with any accompanying unpleasant sensations partially or totally relieved by movement, such as walking or stretching, at least as long as the activity continues
4. The urge to move the legs, with any accompanying unpleasant sensations during rest or inactivity only occurring or worsening in the evening or night as opposed to during the day
5. Occurrence of the above features not solely accounted for as symptoms primary to another medical or a behavioral condition (e.g., myalgia, venous stasis, leg edema, arthritis, leg cramps, positional discomfort, habitual foot tapping)

Specifiers for Clinical Course

A. Chronic-persistent restless legs syndrome/Willis-Ekbom disease (RLS/WED): Symptoms when not treated would occur on average at least twice weekly for the past year
B. Intermittent RLS/WED: Symptoms when not treated would occur on average more than twice weekly for the past year, with at least five lifetime events

Specifier for Clinical Significance

- Symptoms of RLS/WED cause significant distress or impairment in social, occupational, educational, or other important areas of functioning by their effect on sleep, energy and vitality, daily activities, behavior, cognition, or mood

Nonessential but Common Features

1. Periodic limb movements: presence of periodic leg movements during sleep or resting wake at rates or intensity greater than expected for age or medical and medication status
2. Dopaminergic treatment response: reduction in symptoms at least initially with dopaminergic treatment
3. Family history of RLS/WED among first-degree relatives
4. Lack of profound daytime sleepiness

*All 5 must be met to make a diagnosis.

Modified from Allen RP, Picchietti DL, Garcia-Borreguero D, et al. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria—history, rationale, description, and significance. *Sleep Med* 2014;15(8):860–73.

symptom's circadian rhythm correlated with that of subjective vigilance, core body temperature, and salivary melatonin secretion.

The primary motor sign of RLS/WED is periodic limb movements during sleep (PLMS). These occur mostly as physiologic flexor leg movements repeated about every 20 to 40 seconds during NREM sleep and also sometimes when lying down resting awake, called periodic leg movement during wake (PLMW). Almost all RLS/WED patients are found to have these movements but only when observed over three to five nights.[9] PLMS are a sensitive but nonspecific motor sign of RLS/WED. Video 95-4 shows classic periodic leg movement flexion at the ankle with extension of the large toe.

RLS/WED symptom intensity ranges from mild and annoying to a disabling compulsion to keep moving. Symptom frequency also varies from less than yearly to daily. A large population-based study in Europe and the United States found that 37% of patients with RLS/WED symptoms had moderate to severe disease (bothersome symptoms two or more times a week).[10] Videos 95-5, 95-6, 95-7 and 95-8 show the wide range of RLS/WED severity from mild foot to whole body movements.

Most patients with moderate to severe RLS/WED complain of poor sleep. In a study of 133 patients, 85% often experienced difficulty falling asleep at night because of RLS/WED, and 86% were frequently woken by symptoms.[11] Sleep laboratory studies confirm the poor sleep of RLS/WED patients compared with controls, characterized by increased wake after sleep onset and decreased sleep time. Most often, stage 2 sleep is decreased with little or no change in rapid eye movement (REM) sleep.

Daytime fatigue is commonly reported, but usually not sleepiness. Most RLS/WED patients do not report the level of sleepiness expected for their degree of sleep loss.[12] Thus RLS/WED appears to occur with some increased arousal process counteracting the sleep loss effects.

Clinical Course

RLS/WED can start at any age from childhood into late adult life. Familial cases of RLS/WED have an earlier age of onset, typically before the age of 30 years.[13-15] Symptom severity and frequency often fluctuate throughout life. Some report remissions, lasting for months or even years, and also relapses, both without any apparent reason.[16] More severely affected patients commonly report a progressive increase in symptom severity with advancing age.[16]

EPIDEMIOLOGY AND BURDEN OF DISEASE

Epidemiology

RLS/WED is one of the most common sleep-related disorders. The prevalence in European and American populations is about 7%[10] for any RLS/WED symptoms during a year and 2.7% for moderate to severe symptoms.[10] Physician-identified, medically significant RLS/WED occurs in 2.7% of patients seen in general medical practices in Europe.[17] The prevalence of moderate to severe RLS/WED symptoms increases with age from about 0.5% for children to 5% for those older than 70 years.[10] In adults older than 40 years, RLS/WED occurs about twice as often in women than men, but there is no gender difference for children[18] or young adults.[10] The gender difference appears to relate to pregnancy because nulliparous women have the same rate of RLS/WED as men.[19]

Burden of the Illness

Mild to minimal RLS/WED does not appear to have significant social or medical impact.[17] In contrast, moderate to severe RLS/WED significantly adversely affects work performance, quality of life, cognition, and health. Work productivity is diminished by 20% for moderate to severe RLS/WED and by 50% for very severe RLS/WED.[20] Quality of life is similarly significantly diminished as shown by marked

impairment in SF-36 quality-of-life dimensions of vitality, role physical, pain, physical functioning, and general health and somewhat less impairment in social functioning, role emotional, and mental health.[10] This disruption in quality of life is at least as great as the disruption caused by other chronic medical conditions, such as diabetes mellitus, depression, osteoarthritis, or hypertension.[10] RLS/WED patients also show impaired cognitive function involving mostly prefrontal cognitive tasks sensitive to sleep loss.[21,22]

Cardiovascular Disease and Stroke

Two large epidemiologic studies, the Wisconsin Sleep Cohort[23] and the Sleep Heart Health Study,[24] reported a significant relationship between RLS/WED and cardiovascular disease even after controlling for other significant factors. A prospective study showed that among 1986 men followed for 10 years there was a 67% increase in the adjusted relative odds for stroke in those with compared with those without RLS/WED (95% confidence interval [CI] = 1.07 to 2.60; $P = .024$).[25] The prospective Nurses Health Study of 70,977 women found that women who had RLS/WED for more than 3 years had an increased hazard ratio for development of coronary heart disease after controlling for arthritis or diabetes (hazard ratio = 1.94; 95% CI = 1.11 to 3.37) and snoring (hazard ratio = 1.80; 95% CI = 1.09 to 2.97).[26] The prospective study by Szentkiralyi and colleagues, however, showed the converse result of cardiovascular disease increasing the risk for developing RLS.[27] Another prospective study showed no relation of RLS/WED to development of cardiovascular disease.[28]

Observational sleep laboratory studies have found that PLMS or RLS/WED was related to increased nocturnal hypertension, including a nondipping pattern,[29,30] and increased daytime hypertension.[31] A higher prevalence of RLS/WED and periodic limb movements is reported after stroke[32-35] and in patients with congestive heart failure.[36-38] High rates of PLMS have been related to the subsequent development of atrial fibrillation[39] and of cardiovascular disease.[40]

These epidemiologic studies suggest an association among PLMS, RLS/WED, and risk for cardiovascular disease. The underlying mechanisms remain unclear[41-43] but may include the large pulse and blood pressure rises with frequent PLMS in RLS/WED.[44-46]

DIAGNOSIS AND SEVERITY ASSESSMENT

Clinical Diagnosis

No biologic assay is available for diagnosing RLS/WED; thus the diagnosis is based on the clinical evaluation of the patient. In 1995, the IRLSSG established four essential criteria for the diagnosis of RLS/WED. These four criteria were further refined at a National Institutes of Health RLS/WED consensus workshop.[47] They have provided reliable diagnostic standards for clinical practice. Nevertheless, when participants in population-based surveys were asked only about these four criteria, about half of those clinically identified with RLS/WED seemed to lack the disease.[17] Several conditions can produce symptoms very close to those of RLS/WED, and these mimics need to be excluded by a careful differential diagnosis. The revised 2012 IRLSSG official diagnostic criteria, therefore, include a fifth criterion requiring differential diagnoses to exclude mimicking conditions. The current five essential RLS/WED diagnostic criteria[4] are listed in Box 95-1. These also address clinical significance and clinical course, emphasizing heterogeneity of RLS/WED manifestations and the need to code these important disease dimensions. These criteria take care not to confound diagnosis of a neurologic disease with the decision about clinical significance of its symptoms. In addition, diagnostic uncertainty can also be reduced by supportive clinical features such as a family history of RLS/WED and a therapeutic response to dopaminergic medications.

Medical Evaluation

The medical evaluation must include a fasting morning iron evaluation. The patient should also be evaluated for clinical indications of commonly associated medical conditions of neuropathy and uremia. Medication history should be reviewed for any temporal relation to exacerbation or onset of RLS/WED, particularly for antidepressants, dopamine antagonists, and antihistamines.

Sleep Laboratory Test

Although not routinely indicated, polysomnography can be used to measure PLMS, the motor sign of RLS/WED. PLMS provide useful objective support for both diagnosis and assessment of disease severity. The World Association of Sleep Medicine and the IRLSSG established the standards for recording and scoring PLMS.[48] Leg movements are defined as anterior tibialis electromyography (EMG) signals greater than 8 mV above baseline that last 0.5 to 10 seconds and are separated by at least 0.5 second of EMG activity within 2 mV of baseline. Periodic leg movements occur as four or more consecutive leg movements with onsets separated (intermovement interval) by 5 to 90 seconds. Recent studies indicate that an intermovement interval of 10 to 90 seconds is preferred for future studies. A PLMS index (number of periodic leg movements per hour of sleep) greater than 5 for the night was considered pathologic and is still used for children and younger persons, but an index greater than 15 is now often used as a cutoff for adults.

The number of periodic leg movements varies over nights. Periodic leg movements occur mostly in non–rapid eye movement (NREM) sleep and cluster into episodes lasting several minutes or hours. The percentage of all leg movements that are PLMS with an intermovement interval of 10 to 90 seconds, the periodicity index, is 6.5 times less variable than the periodic leg movements index in RLS/WED patients.[49] Diagnostic polysomnography includes central electroencephalography (EEG), electrooculography (EOG), submental EMG, and bilateral EMG of the anterior tibialis muscles. The electrographic picture of a leg movement can vary from one sustained contraction to a polyclonic burst with a frequency of approximately 5 Hz. In RLS/WED patients, approximately one third of all PLMS are associated with EEG signs of arousal.[50] Movements can be seen on Videos 95-9, 95-10, 95-11, 95-12, 95-13, 95-14 and 95-15.

Regardless of the presence of EEG arousals, almost every periodic leg movement is associated with transient tachycardia for 5 to 10 beats followed by a bradycardia[50] plus increased systolic and diastolic blood pressure (on average, 22 mm Hg and 11 mm Hg) (Figure 95-1).[44,45] These are substantial changes that are greater in older patients and when associated with EEG arousals.

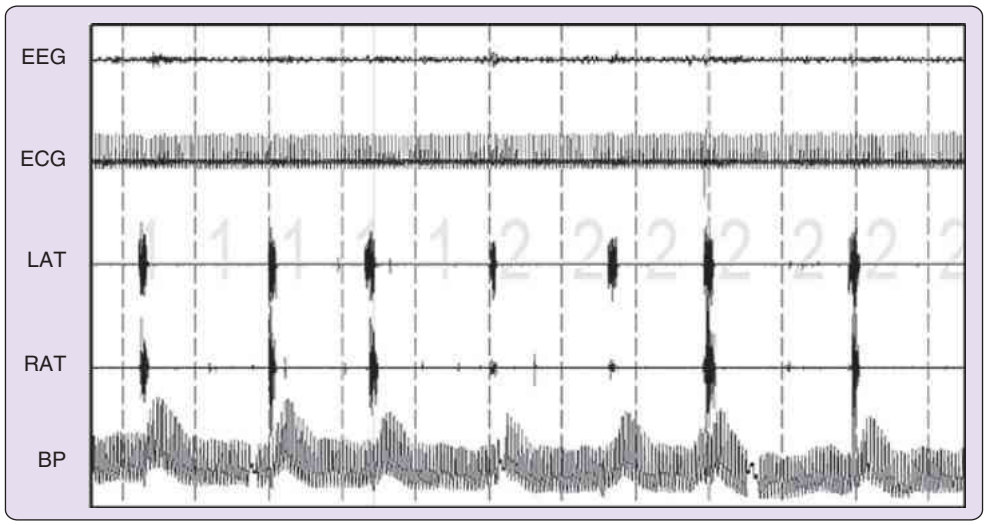


Figure 95-1 Polysomnogram of a patient with restless legs syndrome. Polysomnogram of a patient with restless legs syndrome and periodic limb movements in sleep. The figure shows the periodicity of leg movements and reveals significant increases in blood pressure associated with every periodic limb movement. BP, Blood pressure; ECG, electrocardiogram; EEG, left central electroencephalogram; LAT, left anterior tibialis electromyogram; RAT, right anterior tibialis electromyogram.

Periodic leg movements were first polygraphically documented in RLS/WED.[51] However, they also occur in a wide range of sleep disorders. Five or more PLMS per hour occur in 80% to 90% of patients with RLS/WED, about 70% with RBD, and 45% to 65% with narcolepsy. In RBD, periodic leg movements are often present without arousals during NREM sleep but can also continue into REM sleep, which is unusual for all other conditions. Periodic leg movements were also reported in subjects without any sleep complaint. Although they are rare in young persons, they are relatively common in older adults.[52]

PLMS in patients who complain of primary sleep-onset or sleep-maintenance insomnia or of primary hypersomnia are called periodic limb movement disorder (PLMD) if they have no other cause for the PLMS. Previously, it was assumed that periodic leg movements are responsible for nonrestorative sleep and daytime somnolence in these patients. Although some studies suggested that periodic leg movements may be associated with sleep-wake complaints, most authors have now concluded that periodic leg movements with or without EEG arousals have little effect on nocturnal sleep or daytime vigilance.[53-56] Thus the diagnosis of PLMD should be used with caution given the limited data indicating that PLMS per se disrupt sleep or waking. Despite controversy about their functional significance, PLMS have been useful for sleep laboratory assessments of RLS/WED. In one study, the PLMS index was greater than 5 for 84% and greater than 10 for 70% of RLS/WED patients compared with 36% and 18%, respectively, for controls.[57]

Suggested Immobilization Test

The suggested immobilization test (SIT) was designed to quantify both sensory and motor manifestations of RLS/WED in wakefulness.[57] During the test, patients remain in bed, reclined at a 45-degree angle with their legs outstretched and eyes open. They are instructed to avoid moving voluntarily for the entire duration of the test, which is designed to last an hour and takes place in the evening before bedtime (Video 95-16). Periodic leg movements are measured during waking with the same criteria for PLMS, except the maximal duration of EMG activity is 15 seconds. In addition, every 5 minutes during the test, the patient gives a leg discomfort score. A mean of the 12 discomfort scores has been found to discriminate RLS/WED patients from controls with a sensitivity of 82% and a specificity of 84%. The PLMS index during the SIT was less sensitive.[58]

A possible limitation of the SIT is the single administration time (9:00 PM) because RLS/WED symptoms fluctuate over the course of the afternoon and evening. A multiple SIT (mSIT) consisting of four 60-minute SITs spaced 2 hours apart starting from 6:00 PM helps capture the symptom variability.[59] The mSIT was designed for clinical trials, but its standardized methods may have some use in clinical laboratory assessment of RLS/WED.

Severity Assessments

The 10-question IRLSSG RLS severity scale is a well-validated assessment of RLS severity.[32] (For a sample copy, go to IRLSSG.org.) It has a total score (range 0-40) and two subscales: symptoms and symptom-impact scales.[60] Scores of less than 10 indicate minimal symptoms, and scores of more than 24 indicate moderate to severe symptoms.[61] Scores greater than 15 are required for entry into most clinical trials. This scale and the standard clinical global impression of change are used in pharmacologic clinical trials. Another standard for clinically significant disease often used in surveys is symptoms at least twice a week described as at least moderately bothersome.

RESTLESS LEGS SYNDROME IN CHILDHOOD

Diagnostic Criteria

RLS/WED and PLMS occur in children,[62-64] as shown in Video 95-17. The pediatric diagnosis of RLS/WED[64] requires

meeting the adult criteria (see Box 95-1). Older children should describe the leg discomfort in their own words. Like adults, children with RLS/WED respond to dopaminergic therapy.[65]

Prevalence

The Peds REST study of 10,523 families in the United States and Great Britain found that 1.9% of 8- to 11-year-olds and 2% of 12- to 17-year-olds met criteria for RLS/WED. At least moderately distressing symptoms occurred two or more times a week in 0.5% and 1% of these age groups, respectively.[18]

Relationship to Attention Deficit/Hyperactivity Disorder and Other Medical Conditions

A large meta-analysis confirmed co-occurrence of PLMS and attention deficit/hyperactivity disorder (ADHD).[66] A PLMS index greater than 5 has been reported for 26% to 64% of children with ADHD.[67,68] Those with both ADHD and PLMS showed increased incidence of both personal and family history of RLS/WED.[67,68] Children with ADHD have more PLMS and RLS/WED,[69] and conversely about 44% of children with PLMS are reported to have symptoms of ADHD.[70] Dopaminergic agents improve symptoms of RLS/WED with PLMS but do not clearly reduce ADHD symptoms.[65,71] ADHD occurs with lower serum ferritin and is also responsive to iron therapy for some.[72]

Growing pains, characterized by intense leg pains children experience mostly during the night, are also comorbid with RLS/WED. Growing pains occurred in 80.6% of children with RLS/WED versus 63.2% of those without RLS/WED in the Peds REST study.[18] RLS/WED in children, like in adults, is comorbid with migraine headaches[73] and chronic kidney disease.[74]

RESTLESS LEGS SYNDROME COMORBID CONDITIONS

RLS/WED and PLMS have been related to several other medical and psychiatric conditions but in only a few cases is the association well documented. Three of these conditions have a very strong comorbid relation indicating that development of RLS/WED symptoms may be caused by the other condition. Resolution of these conditions often leads to complete resolution of the RLS/WED. The relation between RLS/WED and the other comorbid conditions is less clear.

Strong Comorbid Relation: Uremia

RLS/WED occurs with uremia, and 15% to 40% of patients on hemodialysis have RLS/WED. Several factors can predispose uremic patients to RLS/WED, particularly anemia and compromised iron regulation. The iron status appears to be critical because intravenous (IV) iron can reduce RLS/WED symptoms.[75] The presence of RLS/WED relates to increased mortality rate in patients with end-stage renal disease.[76] Kidney transplantation often leads to complete resolution of RLS/WED, and the return of RLS/WED symptoms can be an early indicator of transplant failure.[77]

Strong Comorbid Relation: Anemia

RLS/WED occurs in about 35% of patients with iron deficiency anemia[78,79] and at only a slightly lower rate in patients with iron deficiency without anemia.[80] IV iron treatment resolving the anemia also completely resolves the associated RLS/WED in many patients.[81]

Strong Comorbid Relation: Pregnancy

RLS/WED occurs in about 15% to 30% of pregnant women, mostly during the last trimester. It may abate shortly before delivery and generally resolves quickly afterward.[82-84] Iron deficiency is common during pregnancy and has been related to the co-occurrence of RLS/WED.[85,86] IV iron treatment reduced or resolved the RLS/WED during pregnancy for two women with low serum ferritin (≤50 mcg/L),[87] further indicating that iron compromise during pregnancy is a primary factor driving the development of the associated RLS/WED.

Other Comorbid Medical Conditions

Neuropathy appears to increase the risk for RLS/WED. The studies, however, are complicated by problems of both differential diagnosis of neuropathy and RLS/WED and also by the limited use of controls for possible confounding pain effects. One study using a pain control group and carefully validated diagnostic procedures found that RLS/WED occurred in 8% of patients with diabetic neuropathy compared with 3.9% with osteoarthritis.[88] It is, however, unclear whether these results would generalize to all neuropathies. One study showed a high rate (32%) of subclinical small-fiber neuropathy in RLS/WED patients that significantly related to no family history and later age of symptom onset of RLS/WED.[89] Another study reported that RLS/WED patients with neuropathy compared with those without neuropathy had a later age of onset of RLS/WED and less family history.[89] Thus it seems there is some as yet unexplained relation between neuropathy and RLS/WED.

Parkinson disease (PD) deserves special mention because RLS/WED occurs commonly in patients with PD and the major dopaminergic drugs for RLS/WED are also used for PD. RLS/WED in PD often appears after starting dopamine treatment for PD, and the prevalence of RLS/WED is not increased in untreated PD.[90] Thus RLS/WED is not comorbid with PD; rather, treatment of PD with dopaminergic agents will often engender RLS/WED.

Several other comorbid medical conditions have been found to be associated with RLS/WED at a significantly high rate. All conditions associated with iron deficiency appear to show increased risk for RLS/WED, including celiac disease,[91] frequent blood donation,[80] and irritable bowel syndrome.[92] RLS/WED has also been associated with a wide range of other medical conditions, such as hypothyroidism and hyperthyroidism, chronic lung disease, leukemia, Isaac syndrome, stiff-man syndrome, Huntington chorea, multiple sclerosis, essential tremor, migraine, and amyotrophic lateral sclerosis. Considering the high prevalence and the differential diagnostic uncertainty of RLS/WED in the methods used for many of these studies, these associations should be interpreted with caution. However, the rate of occurrence of RLS/WED may increase in patients with multiple comorbid medical conditions,[93] suggesting multiple, complicated interactions increasing the expression of RLS/WED.

Comorbid Psychiatric Conditions

One population-based sample of 1024 participants using excellent diagnostic procedures found the odds ratios (95%

confidence range) of psychiatric conditions occurrence in the past year for those with RLS/WED compared with those without RLS/WED was 2.0 (0.6 to 7.3) for generalized anxiety disorder, 2.7 (1.1 to 6.7) for major depressive disorder, 5.6 (1.4 to 21.9) for obsessive compulsive disorder, and 5.3 (2.0 to 14.0) for panic disorder.[94] In these cases, it appears that RLS/WED increases the risk for having these other disorders rather than the reverse, possibly because of some shared biologic features or a response to the stress of living with RLS/WED.

Substances Associated with Disease

Several substances or medications can induce or worsen RLS/WED or PLMS. These include tricyclics or other antidepressants, lithium carbonate, dopamine D_2 receptor–blocking agents, (neuroleptics, antiemetics), antihistamines, and alcohol.

MEDICAL INVESTIGATION

Given the aforementioned associations between anemia or iron deficiency, iron status should be evaluated in every patient. Iron deficiency cannot be determined by history and can occur with normal hemoglobin, so the standard complete blood count does not suffice. Every patient should have a full iron assessment (i.e., morning fasting serum values for iron, ferritin, total iron-binding capacity, and transferrin saturation). When these levels are abnormal, further medical evaluation is recommended to determine any possible cause, usually involving blood loss. An iron-poor diet may in some cases be a problem.[95] In a patient with RLS/WED, either a serum ferritin level of less than 75 mcg/L or transferrin saturation of less than 17% indicates possible iron deficiency, and iron treatment should be considered.[96]

Medications should be reviewed to reduce, if possible, substances associated with RLS/WED. Because a significant number of RLS/WED patients have peripheral neuropathy,[89,97] a careful clinical examination of sensory and motor functions should be performed with further evaluation if indicated for neuropathy. When taking the history, RLS/WED and PLMS should be differentiated from other state-dependent sensorimotor disorders, such as positional discomfort, hypnic myoclonus, painful legs and moving toes syndrome, nocturnal leg cramps, neuroleptic-induced akathisia, and vascular or neurogenic intermittent claudication. Whenever the diagnosis is doubtful, a polysomnographic recording should be considered, particularly if there is a report of somewhat atypical features such as daytime sleepiness. Leg activity meters provide an alternative for evaluation of PLMS when sleep-disordered breathing is not an issue. Leg activity meters have the major advantage of providing data over 3 to 5 days needed to correct for the large internight variability of PLMS.[9]

ETIOLOGY AND PHYSIOPATHOLOGY

Genetics

Substantial evidence suggests a genetic contribution to RLS/WED. More than 50% of idiopathic cases report a positive family history of RLS/WED.[3,13,16,98,99] In most pedigrees, it segregates in an autosomal dominant fashion, with a penetrance rate of more than 90%.[13]

Multiple linkage studies of familial RLS/WED have, however, identified mostly marginal associations over a wide range of chromosomes. Thus far, none has identified any specific gene associated with RLS/WED. In contrast, modestly large, genome-wide association studies have discovered specific common allelic variants strongly associated with increased risk for RLS/WED occurring at loci for specific genomic regions of *MEIS1, BTBD9, PTPRD, MAP2k/SKOR1,* and *TOX3/BC034767* and at an intergenic region on chromosome 2 (rs6747972). Variants in *MEIS1* appear to carry the largest risk for RLS/WED, and this may indicate an important role for developmental differences in RLS/WED consistent with the role of this gene in early development. The *BTBD9* variants are also strongly associated with PLMS independent of RLS/WED,[100] and most of the variants associated with RLS/WED also relate to some degree to PLMS.[101] The *BTBD9* variant is also associated with reduced serum ferritin[100] for RLS/WED, suggesting a role in iron management consistent with the role of iron deficiency in RLS/WED. The known common variants associated with RLS/WED, however, account for only 7% of RLS heritability.[102,103] The missing genetic factors may involve as yet undetected common alleles, rare variants, or epigenetic features. For example, rare allelic variants of *MEIS1* altering transcription have been associated with RLS/WED.[103]

Neural Substrates

Thus far there is no evidence of significant cell loss or degeneration in the central or peripheral nervous system in RLS/WED. Autopsy studies show no overall histopathologic abnormalities for RLS/WED brains[104] and, in particular, none for the major dopaminergic areas thought to be important for RLS/WED (substantia nigra[104] and A11[105]). Magnetic resonance imaging studies generally find no morphologic abnormalities or loss of gray matter in RLS/WED.[106,107] White matter, however, may be decreased in patients with RLS/WED,[108] particularly in the corpus callosum, anterior cingulum, and precentral gyrus,[109] and there may be an abnormal loss of myelination consistent with that seen in iron-deprived animals.[109]

Abnormalities in central nervous system (CNS) functioning appear to be widespread in RLS/WED. Changes in cortical[110,111] and spinal[112] excitability have been reported, although the latter may result from loss of cortical inhibition.[113] Thalamic connectivity in the resting state seems to be altered.[112] Little is known about the cause of these diffuse functional abnormalities or their significance for RLS/WED.

Neurotransmitter Dysfunction

Dopamine

The therapeutic effects of levodopa and dopamine agonists on RLS/WED and PLMS support the hypothesis that alterations in dopamine signaling may be involved in the pathophysiology of these conditions. Dopamine antagonists generally exacerbate RLS/WED symptoms, and in one study, they precipitated an increase in PLMW in all but one of the patients evaluated.[114] Brain-imaging studies appear to have somewhat conflicting results, but recent analyses have indicated overall consistent findings. Two positron emission tomography (PET) studies reported in the same article confirm decreased membrane-bound dopamine transporter in RLS/WED,[115] whereas overall dopamine transporter

(DAT) evaluated by single-photon emission computed tomography is not changed.[116,117] Iron-deprived animals similarly show a decrease mostly in membrane-bound DAT.[118] Raclopride binding for dopamine receptors is decreased in PET studies of patients with more severe RLS/WED,[119] and in an autopsy study, striatal D_2 receptors had a significant decrease correlated with RLS/WED severity.[120] The decreases in DAT and the D_2 receptors are most consistent with increased striatal dopamine.[119] Cerebrospinal fluid (CSF) studies showed increased 3-orthymethyldopamine, indicating possible increased dopamine production.[121] Thus, overall, RLS/WED appears to have increased, not decreased, presynaptic dopaminergic activation. Increased presynaptic activity leads to a compensatory decreased postsynaptic response that appears to be adequate for most of the usual daily variation in dopamine activity, except at the low point of dopamine activity in the evening and night. When the daily dopamine signal is low, the postsynaptic decreased responsiveness overcompensates for the increased dopamine, producing a failure to respond adequately and accordingly expression of the RLS symptoms. Adding a small amount of dopamine can then increase the stimulation enough to compensate for the excessive postsynaptic decreased responsiveness and reduce the RLS symptoms. Overall, RLS/WED appears best characterized as an increased dopaminergic state, with decreased response to dopamine.[122]

Opioid

The positive pharmacologic response to opioid treatment in RLS/WED and the reversal of that treatment with the opiate receptor blocker naloxone have also been used as an argument in favor of the hypothesis of an endogenous opiate system dysfunction in RLS/WED and PLMS.[123] Pharmacologic data suggest that opiates affect dopamine signaling. An opiate receptor PET scan study showed no difference in postsynaptic opiate receptor binding between RLS/WED patients and controls, but there was an inverse correlation between the degree of opiate receptor binding and the severity of RLS/WED.[124] A preliminary autopsy study showed decreased levels of the endogenous opioids β-endorphin and met-enkephalin in RLS/WED patients compared with controls.[125]

Other Neurotransmitters

Hypocretin (orexin) has been reported to be significantly increased in the CSF of early-onset RLS/WED patients in one study,[126] but this increase was not significant in a second study.[127] Hypocretins increase histamine signaling, and it was found that RLS/WED symptoms can be severely exacerbated by antihistamines, suggesting a histaminergic abnormality that might be related to increased hypocretin.

Iron

Ekbom was among the first to note that RLS/WED commonly occurs with iron deficiency anemia.[85] All conditions producing iron deficiency also engender RLS/WED. Treatment of iron deficiency anemia can completely resolve all RLS/WED symptoms for many patients.[128-130]

One study showed no significant differences in serum ferritin or iron, but the CSF ferritin was reduced and transferrin was increased, consistent with a CNS iron deficiency despite apparently normal peripheral iron status.[131] Magnetic resonance imaging and ultrasound studies of regional brain iron content have consistently shown reduced brain iron, particularly in the substantia nigra of subjects with RLS/WED compared with age-matched control subjects.[132-134] Autopsy analyses of substantia nigra tissue from patients with RLS/WED compared with age-matched control subjects have revealed a complex pattern of iron-related abnormalities.[104] Recent studies with more sensitive methods for measuring iron have found the brain iron deficiency in RLS/WED occurs in other areas beside the substantia nigra and the striatum and in particular occurs in the thalamus.[135] The complex nature of brain iron metabolism in RLS/WED is shown by increased mitochondria ferritin despite decreased cytosolic ferritin,[136] decreased transferrin receptor despite iron deficiency, altered regulation at the blood-brain interface,[137] and association with increased hypoxia inducible factor, indicating activation of hypoxic responses occurring even without actual hypoxia..[138]

These findings, combined with the success of intravenous iron treatments, support the putative concept that a brain iron deficiency causes RLS/WED in many patients. Of interest is the role of iron in dopaminergic transmission in the CNS. Iron deficiency is associated with increased extracellular dopamine, decreased dopamine transporter, and decreased D_2 and D_1 receptors in the striatum of rats similar to the findings in RLS/WED.[122] Thus abnormalities in iron metabolism or environmental factors producing brain iron deficiency may be a primary cause of RLS/WED symptoms or they may be factors contributing to the development of RLS/WED. In this regard, the striking finding that pregnancy increases the risk for RLS/WED in later life raises the interesting concept that this and other environmental factors with significant iron deficiency may alter genetic factors to increase susceptibility to RLS/WED.

TREATMENT

Nonpharmacologic

It is important that patients maintain good sleep hygiene to prevent the development of psychophysiologic insomnia, which is frequently encountered in RLS/WED. Patients should also refrain from drinking alcohol in the evening because it aggravates symptoms in most patients. Some patients report staying active in the evening and delaying their sleep times to later in the night and early morning when the RLS/WED symptoms have abated.[139]

Pharmacologic

Treatment of moderate to severe RLS/WED generally involves many years of continuing essentially palliative treatments. Unfortunately, most of the controlled clinical treatment studies for RLS/WED last only 3 months, with a few lasting 6 months, but some of the significant RLS/WED treatment complications arise after the first 6 months of treatment, particularly as noted later for the dopaminergic treatments.[140] The following emphasizes medication choices assuming long-term treatment.

Four categories of medications are commonly prescribed to treat RLS/WED: dopaminergic agents, alpha-2-delta (α2δ) ligands (a class of anticonvulsants acting upon calcium channels), opioids, and benzodiazepines.

Dopaminergic Medications

Dopaminergic medications have been the most commonly prescribed medications for RLS/WED. The extensive, excellent responses seen with short-term clinical experience and in short-term clinical studies have produced a false impression that these are the most effective medications for RLS/WED. However, one blinded, controlled treatment trial comparing standard dopaminergic treatment to an α2δ ligand showed that an α2δ ligand was as effective in reducing RLS/WED symptoms.[140] Although dopaminergic medications generally produce an excellent immediate and short-term treatment response, clinicians need be aware of the potential for adverse effects emerging with long-term treatment.

Augmentation and Other Long-Term Complications with Dopaminergic Treatments

Three major problems can occur with long-term dopaminergic treatment: (1) Augmentation (worsening of the underlying RLS/WED symptoms) develops in most patients, usually insidiously and often longer duration of treatment.[141] (2) Compulsive behaviors and profound sleepiness develop in a few patients, mostly with higher doses.[142] (3) Stopping the medication is difficult for many because of the strong withdrawal symptoms and significant sleep loss that occur when a dose is missed.

RLS/WED augmentation is a phenomenon in which RLS/WED symptoms become worse than they were before treatment: Symptoms can occur earlier in the day, have greater intensity when present, involve more of the body (e.g., more of the legs, arms as well as legs), have increased periodic leg movements, and have shorter periods of relief after taking the medication.[143] Augmentation leads to a need for increased dose to control symptoms and doses earlier in the day to block the earlier onset of symptoms. Thus, in clinical practice, the most common symptoms of augmentation are increasing and earlier in the day dosing (afternoon, morning) to block the augmentation effects.[141] Augmentation can become a serious medical problem, producing RLS/WED symptoms 24 hours a day even with high doses of dopaminergic medications. The extreme distress with severe RLS/WED augmentation is shown in Video 95-8 for sensory-motor symptoms and in Video 95-7 for in-bed continuous movement. The rate of augmentation varies by medications and is generally higher with shorter acting medications. The augmentation process can develop over at least 10 years.[144] Thus, over 1 to 8 years of treatment using the short- to intermediate-acting dopaminergic agents, about 75% of patients in one large treatment-population study eventually developed symptoms of augmentation.[141]

The risk for augmentation can be reduced by keeping the dose low and in particular by not exceeding the approved dose levels and also not increasing the dose after stable treatment has been established. A clinical need to increase the dose should be considered a warning sign of possible augmentation, and increasing the dose should be done carefully and preferably not more than once.

The recommended treatment for augmentation is reducing the dose, discontinuing the dopaminergic medication, and then switching to an alternate treatment. Switching to a nondopaminergic drug is preferred, although some have used higher doses of a very long-acting dopaminergic, such as rotigotine. It may take several weeks to months to fully recover from augmentation.[145] After augmentation has occurred with long-term treatment, it may reoccur more rapidly if the dopaminergic agent is restarted after withdrawal,[146] suggesting augmentation may produce long-term changes in sensitivity to dopaminergic stimulation.

Compulsive behavior and sleepiness can occur with long-term use of the higher doses of dopamine agonists, presumably related to overstimulation of D_2-like receptors by the currently used agonists.[142,147] All patients being treated for RLS/WED with dopaminergics should be cautioned about these problems, especially because they may not be aware of compulsive behavior or sleepiness. Dopamine agonists have produced excessive gambling with significant financial loss and also inappropriate sexual behaviors. The patient, however, will generally not recognize these as abnormal behaviors. Profound sleepiness appears to occur only rarely and at higher dopaminergic doses.[142]

Abrupt discontinuation of dopamine agonists can produce insomnia and increased RLS/WED symptoms for at least the first 2 to 4 days off the medication. This can occur after missing a single dose and even when the dose is gradually tapered before stopping the medication. This increase in RLS/WED symptoms complicates efforts to consider changing treatments. Withdrawal symptoms appear less common with nondopaminergic medications.

Commonly Used Dopaminergic Medications

Levodopa. Several open-label and placebo-controlled studies have documented the benefits of levodopa, given with a dopa-decarboxylase inhibitor, either benserazide or carbidopa, in idiopathic RLS/WED and RLS/WED associated with uremia. Several adverse effects were reported in patients treated with levodopa, including nausea, vomiting, tachycardia, orthostatic hypotension, hallucinations, insomnia, daytime fatigue, and daytime sleepiness. Morning rebound of RLS/WED symptoms can occur, with symptoms presenting in the morning at the end-of-dose efficacy. Augmentation is common (60% to 82%) and may be severe, requiring medication adjustment.[143,148] It can occur within the first 6 months of treatment even with a low dose (50 mg levodopa).[148] Given these problems, use of levodopa is often limited to as needed not to exceed two or three lower doses (50 to 200 mg) a week.

Dopamine Agonists. Because they are more effective and produce fewer adverse effects (especially augmentation) than L-dopa, dopamine agonists have become one of the first-line treatments for RLS/WED, although now they are used more cautiously at lower doses to avoid long-term treatment complications. Several agonists have been studied in RLS/WED. An older long-acting ergoline-derivative drug, cabergoline, has been found effective for treatment of RLS.[149,150] Cabergoline and the other ergot dopamine agonists have been associated with rare but serious retroperitoneal and pleuropulmonary fibrosis.[151] They are not recommended for RLS/WED treatment except when the lack of alternatives justifies the fibrosis risk.

Two intermediate-duration non–ergoline-derivative agonists, pramipexole and ropinirole, have been extensively studied for the treatment of RLS/WED. Pramipexole, a full agonist for the D_2 subfamily of receptors with preferential

affinity for the D_3 receptor subtype, was found very effective in treating RLS/WED and in suppressing PLMS in double-blind, placebo-controlled studies with sustained efficacy for 3 to 12 months. [152-155] Pramipexole is approved for treatment of moderate to severe RLS/WED at doses of 0.25 to 0.5 mg by the U.S. Food and Drug Administration (FDA) and at doses of 0.25 to 0.75 mg by the European Medicines Agency (EMEA), both with doses to be taken about 2 hours before bed. Augmentation was found in 32% of patients over 3 years of treatment[154,156] and has an annual rate of occurrence of 7% to 8% for at least 8 to 10 years, eventually leading to as many as 75% having experienced some indication of agumentation.[141,144]

Ropinirole, a dopaminergic agonist with a pharmacologic profile similar to that of pramipexole, was also found effective in treating RLS/WED in placebo-controlled studies,[157,158] with sustained efficacy for 12 months.[159] Like pramipexole, ropinirole is well tolerated and rarely requires the adjunctive use of domperidone for nausea. Ropinirole is approved by the FDA and EMEA for treatment of moderate to severe RLS/WED at doses up to 4 mg taken before bed. The rate at which patients develop augmentation with ropinirole is similar to that of pramipexole.[141]

The dopamine agonist rotigotine, with 24-hour transdermal delivery, provides continuous dopamine stimulation and is very effective for treatment of RLS/WED.[160-163] It has high affinities for the D_2 receptor family and also, unlike pramipexole and ropinirole, has significant affinity for D_1, D_5, and α_{2B}-adrenergic receptors. Rotigotine is approved by the FDA and EMEA for treatment of moderate to severe RLS/WED at doses of 1 to 3 mg per 24 hours. This is the only medication that has been prospectively evaluated for very long-term treatment. In that study rotigotine produced an excellent response over 5 years of treatment for the 43% of the patients who continued on the drug, and 39% of these patients were considered free of all RLS/WED symptoms. Nineteen percent of patients discontinued rotigotine for application site reactions. The annual rate of clinically significant augmentation over the 5 years with rotigotine was 2.9%,[164] less than for pramipexole or rotigotine in shorter term studies.

A2δ Ligands (Anticonvulsants)

These ligands act on the α2δ subunit of voltage-dependent calcium channels to reduce the influx of calcium ions into the neuron, thereby reducing release of some neurotransmitters, particularly glutamate, norepinephrine, and substance P. Unlike the dopaminergics, these drugs do not appear to cause augmentation with long-term treatment. They are a first-line treatment for RLS/WED when their adverse effects are not a problem. The primary adverse effects for these drugs are dizziness, somnolence, nasopharyngitis, and weight gain. This class of drugs has been marked by the FDA for possible development of suicidal ideation. This has not been documented as a problem in the RLS/WED studies, but caution is advised when prescribing these drugs for patients with major depression. In general, these drugs should be started at low doses with a gradual increase in dose about once a week until therapeutic levels are reached. Unlike the dopamine agonists, the response to these drugs may not occur on the first few nights of treatment; rather, the clinical response may be delayed by the need to gradually increase the dose to therapeutic levels.

Three α2δ agents are currently used to treat RLS/WED. Gabapentin enacarbil is an oral prodrug for gabapentin that provides better absorption and allows for obtaining a higher continuous blood level compared with oral gabapentin. It has been found to be very effective for treatment of RLS/WED over a full 24-hour period, reducing daytime as well as evening and nighttime symptoms.[165-169] It is approved by the FDA for treatment of RLS/WED at a dose of 600 mg daily. Pregabalin is commonly used to reduce pain. Pregabalin 300 mg is as effective as 0.5 mg pramipexole and more effective than 0.25 mg pramipexole over 12 months of treatment.[140] It has an augmentation rate over a full year of 1.7%, significantly less than the 9.0% for pramipexole 0.5 mg[140] and about the same as the placebo rate of 1% to 2% per year.[170] Thus this low rate may represent a nonpharmacologic process in the natural course of the disease, either symptom fluctuation or gradual progression of the disease rather than augmentation (Figure 95-2).

Gabapentin has been widely used to reduce neuropathic pain, and several very small open-label trials[171-175] and one small placebo-controlled, crossover study[176] showed it is also effective for treatment of moderate to severe RLS/WED. Its efficacy is limited by problems with absorption that limit the actual doses that can be achieved, and serum levels can be variable for a given dose.[177] Neither pregabalin nor gabapentin have been approved by the FDA or EMEA for treatment of RLS/WED.

Opioids

The therapeutic effects of opioids were noted by Ekbom[3] and confirmed in several open-label and controlled clinical trials.[178-181] Persisting efficacy of opioids was demonstrated in long-term follow-up studies.[179,180] Opioids are often prescribed for severe cases of RLS/WED, especially in patients unresponsive to other treatments. A prolonged-release oxycodone combined with naloxone (doses of 5 to 40 mg oxycodone and 2.5 to 20 mg naloxone, twice daily) was found in a placebo-controlled study to provide significant benefit for RLS/WED patients not responding well to their current, mostly dopaminergic, medications.[181] This medication is approved for use in Europe as a second-line treatment. Methadone has also been used for treatment of severe RLS/WED in patients with inadequate responses to other medications. The dose of methadone for RLS/WED is usually 2.5 to 20 mg a day, lower than the usual dose for analgesia.

Although there is little evidence of tolerance or addiction to opioids in the RLS/WED literature, the data are sparse, and therefore prescription of opioids should be restricted to patients without a previous history of substance abuse. Opioids should also be used cautiously in patients who snore and are at risk for having sleep apnea syndrome. Constipation, lethargy, and sleepiness are common adverse effects of opioid treatments.

Benzodiazepines

Several studies have shown that benzodiazepines, including clonazepam,[182] nitrazepam,[183] and temazepam,[46] improve sleep quality and reduce arousals with PLMS in patients with RLS/WED and PLMS. However, the therapeutic effects of benzodiazepines on subjective ratings of RLS/WED symptoms were either modest or not significant. Therefore

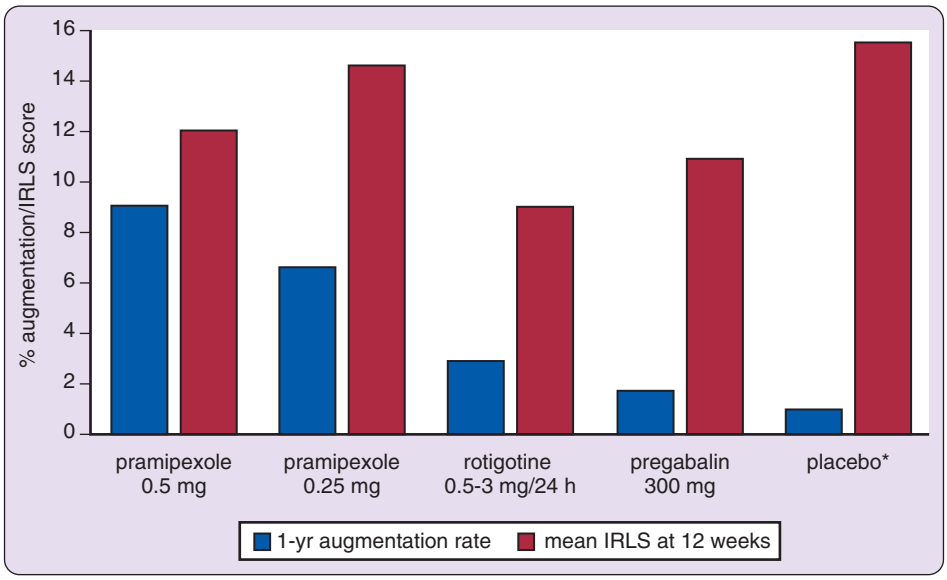


Figure 95-2 Efficacy and augmentation with restless legs syndrome treatments. Restless legs syndrome (RLS) augmentation rate per year and efficacy after 1 year of treatment (mean International Restless Legs Scale [IRLS] score) showing that pregabalin 300 mg is as or more effective than pramipexole (0.5 or 0.25 mg) and that both pregabalin and rotigotine have augmentation rates close to that of placebo, indicating natural progression rather than augmentation. Data are from blinded, controlled 1-year studies of pregabalin,[140] pramipexole,[140] and rotigotine.[164] The data for placebo are prorated from a blinded 6-month study.[170]

benzodiazepines are mostly used to improve sleep continuity in RLS/WED patients.

Iron: Oral and Intravenous

Oral iron as supplemental treatment is indicated when the patient's ferritin level is less than 75 mcg/L.[96] Ferrous sulfate 325 mg, or its equivalent, with vitamin C 100 to 200 mg can be taken twice a day, preferably on an empty stomach, depending on how well the iron is tolerated.

IV iron formulations with tightly bound iron appear to be very effective treatment for 40% to 60% of RLS/WED patients in open-label studies using iron dextran[184-186] and in one placebo-controlled study using ferric caboxymaltose.[187] In these studies, about 20% of patients reported essentially complete remission of their RLS/WED lasting several months. In two controlled studies, iron sucrose, a formulation that more rapidly releases iron for cellular uptake, did not provide the same level of benefit.[188,189] The duration of benefit from IV iron vary considerably, and repeated doses have been found to be useful in a small sample of patients,[190] but care has to be taken to avoid iron overload. IV iron is tolerated well with few adverse effects, except that the high-molecular-weight dextran can produce severe anaphylaxis. Anaphylaxis seems to be related to the iron-dextran formulation and almost exclusively to the high-molecular-weight iron dextran. There appear to be no significant risks for anaphylaxis with the other iron formulations,[191] but given the history of this problem with IV iron, standard treatment options should be readily available for rapid intervention if anaphylaxis occurs.

Clinical Management

In summary, α2δ ligands are considered a good first choice for treatment because they provide excellent long-term treatment when they are tolerated. They are specifically preferred for patients with anxiety, sleep disturbance, and pain, either RLS or neuropathic. They should also be considered if the RLS/WED course is intermittent and future trials off medication are contemplated. Among these medications, only gabapentin enacarbil has been approved by the FDA for RLS/WED treatment. Because of gabapentin absorption problems some patients who fail to respond to gabapentin may respond to gabapentin enacarbil or pregabalin. Dopamine agonists are also a first-line treatment and should be considered particularly for patients with depression and increased risk for falls. They have the major advantage of often producing a rapid therapeutic response with relatively limited immediate adverse effects. Ropinirole, pramipexole, and rotigotine are the only dopamine agonists with FDA and EMEA approval. Some patients who respond poorly to the shorter acting dopamine agonists may respond better to a longer-acting medication. The significant problems with long-term use of dopamine agonists can be reduced by keeping the dose low and not exceeding the FDA-approved daily doses (0.5 mg pramipexole, 4 mg ropinirole, and 3 mg rotigotine). Rather than increasing dose beyond this range, consideration should be given to adding another class of medication or switching medications.

RLS/WED symptoms can spontaneously remit for weeks or even months. Pharmacologic treatments may permit PRN use of medications during symptomatic periods. Daily pharmacologic treatment should be considered if patients complain of RLS/WED occurring at least two nights per week and find the symptoms distressing and affecting their functioning. An occasional drug holiday can be considered to evaluate the need for continuing the treatment. All pharmacologic treatments, except possibly iron, are palliative and do not reduce the underlying disease process. Iron appears in some patients to possibly alter the disease process by improving the well-established RLS abnormality of decreased brain

Table 95-1 Management of Restless Legs Syndrome LS/WED

Agent and Daily Dosage	Side Effects	Countermeasures
Step 1: A2δ Agents		
First-line treatment, particularly if sleep disturbance, pain, or anxiety is present		
Gabapentin enacarbil, 300-600 mg	Dizziness	Reduce dose and add alternate medication class as needed. If fall risk, then discontinue and change to alternate medication class.
Pregabalin, 50-450 mg*	Somnolence, daytime fatigue	Reduce dose and add alternate medication class as needed.
Gabapentin, 100-1800 mg*		If significant, discontinue and change to alternate medication class.
	Tolerance	Discontinue, take drug holiday with return to medication. Switch to alternate medication class.
	Weight gain	Reduce dose and add alternate medication class as needed. If significant, discontinue and change to alternate medication class.
Step 2: Dopamine Agonists		
Alternative first line treatment if depression is present and dose kept low.		
Pramipexole, 0.125-0.5 mg* (0.75 mg in Europe)	Nausea and orthostatic hypotension	Slowly increase dosage or use domperidone if available (10-30 mg).
Ropinirole, 0.5-4.0 mg*	Insomnia	Add or switch to α2δ agent.
Rotigotine, 1-3 mg/24 h		Use a small dose of benzodiazepines in association with dopamine agonists.
	Daytime fatigue and somnolence	Reduce dosage or discontinue dopamine agonists.
	Compulsive or impulsive behavior	Reduce dose and add alternate medication class as needed. If significant, discontinue and change to alternate medication class.
	Tolerance	Discontinue and switch to longer acting dopamine agonist or alternate medication class.
	Augmentation	Discontinue and switch to alternate medication class or longer acting dopamine agonist.
Step 3: Dopamine Precursors		
Useful for intermittent treatment, such as twice a week		
Levodopa-benserazide or levodopa-carbidopa (regular or slow release), 100/25 or 200/50 mg†	Same as for dopamine agonists	See countermeasures for dopamine agonists.
	Morning rebound or augmentation of restless legs syndrome in early evening	Use small extra dose of levodopa during daytime or reduce dosage or combine levodopa with dopamine agonists or benzodiazepines or discontinue levodopa (if severe and persistent).
	Augmentation	Do not used daily. Discontinue and switch to dopamine agonists or a nondopamine medication.
Benzodiazepines		
Useful for sleep promotion		
Clonazepam, 0.5-2.0 mg‡	Daytime somnolence	Reduce dosage.
Temazepam, 15-30 mg‡	Tolerance	Take drug holiday for 2 weeks then return to lower dosage.
Nitrazepam, 5-10 mg*		
Opiates		
Second-line treatment		
Oxycodone-naloxone, 10/5-40/20 mg/day	Constipation	Use for symptom treatment.
Methadone, 2.5-20 mg	Dependency	Take a drug holiday. Discontinue and switch to alternate medication.
Oxycodone, 5-40 mg		
Oral Iron		
Always consider if serum iron ≤75mcg/L *or* transferrin saturation ≤17%		
Ferrous sulfate, 650 mg (325 mg with vitamin C, 100 mg twice a day)	Constipation, stomach upset and pain	Reduce dose, discontinue, take with food.
	Diarrhea, nausea, vomiting	Reduce dose, discontinue, take with food.

*One hour before onset of symptoms in the evening or 1-2 hours before bedtime if symptoms are not present in the evening.
†Considered most appropriate for PRN dosing not more than 3 times a week rather than daily use.
‡Before bedtime usually to promote sleep with restless legs syndrome.

iron insufficiency. This effect of IV iron has been demonstrated in preclinical studies.[192] The dopaminergic medications, in contrast, may produce RLS/WED augmentation, making the underlying RLS symptoms worse than before treatment. Therefore the clinician should carefully assess the therapeutic benefit of the treatments for long-term care versus the severity of adverse effects. Choice of treatment should also consider cost differences.

A therapeutic flowchart is shown in Table 95-1. Each drug is presented with its commonly used therapeutic dosage, its most common side effects, and the appropriate countermeasures to adopt. Higher doses may be considered with caution for all drugs but particularly for dopaminergic medications, for which the risk for augmentation increases with dose.[140] If augmentation develops on a dopaminergic medication, the clinician should consider switching to a very long-acting dopamine agonist, an α2δ ligand, or an opiate. It is still unclear whether the very long-acting dopamine agonists will also in time produce significant augmentation, but their rate of augmentation appears somewhat lower than with the shorter acting dopaminergics. Prolonged-release oxycodone-naloxone,[181] methadone,[180] and IV iron[186] are considered good second-line treatments.

CLINICAL PEARLS

- RLS/WED is a common and underdiagnosed condition with a strong genetic component.
- α2δ Agents provide effective long-term treatment with less risk for augmentation than with levodopa and dopamine agonists.
- Dopamine agonists are effective, and the risk for RLS/WED augmentation with long-term treatment can be reduced by not exceeding regulatory approved dose levels and using longer acting agonists.
- Good second-line treatment options include prolonged-release oxycodone-naloxone, methadone, and IV iron.
- Iron deficiency engenders or exacerbates RLS/WED. Iron deficiency can occur with a normal hemoglobin. Check for serum ferritin of 75 mcg/L or less or percent transferrin saturation of 17% or less, and if either is the case, consider oral iron treatment while also providing any other treatment needed. In some patients, oral iron can lead to complete remission from the RLS/WED.

SUMMARY

RLS/WED is a neurologic disease defined by its primary clinical features of an urge to move the legs, often with abnormal sensations that occur during resting in the evening and night more than the morning or day. The diagnosis is based on clinical history and can be supported by PLMS and a family history. Its severity ranges from annoying to disabling. Moderately severe RLS/WED reduces sleep, quality of life, and work productivity. It is associated with increased risk for cardiovascular disease. Moderate to severe RLS/WED occurs in about 1% to 3% of adults and 0.5% of children.

RLS/WED occurs with many medical conditions. It is closely related to iron status, as documented by epidemiology, biology, and iron treatments. RLS/WED often occurs in families. Genome-wide association studies have identified common allelic variations that increase the risk for RLS/WED.

Short- to intermediate-acting dopamine medications provide excellent initial treatment but with long-term use often produce augmentation (worsening) of the disease. α2δ Agents provide as effective initial treatment as dopaminergics with less risk for augmentation. Longer-acting dopamine agonists also have reduced augmentation. Various iron treatments appear for some to improve the disease process.

Selected Readings

Allen RP, Chen C, Garcia-Borreguero D, et al. Comparison of pregabalin with pramipexole for restless legs syndrome. *N Engl J Med* 2014;**370**(7):621–31.

Allen RP, Ondo WG, Ball E, et al. Restless legs syndrome (RLS) augmentation associated with dopamine agonist and levodopa usage in a community sample. *Sleep Med* 2011;**12**(5):431–9.

Allen RP, Picchietti DL, Garcia-Borreguero D, et al. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria—history, rationale, description, and significance. *Sleep Med* 2014;**15**(8):860–73.

Allen RP, Walters AS, Montplaisir J, et al. Restless legs syndrome prevalence and impact: REST general population study. *Arch Intern Med* 2005;**165**(11):1286–92.

Cho YW, Allen RP, Earley CJ. Lower molecular weight intravenous iron dextran for restless legs syndrome. *Sleep Med* 2013;**14**(3):274–7.

Earley CJ, Connor J, Garcia-Borreguero D, et al. Altered brain iron homeostasis and dopaminergic function in Restless Legs Syndrome (Willis-Ekbom Disease). *Sleep Med* 2014;**15**(11):1288–301.

Hening WA, Walters AS, Wagner M, et al. Circadian rhythm of motor restlessness and sensory symptoms in the idiopathic restless legs syndrome. *Sleep* 1999;**22**:901–12.

Li J, Moore H 4th, Lin L, et al. Association of low ferritin with PLM in the Wisconsin Sleep Cohort. *Sleep Med* 2015;**16**(11):1413–8.

Ondo WG. Methadone for refractory restless legs syndrome. *Mov Disord* 2005;**20**(3):345–8.

Schlesinger I, Erikh I, Nassar M, Sprecher E. Restless legs syndrome in stroke patients. *Sleep Med* 2015;**16**(8):1006–10.

Trenkwalder C, Benes H, Grote L, et al. Prolonged release oxycodone-naloxone for treatment of severe restless legs syndrome after failure of previous treatment: a double-blind, randomised, placebo-controlled trial with an open-label extension. *Lancet Neurol* 2013;**12**(12):1141–50.

A complete reference list can be found online at ExpertConsult.com.

Alzheimer Disease and Other Dementias

Dominique Petit; Jacques Montplaisir; Erik K. St. Louis; Bradley F. Boeve

Chapter 96

Chapter Highlights

- This chapter gives an overview of sleep disturbances, characteristics of sleep architecture and microstructure, and quantitative analyses of wakefulness and sleep electroencephalography (EEG) in several dementias, including Alzheimer disease, progressive supranuclear palsy, Parkinson disease, dementia with Lewy bodies, vascular dementia, Huntington disease, Creutzfeldt-Jakob disease, and frontotemporal dementia.
- To identify patients at risk for developing an overt neurodegenerative disorder, recent research has focused on EEG markers for phenoconversion to dementia in patients with mild cognitive impairment (MCI).
- Certain sleep disorders, including rapid eye movement (REM) sleep behavior disorder (RBD) combined with MCI, substantially increase the risk for developing dementia and parkinsonism.
- Treatment of sleep disturbances in patients with dementia should focus on managing specific symptoms, including insomnia or fragmented sleep; excessive daytime sleepiness; alterations in the sleep-wake circadian rhythm; and excessive motor activity during the night, including RBD, periodic leg movements during sleep, and nocturnal agitation or wandering.

With the steady increase in life expectancy, the prevalence of dementia and neurodegenerative diseases is rising at a rapid pace. It is estimated that the prevalence of dementia doubles every 5 years after age 65 years. With neurodegeneration comes myriad sleep disorders that, if left untreated, can aggravate cognitive deficits, worsen the patient's and caretaker's quality of life, and lead to premature institutionalization. On the other hand, sleep and quantitative electroencephalography (EEG) can sometimes help in making a differential diagnosis or in predicting who is at short-term risk for developing dementia.

ALZHEIMER DISEASE

Alzheimer disease (AD) is a neurodegenerative disorder characterized by progressive decline in memory and other cognitive domains. AD is the primary cause of irreversible dementia in old age. Diagnostic criteria, first established by the National Institute of Neurological Disorders and Stroke—Alzheimer's Disease and Related Disorders Association Work Group, were revised in 2011[1] in light of the discovery of new markers, brain imaging findings, and analyses of amyloid B and tau proteins in cerebrospinal fluid. Accumulation of abnormal tau proteins appears to mediate neuronal dysfunction more so than amyloid deposition. The most highly affected structures early in AD include the entorhinal cortex, hippocampus, amygdala, and nucleus basalis of Meynert, with eventual more diffuse involvement of the suprachiasmatic nucleus (SCN), intralaminar nuclei of the thalamus, locus coeruleus, raphe nuclei, central autonomic systems, and the neocortex. Neuropathology findings include neurofibrillary tangles, neuritic plaques, and neuronal loss. Interestingly, sleep may play a protective role in warding off toxic protein accumulation, with new evidence demonstrating that the brain's "glymphatic" system, the interstitial space surrounding glia and neurons, has substantially higher clearance of amyloid beta during sleep and drug-induced anesthesia compared with wakefulness.[2] Additional animal and human research will be necessary to demonstrate whether this "housekeeping" role of the brain's glymphatic system during sleep reduces potentially toxic protein accumulation and the eventual evolution of neurodegeneration.

Sleep Problems

Sleep disturbance may occur in up to 25% of patients with mild to moderate AD and in about 50% with moderate to severe AD. Several types of sleep problems can be seen, with likely multifactorial causes, including both excessive daytime sleepiness and insomnia with difficulties in both falling and maintaining sleep due to frequent nocturnal awakenings and premature morning awakening (as described in a recent review[3]). These sleep problems pose additional functional consequences; daytime sleepiness is associated with greater impairments in AD patients independent of the level of cognitive impairment.[4] Sleep disturbances are also a major cause of early institutionalization.

Perhaps the most burdensome sleep problem of patients with AD is sundowning, a delirium-like state characterized by confusion, agitation, anxiety, and frequent aggression in the evening or night, with potentially injurious nocturnal wandering. This phenomenon can be explained, at least in part, by an alteration in the biologic clock due to impaired functioning of the hypothalamic SCN.[5] The secretion rhythm of many hormones is affected in elderly people, but it is even more disrupted in AD patients.[6] The timing of the biologic clock is

shifted earlier, as evidenced by two common markers: core body temperature and plasma melatonin.[6]

A series of pathophysiologic findings has shed some light on the circadian rhythm disorder of AD patients. Melatonin production and rhythm are disrupted,[7] even in the early preclinical stages of AD.[8] This might be caused by dysfunction in the sympathetic regulation of pineal melatonin secretion by the SCN.[7] The SCN is under the modulatory influence of the nucleus basalis of Meynert,[9] which degenerates early in AD. In addition, sundowning could result from defective nucleus basalis control of arousal signal processing to the neocortex.[9]

Obstructive sleep apnea (OSA) is more frequent in AD patients than in the general population.[10] A relationship between AD and apolipoprotein E (ApoE), a lipoprotein made in the liver and brain and involved in cholesterol transport and deposition, was first noted in the early 1990s.[11] The risk for developing AD is associated with ApoE-ε4 allele homozygosity and heterozygosity. An association has been also found between the ApoE-ε4 allele and OSA.[12]

Finally, rapid eye movement (REM) sleep behavior disorder (RBD) has also rarely been reported in association with AD. RBD has been reported in one case of mixed AD and Lewy body dementia (DLB) confirmed on autopsy[13] and in 1 out of 15 consecutive patients with AD.[14] The latter study also showed that 3 more patients with AD had REM sleep without atonia. Whether a minority of patients with pure AD truly have RBD, or some AD cases are misdiagnosed, or some AD cases have coexisting DLB (at least in the brainstem) remains unclear. Recent evidence suggests that most RBD cases are associated with synucleinopathy neurodegeneration rather than other proteinopathies.[15]

Polysomnography Findings

Sleep architecture is often abnormal in patients with AD. Certain sleep changes seem to be an exaggeration of those that normally appear with aging. Patients with AD show an increased number and duration of arousals and a concomitant increased percentage of stage N1. Compared with elderly controls, AD patients also show a reduced percentage of slow wave sleep (SWS),[16] the most consistently reported change in mild to moderate AD. Sleep disturbances tend to worsen with increasing severity of AD.[16]

Another change in sleep architecture that looks like accelerated aging in AD is a loss of the specific EEG features of stage N2. Sleep spindles and K-complexes are poorly formed and of lower amplitude, shorter duration, and lower number than those seen in age-matched controls.[17,18] This reduction in fast spindles has been linked to worse immediate recall in AD.[19] With advancing AD severity, it becomes more difficult to distinguish stage N2 from N1 given the absence of characteristic EEG features. The proportion of indeterminate non–rapid eye movement (NREM) sleep increases even further with the disappearance of high voltage (>75 uV) delta waves of SWS.

Conversely, other sleep changes observed in AD are not consistent with the manifestations of accelerated aging. The percentage of REM sleep remains stable in normal aging but is reduced in AD patients compared with controls, mostly because of decreased mean REM sleep episode duration.[17] Variables pertaining to initiation of REM sleep and to its characteristic features are usually unchanged in mild AD, such as REM density, number of REM sleep episodes, REM sleep latency, muscle atonia, and phasic electromyographic activity in REM sleep in most patients without RBD,[17,20] probably because these aspects of REM are under the control of cholinergic neurons in the mesopontine tegmentum that are relatively spared in mild AD. However, lower REM sleep percentage could be due to degeneration of the cholinergic nucleus basalis of Meynert, which normally exerts an inhibitory influence on the thalamic nucleus reticularis.[21] Without strong long-lasting inhibition, the rhythm generator of the thalamus can trigger spindle oscillations, thus curtailing the REM periods (see Chapters 7 and 8).

Quantitative Electroencephalography in Alzheimer Disease

In AD, waking EEG activity is characterized by a slower dominant occipital rhythm than in healthy older people, with increased theta and delta activity compared with age-matched controls. Several studies have attempted to correlate quantitative waking EEG with clinical severity of AD, with variable results. Some demonstrated that relative theta power could aid in distinguishing healthy older adults from those with mild, moderate, and severe dementia, and that decreased alpha and beta band power and increased delta band power were correlated with AD severity.[22,23]

EEG slowing (as indexed by the [(delta + theta)/(alpha + beta)] power ratio) is more apparent during REM sleep than during wakefulness in AD patients.[24,25] Moreover, the distinctive topographic pattern of temporoparietal and frontal regional REM sleep EEG slowing in AD patients[25] parallels findings from neuroradiologic[26] and neuropathologic[27] studies, a pattern not observed for the waking EEG. The REM sleep EEG power ratio, but not the waking EEG power ratio, was also correlated with the Mini-Mental State Examination and with a measure of interhemispheric asymmetry of regional cerebral blood flow in AD.[28] Possibly, the superiority of REM sleep EEG over wakefulness EEG to identify AD-related neurodegeneration is due to the fact that the basal forebrain cholinergic neurons (among the first to degenerate in AD) are usually more active during REM sleep than during wakefulness[29] and cholinergic activity is not as masked by other activating neurotransmitter systems as it is during wakefulness.

Because EEG delta activity is prominent throughout sleep, distinguishing pathologic from normal physiologic delta waves during NREM sleep is particularly challenging. Patients with AD have shown less frequent and lower amplitude elicited K-complexes in response to auditory stimulation than age-matched controls,[30] suggesting that AD patients may have an impaired ability to generate normal physiologic high-amplitude slow waves during NREM sleep. In these patients, the probability of eliciting a K-complex was also correlated negatively with dementia.[30]

Mild Cognitive Impairment as a Prodrome of Alzheimer Disease

One of the major advances in AD research during the past 20 years is an improved understanding of the prodromal phase known as mild cognitive impairment (MCI). The conversion rate from MCI to AD is about 50% over 3 years.[31] The amnestic MCI subtype is more likely to evolve toward AD, whereas nonamnestic MCI more frequently converts to DLB. Sleep problems are one of the four most common neuropsychiatric symptoms of MCI. A meta-analysis confirmed that 15% to

59% of MCI subjects report sleep disturbances.[32] MCI subjects, especially in ApoE-ε4 carriers, have a higher density of SWS arousals and shorter REM sleep duration than controls.[33] Amnestic MCI patients also show fewer sleep spindles and spend less time in SWS.[34] MCI subjects also have greater wake after sleep onset, increased REM sleep latency, and earlier dim-light melatonin onset relative to controls (but similar levels of total melatonin secreted).[35] This advanced melatonin onset is associated with poorer memory in MCI patients.[35]

Several studies have explored resting-state EEG markers for identifying MCI subjects at risk for progression to AD. Four specific biomarkers were recommended for the early diagnosis of AD by the National Institute on Aging–Alzheimer's Association workgroup in 2011: cerebrospinal fluid A-beta amyloid–to-tau ratio, positron emission tomography amyloid, positron emission tomography–fluorodeoxyglucose hypometabolism in specific brain regions, and magnetic resonance imaging–determined atrophy of the hippocampus.[36] However, quantitative EEG (qEEG) markers have several advantages over those measures because qEEG is a noninvasive, low-cost, and widely accessible method to test a large number of at-risk patients. Two basic study designs of qEEG markers have been undertaken in AD. First, studies have compared subjects at different stages of cognitive functioning (controls, MCI, and AD subjects), correlating EEG measures with known AD degenerative markers. Second, longitudinal studies have determined EEG markers at baseline in MCI subjects who subsequently converted to AD at follow-up, in comparison to nonconverters. In the cross-sectional studies, theta power during wakefulness has been correlated with cognitive decline,[37,38] hippocampal body volume,[39] and cerebrospinal fluid P-tau and T-tau, the earliest markers of neurodegeneration in AD.[40] Higher high-to-low alpha ratio correlated with temporoparietal cortical thinning,[41] hippocampal atrophy,[42] atrophy of the ventral basal ganglia and pulvinar,[43] and theta frequency activity.[44] Total coherence of low alpha activity was also negatively correlated with white-matter damage in the cholinergic basal forebrain of amnestic MCI subjects.[45] Longitudinal qEEG studies identified increased theta power, sometimes coupled with decreased alpha power, at baseline in MCI subjects who later converted to AD; these changes were not seen in nonconverters.[46-48] However, increased theta-to-gamma ratio was predictive of developing either AD or non-AD dementia, whereas the ratio of the 2-Hz band just above the individual alpha frequency over the 2-Hz band just below the individual alpha frequency was found to be specifically associated with AD conversion.[48] Reduced alpha power over posterior leads at baseline differentiated stable from progressing MCI.[49] One study found increased theta power only at 21-month follow-up in progressing compared with stable MCI patients.[50] In general, proposed predictive electrophysiologic tools do not yet allow confident classification of subjects at the individual level. One study recommended combining six EEG biomarkers for optimal sensitivity and specificity in predicting which MCI subjects would develop AD within a 2-year period.[51]

PROGRESSIVE SUPRANUCLEAR PALSY

Progressive supranuclear palsy (PSP), also called Steele-Richardson-Olszewski syndrome, is a neurodegenerative tauopathy characterized by progressive axial rigidity, postural instability, and supranuclear gaze palsy. Dementia in PSP primarily reflects dysfunction in the frontal subcortical neural networks.[52]

Excessive daytime somnolence is common in PSP. Hypocretin 1 (orexin A) levels were found to be low in PSP, and these levels were inversely correlated with the duration of PSP.[53] Moreover, sleep efficiency is greatly reduced in PSP, even more so than in PD.[54] RLS is frequent in PSP and may contribute to reduced sleep efficiency and duration.[55] REM sleep may be severely reduced or absent in PSP patients, especially with disease progression and decline in cognitive functioning.[56,57] RBD and REM sleep without atonia has been reported in 15% to 20% of patients with PSP despite reductions in REM sleep duration,[58,59] therefore less frequently than in patients with PD.[54,60] RBD is, however, much more prevalent (78%) in a Guadeloupean form of PSP.[61]

Cognitive decline is, in turn, reflected by a slowing of the EEG. One qEEG study of PSP found predominant frontal slowing during wakefulness compared with controls, but no significant between-group differences during REM sleep for any of the 16 regions studied.[56] Slowing of the frontal EEG during waking is consistent with numerous neuropsychological studies showing frontal lobe functional deficits. Absence of EEG slowing during REM sleep suggests that waking EEG slowing is likely not due to a cholinergic deficit, consistent with findings of normal neocortical and hippocampal choline acetyltransferase activity in PSP.[62] However, dopamine levels are severely reduced in the caudate, putamen, and substantia nigra in PSP patients.[62] Frontal deafferentation resulting from striatopallidal complex dopaminergic deficiency may be responsible for PSP impairments because there are extensive fiber connections between these deep nuclei and the prefrontal region, as suggested by positive correlations between the degree of frontal neuropsychological task impairments and EEG slowing observed in PSP patients.[56]

PARKINSON DISEASE WITH DEMENTIA

Parkinson disease (PD) is a progressive neurologic disorder characterized by rigidity, resting tremor, bradykinesia, and an impairment of postural reflexes and gait that is caused in part by degeneration of dopaminergic neurons in the substantia nigra. Sleep alterations experienced by patients with PD are discussed in Chapter 94; therefore only information relevant to dementia associated with PD is reviewed here.

The incidence of overt dementia in PD is relatively high. In a population-based study of dementia in PD, approximately 80% of nondemented PD patients developed dementia within 8 years.[63] Risk factors include advanced age at onset of symptoms, severe motor symptoms (particularly bradykinesia), levodopa-related confusion or hallucinations, speech and axial involvement, depression, and atypical neurologic features, such as modest response to dopaminergic agents or early autonomic dysfunction.[64]

Demented patients with PD often experience hallucinations. One study found that patients with REM sleep anomalies have more hallucinations than patients without such anomalies.[65] Sleep reduction, particularly REM sleep reduction, could trigger hallucinations due to emergence of REM sleep features during wakefulness. Hallucinations are significantly correlated with the presence of RBD and the amount

of dopaminergic medication, independently of age, gender, disease duration, or Unified Parkinson's Disease Rating Scale score.[66] There is growing evidence that RBD is an early manifestation of a neurodegenerative disorder, particularly the synucleinopathies (e.g., DLB, PD, and multiple system atrophy).[67,68] The occurrence of RBD in PD was estimated at 15% with a structured questionnaire[69] but at 33% using polysomnographic recordings[70]; only half of these cases had been detected at the clinical interview. The phenomenon of REM sleep without atonia could explain the reduction in REM sleep reported in patients with PD when sleep staging has been performed according to the standard criteria because specific alternative criteria need to be used to permit the scoring of REM sleep in the context of REM sleep without atonia.[71]

Approximately one third of patients with PD have EEG slowing regardless of the presence of dementia.[66] Focal temporal-occipital and frontal regional EEG slowing occurs even in some nondemented PD patients.[72] However, only PD-RBD patients have EEG slowing of the dominant occipital frequency.[73] PD-RBD patients have higher theta power during wakefulness in frontal, temporal, parietal, and occipital regions compared with patients who have PD without RBD and control subjects. EEG slowing in patients with PD-RBD might be related to the presence of RBD itself or represent a different synucleinopathy phenotype such as prodromal DLB, rather than being an evolutional stage of PD. In support of that hypothesis is that idiopathic RBD patients without PD have higher theta power in the frontal, temporal, and occipital regions during wakefulness.[74] PSG EEG characteristics have also been found to be altered with PD and dementia. Sleep spindle density is reduced in most studies of PD,[75-77] except for one study.[78] Sleep spindle parameters also predict conversion to dementia in PD patients. Sleep spindle density and amplitude are lower in baseline PSG recordings (parietal and occipital regions) of PD patients who later develop dementia than in PD patients who do not develop dementia.[79]

Patients with PD-RBD also have significantly worse performance on standardized tests of episodic verbal memory, executive functions, and visuospatial and visuoperceptual processing compared with both patients with PD without RBD and control subjects.[80] Interestingly, patients with idiopathic RBD also have worse performance on tests of executive function and verbal memory compared with control subjects.[81]

The association of RBD with dementia in PD was more directly demonstrated by one study.[82] Of 65 patients with PD, 24 had RBD. The frequency of RBD was significantly higher in the PD with dementia group compared with the PD without dementia group (77% vs. 27%). Patients who had PD without RBD had a lower rate of dementia (7.3%) compared with patients with PD-RBD (42%).

There are compelling reasons to believe that RBD may often represent prodromal PD with dementia or DLB. The topography of EEG slowing observed in PD-RBD is similar to that of DLB-associated functional neuroimaging hypoperfusion and hypometabolism.[83,84] Also, the profile of cognitive impairments noted in PD-RBD resembles that of DLB,[85] and many RBD patients later develop DLB.[86,87] Thus, the presence of RBD in patients with PD may be an early sign of an evolution toward dementia. Some evidence suggests that cortical Lewy body–type degeneration is the main source of dementia in PD.[88] Other studies have suggested that α-synuclein–positive cortical (especially frontal) Lewy bodies are associated with cognitive impairment, independent of AD-type pathologic process.[89,90] More studies are necessary to determine the pathologic and neurochemical underpinnings of dementia in PD.

DEMENTIA WITH LEWY BODIES

Dementia with Lewy bodies (DLB) is the second most common neurodegenerative cause of dementia in old age. The core clinical features of DLB are progressive cognitive decline, spontaneous parkinsonism, recurrent visual hallucinations, and fluctuating cognition and vigilance.[91] Autonomic dysfunction is often present.[91] DLB is characterized by the presence of Lewy bodies in limbic and neocortical structures.

A questionnaire study showed that patients with DLB had more overall sleep disturbances, more movement disorders while asleep, and more daytime sleepiness than AD patients.[92] Based on results from the Epworth Sleepiness Scale, 50% of patients with DLB experience excessive daytime sleepiness.[93] However, normal hypocretin-1 levels were found in the cerebrospinal fluid of DLB patients with daytime sleepiness, suggesting that sleepiness in DLB is not related primarily to dysfunctional hypocretin neurotransmission.[53,94] Moreover, contrary to popular belief, fluctuating daytime alertness as measured by objective methods was not found to be related to fluctuations in cognition.[95] A polysomnographic study[96] found that 72% of patients with DLB had a sleep efficiency lower than 80%. Sleep efficiency was not, however, correlated with dementia severity. A high proportion of these patients had pathologic respiratory disturbances indices (70.5%) or periodic leg movements during sleep (PLMS) associated with arousal (45%).[96] As in PD, restless legs syndrome (RLS) and PLMS are indeed common in DLB and can play a part in sleep-onset insomnia, nocturnal arousals, and awakenings.[97] However, a significant proportion of DLB patients have high arousal indices not accounted for by PLMS or respiratory disturbances.[96]

A number of studies or review papers have reported that RBD is common in DLB.[86,87] In a large cohort of 78 DLB patients, 96% had a history of recurrent dream-enactment behaviors, and REM sleep without atonia with or without behavioral manifestations was confirmed in 83% of patients.[96] Inclusion of RBD in the list of core criteria improves sensitivity and specificity of DLB diagnosis.[85,87,98] RBD now figures as a suggestive feature for the diagnosis of DLB in the third report of the DLB consortium.[91] Interestingly, the presence or absence of concomitant RBD was found to be associated with distinct clinical and pathologic characteristics of DLB; patients with concomitant RBD had earlier onset of parkinsonism and visual hallucinations, shorter duration of dementia, lower Braak stage, and lower neuritic plaque scores.[99]

A few quantitative EEG studies have reported EEG slowing during wakefulness in DLB, expressed as a loss of alpha combined with slower dominant[100] and nondominant rhythms,[101] or increased theta activity,[100,102,103] which correlates with the severity of dementia.[100,102] Frontal intermittent rhythmic delta activity has also been reported.[103] Fluctuating cognition, one of the core features of DLB, was also shown to be reflected by the variability of the mean EEG power in DLB patients compared with both AD patients and elderly control subjects.[100,104]

VASCULAR DEMENTIA

The term *vascular dementia* covers a range of problems of various etiologies and includes the entities known as multi-infarct dementia (MID; now a rather outdated term), subcortical ischemic vascular dementia, and Binswanger disease. The most studied form of vascular dementia in sleep medicine is probably MID. An actigraphy study found that patients with MID had a significantly greater disruption of sleep-wake cycles associated with poor sleep quality than AD patients.[105] There was no correlation, however, between the degree of sleep disruption and the severity of intellectual deterioration. OSA was more strongly associated with MID than with AD or other dementias.[106,107] Sleep apnea is considered a risk factor for vascular dementia.[108] Patients with vascular dementia were found to have twice the risk for suffering from insomnia compared with AD patients.[107] A population cohort also demonstrated that, compared with men without sleep disturbances, elderly men with daytime sleepiness at baseline had 4.44 times the risk for developing dementia of vascular origin 10 years later (even after adjustment for possible confounding factors, including cognitive function).[109]

Spectral analysis of the waking EEG of patients with vascular dementia revealed a lower dominant occipital frequency,[110,111] higher theta[110,112,113] and delta power,[112] and lower alpha[110,112] and beta power[112] compared with control subjects. Lower beta power was found to correlate with cognitive performance as measured by the Rey auditory verbal learning test and the Mini-Mental State Examination,[112] but no correlation between qEEG variables and dementia severity was found in another study.[113] In addition, in patients with MID, alpha had migrated to more anterior regions. There was no gender difference in the qEEG of patients with MID.[110,111] Compared with AD patients, patients with vascular dementia had lower delta power,[113] the mean frequency in temporal derivations discriminated the two groups in patients with mild to moderate dementia, and the alpha-to-delta ratio distinguished the two groups in patients with moderate dementia.[114] Larger EEG source fluctuations were found in patients with vascular dementia compared with AD patients and controls, possibly reflecting decreased vigilance control and increased cognitive fluctuations.[115]

HUNTINGTON DISEASE

Huntington disease (HD) is an autosomal dominant hereditary condition associated with atrophy of basal ganglia structures, especially the caudate nucleus, and characterized by choreic movements and progressive dementia associated with psychotic features. A CAG trinucleotide repeat in the *HTT* gene located on the short arm of chromosome 4 is the cause of this condition.[116]

Patients with HD have a disrupted night-day activity pattern, and similar patterns are seen in animal models. Transgenic mice carrying the HD mutation showed a disruption of night-day activity, which worsened as the degeneration progressed, but also showed a marked reduction in the expression of *mPer2* and disrupted expression of *Bmal1* in the SCN, the motor cortex, and the striatum.[117] Ubiquitin-proteasome dysfunction has been suggested to play a role in the pathogenesis of HD,[118] and such inclusions have also been found in the SCN of HD transgenic mice.[119] In humans, a postmortem study found that patients with HD had 85% fewer vasoactive intestinal peptide neurons and 33% fewer arginine-vasopressin neurons in the SCN.[120] Dim-light melatonin onset was delayed by 1½ hours in HD patients, and daytime melatonin levels were correlated with severity of functional impairment.[121] Delayed sleep phase was also associated with depression and poorer cognitive performance in HD patients.[122] Daily treatment with alprazolam in HD transgenic mice reversed dysregulated expression of Per2 and Prok2, an output factor of the SCN that controls behavioral rhythms, and also markedly improved cognitive performance in a visual discrimination task.[123] The combination of bright-light treatment and restricted periods of voluntary exercise was also shown to improve the behavioral synchronization to the light-dark cycle and to delay the disintegration of the rest-activity rhythm in the transgenic mouse model.[124] In an HD sheep model, the circadian behavioral abnormalities were among the earliest symptoms.[125] Restoring circadian rhythms in HD patients might thus improve cognitive dysfunction, the most devastating feature of HD.

Disturbed sleep architecture has also been reported in HD patients. Specifically, they showed longer sleep latency, lower sleep efficiency, frequent nocturnal awakenings, and less SWS than age-matched controls.[126,127] A polysomnography study of a large number of HD patients and controls replicated most of these findings, except for sleep latency, which was unchanged, and REM sleep latency, which was found to be longer instead of shorter.[128] REM sleep duration was significantly reduced in presymptomatic carriers of abnormal CAG repeat expansion in the *HTT* gene and decreased as disease severity increased. Three out of 25 patients with HD had RBD. Finally, patients with HD did not have more daytime sleepiness but had more PLMS than controls. Contrary to patients with other neurodegenerative diseases, HD patients showed a higher density of sleep spindles compared with healthy controls.[76,127] Sleep disturbances, including less SWS and more time spent awake, correlate with the degree of caudate atrophy and the severity of clinical symptoms.[127] In fact, one study found sleep disturbances only in moderate to severe cases of HD, and none in mild cases.[126] However, no correlation was found between CAG repeat length and sleep disturbances.[128] Finally, no difference was found on sleep respiratory variables between patients with HD and control subjects.[128]

On visual inspection, waking EEG exhibits gradual slowing and diminished amplitude as the disease progresses. Waking qEEG in HD demonstrated increased theta and decreased alpha activity in HD compared with controls; patients with HD were similar to patients with AD.[129]

CREUTZFELDT-JAKOB DISEASE

Creutzfeldt-Jakob disease (CJD) is a prion-related transmissible spongiform encephalopathy causing extensive neuronal degeneration and pathologic changes, especially in the cortex, resulting in myoclonic jerks and rapidly evolving dementia and leading to death. CJD typically develops between the fifth and the seventh decades of life. Younger age of onset is associated with an increased likelihood of sleep disturbances and other symptoms.[130] The mean survival duration is 4 to 8 months,[131] although 5% to 10% of patients have a clinical course that spans 2 years or more.

Two large-sample studies[132,133] reported that half of CJD patients experienced sleep disturbances, mostly severe insomnia. In some patients, sleep disturbances were a prodromal or presenting symptom.[132] In fact, there is a continuum between CJD and fatal familial insomnia (FFI).[134-136] A mutation of the prion protein at codon 178 is present in both CJD and FFI. A concomitant polymorphism at codon 129 (valine versus methionine) appears to determine whether CJD or FFI will ensue. However, the 129 polymorphism alone (common in the general population) does not appear to be associated with important changes in polysomnogram variables or insomnia complaints.[137] Predominant thalamic pathology with little cortical involvement is characteristic of FFI, whereas progressive cortical dysfunction without major thalamic changes is usually seen in CJD. However, this is not always the case.[138] Polysomnographic studies of CJD reveal disorganized sleep patterns with sudden transitions between sleep stage, few sleep spindles and K-complexes in N2 (which may be difficult to distinguish from N3), decreased slow waves, and lower REM sleep percentage and REM density.[134,139,140] Episodes of nocturnal oneiric, sometimes aggressive behavior with dream-reality confusion, have been reported in some patients.[132,140] An indeterminate state (neither clear wakefulness nor clear sleep) has also been observed in these patients[134,140] and in patients with FFI.[141] Sleep apneas, central or obstructive, are prevalent in this condition.

The hallmark awake EEG feature in patients with CJD is periodic sharp wave complexes within a background of generalized slow but low-voltage EEG, consistent with diffuse cerebral pathology.[131] Periodic sharp wave complexes, usually generalized biphasic complexes, are invariably present by the time patients evolve clinically evident myoclonus, and typically present by 3 months after symptom onset.[142] Periodic sharp wave complexes are part of the diagnostic criteria for probable CJD[143] because they are present in about two thirds of patients and show a high specificity (occurring in only 9% of patients with another neurodegenerative disorder).[144] Sleep EEG studies also report the presence of periodic sharp wave complexes as early as 1 to 3 months after the onset of symptoms.[145,146] Cyclic changes with periodic complex phases alternating with semirhythmic theta-delta activities have been described.[145,146]

FRONTOTEMPORAL DEMENTIA

Frontotemporal dementia (FTD) is a neurobehavioral syndrome associated with accumulation of tau (e.g., Pick disease, corticobasal degeneration, progressive supranuclear palsy) or TAR DNA binding protein molecular weight 43. FTD is a progressive, degenerative condition characterized by loss of executive or language abilities and several other neurobehavioral features, such as loss of insight, overactivity, lack of social awareness, disinhibition, and lack of personal hygiene.[147] Approximately 5% to 15% of patients with dementia have a disorder within the FTD spectrum. FTD is likely underdiagnosed because of its similarities with AD, especially later in the progression of the disease. However, unlike AD, FTD initially manifests with progressive aphasia and personality changes, whereas memory tends to remain relatively intact. Structural and functional brain imaging shows atrophy, reduced cerebral blood flow, or diminished glucose metabolism in frontal and anterior temporal areas.[147]

As in AD, FTD is generally accompanied by a disturbance of the alpha rhythm and of the sleep-wake rhythm, which worsens with the progression of the disease.[148] Patients with FTD show more nighttime activity,[149] particularly the behavioral-FTD variant (85%) compared with the semantic variant (3%),[150] less morning activity, and lower sleep efficiency compared with control subjects.[149] For a comparable level of cognitive impairment, patients with FTD showed more sleep disruption and a worse sleep macrostructure than patients with AD, regardless of sleep apnea or other primary sleep disorders.[151] Sleep-disordered breathing seems as prevalent in FTD as in AD.[152] REM sleep parameters, however, were found to be more altered in AD than in FTD.[153]

Electroencephalographic slowing during wakefulness is observed in FTD. One study[154] found increased theta power, and another[155] reported increased delta and theta power more prominently in anterior regions in most patients. However, 31% had normal EEGs, and most patients had a preserved dominant occipital frequency.[155] EEG measures of functional connectivity were found normal in patients with mild to moderate FTD.[156]

TREATMENT OF SLEEP DISORDERS IN PATIENTS WITH DEMENTIA

One useful approach to address sleep disorders in the demented patient involves considering symptoms within four major categories: insomnia or fragmented sleep, excessive daytime sleepiness, alteration in the sleep-wake circadian rhythm, and excessive motor activity during the night, including RBD, PLMS, and nocturnal agitation or wandering.[157] One should keep in mind that some of these disturbances could be due to RLS, OSA, malnutrition, infections, medication effects (often polypharmacy), depression, bladder catheterization, fecal impactions, or disturbing environmental factors. Management frequently requires identification and treatment of underlying medical or psychiatric disorders. For each of these four categories of sleep disorders, we review the appropriate pharmacologic and nonpharmacologic treatment strategies. A summary of selected medications with suggested dosage and titration schedule also appears in Table 96-1.

Insomnia

Insomnia may affect as many as 40% to 60% of patients with dementia.[152] Insomnia that is comorbid with other sleep disturbances is a common occurrence in patients with dementia (described in a recent review[158]). Patients with cognitive impairment are often unable to explain why they are unable to sleep through the night, so caregivers and physicians should carefully investigate possible sources for insomnia. Evaluations of pain, concomitant medical conditions, and medications are essential to treat the patient successfully. For example, both untreated depression and some antidepressant medications (venlafaxine, fluoxetine, and bupropion) can lead to insomnia. Cholinesterase inhibitors such as donepezil, which can improve cognitive and behavioral symptoms in AD patients, can also cause insomnia. "Activating" antidepressants, cholinesterase inhibitors, and stimulants may cause or aggravate insomnia, especially when given at night. This problem usually can be avoided if the medication is administered no later than the evening meal, although in some patients these medications must be given no later than lunchtime. Melatonin

Table 96-1 Sleep Disorders and Disturbances in Dementia: Selected Medications with Suggested Dosing Schedules*

Initial Medication	Starting Dose	Suggested Titrating Schedule	Typical Therapeutic Range
Insomnia			
Trazodone	25 mg qhs	Increase in 25-mg increments q3-5 d	50–200 mg/night
Chloral hydrate	500 mg qhs	Increase in 500-mg increments q5-7d	500–1500 mg/night
Melatonin	3 mg qhs	3–6 mg nightly	3–12 mg/night
Quetiapine	25 mg qhs	Increase in 25-mg increments q3d	25–100 mg/night
Zolpidem	5 mg qhs	Increase to 10 mg qhs if necessary	5–10 mg/night
Restless Legs Syndrome, Periodic Limb Movements in Sleep			
Pramipexole	0.125 mg qhs	Increase in 0.125-mg increments q2-3d	0.25–0.75 mg/night
Gabapentin	100 mg qhs	Increase in 100-mg increments q2-3d	300–1800 mg/night
Pregabalin	25–50 mg qhs	Increase in 25–50 mg increments q3-7d	100–600 mg/night
Excessive Daytime Somnolence			
Methylphenidate	2.5 mg q AM	Increase in 2.5- to 5-mg increments q3-5d in bid dosing (AM and noon)	5 mg q AM to 30 mg bid
Modafinil	100 mg q AM	Increase in 100-mg increments q5-7d in bid dosing (AM and noon)	100 mg q AM to 400 mg/day (400 mg q AM or 200 mg bid)
Armodafinil	50 mg q AM	Begin with 50 mg q AM, increase gradually up to 250 mg q AM	50–250 mg q AM
Amphetamine/ dextroamphetamine	5 mg q AM	Increase in 5-mg increments q7d in qd-bid dosing (AM and noon)	5 mg q AM to 20 mg bid
REM Sleep Behavior Disorder			
Clonazepam	0.25 mg qhs	Increase in 0.25-mg increments q7d	0.25–0.75 mg/night
Melatonin	3 mg	3–6 mg/night	3–12 mg/night
Psychotic Features, Behavior Dyscontrol, Nocturnal Agitation, Nocturnal Wandering			
Donepezil	5 mg q AM	Increase to 10 mg q AM 4 wk later	5–10 mg q AM
Rivastigmine†	1.5 mg bid	Increase in 1.5-mg increments q4wk in bid dosing (AM and hs)	3–6 mg bid
Galantamine†	4 mg bid	Increase in 4-mg increments q4wk in bid dosing (AM and hs)	4–12 mg bid
Risperidone	0.5 mg qhs	Increase in 0.5-mg increments q7d in bid dosing (AM and hs)	0.5 mg qhs to 1.5 mg bid
Olanzapine	5 mg qhs	Increase in 5-mg increments q7d in bid dosing (AM and hs)	5 mg qhs to 10 mg bid
Clozapine‡	12.5 mg qhs	Increase in 12.5-mg increments q2-3d	12.5–50 mg qhs
Quetiapine	25 mg qhs	Increase in 25-mg increments q3d	25–100 mg qhs
Valproic acid‡	125 mg qhs	Increase in 125-mg increments q3-7d in bid to tid dosing	250 mg qhs to 500 mg tid
Carbamazepine‡	100 mg qhs	Increase in 100-mg increments q3-7d in bid to tid dosing	200 mg qhs to 200 mg tid

**Disclaimer:* The choice of which agents to use and which dosing schedules to recommend must be individualized. It is the responsibility of the clinician to consider potential side effects, drug interactions, allergic response, life-threatening reactions (e.g., leukopenia with clozapine), dosing changes due to renal or hepatic dysfunction, etc., before administering any drug to any patient, including those listed above. Drs. Petit, Montplaisir, St. Louis, Boeve, their respective institutions, and Elsevier will not be held responsible for any adverse reactions of any kind to any patient regarding the content of this information.

†If insomnia is problematic, the second dose should be given no later than the evening meal.

‡Requires periodic laboratory monitoring; refer to the manufacturer's instructions for laboratory monitoring.

Modified from Boeve BF. Update on the diagnosis and management of sleep disturbances in dementia. *Sleep Med Clin* 2008;3(3):347–60.

was ineffective in treating insomnia in patients with AD in multicenter, placebo-controlled trials.[159,160]

Simple but effective interventions should be tried first, such as instituting proper sleep hygiene (e.g., regular schedule, bedtime routine), limiting caffeine and alcohol intake, and increasing activity and exercise during the daytime with avoidance of prolonged daytime napping. One recent randomized controlled trial of walking, bright-light exposure, or combination therapy in community-dwelling adults with possible or probable AD found that all three active interventions decreased night time spent awake.[161] Before prescribing medication to treat insomnia, the clinician should keep in mind that many hypnotic agents (especially benzodiazepines) can exacerbate cognitive deficits and OSA and can even induce daytime sleepiness because of carryover effects. Use of sedative-hypnotics is associated with longer hospital stays and significantly increased fall risk in the acute care setting.[162] If no cause is found for the insomnia, trazodone or chloral hydrate may be considered, although some studies have shown little alteration in actigraphic sleep parameters in institutionalized patients with dementia receiving sedative-hypnotic medications.[159,163]

In some cases, insomnia can result from unrecognized or untreated RLS. In two recent studies, RLS occurred in approximately 4% to 5% of patients with dementia, and RLS is probably at least as common as in the general population.[152,164] RLS may be difficult to diagnose in patients with cognitive impairment, and in one recent study analyzing probable RLS in 59 patients with dementia, two expert raters found that probable RLS occurred in 24% of patients and that probable RLS was associated with nocturnal agitation behaviors.[165] In another recent study, patients with RLS and early dementia often showed repetitive mannerisms and restlessness, and RLS behaviors were associated with selective serotonin reuptake inhibitor use and a polysomnographic PLMS index greater than 15 per hour.[166] Several medications, especially dopamine agonists, have been proved efficacious and well tolerated in nondemented persons with RLS (for review, see Chapter 97). However, to our knowledge, no study has assessed the efficacy and safety of these agents in demented patients. In some patients, dopaminergic agents can cause insomnia because of their stimulating effects or can trigger or exacerbate psychosis.

Excessive Daytime Sleepiness

Excessive sleepiness during the day has been reported mainly in PD. Daytime sleepiness may result from poor sleep, dopaminergic therapy, or comorbid OSA, or it may be due to PD itself.[167] Somnolence not resulting from another primary sleep problem can also affect patients with AD, DLB, and FTD. In such cases, methylphenidate (at a low dose), modafinil, or armodafinil can be effective in improving alertness without producing significant adverse effects, although careful monitoring of blood pressure is indicated to ensure hypertension does not evolve during stimulant medication administration.

Hypersomnolence can also result from OSA, a condition often associated with degenerative disorders, especially AD and MID. The relationship between sleep apnea syndrome and dementia is complex. On one hand, it is known that sleep apnea is associated with cognitive deficits, some of which may be improved with continuous positive airway pressure (CPAP) therapy.[168] There have been cases of patients with OSA in whom dementia was subsequently diagnosed and whose dementia subsided with CPAP therapy.[169] One study showed that long-term CPAP treatment succeeded in slowing the cognitive deterioration and improving sleep and mood in patients with AD and OSA.[170] However, our clinical experience indicates that only a minority of patients with a dementing illness significantly improve both functionally and on psychometric testing with CPAP therapy, but a significant proportion of patients tolerate CPAP and use it nightly, and spouses enjoy a more consolidated sleep when their bed partners with dementia are on CPAP therapy.[170]

Circadian Rhythm Disorders

Several studies have demonstrated disruption of the sleep-wake rhythm in patients with dementia, especially in AD and FTD. In fact, insomnia and excessive daytime somnolence can be the manifestation of a primary circadian disorder. Degenerative changes in the biologic clock, the SCN of the hypothalamus, and the pineal gland, resulting in reduced melatonin production, may be responsible for the disorganization and flattening of the circadian rhythms.[6,8] Melatonin can be helpful for sleep-wake cycle disturbances in patients with dementia, improving sleep, reducing sundowning, and slowing the progression of functional impairment in AD.[171,172] Bright-light therapy administered in the evening has been found effective to alleviate sleep-wake cycle disturbances in patients with dementia and to better consolidate their nighttime sleep.[173-175] In some patients, regular daylight exposure is also effective for day-night reversal problems.

Excessive Motor Activity during the Night

RBD is prevalent in various degenerative conditions, especially in the synucleinopathies, compared with AD, FTD, and PSP.[15] There is a high interpatient variability in the severity of RBD, but the symptoms generally decrease with disease progression. It is important to differentiate RDB from nocturnal wandering by taking a careful history. When diagnosis is uncertain and the potential for injury is present, a polysomnographic and video recording is justified. The first step in the management of RBD is to ensure the safety of the patient by removing potentially dangerous objects from the bedroom, placing a soft mattress on the floor next to the bed, and removing any firearms from the bedroom. Clonazepam, the traditional treatment of choice for RBD in nondemented persons, can potentially worsen cognition and can aggravate OSA. Before prescribing this agent, it is essential to ensure that the patient does not experience OSA or that patients with OSA are adherent to nasal CPAP therapy at an effective treatment pressure. Clinical experience suggests that clonazepam is well tolerated and produces few or no cognitive side effects in most patients with dementia and concomitant RBD. Melatonin has also been shown to improve RBD symptoms.[176-179] If depression is also present, treatments other than nefazodone should be considered because this drug increases REM sleep, contrary to most other antidepressants, and can therefore potentiate RBD; agomelatine, a melatonergic antidepressant, was recently reported to improve RBD symptoms and may be promising for use in those with RBD and comorbid depression.[180] The cholinesterase inhibitor rivastigmine also improved RBD symptom frequency in a pilot treatment trial in patients with PD.[181] A higher level of evidence from well-powered, definitive randomized controlled

treatment trials is necessary to support specific RBD treatments.[182]

The prevalence of PLMS in dementia has not been estimated exactly. However, it is known to be elevated, especially in the synucleinopathies.[183] Without a polysomnographic recording, the severity and clinical significance of PLMS are difficult to assess. If they are bothersome to the patient or cause daytime sleepiness as a result of sleep fragmentation, treatment with dopaminergic agonists can be considered. Dopaminergic agents should be used particularly cautiously in patients with psychotic features.

One of the heavier burdens on families of elderly demented patients and a primary cause of institutionalization is the lack of sleep because of nocturnal agitation or nocturnal wandering. Nocturnal agitation could be the result of discomfort (constipation, full bladder, clothing, heat, cold), pain (pressure sores, infection), or environmental interruptions (staff noise, light); hence, verifying potential sources of discomfort and pain is crucial. As for insomnia management, eliminating alcohol and restricting caffeine intake to the morning can improve nocturnal agitation. Behavioral techniques should be tried before resorting to psychotropic or sedative-hypnotic medications. However, if necessary, medication options including atypical neuroleptics (risperidone, olanzapine, clozapine, quetiapine), antiepileptic drugs (carbamazepine, lamotrigine, valproic acid, gabapentin), benzodiazepines (clonazepam, lorazepam), trazodone, or chloral hydrate can be effective in treating nocturnal agitation (see Table 96-1 for dosages and titration schedule). Cholinesterase inhibitors can significantly reduce hallucinations for patients who are frightened or significantly bothered by them. In these patients, medications with hallucinatory side effects (levodopa, dopamine agonists, anticholinergics, amantadine, selegiline) should be decreased or eliminated.

CONCLUSIONS

Sleep disturbances are frequent in patients with dementia. Although a common pattern of sleep impairment can be observed in dementia, the study of sleep variables and quantitative EEG in different states are valuable tools in aiding diagnosis and evaluating for pharmacologic treatment.

CLINICAL PEARLS

- Patients with dementia often have important and disturbing sleep problems that can lead to premature institutionalization.
- When managing sleep disturbances in patients with dementia, clinicians should carefully look for underlying causes.
- Nonpharmacologic treatments and basic sleep hygiene principles should be undertaken before considering psychotropic or sedative-hypnotic medications that could exacerbate cognitive deficits or OSA, which should also be promptly diagnosed and treated when identified.
- When a pharmacologic treatment is necessary, one useful approach is to directly target the symptoms, which can be grouped in four categories: (1) insomnia or fragmented sleep, (2) excessive daytime sleepiness, (3) alteration in the sleep-wake circadian rhythm, and (4) excessive motor activity during the night, including RBD, PLMS, and nocturnal agitation or wandering.

SUMMARY

Sleep disturbances are frequent in patients with dementia. Sleep is usually more fragmented, with more frequent awakenings and a longer duration of time awake; SWS is decreased; sleep spindles and K-complexes are less well formed or less numerous, so sleep stages are more difficult to distinguish; and REM sleep may be reduced. Quantitative analyses often show slowing of the EEG during wakefulness and, in AD, during REM sleep. Patients with dementia frequently present with comorbid sleep disorders, including OSA, PLMS, and RBD. As dementia advances, sleep disturbances and EEG slowing typically worsen in parallel with progression of neurodegeneration.

ACKNOWLEDGMENTS

Dr. Montplaisir is supported by grants from the Canadian Institutes of Health Research. Dr. Boeve is supported by the Alzheimer's Association, Mangurian Foundation and by grants AG006786, AG015866, and AG016574 from the National Institute on Aging. Dr. St. Louis is supported by the Mayo Clinic CCaTS grant 1 UL1 RR024150-01.

Selected Readings

Arnulf I, Nielsen J, Lohmann E, et al. Rapid eye movement sleep disturbances in Huntington disease. *Arch Neurol* 2008;**65**(4):482–8.

Boeve BF. Update on the diagnosis and management of sleep disturbances in dementia. *Sleep Med Clin* 2008;**3**(3):347–60.

Boeve BF, Silber MH, Ferman TJ, et al. REM sleep behavior disorder and degenerative dementia: an association likely reflecting Lewy body disease. *Neurology* 1998;**51**(2):363–70.

Gagnon JF, Fantini ML, Bedard MA, et al. Association between waking EEG slowing and REM sleep behavior disorder in PD without dementia. *Neurology* 2004;**62**(3):401–6.

Hita-Yanez E, Atienza M, Gil-Neciga E, Cantero JL. Disturbed sleep patterns in elders with mild cognitive impairment: the role of memory decline and ApoE epsilon4 genotype. *Curr Alzheimer Res* 2012;**9**(3):290–7.

Kundermann B, Thum A, Rocamora R, et al. Comparison of polysomnographic variables and their relationship to cognitive impairment in patients with Alzheimer's disease and frontotemporal dementia. *J Psychiatr Res* 2011;**45**(12):1585–92.

Landolt HP, Glatzel M, Blattler T, et al. Sleep-wake disturbances in sporadic Creutzfeldt-Jakob disease. *Neurology* 2006;**66**(9):1418–24.

Montplaisir J, Petit D, Lorrain D, et al. Sleep in Alzheimer's disease: further considerations on the role of brainstem and forebrain cholinergic populations in sleep-wake mechanisms. *Sleep* 1995;**18**(3):145–8.

Osorio RS, Gumb T, Pirraglia E, Alzheimer's Disease Neuroimaging Initiative. Sleep-disordered breathing advances cognitive decline in the elderly. *Neurology* 2015;**84**(19):1964–71.

Pao WC, Boeve BF, Ferman TJ, et al. Polysomnographic findings in dementia with Lewy bodies. *Neurologist* 2013;**19**(1):1–6.

Peter-Derex L, Yammine P, Bastuji H, Croisile B. Sleep and Alzheimer's disease. *Sleep Med Rev* 2015;**19**:25–38.

A complete reference list can be found online at ExpertConsult.com.

Chapter
97

Epilepsy, Sleep, and Sleep Disorders

Milena K. Pavlova; Sanjeev V. Kothare

Chapter Highlights

- Seizures may disrupt sleep, whereas sleep loss and sleep disorders may worsen epilepsy. This chapter describes the complex and bidirectional interactions of sleep and epilepsy.
- Epilepsy is a chronic disease, and seizures often occur at specific times of day and in certain stages of sleep.
- This chapter describes epilepsy syndromes that are manifested with predominantly nocturnal seizures.
- The differentiation between nocturnal seizures and parasomnias can be challenging, and this chapter provides clinical pearls that help better identify seizures.

Epilepsy and sleep disorders are considered by many to be common bedfellows. Sleep can affect seizure occurrence, threshold, and spread, whereas epilepsy can have a profound effect on the sleep-wake cycle and sleep architecture. Many factors can contribute to sleep disruption in patients with epilepsy, including inadequate sleep hygiene, coexisting sleep disorders, circadian rhythm disturbances, epilepsy per se, seizure frequency, and effect of antiepileptic medications.

WHAT IS EPILEPSY?

Epilepsy is characterized by the tendency to have repeated, unprovoked seizures. The seizures can occur spontaneously or be triggered reflexively by flashing lights or other sensory stimuli. The diagnosis of epilepsy[1] requires one of the following:

- At least two unprovoked (or reflex) seizures occurring more than 24 hours apart
- One unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring during the next 10 years
- A diagnosis of an epilepsy syndrome

Epilepsy may result from genetic, structural, metabolic, traumatic, infectious, or unknown causes.

An individual seizure is the transient occurrence of signs or symptoms due to abnormal, excessive, or synchronous neuronal activity in the brain. A seizure can be focal (or partial) if it originates from a process within one hemisphere or generalized if it arises from both hemispheres. The seizure can be further classified by the type of motor activity observed as hypomotor, hypermotor, or automotor; it may include tonic or clonic movements; or it may be atonic (sudden loss of muscle tone) or myoclonic. When the seizure is focal, a variety of activity can be observed, depending on the brain region that leads to the seizures (epileptogenic zone), such as a stereotypic phrase or movement; stereotypic sensation (e.g., a specific sense of smell or hallucination); versive movement of head, eyes, or extremities; automatisms; or sometimes an abrupt cessation of activity. These abnormalities may be associated with alteration of consciousness.

Seizures can be identified electrographically by distinct, rhythmic, epileptiform patterns (Figure 97-1) that disrupt the normal electroencephalographic (EEG) background, evolve in amplitude and frequency, often spread to involve other brain regions, and end abruptly, often followed by slowing or suppression of the EEG rhythms of the affected area. In between seizures, patients with epilepsy frequently have EEG abnormalities called spikes and sharp waves, which are sharply contoured waves that stand out from the background and are often followed by a slow wave (Figure 97-2). Spikes are usually 30 to 70 microseconds in duration, whereas sharp waves are 70 to 200 microseconds in duration. Both have similar clinical significance.

Epilepsy Syndromes

Many patients exhibit a constellation of specific clinical and electrographic characteristics that allow identification of a specific epilepsy syndrome. In others, the epilepsy is nonsyndromic. More than 50 distinct epilepsy syndromes have been defined, and the next sections review several of the syndromes that have a consistent association with sleep.

SLEEP-RELATED EPILEPSY SYNDROMES

Syndromes in Childhood

Benign Rolandic Epilepsy

Also called benign epilepsy with centrotemporal spikes, benign rolandic epilepsy is the most common partial epilepsy syndrome in children, with an onset between the ages of 3 and 13 years and remission in adolescence.[2] The typical presentation is a partial seizure with paresthesias and tonic or clonic activity of the lower face, associated with drooling and dysarthria. The seizures are mostly nocturnal, with 55% to 59% of patients having seizures exclusively during sleep.[3] The electroencephalogram shows characteristic central and temporal spikes that occur bilaterally but independently and are potentiated during non–rapid eye movement (NREM) sleep. The discharge rate is increased during drowsiness and light sleep compared with the waking record, with no change in spike morphology. Despite the increased frequency of seizures and spikes during sleep, the sleep architecture is unaffected,

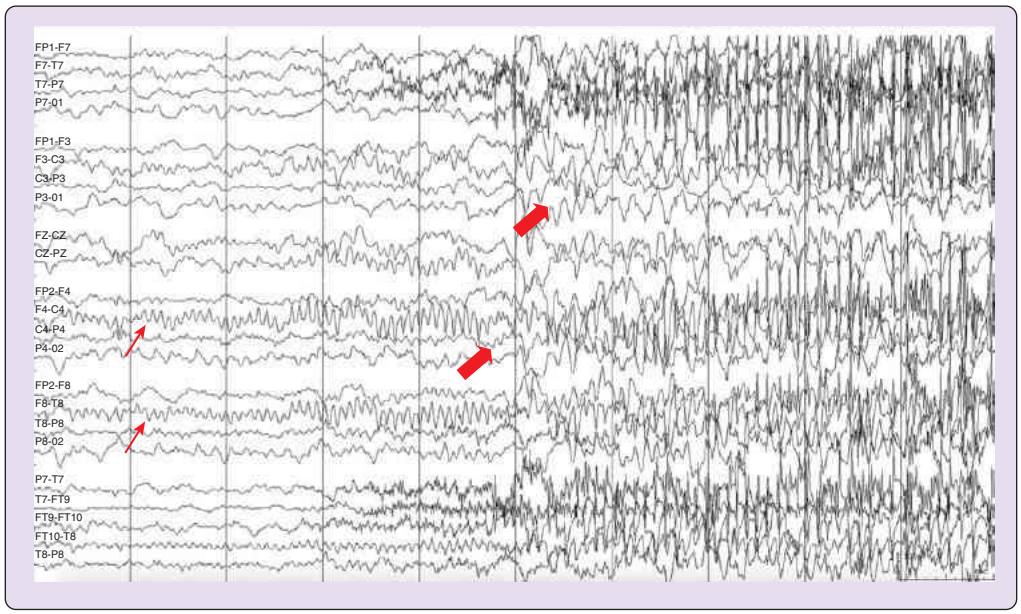


Figure 97-1 Electroencephalogram depicting a focal seizure with secondary generalization. Rhythmic activity starts in the centrotemporal area (F4/C4-F8/T8, *small arrows*), and in the next several seconds, it spreads to both hemispheres (*large arrows*).

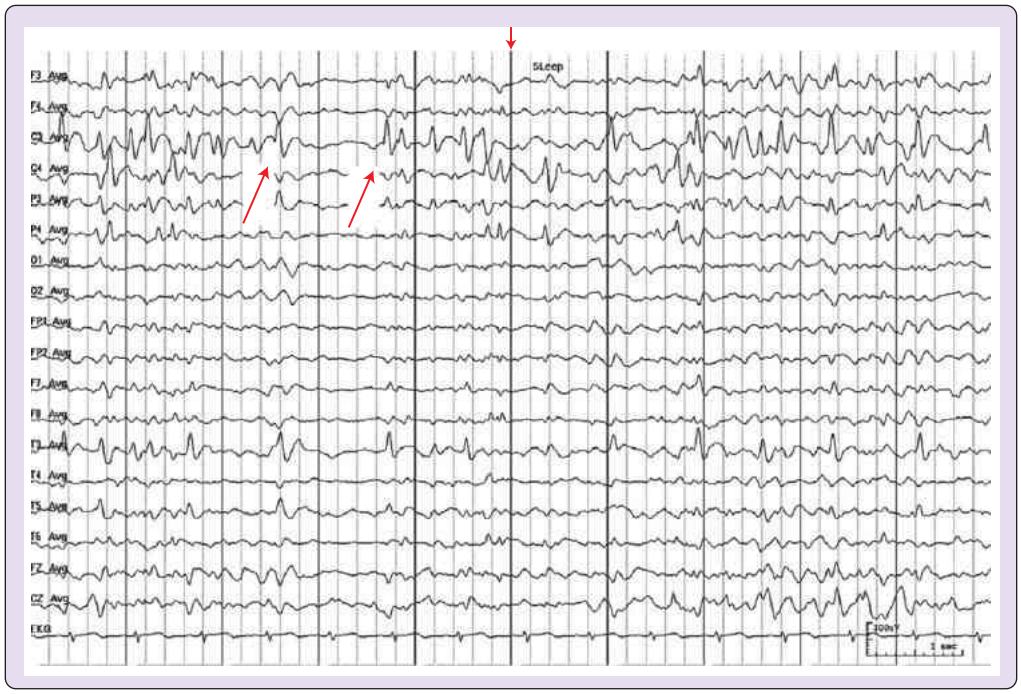


Figure 97-2 Spikes—abnormalities that are commonly seen on the electroencephalogram of patients with epilepsy. Multiple spikes (*arrows*) are seen in a patient with benign epilepsy with centrotemporal spikes. The spikes are maximal in the centrotemporal areas bilaterally, but they occur independently (channels C3-T3-C4-T4).

and sleep is not disrupted. The response to medications is excellent, and the prognosis is universally benign from an epilepsy perspective. However, these children often have deficits in visuospatial short-term memory, attention and cognitive flexibility, picture naming, visuoperceptual skills, and visuomotor coordination. These deficits may be related to nocturnal spiking. Reducing the nocturnal spike index has resulted in improved cognition, albeit at the cost of significant side effects.[4] These spikes are often seen incidentally in patients undergoing sleep studies to rule out obstructive sleep apnea (OSA) who have never had overt seizures.[5]

Benign Occipital Lobe Epilepsy

This infantile variant of benign epilepsy of childhood with occipital paroxysms (also known as Panayiotopoulos syndrome) is another benign epilepsy syndrome seen in children aged 2 to 6 years. It is characterized by prolonged periods of eye deviation and autonomic instability (temperature,

heart rate, respiration, blood pressure), hemiconvulsive and generalized tonic-clonic seizures in sleep, and vomiting on awakening. Interictal electroencephalograms show occipital spikes, whereas ictal electroencephalograms show electrographic seizures emanating from the occipital region during sleep.[6] Almost always, the epilepsy goes into remission within 2 years of onset.

Electrical Status Epilepticus in Slow Wave Sleep

Electrical status epilepticus in slow wave sleep (ESES) is characterized by spike wave complexes "continuously" during NREM sleep but not during wake or REM sleep.[7] The term continuous is applied only to EEG abnormalities with spikes occurring frequently (≥85% or epochs) during NREM sleep, persisting on three or more recordings during a period of 1 month.

Seizure onset typically occurs at 4 to 5 years of age. These seizures are partial or generalized and occur predominantly during sleep, with staring spells (atypical absence seizures) when awake, along with behavioral and language regression. Cognitive decline and mental retardation are noted in 50% of patients. Aggressive treatments to abolish paroxysmal EEG changes include corticosteroids, intravenous gamma globulins, and high-dose antiepileptic medications.

Landau-Kleffner Syndrome

Landau-Kleffner syndrome (LKS) is an acquired disorder with epileptic aphasia in which children, usually 3 to 8 years of age and who have developed age-appropriate speech, experience language regression with verbal auditory agnosia, epileptiform activity during sleep, behavioral disturbances, and sometimes overt seizures, more often in sleep.[7] Seizures arise out of sleep (focal clonic, atypical absences) and are less frequent and less severe than in ESES (absent in 20% to 30%). Behavioral problems are also less severe than in ESES.

There are several similarities between ESES and LKS. Both conditions demonstrate a normal EEG background during wakefulness, with rare focal or generalized spike wave discharges. In ESES, however, discharges during sleep are generalized, whereas in LKS, spike wave activity is mainly in the temporal channels. In ESES, epileptiform activity becomes virtually continuous during NREM sleep, such that it may be impossible to distinguish sleep stages.

Infantile Spasms

This is a catastrophic epilepsy syndrome characterized by a triad of epileptic flexor-extensor spasms of the body, variable intellectual disability, and chaotic (otherwise termed hypsarrhythmic) EEG pattern, with onset between the ages of 3 and 18 months. Interestingly, these spasms tend to cluster on awakening in the morning.[8]

Syndromes Predominantly in Adulthood

Acetylcholine Receptor Mutations

Autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) is an adult epilepsy syndrome that is manifested with nocturnal seizures. Typical onset is in young adulthood (in the 20s), but it may also start in childhood and teenage years.[9] The clinical manifestations vary among individuals, although within the same individual, the seizures are stereotypic. These behaviors can include sudden awakenings with dystonic or dyskinetic movements (in 42% of patients from recent literature),[10] complex behaviors (13%), and sleep-related violent behavior (5%). The corresponding EEG findings include ictal epileptiform abnormalities predominantly over frontal areas in 31% of patients or rhythmic ictal slow wave activity over larger anterior cortical areas in another 47%. The disorder likely results from a mutation in the genes coding for the $alpha_4$ and $beta_2$ subunits of the nicotinic acetylcholine receptor (*CHRNA4* or *CHRNB2*). Currently, this is the only epilepsy syndrome in which the identified cause is an abnormal receptor that is also involved in the regulation of sleep. A third of patients also have associated NREM parasomnias.

Other Epilepsies with Relation to Sleep

Distinct, syndrome-specific patterns of seizures have been described with other syndromes as well. For example, as the name suggests, generalized tonic-clonic seizures on awakening include generalized seizures that occur in the morning. Juvenile myoclonic epilepsy is characterized by myoclonic, absence, and generalized tonic-clonic seizures, and the myoclonic seizures tend to occur in the morning. A common symptom is myoclonus (a jerk of an extremity) soon after awakening, often before breakfast. In addition to ADNFLE, frontal lobe seizures generally occur at night and during sleep, as described in further detail later.

PATTERNS OF SEIZURE FREQUENCY

The timing of seizures from any cause (e.g., mesial temporal sclerosis, tumor, vascular malformations) often follows a pattern. It depends on time of day and stage of sleep and varies by the epileptogenic onset zone (the part of the brain that leads to the individual seizure).

Effect of Sleep Stage on Seizures

Many studies have examined the frequency of seizures in specific sleep stages and in wakefulness. The most striking and consistent finding is that seizures are extremely rare in REM sleep. A review of 42 studies (scalp and intracranial EEG recordings) that included a total of 1458 patients reported that the lowest number of seizures are seen in REM sleep compared with all other states.[11] More specifically, compared with REM sleep, in wakefulness there are eight times more focal seizures. The highest proportion of seizures occurs in N1 and N2 stages of NREM sleep (respectively, 87 and 68 times more than in REM sleep), whereas in N3 sleep, seizures are slightly less frequent (51 times more than in REM sleep).

It is unclear what characteristics of REM sleep physiology are responsible for this unusual phenomenon. Some researchers hypothesize that the EEG desynchronization of REM sleep may reflect a unique pattern of neuronal connectivity that provides some protection against seizures.

Effect of Circadian Rhythms on Seizures

Even within the same state, the frequency of seizures varies with time of day, possibly because of the effects of endogenous circadian rhythms on brain activity. Several studies have described seizures captured in the hospital during continuous EEG monitoring. Early studies have reported a midafternoon peak in the frequency of temporal lobe seizures[12] with a similar cosinor distribution of seizures in adults with temporal lobe epilepsy, as also observed in an animal model.[13] Seizures originating from different brain locations

had different times of occurrence[13]; 50% of all temporal lobe seizures occurred between 15:00 and 19:00, whereas extratemporal seizures had a different distribution, suggesting that the peak time of seizure frequency varies by epileptogenic region. Further studies in adults using more precise localization techniques (performed with intracranial electrodes) confirmed consistent peaks in the timing of seizures by their location: occipital, between 16:00 and 19:00; parietal, between 4:00 and 7:00; frontal, between 4:00 and 7:00; and mesial temporal, between 16:00 and 19:00, with a smaller peak in the morning between 7:00 and 10:00.[14,15]

In children, the patterns are slightly different. Clonic, atonic, hypomotor, and myoclonic seizures are more common in the daytime, whereas nighttime predominance was noted for automotor and hypermotor seizures, especially during sleep.[16] Generalized and occipital seizures were more common during the day (6:00 to 18:00), whereas a nighttime pattern (18:00 to 6:00) was noted for temporal and frontal seizures, which tend to arise from wakefulness and sleep, respectfully.

All of these studies have significant limitations. All were performed within the hospital, where many activities occur at regular intervals (vital signs, scheduled examinations). Light levels are generally higher, and this may affect circadian rhythms by suppressing melatonin secretion or alter its pattern of secretion. In addition, weaning of antiepileptic medications (often done to facilitate recording of seizures) may also affect the timing of seizures. One study addressed this limitation by including patients with continuous home EEG recording and a diary of symptoms. It revealed a nocturnal pattern for frontal lobe seizures and an evening predominance for temporal lobe seizures.[17]

Dosing of Antiepileptic Medications to Reduce Nocturnal Seizures

The knowledge of this pattern of seizure frequency may be used to optimize therapy. A study[18] reported treatment of 17 children with nocturnal or early-morning seizures who were switched to a proportionally higher dose of antiepileptic medications in the evening and were retrospectively reviewed for seizure outcome and side effects. This differential dosing improved the patient's health, with seizure freedom in 65% of patients (11 of 17) and more than 50% reductions in seizures in 88% (15 of 17).

EFFECTS OF EPILEPSY ON SLEEP

Patients with epilepsy frequently have fragmented sleep as well as excessive daytime somnolence.[19-21] Causes include primary sleep disorders that disrupt sleep (e.g., sleep apnea, limb movements), nocturnal seizures that cause sleep fragmentation, and effects of medications. Insomnia is reported by 40% to 51% of epilepsy patients.[22,23] Furthermore, epilepsy patients with insomnia also have a higher frequency of depressive symptoms as well as poorer quality of life.[22]

Sleep fragmentation is a common complaint, and in many, the cause of sleep disruption may be nocturnal seizures. An early publication included a visual example of individual seizures causing awakenings during polysomnography of a patient with epilepsy who was not treated with antiepileptic medication.[24] After treatment, the patient's sleep became more continuous.

Sleep architecture may also be altered in patients with epilepsy. A decreased amount of REM sleep has been reported by several researchers.[25,26]

Primary sleep disorders are relatively common in patients with epilepsy. A report included the results from a cohort of 40 children with epilepsy[27] who underwent a sleep study because of various sleep complaints. Thirty-three patients (83%) exhibited snoring (42.5%), sleep disordered breathing (obstructive hypoventilation in 12.5%, OSA in 20%, and upper airway resistance syndrome in 7.5%), or periodic limb movements of sleep (10%). Children with poor seizure control had significantly lower sleep efficiency, a higher arousal index, and a higher percentage of REM sleep compared with children who were seizure free or exhibited good seizure control. Patients with epilepsy and OSA had significantly higher body mass index, longer sleep latency, higher arousal index, and lower apnea-hypopnea index but significantly more severe desaturation compared with patients with uncomplicated OSA. A significant proportion of children with epilepsy referred for polysomnography with diverse sleep problems manifested sleep disordered breathing, including OSA, and adults with epilepsy also have a higher prevalence of OSA.[28]

EFFECTS OF SLEEP ON EPILEPSY

Effects of Sleep Deprivation on Seizures

Epilepsy patients frequently identify sleep loss as a major factor that provokes seizures. In a recent study,[29] more than 97% of patients with epilepsy reported at least one factor that provokes seizures, and the top three are acute and probably also chronic sleep loss, fatigue, and stress.

Sleep deprivation is often used in epilepsy monitoring units to increase the frequency of seizures. In addition, interictal epileptiform discharges are also more apparent after sleep deprivation.[30]

Effects of Obstructive Sleep Apnea on Epilepsy

When polysomnography was performed on patients with medication-resistant epilepsy,[31] one third of the patients were found to have OSA (apnea-hypopnea index ≥ 5). Furthermore, older adults with poorly controlled seizures have more frequent OSA.[32] A pilot placebo-controlled trial of the effectiveness of continuous positive airway pressure (CPAP) to help seizure control revealed that among epilepsy patients with OSA, a 50% reduction in seizure frequency was seen more frequently among those who were treated with therapeutic CPAP (32%) than among those receiving sham CPAP (15%). Some patients in this study with large reductions in seizure frequency had only mild OSA.[33] Although this study was not powered to detect subtle differences or for stratification by apnea severity or patient characteristics, the study's overall findings support the notion that treatment of the sleep apnea leads to better seizure control.

In a study of CPAP compliance in adults with epilepsy and OSA, 28 patients were CPAP compliant and 13 were not CPAP compliant.[34] In the compliant group, CPAP use reduced seizure frequency from 1.8 per month to 1 per month ($P = .01$). In the noncompliant group, no significant difference in seizure frequency was noted between baseline (2.1 per month) and follow-up at 6 months (1.8 per month; $P = .36$). Sixteen of the 28 CPAP-compliant subjects became seizure

free, whereas only 3 of 13 non–CPAP-compliant subjects were seizure free (relative risk, 1.54; $P = .05$). Thus, good CPAP compliance in patients with epilepsy and OSA can reduce the frequency of seizures.

Similar findings have been seen in children. A study observed 27 children with epilepsy and OSA who were treated with adenotonsillectomy.[35] Three months after the surgery, 10 patients (37%) became seizure free, 3 (11%) had more than 50% seizure reduction, and 6 (22%) had smaller reductions in seizure frequency, whereas 2 (7%) demonstrated unchanged seizure frequency and 6 (22%) manifested a worsening of seizure frequency. The median seizure frequency per month was 8.5 (interquartile range, 2–90) before surgery and 3 (interquartile range, 0–75) after surgery, with a 53% median seizure reduction. Multivariate analysis demonstrated a trend toward seizure freedom with each percentile increase in body mass index and early age at surgery. Thus, adenotonsillectomy for OSA in children may decrease seizure frequency, especially in children with elevated body mass index scores and younger age at time of surgery.

EFFECT OF EPILEPSY TREATMENTS ON SLEEP

Antiepileptic treatment may affect sleep.[36-38] Effects vary by type of medication and comorbidities. In general, with improvement of seizure control, the regularity of the sleep cycle improves and sleep becomes more consolidated. However, some antiepileptic medications have been associated with insomnia, and others with excessive daytime sleepiness.[36]

Vagus nerve stimulation (VNS) is a nonpharmacologic therapy approved for use in patients with refractory epilepsy. An implanted device is programmed to stimulate the vagus nerve with electrical impulses and is typically used when pharmacologic therapy has been unsuccessful. VNS may also affect sleep, specifically sleep architecture and breathing. Some studies have found improvement of sleepiness and sleep architecture with VNS treatment.[39] However, VNS may also increase the incidence of sleep-related breathing disturbances, leading to OSA.[40] Therefore, evaluation and treatment of OSA, if it is present, may be warranted for patients undergoing VNS therapy.

DIFFERENTIAL DIAGNOSIS OF NOCTURNAL SEIZURES FROM OTHER EVENTS

Clinically, it can be a challenge to distinguish nocturnal epileptic seizures from movement disorders, psychogenic nonepileptic seizures, and parasomnias. Unlike in periodic limb movement disorder, the rhythmic movements seen during an epileptic convulsion are much faster in frequency and have a distinct beginning, evolution, and end. Psychogenic nonepileptic seizures rarely arise during sleep, although individual patients have been described.[41] Psychogenic events are usually longer, have a waxing-waning pattern, and usually occur when the event can be witnessed by others (Table 97-1).

Most critical is distinguishing nocturnal seizures from a NREM parasomnia. As the events occur at night, history is often sparse, and the immediately available test results can be difficult to interpret, which can lead to incorrect diagnosis. For example, several authors reported that more than half of the patients with ADNFLE have been incorrectly diagnosed with a parasomnia.[10,41] The distinction is difficult because both disorders occur at night, impair sleep, and can be worsened by stress or sleep fragmentation. In addition, both types of events can be associated with amnesia for the event, and the interictal epileptiform abnormalities, which are so helpful in the positive identification of epilepsy, are uncommon in patients with any form of frontal lobe epilepsy and thus may be absent on a routine 30-minute electroencephalogram or even overnight recording.

Helpful elements in the presenting history include age at onset, duration, occurrence of multiple events in the same

Table 97-1 Differential Diagnosis of Nocturnal Seizures Versus Parasomnia

Characteristic	Seizure	Parasomnia
Age at onset	Variable	Usually childhood onset
Course over time	Stable	Typically disappears in adulthood
Stage of sleep	NREM, more frequently N2	Slow wave sleep
Time of night	Any time, often first half	Typically first third of the night
Duration	~30 seconds to 2–3 minutes	Few minutes, may last as long as ~30 minutes
Course of the individual event	Beginning, evolution, and end	Can be waxing-waning
Type of behavior	Stereotypic, nonpurposeful versive movements and dystonic posture can be seen	Complex, variable from event to event; can appear purposeful
End of the event	Abrupt	Gradual emergence of consciousness
Number of events in the same night	Often multiple (>3)	1–2
Electroencephalography during the event	When visible, focal rhythmic activity with evolution in time and space and abrupt end can occur; muscle artifact may often obscure the recording	Normal; muscle artifact may often obscure the recording
Electroencephalography between events	Interictal discharges are highly specific; however, a normal recording does not rule out epilepsy	Normal

night, event frequency, and any description of the events. Derry et al[42] created a standardized instrument, the Frontal Lobe Epilepsy and Parasomnias scale, which systematizes the approach to history. Typically, characteristics that indicate a higher likelihood of nocturnal seizures include a relatively short duration (<2 minutes), stereotypic behaviors, clustering (multiple events in the same night), and prominent dystonic posturing or tonic limb extension. This scale has been successfully used in clinical practice.[43]

Capturing an event on a video or EEG recording is extremely helpful. However, the absence of typical EEG features does not completely rule out seizures. Muscle artifact from movement can obscure the EEG findings, and a focal seizure may simply not be visible with a standard polysomnogram montage.[44] Using an extended EEG montage better detects focal seizures as well as interictal discharges.[45]

Regarding analysis of the video recording, Derry et al[46] proposed a decision tree to distinguish seizures from NREM parasomnias, based on the following characteristics:

- Whether there is a clear arousal to full consciousness and how this occurs: if the patient does arouse, whether this occurs abruptly or gradually and whether the patient remains supine or prone or engages in more complex behaviors with sitting or getting up from bed after the event.
- Presence of any versive movements or dystonic posturing during the event. Versive movements or dystonic posture, or an abrupt ending with the patient remaining prone, would suggest a seizure. Complex behaviors and getting up from bed would point toward parasomnia.

Capturing the individual events on a single-night polysomnogram is difficult as these events are relatively rare, and they may not occur on the night of recording. Extending the length of the recording may be helpful in many clinical situations. The most reliable diagnosis is achieved by hospital admission with continuous video-EEG monitoring with the goal of capturing one or more events. This allows a long recording, which increases the likelihood of capturing interictal abnormalities; examination of the EEG recording for ictal abnormalities; and review of the video for any typical ictal phenomena or, if multiple events are captured, for stereotypy. In situations in which this is not economically or logistically feasible, an outpatient continuous EEG recording can be considered. The yield of outpatient testing is limited by lack of video as well as by the potential for loss of EEG signal from disconnected electrodes. The typical length of ambulatory EEG recordings is 48 to 72 hours. Longer recordings are technically possible, but with the multitude of activities and movements the patients engage in, the EEG signal progressively deteriorates and eventually becomes difficult to interpret.

CONSEQUENCES OF EPILEPSY

Cardiorespiratory Abnormalities During Seizures

Epilepsy is associated with increased morbidity and mortality. The most devastating consequence is sudden unexpected death in epilepsy patients, a leading cause of death in young and otherwise healthy patients with epilepsy. Sudden death is at least 20 times more common in epilepsy patients than in patients without epilepsy. A significant proportion of patients with epilepsy experience cardiac and respiratory complications during seizures that may contribute to sudden unexpected death.

Respiratory changes often occur with generalized as well as with focal seizures, especially those arising out of the mesial temporal structures. These include central and obstructive apneas; hypoventilation with hypercapnia and oxygen desaturation; and respiratory and metabolic acidosis, bradypnea, and tachypnea[47] (Figure 97-3). Cardiac abnormalities include tachycardia, bradycardia, hypotension, hypertension, tachyarrhythmias, bradyarrhythmias including asystole, and prolongation of the QTc interval.[48]

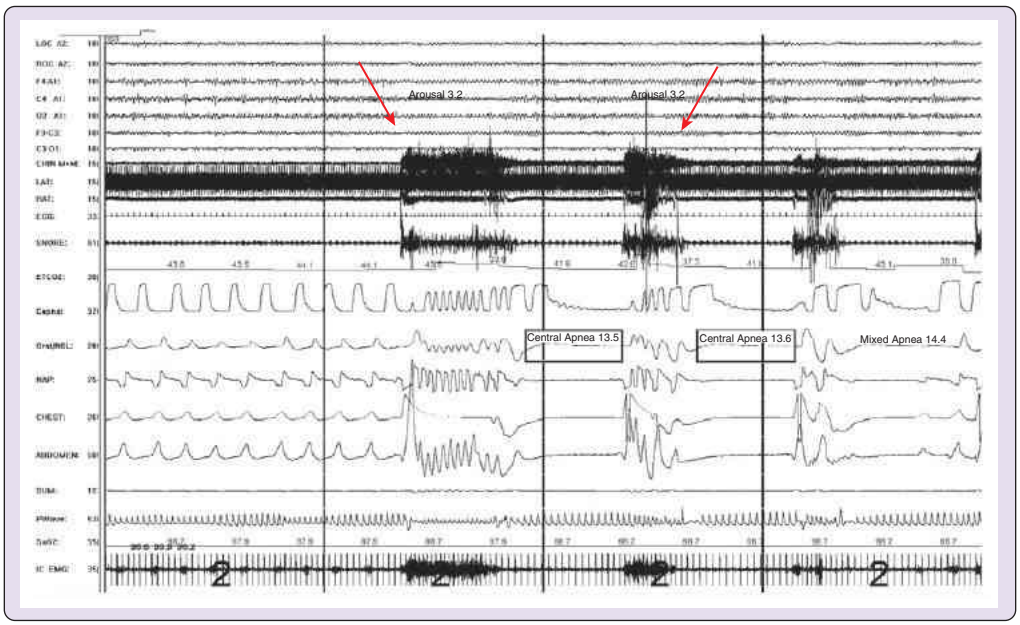


Figure 97-3 Tonic seizures (*arrows*) with tachypnea and tachycardia during the events followed by central apneas.

CLINICAL PEARLS

- Patients with epilepsy have frequent sleep complaints, and sleep disorders may worsen epilepsy.
- Treatment of sleep disorders may improve control of epilepsy.
- Seizures occur in patterns that depend on sleep stage and circadian factors. Integrating the treatment of seizures with knowledge of the chronobiologic pattern may improve treatment.
- Distinguishing seizures from NREM parasomnias is sometimes difficult. Clustering (multiple events in the same night), stereotypic behaviors, and dystonic posture or versive movements suggest seizures, whereas longer duration and complex behaviors are more common with parasomnias.

SUMMARY

Epilepsy is a chronic disease characterized by recurrent seizures. The frequency of seizures can be influenced by sleep stage and circadian rhythms, depending on the seizure focus. Several epilepsy syndromes are manifested with predominantly or exclusively nocturnal seizures. These include syndromes with benign prognosis in childhood, such as benign epilepsy with centrotemporal spikes and benign occipital epilepsy; some adult epilepsies, such as ADNFLE; and other syndromes with poorer prognosis, such as ESES. Other seizures, such as those from temporal lobe origin, tend to occur during wakefulness, most often in the mid to late afternoon.

Nocturnal seizures should be differentiated from movement disorders as well as from parasomnias. Behaviorally, seizures have stereotypic presentation, short duration (30 seconds to 2 minutes), and a tendency for clustering (multiple events in the same night), whereas parasomnias have a more fluctuating course during the night and may include complex, nonstereotypic behaviors. Adequate diagnosis is established by identifying typical EEG patterns (ictal or interictal) as well as by careful analysis of any events captured on video for stereotypic features.

Selected Readings

Bazil CW. Nocturnal seizures and the effects of anticonvulsants on sleep. *Curr Neurol Neurosci Rep* 2008;**8**:149–54.

Delil S, Senel GB, Demiray DY, Yeni N. The role of sleep electroencephalography in patients with new onset epilepsy. *Seizure* 2015;**31**:80–3.

Derry CP, Harvey AS, Walker MC, et al. NREM arousal parasomnias and their distinction from nocturnal frontal lobe epilepsy: a video EEG analysis. *Sleep* 2009;**32**:1637–44.

Ferlisi M, Shorvon S. Seizure precipitants (triggering factors) in patients with epilepsy. *Epilepsy Behav* 2014;**33**:101–5.

Fisher RS, Acevedo C, Arzimanoglou A, et al. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 2014;**55**:475–82.

Hollway JA, Aman MG. Sleep correlates of pervasive developmental disorders: a review of the literature. *Res Dev Disabil* 2011;**32**:1399–421.

Kothare SV, Zarowski M. Sleep and epilepsy: common bedfellows. *J Clin Neurophysiol* 2011;**28**:101–2.

Malow BA, Weatherwax KJ, Chervin RD, et al. Identification and treatment of obstructive sleep apnea in adults and children with epilepsy: a prospective pilot study. *Sleep Med* 2003;**4**:509–15.

Manni R, Terzaghi M. Comorbidity between epilepsy and sleep disorders. *Epilepsy Res* 2010;**90**:171–7.

Pavlova M, Singh K, Abdennadher M, et al. Comparison of cardiorespiratory and EEG abnormalities with seizures in adults and children. *Epilepsy Behav* 2013;**29**:537–41.

Pavlova MK, Woo Lee J, Yilmaz F, Dworetzky BA. Diurnal pattern of seizures outside the hospital: is there a time of circadian vulnerability? *Neurology* 2012;**78**:1488–92.

van den Munckhof B, van Dee V, Sagi L, et al. Treatment of electrical status epilepticus in sleep: a pooled analysis of 575 cases. *Epilepsia* 2015;**56**:1738–46.

A complete reference list can be found online at ExpertConsult.com.

Chapter 98

Other Neurologic Disorders

Antonio Culebras

Chapter Highlights

- Many neurologic conditions can cause sleep abnormalities, especially if they produce lesions in strategic locations such as the brainstem, diencephalon, or thalamus that extensively disturb brain function. Acute processes include stroke, head trauma, diffuse encephalopathies, and the encephalitides, and chronic diseases include multiple sclerosis and neurodegenerative disorders.
- Some chronic neurologic disorders not associated with structural brain alterations, such as headache syndromes, show a peculiar but well-established association with sleep anomalies. Stage-specific headaches include migraine, cluster headache, chronic paroxysmal hemicrania, and hypnic headache.
- The occurrence of stage-specific headaches and the linkage between rapid eye movement sleep and acute headache attacks point to the influence of circadian rhythms and sleep on headaches.
- Severe obstructive sleep apnea may be associated with a high burden of periventricular white matter disease and contribute to vascular cognitive impairment.
- Upper cervical cord lesions may be associated with hypoventilation and obstructive sleep apnea syndrome.
- Brain tumors may be associated with a variety of sleep disorders, depending on their location.

Sleep is a function of the brain. In consequence, sleep disorders can occur when the brain falters. Many structural brain lesions like stroke, multiple sclerosis, tumors, trauma, and the like, as well as dysfunctions without obvious structural alteration, like migraine and cluster headache, can cause sleep alterations that complicate or aggravate the clinical course. Thus it is important for the sleep specialist to be acquainted with neurologic conditions that can contribute to or be associated with sleep pathology.

SLEEP-RELATED HEADACHE

There is an intimate relationship between some headache syndromes and sleep.[1] Practitioners and patients have been aware of the peculiar effect of sleep on terminating attacks of headache,[2] and patients have many times complained of being awakened at night or of waking up in the morning with a headache. The discovery of stage-specific headaches and the intriguing linkage between rapid eye movement (REM) sleep and acute headache attacks[3] highlight the influence that circadian rhythms and sleep can have on the triggering of headaches. In addition, several well-defined sleep disorders are commonly associated with headaches. These include sleep apnea syndrome with headache on awakening, sleep stage–related headache, and parasomnias and headache.

Epidemiologic studies conducted in headache clinics have noted that 17% of the total headache group reports headaches at night or in the early morning before the final awakening period.[4] Up to 55% of patients in the sleep-related headache subgroup had a specific sleep disorder identified by polysomnographic monitoring in a sleep center.

The *International Classification of Sleep Disorders*, third edition,[5] recognizes the category of sleep-related headaches, classified under other sleep disorders. The common expression is the occurrence during sleep or immediately on awakening.

Stage-Specific Headaches

Migraine Headaches

Migraine headaches can occur at night during N3 or REM sleep. Sleep-related migraine attacks are characterized by unilateral throbbing head pain in association with nausea, vomiting, scotomas, visual field defects, photophobia, paresthesias, and even hemiparesis and aphasia. Not all symptoms may be present given the idiosyncratic nature of migraine headaches. Attacks can last for hours to several days. Migraine attacks in children younger than 8 years often resolve after an interval of sleep.[6] Children with migraine can have an increased incidence of disturbed sleep and parasomnias[7] such as somnambulism, night terrors, and enuresis.[8] Prolonged deep sleep is a risk factor for the provocation of sleep terrors and somnambulism as well as for triggering migraine attacks in susceptible patients at any age.[9] Fifty-four percent (64% women, 35% men) of patients with narcolepsy report migraine with the full complement of the International Headache Society criteria.[10]

Although migraine headaches may be provoked by sleep, the most common association is the onset of sleep following a migraine attack. The therapeutic effect of sleep[11] in some

attacks of migraine may be related to serotonin metabolism, but proof is lacking. The trigeminovascular system, which promotes vasodilation and release of calcitonin gene–related peptide and substance P,[12] has been implicated in the mechanism of migraines because calcitonin gene-related peptide is elevated in the jugular venous blood of migraineurs during the attack.[13]

Migraine attacks have been linked to the release of serotonin (5-HT). During migraine headaches, 5-HT is released from platelets and 5-hydroxyindoleacetic acid, the main metabolite of serotonin, is excreted in excess in the urine following the attack.[14] Sumatriptan (an agonist of the 5-HT_1 receptor found in cerebral arteries, where it has an inhibitory effect) aborts the migraine headache. Methysergide (an antagonist of the 5-HT_2 receptor found mainly in temporal arteries, where it has an excitatory effect) also terminates migraines.

Abnormal patterns of hypothalamic hormone secretion such as decreased nocturnal prolactin peak, increased cortisol concentrations, a delayed nocturnal melatonin peak, and lower melatonin concentrations have been reported in patients with chronic migraine.[15]

Proper sleep hygiene is paramount to help prevent sleep-related migraine headaches. Bruni and coworkers[16] evaluated the effect of improving sleep habits in 70 migraineurs with poor sleep hygiene. In their study, the mean duration and frequency of migraine attacks were significantly reduced when proper sleep hygiene was maintained.

Cluster Headaches

Cluster headaches occur in 0.4% of men and 0.8% of women.[17] They are characterized by severe unilateral, periorbital, malar, and temporal pain with lacrimation, rhinorrhea, nasal mucosa engorgement, forehead perspiration, and flushing of the malar area. Attacks are of abrupt onset and termination, often at the same time each day, generally last 2 hours or less, and can recur several times during a 24-hour period. Seventy-five percent of cluster headaches occur predominantly at night between 9 PM and 10 AM.[18] Cluster headaches have been linked with REM sleep and with sleeping late in the morning, a situation that promotes REM sleep, a possible triggering factor. Spontaneous remissions lasting several months are the norm.

Chronic Paroxysmal Hemicrania

Chronic paroxysmal hemicrania is characterized by attacks of severe pain also associated with conjunctival hyperemia, rhinorrhea, and, more rarely, Horner syndrome. Attacks may appear predominantly at night, usually waking the patient at the same hour, sometimes in close linkage with REM sleep, leading to the term *REM* sleep–locked headache. Chronic paroxysmal hemicrania is considered a variant of cluster headache that features several attacks per day of intense pain of 30 minutes' duration or less. The therapeutic response to the administration of indomethacin is quasi-specific and helps confirm the diagnosis of chronic paroxysmal hemicranias.

Hypnic Headache

Hypnic headache is an idiopathic headache disorder of rare occurrence, observed in older age groups. The mean age of onset is 63 years, with a range of 36 to 83 years.[19] The alternative names—clockwise headache and alarm headache—also attest to its regularity and exclusive occurrence at night. Unlike cluster headache and chronic paroxysmal hemicrania, hypnic headache occurs in a diffuse localization in two thirds of patients. Its intensity varies widely, and only one third of patients complain of severe pain. The pain usually lasts longer than 1 hour, and generally there is only one attack per night. Most patients experience the attack during the middle third of the night and report regularity in its occurrence. Less than 10% of patients have associated autonomic symptoms such as lacrimation, nasal congestion, or rhinorrhea, and these are usually mild when they occur.

Results of laboratory tests, including magnetic resonance imaging (MRI) of the head, electroencephalography, and Doppler ultrasound have been invariably normal. Polysomnography has shown occurrence of headache attacks during REM and non-REM sleep.[20] The appearance of hypnic headaches at the same time during the night has led others to suggest this is a chronobiologic disorder.

Hemicrania Horologica

Hemicrania horologica, or clocklike hemicrania,[21] is a very rare disorder, with headaches lasting 15 minutes. They occur with clocklike precision every 60 minutes, day and night. Hemicrania horologica differs from chronic paroxysmal hemicrania in the lack of autonomic signs, the clocklike regularity over 24 hours, and the response to nonsteroidal antiinflammatory drugs or indomethacin. Unlike hypnic headache, attacks also occur during the day.

Headache on Awakening

Headache on awakening occurs in half of patients with sleep apnea syndrome. In at least one study, the frequency of headaches was unrelated to the severity of the sleep apnea.[22] The headache is generally diffuse and of mild to moderate intensity, with a tendency to disappear as the patient becomes active. Successful treatment of sleep apnea syndrome is associated with significant improvement of the headache in 30% of patients.[23] Prolonged afternoon naps may also be followed by headache.

Headaches on awakening in patients with sleep apnea syndrome have been associated with a variety of mechanisms, including hypoxemia, hypercapnia, altered cerebral blood flow, and depression. Headaches on awakening may occur with other disorders. These are common in children with brain tumors, but they appear in only 5% of adults with brain tumors. They may also appear in relation to bruxism, systemic hypertension, depression, muscle tension, alcohol intoxication, and sinus inflammation.

Bruxism, or clenching and grinding of teeth, is a parasomnia that occurs predominantly in stages 1 and 2 non–rapid eye movement (NREM) sleep and sometimes in REM sleep, occasionally leading to headache on awakening (see Chapter 144).

Insomnia and Headache

Insomnia and headache often occur together. In the postconcussion syndrome, insomnia and headache are prominent symptoms. On the other hand, insomnia and daytime fatigue are commonly reported by patients with chronic headache. Chronic headache sufferers feel more tired (especially women) and do not sleep as well at night (especially men).[24] Fibromyalgia occurs in 35.6% of patients with chronic migraine, also known as transformed migraine; this group of patients has a higher incidence of insomnia.[25]

Headaches in Children

Chronic headaches in children are often associated with sleep alterations. Common sleep disorders are decreased duration of sleep at night, poor sleep hygiene, increased number of nocturnal awakenings, somnambulism, somniloquy, enuresis, and snoring.[26] One study found that children with headaches have a significantly higher prevalence of excessive daytime sleepiness, narcolepsy, and insomnia compared with children without headaches.[27] This contradicts previous work stating that children with headaches have a higher prevalence of sleep apnea, restlessness, and parasomnias, an association that may be more specific for genuine migraines and not for headaches in general. The authors conclude that pediatricians should inquire about daytime sleepiness, narcolepsy, and insomnia in children with headaches.

Drugs and Headache

A variety of drugs used for the treatment of sleep disorders can cause headache as a prominent adverse event. In an analysis of several placebo-controlled studies, modafinil was associated with headache in 34% of subjects compared with 23% of those taking placebo.[28] Headache has also been cited as a common adverse effect of both melatonin and ramelteon for treatment of insomnia.[29,30] Common side effects of amphetamines during long-term treatment in patients with narcolepsy include headache along with irritability, bad temper, and profuse sweating.[31]

Exploding Head Syndrome

Exploding head syndrome is typically considered in the differential diagnosis of sleep-related headaches because although it does not cause head pain, it occurs during sleep and patients localize symptoms in the head. This syndrome is characterized by abrupt flashing lights and noises perceived inside the head during the night.[32,33] The attacks last only seconds and terrify patients despite the absence of pain. The episodes have been shown by polysomnography to appear during any stage of sleep.[34] Generally the exploding head syndrome occurs in persons older than 50 years and has a benign prognosis.[35] Reassurance and administration of clomipramine are curative in most instances of exploding head syndrome.

Differential Diagnosis and Diagnostic Workup

Nocturnal migraine, cluster headache, nocturnal paroxysmal hemicrania, and hypnic headache need to be differentiated from other acute severe headaches, such as those associated with intracranial brain tumors, ruptured aneurysm, and meningitis. Patients with intracranial tumors who are awakened at night by headache report improvement on getting out of bed. Headaches on awakening, as observed in sleep apnea patients, are also seen in patients with severe hypertension, depression, intracranial tumor, muscle-contraction headache, alcohol intoxication, and craniofacial sinus disease. Hypnic headache differs from migraine headache, cluster headache, and chronic paroxysmal hemicrania because the pain is commonly diffuse or bilateral and patients are older. Autonomic symptoms are more prominent in cluster headache and chronic paroxysmal hemicrania. The following characteristics should help make a diagnosis of hypnic headache: older age group, punctuality of attack occurrence, mild or no autonomic symptoms, diffuse location, duration of 1 hour to a maximum of 2 hours, tendency to appear in relation to dreams, and no symptoms characteristic of migraine such as photophobia, phonophobia, or nausea.

Causes for concern are first or worst-ever headache, associated neurologic symptoms or signs, progressive worsening of headache over days or weeks, intractable nausea or vomiting, fever, lethargy, confusion, and stiff neck. Patients who exhibit causes for concern should have a neurology consultation, neuroimaging studies, or lumbar puncture. Nocturnal polysomnography is indicated for the study of patients suspected of having sleep apnea syndrome or recurrent parasomnias. Videotaping should always be included in the polysomnographic study of parasomnias.

Patients with migraine, cluster headache, and hypnic headache can wake up with an acute attack more often during REM sleep than during other stages of sleep, and those with cluster headache and chronic paroxysmal hemicrania can suffer the attack at the same time of the night every night. Attacks of chronic paroxysmal hemicrania may be so closely linked to REM sleep that they have been termed REM sleep locked. Polysomnography has been recommended in patients complaining of early morning and nocturnal headaches.[36] In a study of 25 patients with headache, Paiva and coworkers found 21 patients with disturbed sleep, and in 13 of these, the clinical diagnosis had to be reassessed after polysomnography owing to the finding of obstructive sleep apnea (OSA), periodic limb movements in sleep (PLMS), alpha-delta sleep, and insomnia.

MANAGEMENT

Preventive treatment of migraine, cluster headaches, and chronic paroxysmal hemicrania includes good sleep hygiene, with avoidance of precipitating factors such as sleep deprivation, excessive sleep, stress, trauma, and ingestion of certain idiosyncratic foods, including alcohol. Pharmacologic prevention of migraine includes administration of beta blockers, flunarizine, valproic acid, topiramate, calcium channel blockers, serotonin receptor antagonists (methysergide, only for use in periods not to exceed 4 weeks), and 5-HT_2 antagonists (cyproheptadine and methylergonovine). Prevention may also be achieved with antidepressants that interact with serotoninergic receptors such as tricyclic antidepressants, monoamine oxidase inhibitors, and selective serotonin reuptake inhibitors (fluoxetine and sertraline); anticonvulsants, particularly in children with abnormal electroencephalography; and nonsteroidal antiinflammatory agents.[37] Migraine attacks may be aborted with administration of sumatriptan, a serotonin 5-$HT_{1B/1D}$ receptor agonist, given by subcutaneous injection (6 mg, may repeat after 1 hour; limit is two injections in 24 hours). Other abortive medications include ergotamine tartrate, dihydroergotamine, naratriptan, rizatriptan, zolmitriptan, electriptan, frovatriptan, and isometheptene mucate. Rizatriptan and zolmitriptan come in a melting tablet version that can be taken without water. Nonvasoconstrictive abortive agents are butorphanol tartrate for subcutaneous injection and nasal spray. Emergency department management may include narcotic injections with promethazine or hydroxyzine for nausea. Acetaminophen, corticosteroids, and nonsteroidal antiinflammatory derivatives complete the list of medications used to abort attacks of migraine. Symptomatic treatment for migraine attacks includes nonsteroidal antiinflammatory

derivatives, mixed barbiturate and analgesics, antiemetics (promethazine, 50 mg), and, if pain is severe, meperidine (50 mg) or codeine sulfate (30 mg).

Cluster headaches may be prevented with ergotamine derivatives at bedtime (1 to 3 mg sublingual), amitriptyline (150 mg daily), methysergide (6 to 8 mg daily), prednisone (40 mg daily), and lithium carbonate (initial dose 250 mg). For long-term prophylaxis other medications in isolation or in combination may be used; these include verapamil, topiramate, and divalproex. Acute attacks are terminated with inhalation of oxygen; parenteral triptans and dihydroergotamine by mouth may be effective too. Chronic paroxysmal hemicrania responds specifically to indomethacin (50 mg at bedtime or 25 mg three times a day). Morning headaches related to sleep apnea syndrome generally disappear with successful management of the sleep apnea condition. Hypnic headaches have responded to any of the following regimens at bedtime: coffee; ergotamine tartrate, 0.6 mg; phenobarbital, 40 mg, with belladonna, 0.2 mg; atenolol, 25 mg; aspirin, 325 mg, with caffeine, 40 mg; indomethacin, 25 mg; and flunarizine, 5 mg. Successful prophylaxis with lithium carbonate has also been reported. Postconcussion headaches sometimes respond to tricyclic antidepressants. Drug-related headaches tend to respond to conventional analgesics, gradually improve with continued use of the drug, and disappear with discontinuation of the drug.

MULTIPLE SCLEROSIS

Epidemiology

Sleep disorders are more common in multiple sclerosis (MS) than expected by chance. Reports published in the first half of the twentieth century cite cases of MS associated with sleep attacks.[38-40] Subsequently, cases of narcolepsy with cataplexy in MS, familial or not, were reported.[41,42] Some patients with hypersomnia have lesions on MRI that suggest plaques in the hypothalamus as well as undetectable levels of hypocretin in spinal fluid.[43] In general, sleep disturbance is common in MS and can be caused by multiple factors ranging from depression to specific lesions.[44]

There is coincidence of genetic susceptibility between MS and narcolepsy. The susceptibility to MS is coded by genes within or close to the human leukocyte antigen (HLA) DR-DQ subregion.[45] On the other hand, patients with narcolepsy exhibit the highest known association between the HLA-DR2 and -DQw1 antigens and a disease entity. This has led some authors to postulate a common immunogenetic etiology.[46] Others have postulated that the HLA-Dw2 haplotype in patients with MS and narcolepsy extends to the DRB5 locus 76. An overwhelming portion of genetic risk for narcolepsy with cataplexy may be found at the DQB1 locus.[47,48] Formally, however, narcolepsy does not yet qualify as an autoimmune disease in that pathogenic autoantibodies or T cells have not been found to date.

Restless legs syndrome (RLS) is significantly associated with MS, especially in patients with severe pyramidal and sensory disability. The results of a multicenter study involving 861 patients with MS[49] strengthen the idea that the inflammatory damage correlated with MS can induce a secondary form of RLS. In this study, the prevalence of RLS was 19% in MS compared with 4.2% in control subjects. RLS in patients with MS has a significant impact on sleep quality; therefore it should be searched for, particularly in the presence of insomnia unresponsive to treatment with common hypnotic drugs.

Clinical Manifestations

Chronic fatigue is common in MS and can confound the interpretation of sleep disturbances. Patients report difficulty falling asleep, restless sleep, nonrestorative sleep, and early morning awakenings more often than control subjects.[50] A variety of underlying physical and emotional factors (bladder problems, spasticity, muscle spasms, periodic leg movements, depression, and anxiety) that converge to disturb nocturnal sleep should be considered. Excessive daytime somnolence may be secondary to sleep disruption, which is likely amenable to proper management.

In a study of 28 consecutive patients with MS, 54% reported sleep-related problems,[51] including difficulty initiating or maintaining sleep, frequent awakenings because of leg spasms, habitual snoring, and nocturia. Sleep apnea syndrome occurred in two patients, and three showed episodes of nocturnal desaturation. MRI of the brain was abnormal in 20 of 22 cases studied. OSA, in particular, may be highly prevalent and an underrecognized contributor to fatigue in patients with MS.[52]

A polysomnographic study of 25 patients with definite MS showed significantly reduced sleep efficiency and more awakenings during sleep, suggesting a multifactorial etiology of the sleep disorder.[53] Periodic leg movements were found in 36% of patients compared with 8% of controls. Central sleep apneas were found in two patients. MRI of the brain showed more lesions in the cerebellum and brainstem of MS patients with periodic leg movements.

Management

Fatigue is the most pervasive symptom in MS. Amantadine and modafinil have been suggested to improve chronic fatigue in these patients.[54,55] In a single-blind study involving 72 patients with MS, modafinil (200 mg/day) significantly improved fatigue and was well tolerated.[56] Some authors have reported a regression of symptoms of sleep disturbance with dexamethasone[57] or prednisolone therapy.[58] Dopaminergic agonists may be useful for control of RLS and PLMS.

Acute Disseminated Encephalomyelitis

Acute disseminated encephalomyelitis is an acute inflammatory demyelinating disease of the central nervous system (CNS) following viral illness or vaccination. The disorder is probably mediated by immunologic mechanisms.

There is a report of hypersomnia in a 5-year-old girl who presented with hypersomnia due to acute disseminated encephalomyelitis affecting the posterior hypothalamus, brainstem, and basal ganglia.[59] Her total sleep time and slow-wave sleep were increased, but REM sleep was within the normal range. Hypocretin levels in cerebrospinal fluid (CSF) were not assayed. After treatment with intravenous dexamethasone, her hypersomnia improved and the lesions disappeared.

HEREDITARY NEURODEGENERATIVE AND METABOLIC DISORDERS

The occurrence of sleep disorders in this large group of neurologic conditions is limited to case reports and small series

of patients. The list continues to expand with the addition of new reports from the literature.

Clinical Manifestations

Neurodegenerative disorders may have predominantly cerebellar dysfunction as in some forms of multiple system atrophy, extrapyramidal manifestations as in the synucleinopathies, or a combination of cerebellar and sensorimotor signs as in the spinocerebellar atrophies. Patients with neurodegenerative disorders have many sleep-related complaints, including insomnia, hypersomnia, circadian dysrhythmia, abnormal movements, and abnormal behavior during sleep. They are also at risk for the development of sleep apnea syndrome and PLMS. Excessive daytime somnolence secondary to fragmentation of nocturnal sleep caused by sleep apnea, PLMS, or other factors is a common complaint.

Diagnosis

The sleep evaluation of patients with a neurodegenerative disorder should include overnight polysomnography with video monitoring followed by a Multiple Sleep Latency Test (MSLT) in specific cases of associated hypersomnia. The objectives are to assess the presence of sleep apnea syndrome, PLMS, and REM sleep without atonia and to measure excessive daytime somnolence. In some cases, the MSLT shows REM sleep in daytime naps with short-onset REM sleep latencies suggestive of narcolepsy. Actigraphy has been used in some laboratories to document motor activity during sleep and waking that may unveil a circadian dysrhythmia.

Treatment

Management of sleep disorders in patients with neurodegenerative disorders follows the general guidelines. Sedatives and hypnotics should be administered with caution to this group of patients to avoid aggravation of muscle weakness or gait ataxia, not only during daytime hours but also during nocturnal awakenings with ambulation in the dark.

Illustrative Neurodegenerative Disorders

Machado-Joseph Disease

Machado-Joseph disease[60] is also known as type 3 spinocerebellar ataxia. Increased prevalence of RLS and PLMS has been reported in this condition. The clinical evaluation of patients with spinocerebellar ataxia type 3 should pursue possible presence of sleep apnea syndrome and PLMS. REM sleep behavior disorder, a condition also prevalent in the synucleinopathies, has been reported in association with Machado-Joseph disease and may be related to striatal monoaminergic deficit.[61,62] Vocal cord abductor paralysis and stridor have been described in Machado-Joseph disease.[63]

Charcot-Marie-Tooth Disease

Charcot-Marie-Tooth (CMT) disease is a hereditary motor and sensory polyneuropathy characterized by degeneration of peripheral nerves and roots. Patients exhibit distal muscle weakness, atrophy, and sensory impairment. Phrenic neuropathy can cause dysfunction of the diaphragm, leading to chronic hypoventilation, particularly in REM sleep.[64] Vocal cord dysfunction, possibly owing to laryngeal nerve involvement, is found in association with several CMT types and can often mimic asthma.

Patients with CMT disease have a high incidence of significant sleep apnea events whose severity is highly correlated with the severity of their peripheral neuropathy. In one study of 14 patients with CMT type 1, the apnea-hypopnea index (AHI) correlated with neurologic disability but not with body mass index and age. Researchers have hypothesized that sleep apnea in CMT disease can be caused by pharyngeal neuropathy affecting the function of pharyngeal dilator muscles, a phenomenon that increases the collapsibility of the upper airway, leading to recurring obstructive respiratory events.[65,66] Bilevel positive airway pressure may be more appropriate than continuous positive airway pressure (CPAP) for treating sleep apnea in patients with airway weakness, especially those with concomitant restrictive pulmonary impairment.[67]

Niemann-Pick Disease

Niemann-Pick disease type C is a rare autosomal recessive condition characterized by the accumulation of unesterified cholesterol in many tissues and storage of sphingolipids in liver and brain. Adult patients exhibit ataxia, dystonia, dementia, and vertical supranuclear palsy along with hepatosplenomegaly. Some patients report hypersomnia and cataplexy. Recent investigations have shown reduced CSF levels of hypocretin in this condition, which are likely responsible for sleep abnormalities and cataplexy.[68]

ACUTE ENCEPHALITIDES

Sleeping Sickness

Sleeping sickness is a meningoencephalitis caused by the protozoan *Trypanosoma brucei*. The parasite is transmitted to humans by the sting of the tsetse fly in Africa, where 20,000 new cases are reported each year. Sleeping sickness begins with a phase of systemic disease, after which the parasite invades the CNS with manifestations of insomnia and daytime hypersomnia, followed by psychomotor retardation. The disease can progress to extrapyramidal manifestations, ataxia, gait disorder, seizures, coma, and death. Polysomnography reveals sleep-onset REM sleep periods shortly after CNS invasion by trypanosomes, then high-amplitude slow waves, suggesting a diffuse encephalopathy, with preservation of REM sleep parameters until the final phases of the disease.[69,70] CSF analysis has shown increases in cell count, high protein content, and increased immunoglobulin M levels.[71] Neuropathologic studies have shown demyelinating lesions in cerebral hemispheres and brainstem.[72]

Limbic Encephalitis

There are published case reports of severe sleep alterations, including REM sleep behavior disorder, hypersomnia, and insomnia associated with limbic encephalitis that were in some instances reversed with immunotherapy and steroid administration.[73,74]

Nonspecific Viral Encephalopathies

Transient OSA has been described in association with a case of presumed viral encephalopathy.[75] Paradoxical respiratory efforts during sleep, along with frequent episodes of tachycardia, bradycardia, and asystole, led to suspicion of OSA syndrome, which was documented with portable polysomnography. All apnea events and cardiovascular concomitants resolved with CPAP. This case illustrates the occurrence of

sleep-related transient respiratory obstructions in diffuse acute encephalopathy that, if undetected, can lead to serious cardiovascular consequences.

Encephalitis Lethargica

Encephalitis lethargica was described by von Economo and others in the early part of the twentieth century but has now virtually disappeared.[76] The clinical and neuropathologic study of this form of encephalitis brought to light the correlation between hypothalamic lesions and sleep disorders and contributed to understanding the links between basal ganglia lesions and extrapyramidal movement disorders.[77] The most common presentation was acute fever, severe lethargy, and ophthalmoplegia followed by hyperkinesia.[78] In the subsequent months to years, many of these patients developed a postencephalitic syndrome with prominent parkinsonian features.[76]

BRAIN TUMORS

Brain tumors can disrupt sleep-wake cycles by virtue of their location or indirectly by causing intracranial hypertension or hydrocephalus, or both. For example, symptomatic narcolepsy can occur with craniopharyngiomas compressing the floor of the third ventricle that presumably injure the orexin/hypocretin-producing neurons (Figure 98-1).[79] Symptomatic narcolepsy has also been reported with gliomas and colloid cysts of the third ventricle as well as with pituitary adenomas and midbrain gliomas.[80] Sleepiness following hypothalamic injury in the course of resection of an astrocytoma has been reported in association with a low concentration of hypocretin in CSF.[81] Lower brainstem tumors have been associated with severe hypoventilation and respiratory failure during sleep (Ondine's curse) requiring tracheotomy and assisted ventilation. Increased intracranial hypertension and obstructive hydrocephalus have been associated with subalertness and lethargy. Fatigue and drowsiness are prominent symptoms in patients with systemic cancer and brain metastases.[82]

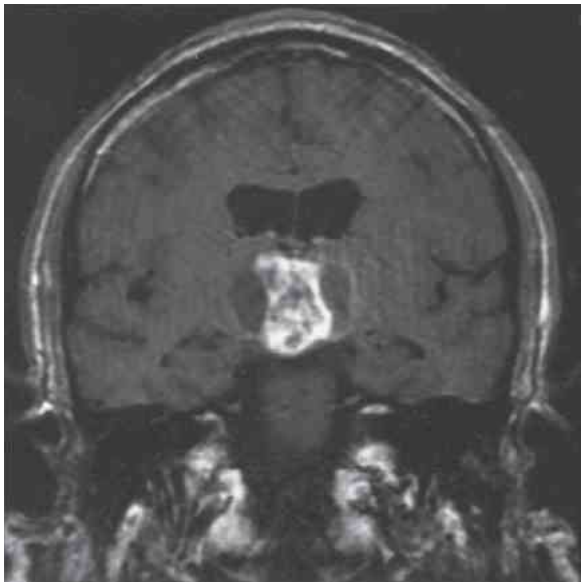


Figure 98-1 Brain magnetic resonance imaging of a 55-year-old man complaining of headaches and excessive daytime somnolence that was initially diagnosed as sleep apnea syndrome. The large cystic mass compressing the diencephalon and floor of the third ventricle was found at operation to be a craniopharyngioma. (From Culebras A. *Clinical handbook of sleep disorders.* Boston: Butterworth-Heinemann; 1996, with permission.)

COGNITIVE IMPAIRMENT AND OBSTRUCTIVE SLEEP APNEA

Clinical evidence suggests that moderate to severe OSA is a risk factor for development of cognitive impairment.

In a prospective study of sleep and cognition in 298 women without dementia, Yaffe and colleagues showed that older women (mean age, 82.3 years) with OSA (defined as an AHI >15 events per hour of sleep) were more likely to develop cognitive impairment.[83] Results were adjusted for age, race, body mass index, education level, smoking status, presence of diabetes, presence of hypertension, medication use, and baseline cognitive scores. The authors found that compared with the 193 women without sleep-disordered breathing, the 105 women with sleep-disordered breathing were more likely to develop mild cognitive impairment or dementia (31% vs. 45%); intermittent hypoxia in apnea or hypopnea also was predictive of mild cognitive impairment or dementia.

Cognitive impairment may be the result of small silent infarctions or subcortical small vessel disease,[84,85] also known as leukoaraiosis. Kim and associates performed polysomnograms in 503 individuals (mean age, 60 years) who were free of previously diagnosed cardiovascular and neurologic diseases.[85] Using MRI, they identified white matter disease in 40%, and regression analysis revealed an association with moderate to severe OSA (AHI >15). The authors concluded that moderate to severe OSA is an independent risk factor for white matter changes in middle-aged and older individuals. Another study showed an association between poor sleep quality and white matter hyperintensity severity as shown by MRI of brain.[85a]

The presence of silent cerebral infarctions in individuals with OSA was investigated in another study.[86] The authors did polysomnograms and brain MRI on 746 participants aged 50 to 79 years and found that 12% of subjects had moderate to severe OSA (AHI ≥15). Subjects with OSA older than 65 years were 2.4-fold more likely to have silent cerebral infarctions and 3.5-fold more likely to have lacunar infarction.

The cause of these vascular lesions is not entirely clear, but intermittent nocturnal hypoxia may contribute to ischemic damage in the cerebral periventricular territory of long penetrating terminal arteries.[87] Intermittent hypoxia adds ischemic burden to this vascular border zone territory with blood flow that may be already precarious as a result of diabetes with vascular autonomic dysregulation and poorly controlled hypertension. Ischemic damage to the cerebral periventricular white matter may disturb the connections of the cortex with the thalamus, leading to a form of subcortical dementia characterized by apathy, decreased executive functions, poor memory, and in advanced cases difficulty walking and urinary incontinence (Binswanger disease).

Treatment of OSA with CPAP may lower cerebrovascular risk by decreasing 24-hour urinary catecholamine excretion, improving arterial stiffness and baroreflex sensitivity, and reducing mean 24-hour ambulatory blood pressure.[88,89] However, CPAP applications will not modify structural

lesions of the brain, and therefore early diagnosis and treatment of sleep apnea before structural brain damage ensues is strongly recommended, particularly in patients with several risk factors for stroke.

SPINAL CORD DISEASE

Pathophysiology

Ventilation depends on the integrity of the spinal cord, and in patients with spinal cord lesions, respiratory dysfunction and consequent sleep disturbances are common. The diaphragm is controlled by the phrenic nerves, which originate in the ventral horn, extending from C3 to the caudal part of C5.

The descending respiratory pathway is formed by crossed fibers situated deep in the anterior white column in the vicinity of the anterior horn projecting mainly from the ventral respiratory group in the medulla. High cervical lesions and phrenic nerve damage can cause unilateral or bilateral paralysis of the diaphragm, depending on the extent of the lesion.

The intercostal muscles receive their innervation through descending pathways located dorsal to diaphragmatic pathways in the vicinity of the lateral spinothalamic tract. Voluntary respiration is mediated by fibers in the lateral pyramidal tract, whereas involuntary automatic respiration is mediated by reticulospinal fibers emerging from the brainstem respiratory centers. Spinal motor nuclei situated in segments T1 to T11 give rise to intercostal nerves that innervate intercostal muscles.

Accessory respiratory muscles receive innervation from cranial nerves XI and nerves C1 to C8. Upper airway muscles involved in nasal, pharyngeal, and laryngeal dilation are innervated by cranial nerves V (tensor veli palatini muscles), VII (levator alae nasi muscles), X (cricothyroid and thyroarytenoid muscles), XII (genioglossus, geniohyoid, sternohyoid, and sternothyroid muscles), and C1 to C4 (geniohyoid, sternothyroid, and sternohyoid muscles). These muscles are unaffected by spinal cord lesions below C5, a lesion compatible with a respirator-free life.

Lesions of the phrenic and intercostal motor neurons in the spinal cord can occur with spinal cord tumors, spinal trauma, spinal surgery (e.g., cervical cordotomy or anterior spinal surgery), and myelitis. Patients with syringomyelia and syringobulbia (fluid-filled cavities in the spinal cord or brainstem, respectively) with dysphonia and dysphagia are particularly prone to severe respiratory disturbances during sleep.[90] In one study of 22 patients with stable spinal cord lesions above T10,[91] 45% had some evidence of OSA syndrome. Cognitive changes in patients with tetraplegia may be related to sleep apnea syndrome.[92] Although excessive daytime somnolence secondary to sleep-related respiratory dysfunction is the most common symptom, patients with spinal cord diseases can complain of insomnia as a result of immobility, neck pain, and central pain syndrome.

Phrenic nerve damage leads to diaphragmatic paralysis. Unilateral paralysis is asymptomatic, but bilateral paralysis is invariably symptomatic and may be life-threatening. Paresis or weakness with partial diaphragmatic dysfunction can cause sleep-related ventilatory insufficiency. In the supine position, patients complain of profound difficulty breathing because of decreased lung volume and increased respiratory effort as the abdominal contents press into the thorax. With bilateral severe or acute phrenic nerve injury, patients present with nocturnal orthopnea, cyanosis, and fragmented sleep followed by morning headaches, vomiting, and daytime lethargy. Phrenic nerve weakness or paralysis is most prominent during REM sleep when the diaphragm is the only functional respiratory muscle. Upper airway resistance is also higher in REM sleep, contributing to decreased ventilatory efficiency. Patients with weak pharyngeal dilator muscles and a weak diaphragm as a result of a diffuse neuromuscular disorder or bilateral phrenic nerve paralysis exhibit the most serious respiratory compromise in REM sleep[93] (see Chapter 96).

Craniovertebral junction malformations or Chiari malformations, with or without syringomyelia and basilar invagination, can produce dysfunction of neurons in the brainstem, cerebellum, cranial nerves, and upper spinal cord. The incidence of OSA is significantly higher in patients with craniovertebral junction malformations, especially if basilar invagination is present.[94] OSA can appear after anterior cervical spine fusion. Guilleminault and coworkers[95] found that placement of anterior cervical plates at the C2 to C4 level reduced the size of the upper airway, causing OSA syndrome. The condition was controlled with positive airway pressure applications.

RLS and PLMS can appear in patients with acute transverse myelitis.[96] PLMS has been reported in patients with syringomyelia.[97]

Treatment

Treatment of sleep-disordered breathing in spinal cord diseases should follow the same general principles as those suggested for neuromuscular disorders. Noninvasive nocturnal ventilation has improved the quality of life and increased survival in many neuromuscular disorders. Patients with significant neuromuscular disorders or high spinal cord disease should be considered candidates for evaluation with polysomnography. In most patients with neuromuscular conditions, the most effective time to introduce noninvasive ventilation is when symptomatic sleep-disordered breathing develops.[98]

A word of caution comes from a study[99] showing that obese patients who have spinal cord injury and are taking antispasticity medications might have a higher risk for developing snoring and OSA. The greatest risk appeared in patients taking diazepam or diazepam and baclofen in combination.

CLINICAL PEARLS

- Headaches on awakening, as observed in sleep apnea patients, are also seen in patients with severe hypertension, depression, intracranial tumors, muscle-tension headache, alcohol intoxication, and craniofacial sinus disease.
- In headache patients, causes for concern are first or worst-ever headache, associated neurologic symptoms or signs, progressive worsening of headache over days or weeks, intractable nausea or vomiting, fever, lethargy, confusion, and stiff neck.
- The association between MS and sleep disorders is more common than expected by chance.
- Sleep apnea in patients with Charcot-Marie-Tooth disease may be a consequence of pharyngeal neuropathy affecting the function of pharyngeal dilator muscles.
- Severe sleep apnea may contribute to vascular cognitive impairment.

SUMMARY

As a function of the brain, sleep is disrupted by neurologic conditions that affect brain mechanisms. Some alterations may be structural, like brain tumors, stroke, and MS, but others are nonstructural like migraine and cluster headaches. Lesions that affect the brainstem, hypothalamus, and thalamus tend to be more disruptive of sleep. There is an intimate relationship between some headache syndromes and sleep. Clinicians have been aware of the peculiar effect of sleep on terminating attacks of headache, and patients often report being awakened at night or of waking up in the morning with a headache. Sleep stage–specific headaches are well known and confirm the intimate relationship between headaches and sleep. The management of sleep-related headaches includes prominently the regularization of sleep patterns and sleep duration. Patients with MS complain prominently of fatigue that may be the consequence of multifactorial sleep alterations. Some sleep disturbances associated with MS like sleep apnea are treatable and should be considered in the overall management of MS. Brainstem disorders of any origin, whether chronic neurodegenerative, acute, or tumoral, can disrupt sleep and wakefulness. There is growing evidence supporting the role of sleep apnea in adding burden to subcortical small vessel disease, a very common condition in old age and a contributor to cognitive decline. In summary, sleep disturbances complicate the course of many neurologic conditions, and their proper management should help alleviate the manifestations of neurologic disease.

Selected Readings

Barloese MC. Neurobiology and sleep disorders in cluster headache. *J Headache Pain* 2015;**16**:562. doi: 10.1186/s10194-015-0562-0.

Botelho RV, Bittencourt LR, Rotta JM, Tufik S. A prospective controlled study of sleep respiratory events in patients with craniovertebral junction malformation. *J Neurosurg* 2003;**99**:1004–9.

Culebras A. Headache disorders and sleep. In: Culebras A, editor. *Sleep disorders and neurologic diseases*. New York: Informa Healthcare USA; 2007.

Del Brutto OH, Mera RM, Zambrano M, et al. Poor sleep quality and silent markers of cerebral small vessel disease: a population-based study in community-dwelling older adults. (The Atahualpa Project). *Sleep Med* 2015;**16**:428–31.

Dziewas R, Waldmann N, Böntert M, et al. Increased prevalence of obstructive sleep apnoea in patients with Charcot-Marie-Tooth disease: a case control study. *J Neurol Neurosurg Psychiatry* 2008;**79**(7):829–31.

Iranzo A, Muñoz E, Santamaría J, et al. REM sleep behavior disorder and vocal cord paralysis in Machado-Joseph disease. *Mov Disord* 2003;**18**:1179–83.

Iseki K, Mezaki T, Oka Y, et al. Hypersomnia in MS. *Neurology* 2002;**59**:2006–7.

Italian REMS Study Group, Manconi M, Ferini-Strambi L, et al. Multicenter case-control study on restless legs syndrome in multiple sclerosis: the REMS study. *Sleep* 2008;**31**:944–52.

Kim H1, Yun CH, Thomas RJ, et al. Obstructive sleep apnea as a risk factor for cerebral white matter change in a middle-aged and older general population. *Sleep* 2013;**36**(5):709–15B.

Kohler M, Stradling JR. Effect of CPAP on stroke risk factors and stroke. In: Culebras A, editor. *Sleep, stroke and cardiovascular disease*. Cambridge, UK: Cambridge University Press; 2013.

Liang JF, Fuh JL, Yu HY, et al. Clinical features, polysomnography and outcome in patients with hypnic headache. *Cephalalgia* 2008;**28**:209–15.

Osorio RS, Gumb T, Pirraglia E, et al. Sleep-disordered breathing advances cognitive decline in the elderly. *Neurology* 2015;**84**:1964–71.

Román GC, Erkinjuntti T, Wallin A, et al. Subcortical ischaemic vascular dementia. *Lancet Neurol* 2002;**1**(7):426–36.

Yaffe K, Laffan AM, Harrison SL, et al. Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA* 2011;**306**(6):613–19.

Zifko UA. Management of fatigue in patients with multiple sclerosis. *Drugs* 2004;**64**:1295–304.

A complete reference list can be found online at ExpertConsult.com.

Sleep Disorders After Traumatic Brain Injury

Chapter 99

Philipp O. Valko; Christian R. Baumann

Chapter Highlights

- Traumatic brain injury (TBI) is among the most frequent causes of chronic disability, especially in younger adults, but the high prevalence and burden of sleep-wake disorders after TBI have only recently been recognized.
- The most frequent sleep-wake disturbances after TBI are conditions with impaired arousal, including excessive daytime sleepiness, hypersomnia (increased sleep need per 24 hours, more than 2 hours/day compared with before TBI), and fatigue. Insomnia is also frequent but probably overdiagnosed because disturbed nighttime sleep due to posttraumatic circadian sleep-wake disorders is often mistaken for posttraumatic insomnia.
- TBI survivors sometimes misperceive their sleep-wake disturbances, and marked underestimation of various sleep-wake disturbances is particularly common after moderate to severe TBI. Because overlooked sleep-wake disturbances may compromise recovery from TBI and impair quality of life, careful examination is mandatory even if TBI patients deny any disturbances.
- Many factors can cause sleep-wake disturbances after TBI, including post traumatic stress disorder and obstructive sleep apnea, and the approach to such patients is therefore challenging and should be multidisciplinary. Current treatment strategies are limited, but recent insights into the underlying pathophysiology give hope for the development of novel and tailored therapeutics.

Each year, roughly 10 million people worldwide sustain a traumatic brain injury (TBI), and men are more often victims than women.[1] TBI represents the leading cause of death and long-term disability among children and young adults.[2,3] Based on a literature review covering the period from 1990 to 2005, the annual incidence of TBI is estimated between 108 and 332 cases per 100,000.[4] Because a large proportion of patients with mild TBI never seek medical attention, the true incidence of TBI is likely far greater; according to a recent epidemiologic study from New Zealand, the incidence may be as high as 749 new TBI cases per 100,000/year.[5] Motor vehicle crashes are the most frequent cause of TBI among young adults, whereas the growing proportion of elderly people has led to an increased frequency of TBI due to falls.[2] Other common causes of TBI are high-contact sports such as American football and boxing and warfare-related events, including blast injuries, blunt force trauma, and penetrating injuries.[6,7]

From a clinical perspective, the acute phase of TBI is characterized by loss of consciousness, anterograde and retrograde amnesia, and other neurologic symptoms. The severity of TBI is clinically classified using the Glasgow Coma Scale (GCS) as mild (GCS 13 to 15), moderate (GCS 9 to 13), and severe (GCS 3 to 8) TBI (Figure 99-1).[8] The core neurologic deficit in acute TBI is a quantitative and qualitative impairment of vigilance, ranging from coma in severe TBI to drowsiness and inattention in mild TBI.

Recovery from TBI may take many months up to several years. Because of the marked heterogeneity in etiology, mechanical impact, secondary injuries (in particular, brain swelling with increased intracranial pressure and subsequent impairment of cerebral perfusion), comorbidities, and individual vulnerability, it is challenging to predict outcome after TBI (Figure 99-2). However, the number of TBI survivors is steadily growing because the incidence of TBI continues to rise, the identification of patients with mild TBI has improved, and mortality after severe TBI has decreased from about 55% to 20% over the past two decades.[9-11] As a consequence, medical and public awareness of the long-term neurologic and neuropsychiatric sequelae has substantially improved over the past years, including better recognition of sleep-wake disturbances.[12-15]

This chapter reviews the current knowledge on sleep-wake disturbances after TBI from a clinical perspective and outlines the management the authors have found useful over the years. For a review of the physiological changes after TBI see Chapter 25.

OVERVIEW OF SLEEP-WAKE DISTURBANCES AFTER TRAUMATIC BRAIN INJURY

In 1949, the British neurosurgeon Sir Hugh Cairns focused his Victor Horsley memorial lecture on disturbances of consciousness and reported on several patients with acute coma induced by TBI.[16] One patient, a young soldier, had sustained severe TBI with prolonged coma and remained hospitalized for more than 5 months. It took him several weeks to regain full consciousness, and after 5 months, he still complained of difficulty "in grasping the meaning of pictures." When observing this gradual recovery of consciousness, Cairns

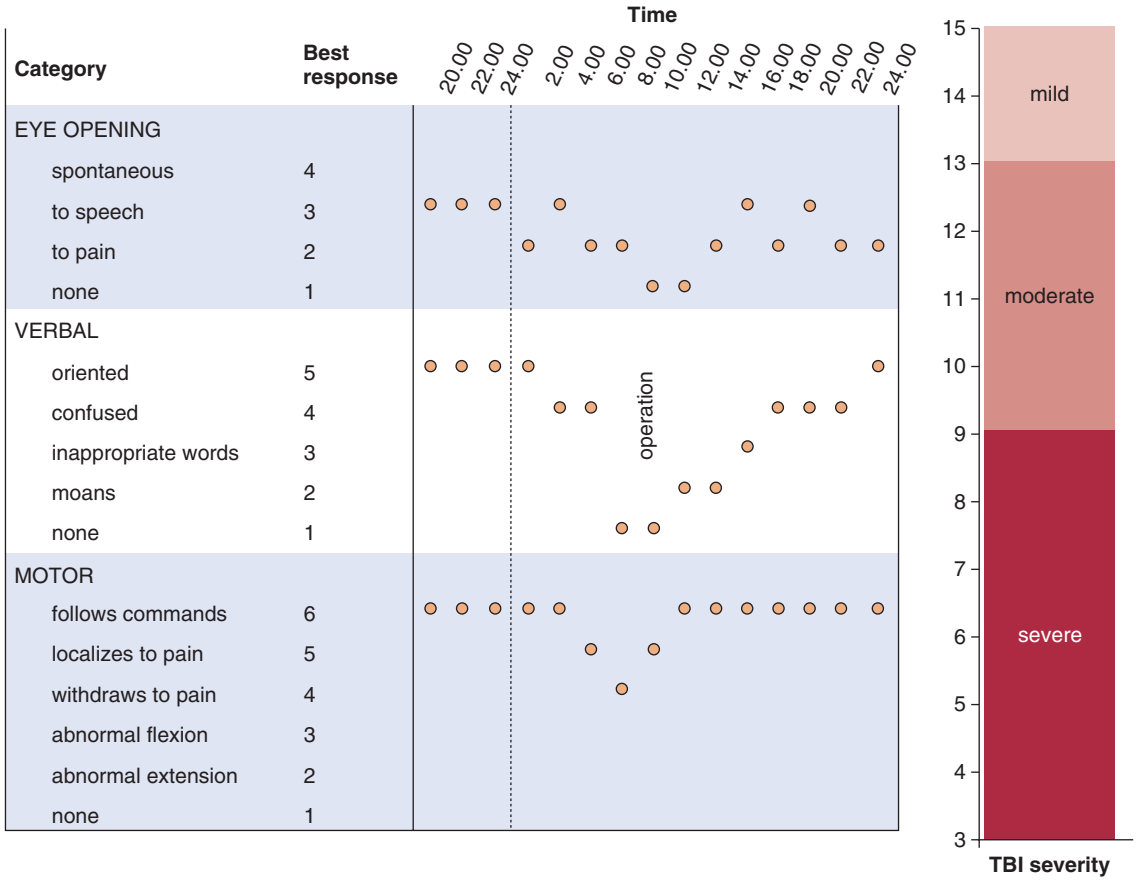


Figure 99-1 Glasgow Coma Scale (GCS), modified from the original 1974 publication of Teasdale and Jennett.[8] The GCS is used for initial classification of traumatic brain injury (TBI) severity and for serial assessment. The figure illustrates how the three items of the GCS in a single patient can be continuously recorded during hospitalization.

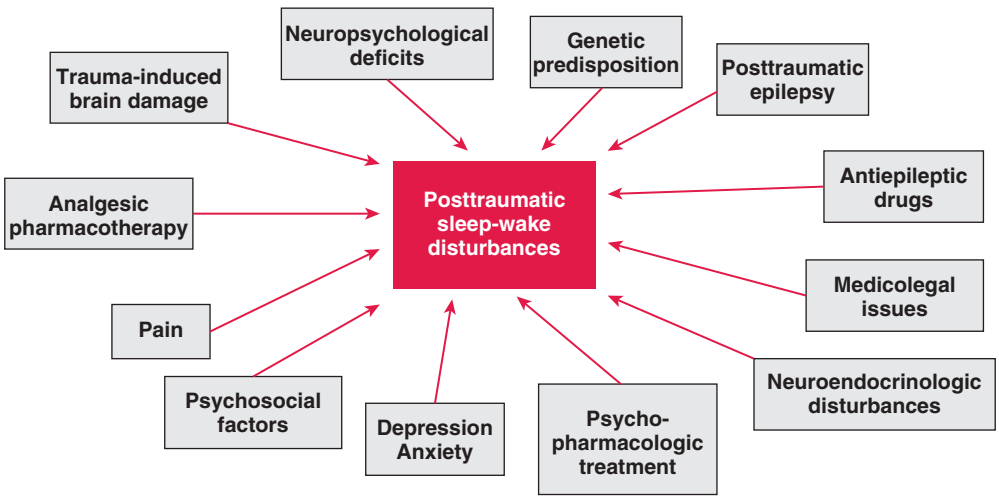


Figure 99-2 Potential contributors to posttraumatic sleep-wake disturbances.

noted: "We have no words to describe these states which lie between coma and full consciousness."[16] This account can be read as an early illustration of the wide range of posttraumatic arousal disturbances. Although Cairns did not use such words as arousal, sleepiness, increased sleep need or fatigue, there is nevertheless no doubt that his patient had all these arousal disturbances at different times. After 5 months, the soldier still had some kind of mental fatigue, but whether he continued to suffer from posttraumatic sleep-wake disturbances is unknown.

In the scientific literature, the problem of posttraumatic sleep-wake disturbances was neglected for a long time. When researchers from Oklahoma some 30 years later observed differences in sleep recordings between controls and TBI patients 6 to 59 months after injury, they wondered, "Do patients continue to show sleep abnormalities more than six months after injury?"[17] Another decade later, a questionnaire-based study from Israel compared frequency and type of subacute and chronic posttraumatic sleep-wake disturbances in 22 inpatients from a rehabilitation center and 77 already discharged patients.[18] Among the hospitalized TBI patients (median, 3.5 months after injury), 73% had sleep-wake disturbances, whereas in the second group, with a median latency since TBI of 29.5 months, still more than half (52%) suffered from sleep-wake disturbances. The two groups, however, differed with regard to the type of their complaints. During the subacute phase, 81.2% of all patients with sleep-wake disturbances had insomnia, namely difficulty initiating and maintaining sleep, whereas excessive daytime sleepiness (EDS) was the predominant symptom in the chronic TBI group, accounting for 72.5% of sleep-wake disturbances.[18] Importantly, the authors also noted a negative effect of sleep-wake disturbances after TBI on occupational outcome and emphasized the need to manage sleep-wake disturbances after TBI early in the rehabilitation process.

The first systematic and prospective studies in this regard appeared only in 2007.[19,20] We now know that sleep-wake disturbances are highly prevalent 3 to 6 months after TBI and may persist for several years.[19-21] Six months after TBI, 47 of 65 consecutive patients (72%) presented with de novo sleep-wake disturbances; that is, the patients deny having similar complaints before the accident.[19] Arousal disturbances were the most common type, with EDS or fatigue in 55% and hypersomnia (defined as needing ≥2 hours more sleep per 24 hours compared with before TBI) in 22% of patients.[19] Furthermore, 43% of patients had no identifiable etiology of the posttraumatic sleep-wake disturbance (e.g., obstructive sleep apnea [OSA] syndrome, depression, insufficient sleep syndrome), suggesting a direct causative role of the trauma-induced brain damage (Figure 99-3). In a follow-up study, the authors reevaluated 51 patients 3 years after TBI, and the prevalence of posttraumatic sleep-wake disturbances was still remarkably high: 67% of patients still had symptoms, including fatigue (35%), hypersomnia (27%), EDS (12%), and insomnia (10%).[21] Another prospective clinical and sleep laboratory study performed 3 months or more after TBI

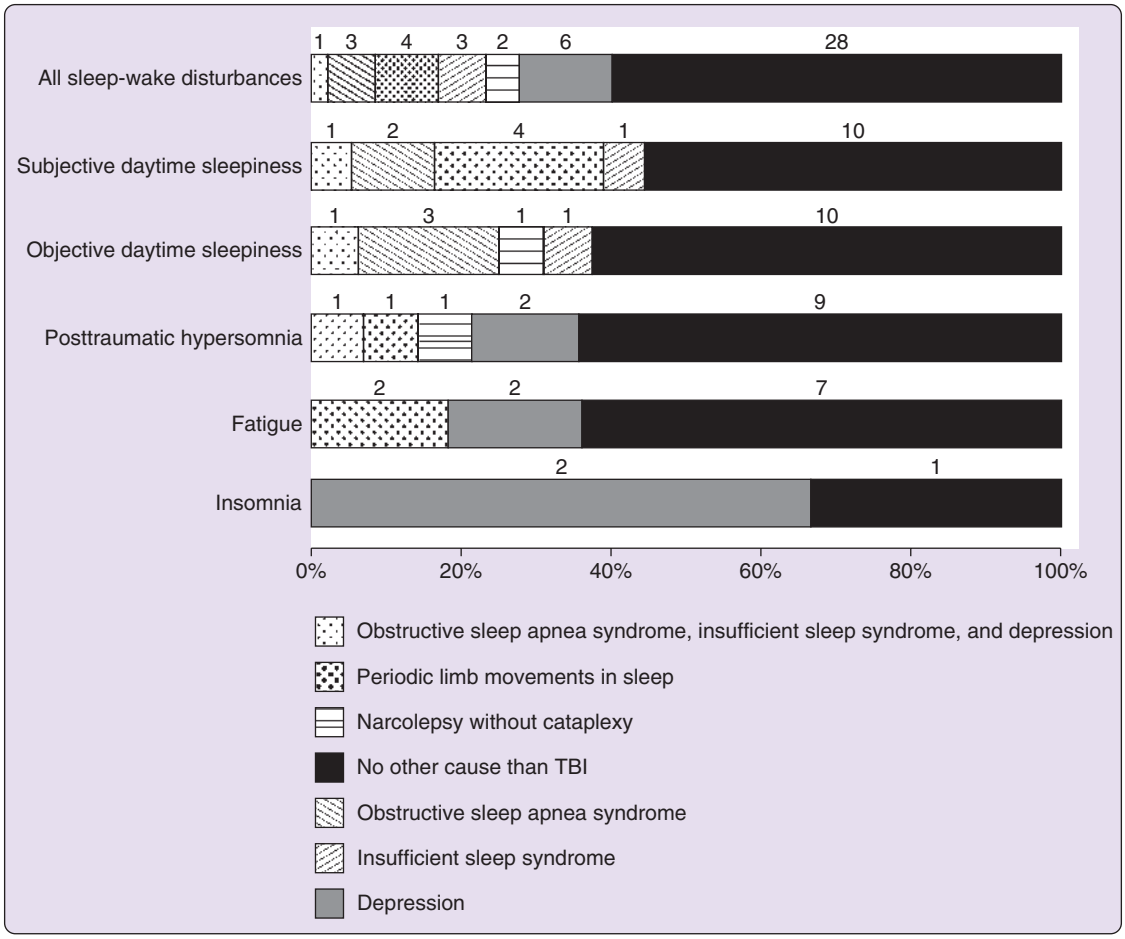


Figure 99-3 Conditions and comorbidities identified as likely causes of posttraumatic sleep-wake disturbances. In a substantial proportion of patients, the etiology remained unclear despite extensive diagnostic workup. In these patients, trauma-induced brain damage was likely the primary cause of the sleep-wake disturbances. TBI, Traumatic brain injury. (From Baumann CR, Werth E, Stocker R, et al. Sleep-wake disturbances 6 months after traumatic brain injury: a prospective study. *Brain* 2007;130:1873–83.)

showed similar results, with sleep-wake disturbances in 46% of patients and EDS in 25%.[20]

CLINICAL SIGNIFICANCE OF SLEEP-WAKE DISTURBANCES AFTER TRAUMATIC BRAIN INJURY

Increasing evidence demonstrates the importance of timely diagnosis and treatment of posttraumatic sleep-wake disturbances. Posttraumatic sleep-wake disturbances compromise health-related quality of life and rehabilitation outcome.[22,23] In one study, TBI patients with disrupted nighttime sleep patterns had worse daytime function and required longer stays in both acute and rehabilitation settings.[24] The presence of posttraumatic sleep-wake disturbances can trigger or worsen other TBI-related symptoms, such as pain, irritability, headaches, fatigue, and cognitive deficits.[25] In addition, long-term functional outcomes are reduced in patients with posttraumatic sleep-wake disturbances.[26]

There is also a meaningful association between reduced sleep quality and posttraumatic psychological distress. In a recent, prospective, longitudinal survey of 29,640 U.S. Navy and Marine Corps men, sleep problems were identified as an early marker of increased risk for developing posttraumatic stress disorder (PTSD) at a later time point.[27] Specifically, many service members with a medical history of TBI reported sleep problems on return from deployment yet manifested depression or PTSD only several months later.[27]

Finally, many TBI survivors struggle with medicolegal issues. More than 50% of patients with posttraumatic EDS may have exhausting discussions with their insurance companies and other financial problems.[28] In addition, patients with sleep-wake disturbances after TBI may exhibit worse occupational outcome.[29]

EXCESSIVE DAYTIME SLEEPINESS AFTER TRAUMATIC BRAIN INJURY

Despite major methodologic differences, most studies identify EDS as one of the most prevalent posttraumatic arousal disturbance. EDS can be defined by subjective (e.g., Epworth Sleepiness Scale [ESS] score ≥10) or objective (e.g., short mean sleep latency on Multiple Sleep Latency Test [MSLT] or Maintenance of Wakefulness Test [MWT]) criteria. The correlation between ESS scores and MSLT findings is generally poor.[30,31] For example, no correlation was apparent between ESS scores and mean sleep latency on MSLT in 71 adults examined 38 ± 60 months after TBI.[32] In this study, 47% had a mean MSLT sleep latency of 10 minutes or less, and 18% had latencies of 5 minutes or less.[32] Of note, 30% had objective EDS but normal respiratory and periodic limb movement indices on the polysomnogram, suggesting that the cause of posttraumatic EDS is not disturbed sleep. This is in contrast to an earlier study that suggested OSA may be the main cause of posttraumatic EDS,[33] whereas other groups have failed to uncover any major sleep-wake or other neurologic disturbances responsible for posttraumatic EDS.[19] Prospective longitudinal studies have suggested a gradual decrease in the frequency of posttraumatic EDS.[21,34] Still, some of these patients who recover from EDS subsequently suffer from persistent fatigue.[21] However, many older studies are not controlled, and large population-based cohort studies suggest that short mean sleep latencies on MSLT are more prevalent in the normal population than formerly suspected.[35] In a controlled prospective approach, we have recently confirmed an increased prevalence of posttraumatic EDS compared with healthy matched controls.[36] In the same study, EDS as assessed by MSLT was more pronounced and more frequent than subjective EDS measured by the ESS.[36] This indicates that patients may underestimate posttraumatic EDS, a fact that must be borne in mind when dealing with medicolegal questions like the capability of TBI survivors to drive motor vehicles.

Posttraumatic EDS can be severe and may sometimes resemble narcolepsy. This diagnostic challenge is discussed later.

HYPERSOMNIA AFTER TRAUMATIC BRAIN INJURY

Increased sleep need is a major problem after TBI and has hitherto been referred to as *posttraumatic hypersomnia*. However, the definition of hypersomnia is inconsistent and is often used interchangeably for *sleepiness*.[34,37] Importantly, increased sleep need after TBI is not always accompanied by EDS, yet significantly prolonged sleep times may represent a debilitating symptom even in the absence of EDS. Affected patients may face difficult limitations on their social, family, and work-related activities. Thus it became necessary to find a more appropriate term for this apparently frequent condition in TBI patients, and we recently proposed the word *pleiosomnia*, originating from the Greek word *pleio* (more, excessive) and the Roman word *somnus* (sleep).[38] We define pleiosomnia as increased sleep need of 2 hours or more per 24 hours compared with pre-TBI conditions.

In a large, prospective study, 22% of our TBI patients had pleiosomnia.[19] In a follow-up study to more specifically characterize posttraumatic pleiosomnia, two in three patients with posttraumatic pleiosomnia had no subjective EDS but had longer sleep times per 24 hours and increased slow wave sleep.[38] Likewise, on MSLT only 42% of the patients had a mean sleep latency of less than 8 minutes, that is, objective EDS (Figure 99-4). One important observation was that actigraphy displayed increased sleep on weekends in several patients with posttraumatic pleiosomnia, indicating insufficient sleep syndrome. Thus patients with posttraumatic pleiosomnia may be at risk for secondary development of EDS as a consequence of insufficient sleep syndrome, especially younger patients returning to work and those parenting small children.

Last, TBI patients may underestimate their actual amounts of sleep; pleiosomnia patients' reports of sleep times per 24 hours on sleep logs are significantly shorter than simultaneous actigraphy measures.[19,36,38]

FATIGUE AFTER TRAUMATIC BRAIN INJURY

Fatigue is usually defined as a subjective feeling of exhaustion, physical and mental tiredness, apathy, and a persistent lack of energy.[39,40] The current literature offers a large variety of self-report questionnaires on fatigue, but there is no objective measure. It is thus unsurprising that the research on posttraumatic fatigue provides inconsistent and often divergent results. The prevalence of fatigue in TBI survivors ranges from 16% to 80%.[41] Posttraumatic fatigue is one of the most

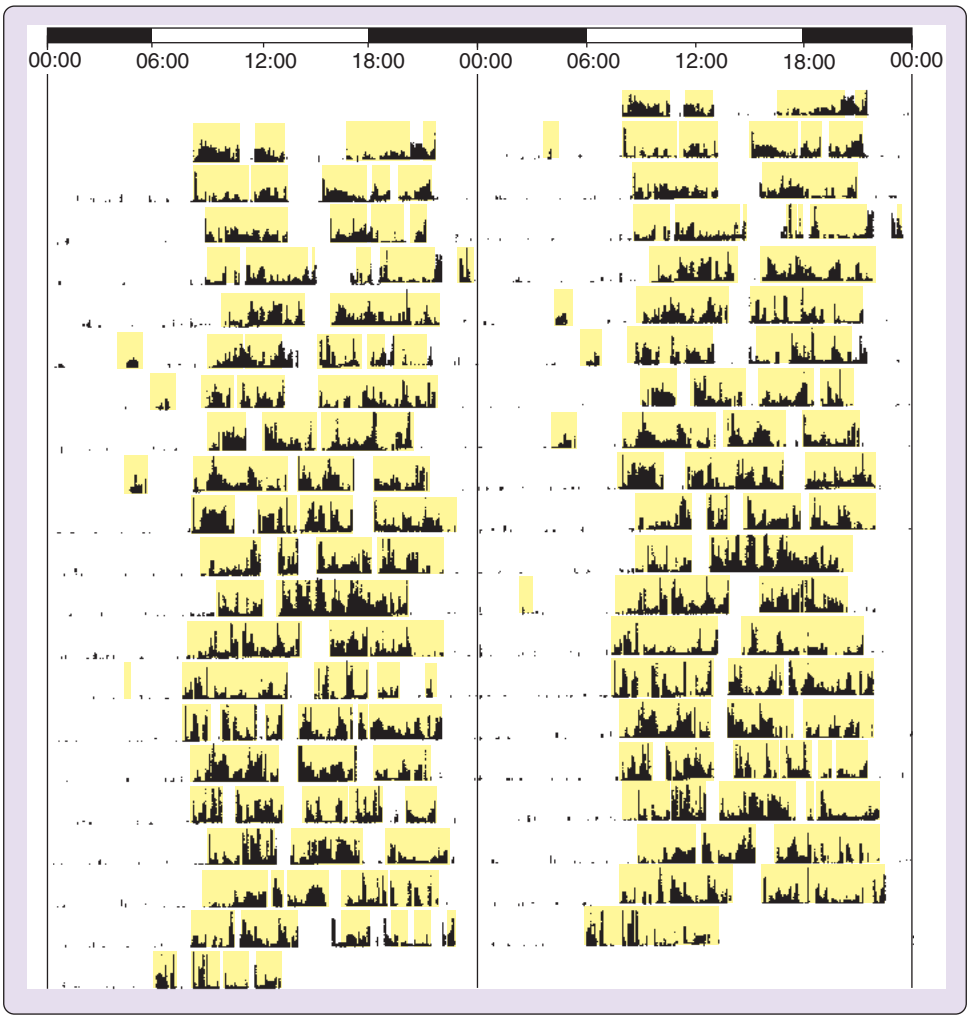


Figure 99-4 Actigraphy in a patient with posttraumatic pleiosomnia but without excessive daytime sleepiness. The actigraphy was performed 12 months after mild traumatic brain injury and demonstrated increased sleep amount per 24 hours (13 hours), with prolonged sleep times during the night and multiple rest periods during the day. Mean sleep latencies were 18 minutes on the Multiple Sleep Latency Test and 40 minutes on the Maintenance of Wakefulness Test.

persistent symptoms after TBI, and in a longitudinal study at 10 years post-injury, fatigue still emerged as the most frequent complaint (together with balance problems).[42-45] Among 22 adults with moderate to severe TBI, fatigue was more prominent than EDS 1 to 11 years after injury.[46] Posttraumatic fatigue is associated with limited daily functioning.[47] Higher levels of posttraumatic fatigue are associated with reduced quality of life.[48,49] The effect of posttraumatic fatigue on other symptoms is unclear. For instance, several groups report a correlation between posttraumatic fatigue and worse cognitive performance,[49,50] whereas other studies fail to detect any robust relationship.[51,52] Among many other causes, neuroendocrine abnormalities have been suggested to influence posttraumatic fatigue; in particular, lower basal cortisol levels have been correlated with higher fatigue scores.[53,54] Depression and anxiety are frequent after TBI, and an association with fatigue has been shown repeatedly.[21,49,55] However, a substantial proportion of patients with posttraumatic fatigue do not present mood disturbances, suggesting another cause and thereby illustrating the multifaceted etiology of this condition.

INSOMNIA AFTER TRAUMATIC BRAIN INJURY

Estimations of the prevalence of posttraumatic insomnia are varied mainly because of methodologic differences and inconsistent definitions. Most reports indicate that insomnia affects 30% to 70% of TBI patients,[22] but one prospective study using objective measures such as actigraphy and polysomnography found insomnia in only 5%.[19] In fact, some studies showed that TBI patients might overestimate insomnia symptoms. Compared with subjective, questionnaire-based measures of insomnia, two nights of polysomnography revealed lower frequencies of disrupted sleep.[56] In a prospective study using the Pittsburgh Sleep Quality Index, insomnia occurred in 30% of 50 consecutive TBI patients.[57] Among 116 consecutive patients with combat-related TBI, 55.2% had insomnia according to the *Diagnostic and Statistical* Manual *of Mental Disorders*, fourth edition criteria.[58] Pain is a common complication of TBI and a frequent cause of disturbed sleep.[59] In a retrospective study, 45% of 184 somnolent TBI survivors reported insomnia, and pain appeared as the predominant

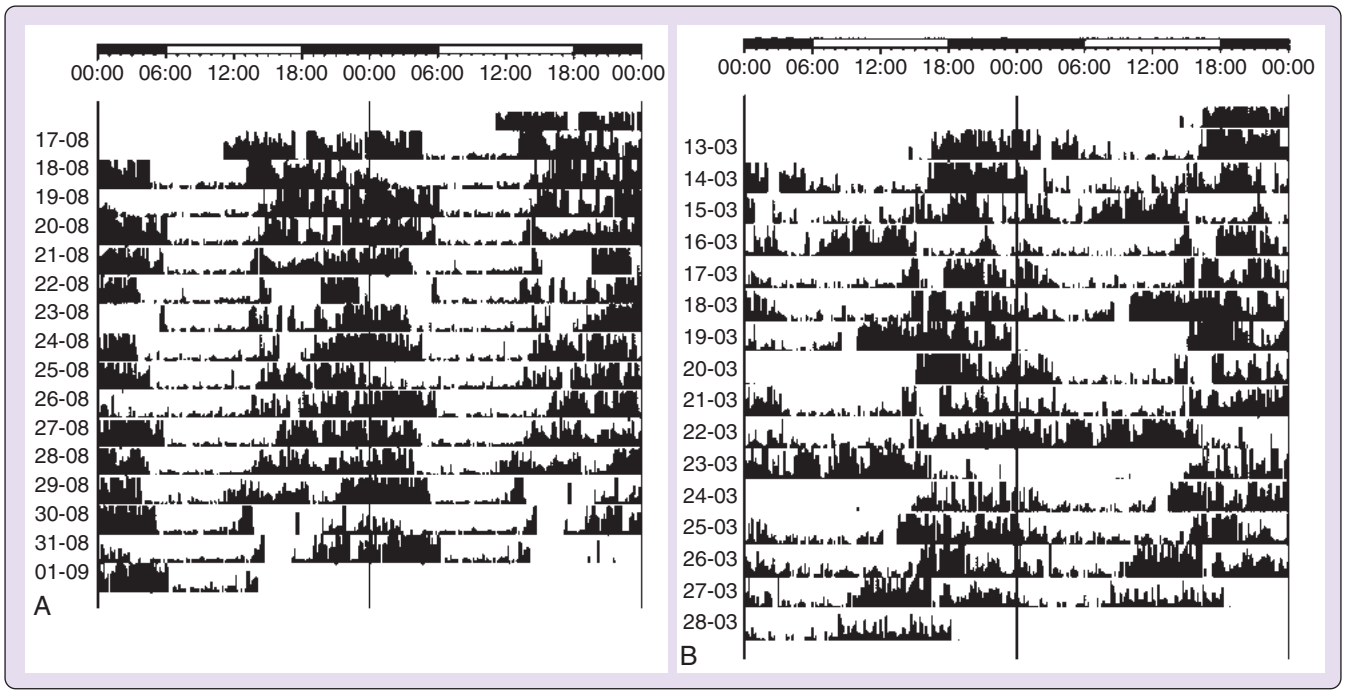


Figure 99-5 Actigraphy in two patients with different posttraumatic circadian sleep-wake disorders. The first patient **(A)** has delayed sleep phase syndrome. The second patient **(B)** has an irregular sleep-wake pattern. (From Ayalon L, Borodkin K, Dishon L, et al. Circadian rhythm sleep disorders following mild traumatic brain injury. *Neurology* 2007;68:1136–40.)

underlying feature.[28] Other contributing factors include dizziness, anxiety, and depression.[60] On the other hand, posttraumatic insomnia can interfere with rehabilitation because it often exacerbates other symptoms, including headache and emotional distress, and it may aggravate cognitive impairment.[61] Indeed, when comparing TBI patients with and without posttraumatic insomnia, sustained attention is significantly worse in the former group.[62]

Studies using polysomnography have revealed reduced sleep efficiency, increased sleep fragmentation, and increased sleep-onset latency among TBI patients.[56,63-65] Sleep fragmentation appears to be particularly pronounced in those patients with mild TBI, anxiety, and depression.[64] In fact, patients with a history of mild TBI more frequently report insomnia symptoms, together with higher levels of fatigue, depression, and pain.[61] On the other hand, repetitive TBI increases the severity of insomnia, as shown in 150 male military patients seen in a TBI clinic in Iraq.[66] Others have not found a correlation between injury variables and occurrence of insomnia.[67]

CIRCADIAN SLEEP-WAKE DISORDERS AFTER TRAUMATIC BRAIN INJURY

Few studies have described circadian sleep-wake disorders after TBI, perhaps because they lacked temperature and melatonin measurements.[14,19] However, posttraumatic circadian disturbances may be underestimated because problems such as initiating or maintaining sleep are easily confused with insomnia. One systematic study of 42 patients with insomnia complaints after mild TBI identified circadian rhythm sleep disorders in 36% of patients using actigraphy, saliva melatonin measurements, and body temperature measurement (Figure 99-5).[68] Other groups have confirmed a high prevalence of circadian rhythm disorders after TBI.[69-73] In addition, evening melatonin production is significantly lower in 23 TBI patients more than 1 year after injury than in 23 age- and sex-matched healthy controls.[74] This finding indicates that TBI may be associated with persistent circadian sleep-wake disorders and impaired melatonin synthesis.[74]

OTHER SLEEP-WAKE DISTURBANCES AFTER TRAUMATIC BRAIN INJURY

Narcolepsy

In the first decades of the 20th century, when the independency of narcolepsy as a nosological entity was still under debate, several reports on so-called posttraumatic narcolepsy emerged in the literature.[75-78] During the past 25 years, additional reports have appeared on this topic.[19,28,79-86]

The name *posttraumatic narcolepsy* suggests that TBI directly destroys the hypocretin-producing neurons in the posterior hypothalamus, thereby leading to narcolepsy symptoms, including EDS, cataplexy, hypnagogic hallucinations, and sleep paralysis. Indeed, acute TBI is associated with severe (but transient) hypocretin deficiency in the cerebrospinal fluid (CSF),[19,87] and postmortem studies of patients with severe TBI reveal mild damage to the hypocretin-producing neurons.[88,89] However, TBI also damages other sleep-wake regulating hypothalamic cell groups, including neurons producing melanin-concentrating hormone or histamine.[88,89] Thus the existence of true posttraumatic narcolepsy has to be questioned because, as opposed to the highly selective hypocretin deficiency in narcolepsy, TBI causes nonselective damage to the hypothalamus and other brain regions. This is

also reflected by the clinical characteristics; most patients with so-called posttraumatic narcolepsy do not have typical cataplexy but usually have hypersomnia (contrary to the normal amount of sleep per 24 hours in narcolepsy). Even in the presence of proven hypocretin CSF deficiency and a positive MSLT, the symptoms of such patients are best described as a *narcolepsy-like phenotype after TBI*. If TBI-induced brain injury damages the hypothalamus, the reported sleep-wake disturbance clearly goes beyond typical narcolepsy with cataplexy, as illustrated by the extensive literature on secondary narcolepsy.[90,91] A further limitation pertains to the low specificity of the MSLT-based diagnosis of narcolepsy without cataplexy, consisting of mean sleep latency lower than 5 minutes and multiple sleep-onset REM sleep periods.[92]

On the other hand, we found a history of TBI in 7 of 37 consecutive narcolepsy patients with clear-cut cataplexy, with proven CSF hypocretin deficiency and positive HLA DQB1*0602 haplotype.[86] The latency between TBI and narcolepsy onset was 2 years or less in all but 1 patient. This unexpectedly high prevalence of TBI (19%) among narcolepsy patients suggests some kind of causal relationship. Perhaps partial loss of the hypocretin neurons due to TBI triggers an autoimmune or neurodegenerative process, leading to overt narcolepsy in susceptible patients.[86]

Kleine-Levin Syndrome

Kleine-Levin syndrome (KLS) is a very rare sleep disorder of possible hypothalamic origin, characterized by recurrent episodes of extreme hypersomnia and behavioral disturbances such as hypersexuality and hyperphagia.[93] First described almost a century ago, the etiology of this disorder remains a mystery (see Chapter 100). Several researchers observed a temporal link between the onset of KLS and TBI.[94-98] Billiard and Podesta recently reviewed most of these cases and concluded that most patients present with genuine KLS, whereas in others the delay between TBI and occurrence of symptoms argues against a causal relationship between the two conditions.[99] The largest review on KLS included 186 patients, and in 9%, TBI was identified as a precipitating factor, compared with 38% of cases triggered by infections.[93]

Obstructive Sleep Apnea

In our prospective study of posttraumatic sleep-wake disturbances, only 11% of patients had OSA,[19] whereas other groups have reported higher frequencies ranging from 23% to 36%.[20,28,58,100] When extensively evaluating pre-TBI behavior, including bed partner interviews, OSA emerges in a large proportion of TBI survivors as a de novo posttraumatic feature.[28] Among 116 patients with combat-related TBI, multivariate analysis identified blunt trauma (as opposed to blast injury) as a significant predictor of OSA.[58]

The link between TBI and OSA is unclear. Preexisting EDS due to OSA increases the risk for TBI, especially while driving. Thus OSA might represent a risk factor for TBI in many patients, rather than a posttraumatic consequence. On the other hand, posttraumatic sequelae such as physical disability, depression, pain-related immobility, and pharmacologic treatment for posttraumatic epilepsy or mood disturbances, could cause weight gain and thus increase the risk for OSA.[59] Partial loss of neurons producing hypocretin in TBI patients might also trigger weight gain.[87-89] Furthermore, neuroendocrinologic disturbances are common after TBI.[101] The apolipoprotein E ε4 allele (ApoE-ε4) is associated with cognitive decline and development of dementia in both healthy subjects and TBI patients, and it seems to increase the risk for OSA in TBI patients.[102] Thus the combination of ApoE-ε4 and OSA may combine to impair cognition in TBI patients.

Recognition of OSA in TBI survivors is important because it is significantly associated with EDS and contributes to cognitive impairment.[33,103] Conversely, posttraumatic neuropsychiatric consequences such as PTSD may adversely affect adherence with continuous positive airway pressure therapy.[104]

PARASOMNIAS, SLEEP PARALYSIS, AND HYPNAGOGIC HALLUCINATIONS

Among 184 patients with head and neck trauma evaluated with interviews, questionnaires, and electrophysiologic examinations, as many as 53% reported hypnagogic hallucinations, 15% had sleep paralysis, and 9% had REM sleep behavior disorder.[28] Rodrigues and Silva reported on a 28-year-old man with severe TBI and slow recovery after an initial 2-month coma, who then developed a transient syndrome characterized by narcolepsy-like EDS, polysomnography-documented REM sleep behavior disorder, and periodic limb movements during sleep.[105] In another retrospective study on 60 TBI patients, one-fourth of all patients had various parasomnias, sometimes in combination, including acting out dreams (8.3%), nightmares (6.7%), sleepwalking (8.3%), sleep paralysis (5%), nocturnal enuresis (5%), cataplexy (3.3%), and nocturnal eating (3.3%).[100] Apparently, none of the patients had any sleep complaints before the TBI,[100] suggesting that trauma-induced damage to brain circuits that regulate sleep-wake behavior contributes to posttraumatic parasomnias. In our own study, we found a low incidence of sleep paralysis (5%) and hypnagogic hallucinations (5%), which was even lower than the prevalence of these features in the normal population.[19] Thus it remains unknown whether parasomnias in TBI survivors emerge as de novo symptoms or simply are undiagnosed before the TBI.

DISTURBED DREAMING

Dreaming can be altered in TBI patients depending on the time since TBI, the presence of other comorbidities, and memories concerning the moments immediately before the TBI. In the acute phase after TBI, dream recall seems decreased and may even be completely lost despite normal REM sleep.[17,106] During recovery, when TBI survivors tend to have longer and more consolidated sleep, dream recall is still reduced, and patients indicate less vivid dreams.[107,108] On the other hand, persistent nightmares with trauma-related content are highly prevalent in TBI survivors and may in turn worsen posttraumatic sleep-wake disturbances, in particular by interrupting sleep.[109,110] Guilleminault and colleagues reported nightmares in up to 41% of TBI patients.[28]

SLEEP-WAKE DISTURBANCES AFTER TRAUMATIC BRAIN INJURY IN CHILDREN

Although TBI is the most common cause of long-term disability in children, there is a striking paucity of data regarding posttraumatic sleep-wake disturbances in this age group.[111] In

light of the evidence that sleep is important for learning, memory, and neural plasticity, more studies on this topic are urgently needed.[112-114]

Among 19 adolescents with sleep disturbances 3 years after TBI, polysomnography and actigraphy revealed significantly lower sleep efficiency with more awakenings and more wake time than in controls.[115] Another prospective and longitudinal study confirmed the persistence of sleep disturbances up to 24 months after TBI.[116] The authors identified mild TBI, frequent pain, and psychosocial problems as risk factors for disturbed sleep. Conversely, sleep disturbances predict poorer functional outcome in children with moderate to severe TBI.[116] Acute TBI also alters sleep patterns in infants and young children.[117]

The few available studies on posttraumatic sleep-wake disturbances mostly focus on sleep problems and less on impaired daytime vigilance.[118-120] A prospective study compared the frequency of sleep problems between children aged 6 to 12 years with a history of moderate TBI ($n = 56$) or severe TBI ($n = 53$) and controls with only orthopedic injuries ($n = 80$) at 6, 12, and 48 months after TBI.[119] They found more sleep problems only in the severe TBI group, whereas children from the moderate TBI and orthopedic groups did not differ. In contrast, another study of children in the same age range considered also the parents' perspective and revealed greater parent reports of sleep disturbances 6 months after mild TBI compared with parent reports of children with only orthopedic injuries.[120] Of note, the affected children with mild TBI did not self-report greater sleep problems, thus indicating the importance of a rigorous search for sleep-wake disturbances in children with TBI. Likewise, using actigraphy and questionnaires in 15 children with moderate to severe TBI, poor sleep with prolonged sleep onset and impaired sleep maintenance was more common, notably without any evidence for circadian rhythm disorders or more frequent daytime napping.[121]

Diagnosis of Sleep-Wake Disturbances after Traumatic Brain Injury

The clinical manifestations of sleep-wake disturbances after TBI are pleomorphic, and the underlying causes are complex and differ from one patient to another. At the same time, sleep-wake disturbances may be obscured by other, more salient clinical consequences of TBI. Because sleep-wake disturbances can impair recovery from TBI, correct and timely recognition is important. In addition to sleep specialists and neurologists, awareness of posttraumatic sleep-wake disturbances is important for neurosurgeons, physicians working in intensive care units and rehabilitation centers, general practitioners, and psychiatrists. Currently there are no useful risk factors or biologic markers heralding the development of sleep-wake disturbances soon after TBI, except for intracranial hemorrhage, which is associated with posttraumatic hypersomnia.[36] Hence we should consider the possibility of sleep-wake disturbances in every TBI patient, and ideally, this screening should be repeated at several time points. A detailed overview on this topic has recently been published.[122]

Assessment of posttraumatic sleep-wake disturbances includes a structured interview, questionnaires for semiquantitative measurement and longitudinal monitoring, and sleep laboratory examinations such as whole-night polysomnography, MSLT, MWT, and actigraphy. Many self-report instruments are available to assess sleep-wake disturbances; most of them, however, have not been developed and validated specifically for TBI patients.[123]

A detailed history and self-report questionnaires are a crucial diagnostic aid, but many studies report a marked discrepancy between subjective estimates and the results of sleep laboratory examinations. In general, patients with moderate to severe TBI may underestimate their sleep-wake disturbances, whereas patients with mild TBI tend to overestimate. Table 99-1 presents an overview of studies reporting discrepant findings between subjective and objective measures of posttraumatic sleep-wake disturbances. Paradoxically, these underestimations and overestimations of sleep-wake disturbances by TBI patients are actually some of the most consistent findings across studies on posttraumatic sleep-wake disturbances.[19,32,38,42,56] The underlying mechanisms for this discrepancy are unknown, but potential explanations include distorted perception because of neurologic or neuropsychiatric comorbidities, deficits in frontal-executive functioning, and impaired self-awareness.[38,74,123-125] The Patient Competency Rating Scale is a common tool to assess self-awareness in TBI patients.[126] Comparing 30 TBI patients with high levels of self-awareness and 32 TBI patients with impaired self-awareness, appropriate perception of symptoms was associated with better neuropsychological function and higher functional independence.[124] The deleterious effects of impaired self-awareness after TBI may include poorer treatment outcome, longer hospital and rehabilitation stay, and poor compliance.[125]

In the diagnostic management of TBI patients, this peculiarity has to be considered. As a consequence, a high degree of clinical suspicion is crucial when searching for potential sleep-wake disturbances in TBI patients. Whenever possible, the clinician should obtain the perspectives and ratings of relatives and caregivers. Finally, the threshold to perform sleep laboratory examinations should be low.

Treatment of Sleep-Wake Disturbances after Traumatic Brain Injury

Sleep-wake disturbances after TBI often have multiple causes that should be considered by the treating physician. For instance, insomnia will likely persist despite successful treatment of sleep-disordered breathing if concomitant features such as pain and depression are not adequately addressed. This might explain why several treatment studies in TBI patients failed to obtain significant improvements of sleep-wake disturbances after TBI.[127,128] On the other hand, successful treatment of sleep-wake disturbances after TBI may improve pain, communication, cognition, and mood disturbances, thereby optimizing recovery and outcomes.[129] Currently, the number of inadequately treated TBI patients appears high. In one study, for instance, 60% of TBI patients with chronic and severe insomnia syndromes were not receiving any treatment.[61]

Treatment of sleep-wake disturbances after TBI includes pharmacologic and nonpharmacologic therapy. Good sleep hygiene should be encouraged in all TBI patients, although this approach alone failed to produce significant improvements in a recent study.[130] Cognitive-behavioral therapy may represent a promising approach in patients suffering from posttraumatic insomnia.[131,132] A randomized, placebo-controlled study revealed positive effects of blue light therapy

Table 99-1 Subjective Estimates and Objective Sleep Laboratory Findings Correlate Poorly in Patients with Traumatic Brain Injury

Type of Sleep-Wake Disturbance	Study	Study Design	Patients (age)	Injury Severity	Subjective Measurement	Objective Measurement	Outcome
Insomnia	Ouellet & Morin, 2006[56]	Prospective	14 TBI patients (30 ± 10 yr) 14 healthy good sleepers (30 ± 10 yr)	Mild: *n* = 4 Moderate: *n* = 5 Severe: *n* = 5	Insomnia Severity Index TBI: 18.3 ± 3.5 Controls: 1.7 ± 1.6	Sleep latency/ efficiency (PSG) TBI: 23 ± 21 min/ 87 ± 7% Controls: 20 ± 9 min/91 ± 3%	Despite showing similar PSG findings, TBI patients had significantly higher subjective insomnia measures.
Hypersomnia	Sommerauer et al., 2013[38]	Retrospective case-control	36 patients with pleiosomnia (36 ± 12 yr) 36 age-/sex-matched controls	Mild: *n* = 13 Moderate: *n* = 7 Severe: *n* = 16	Sleep logs TBI: 9.4 hr (6.8–15.0 hr) Controls: 7.5 hr (6.1–9.3 hr)	Actigraphy TBI: 10.8 hr (8.0–15.6 hr) Controls: 7.3 hr (5.7–9.2 hr)	TBI patients significantly underestimated their sleep need (*P* = .02). Conversely, controls showed good agreement between self-reported and actigraphically measured sleep need.
EDS	Baumann et al., 2007[19]	Prospective	65 patients (39 ± 17 yr), 6 mo post-TBI	Mild: *n* = 26 Moderate: *n* = 15 Severe: *n* = 24	Epworth Sleepiness Scale 7.5 (range, 2–20) ≥10 in 18 patients (28%)	MSLT Mean sleep latency: 9 ± 5 min, ≤5 min in 16 patients (25%)	The combined presence of both subjective and objective EDS was observed in 38%, but only in 9% did subjective and objective EDS concur. Likewise, ESS score and MSLT results did not correlate.
EDS	Masel et al., 2001[32]	Prospective	71 patients (32 ± 11 yr) 38 ± 60 mo after injury Patients with objective EDS (*n* = 33) compared with those without (*n* = 38)	Various head injuries (83% with accidental TBI)	Epworth Sleepiness Scale Without obj. EDS: 6.0 ± 5.3 With obj. EDS: 6.5 ± 4.3	MSLT Mean sleep latency: Without obj. EDS: 14.3 ± 2.4 min With obj. EDS: 6.4 ± 1.9 min	Patients with objective EDS (mean sleep latency on MSLT ≤10 min) estimated their sleepiness similarly to patients without EDS, indicating an inability to perceive posttraumatic EDS.
Amelioration of EDS with modafinil	Kaiser et al., 2010[134]	Double-blind, randomized, placebo-controlled	20 patients with EDS or fatigue (modafinil group: 37 ± 9 yr; placebo group: 43 ± 19 yr)	GCS: 7–8 2 yr after injury	Estimation on vigilance impairment amelioration with 100–200 mg modafinil or placebo	MWT (increase of mean sleep latency) Modafinil: 8.4 ± 9.6 min Placebo: 0.4 ± 6.2 min	Similar subjective estimation of vigilance improvement between modafinil and placebo groups but significantly improved ability to remain awake in the day (*P* = .005)

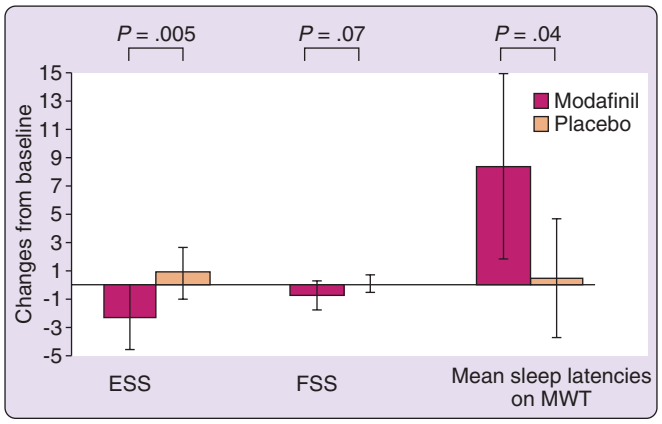


Figure 99-6 Effects of modafinil on posttraumatic excessive daytime sleepiness (EDS) and fatigue. Modafinil (100–200 mg each morning) improves subjective and objective posttraumatic EDS but not fatigue. ESS, Epworth Sleepiness Scale; FSS, Fatigue Severity Scale; MWT, Maintenance of Wakefulness Test. (From Kaiser PR, Valko PO, Werth E, et al. Modafinil ameliorates excessive daytime sleepiness after traumatic brain injury. *Neurology* 2010;75:1780–5.)

on posttraumatic fatigue and EDS.[133] In a double-blind, randomized, placebo-controlled study, 100 to 200 mg/day of modafinil significantly improved ESS scores and mean sleep latencies on MWT (Figure 99-6).[134] In addition, actigraphy suggested an increase in wakefulness of almost 2 hours per day, but without reaching statistical significance. Modafinil does not improve fatigue,[134] but fatigue may improve with treatment of associated pain or depression.

CLINICAL PEARLS

- Sleep-wake disturbances are common among TBI survivors but are often undiagnosed.
- After TBI, arousal disturbances, including hypersomnia, EDS, and fatigue, are the most common type of sleep-wake disturbances.
- PTSD nightmares and sleep disturbance and OSA may complicate the clinical picture.
- Recognition of these problems requires a high degree of clinical suspicion because they are often underreported by TBI patients.
- Untreated sleep-wake disturbances may negatively influence recovery and functional outcomes.

SUMMARY

During the past decade, a large body of studies has expanded our knowledge of the epidemiologic, clinical, and diagnostic aspects of sleep-wake disturbances after TBI. Disturbed arousal represents the most common sleep-wake disturbance after TBI and includes excessive daytime sleepiness, hypersomnia, and fatigue. Recognition of sleep-wake disturbances after TBI is often unsatisfactory because the treating physician either overlooks, underestimates, or incorrectly appreciates them. Sleep-wake disturbances may be obscured by other, more pressing clinical consequences of TBI, or the patients themselves may not adequately recognize their sleep-wake problems. Because sleep-wake disturbances can negatively affect recovery after TBI and significantly reduce quality of life, measures to quickly diagnose and treat the disorder are clearly warranted. In addition, more tailored therapeutics are needed for TBI survivors with sleep-wake disturbances.

Selected Readings

Ayalon L, Borodkin K, Dishon L, et al. Circadian rhythm sleep disorders following mild traumatic brain injury. *Neurology* 2007;**68**:1136–40.

Baumann CR, Werth E, Stocker R, et al. Sleep-wake disturbances 6 months after traumatic brain injury: a prospective study. *Brain* 2007;**130**:1873–83.

Cohen M, Oksenberg A, Snir D, et al. Temporally related changes of sleep complaints in traumatic brain injured patients. *J Neurol Neurosurg Psychiatry* 1992;**55**:313–15.

Collen J, Orr N, Lettieri CJ, et al. Sleep disturbances among soldiers with combat-related traumatic brain injury. *Chest* 2012;**142**:622–30.

Gilbert KS, Kark SM, Gehrman P, Bogdanova Y. Sleep disturbances, TBI and PTSD: implications for treatment and recovery. *Clin Psychol Rev* 2015;**40**:195–212.

Guilleminault C, Yuen KM, Gulevich MG, et al. Hypersomnia after head-neck trauma: a medicolegal dilemma. *Neurology* 2000;**54**:653–9.

Imbach LL, Valko PO, Li T, et al. Increased sleep need and daytime sleepiness 6 months after traumatic brain injury: a prospective controlled clinical trial. *Brain* 2015;**138**:726–35.

Katz DI, Cohen SI, Alexander MP. Mild traumatic brain injury. *Handb Clin Neurol* 2015;**127**:131–56.

Macera CA, Aralis HJ, Rauh MJ, MacGregor AJ. Do sleep problems mediate the relationship between traumatic brain injury and development of mental health symptoms after deployment? *Sleep* 2013;**36**:83–90.

Mollayeva T, Kendzerska T, Colantonio A. Self-report instruments for assessing sleep dysfunction in an adult traumatic brain injury population: a systematic review. *Sleep Med Rev* 2013;**17**:411–23.

Ouellet MC, Savard J, Morin CM. Insomnia following traumatic brain injury: a review. *Neurorehabil Neural Repair* 2004;**18**:187–98.

Shekleton JA, Parcell DL, Redman JR, et al. Sleep disturbance and melatonin levels following traumatic brain injury. *Neurology* 2010;**74**:1732–8.

Sommerauer M, Valko PO, Werth E, Baumann CR. Excessive sleep need following traumatic brain injury: a case-control study of 36 patients. *J Sleep Res* 2013;**22**:634–9.

A complete reference list can be found online at ExpertConsult.com.

Chapter 100

Kleine-Levin Syndrome

Isabelle Arnulf

Chapter Highlights

- Kleine-Levin syndrome (KLS) is a rare, relapsing-remitting disease affecting mostly adolescents. It is characterized by episodes of hypersomnia lasting one to several weeks, along with cognitive, behavioral, and psychiatric disturbances. Patients are normal between episodes.
- The sudden, severe (characterized by more than 18 hours of sleep per day), and recurrent hypersomnia helps differentiate KLS from other psychiatric mimics.
- Older case reports emphasized symptoms of hypersexuality, hyperphagia, male sex, and teenage onset, but recent large series now give a more precise and different picture.
- Derealization (a striking feeling of unreality), confusion, and apathy occur in all patients; disinhibited behavior is less frequent.
- Approximately one third of affected patients are female. KLS begins before the age of 12 years in approximately 10% of patients and after age 20 in 10%.
- Hypersomnia episodes usually become less severe and less frequent over time and eventually disappear with advancing age. Up to 28% of patients, however, experience prolonged episodes (lasting more than 30 days), and some 15% show no sign of recovery after more than 20 years of the disease.

Kleine-Levin syndrome (KLS) is a rare disease characterized by recurrent episodes of severe hypersomnia as well as cognitive, behavioral, and psychological disturbances.[1] The typical presentation is in an otherwise completely normal teenager, who suddenly (usually in a context of flu, alcohol use, or sleep deprivation) looks exhausted, needs to sleep 18 to 20 hours a day, becomes mute or answers just "yes" or "no" to questions, and reports feeling "unreal." The patient sleeps in a darkened room for 1 or 2 weeks, comes out only for brief meals and use of the toilet, and then immediately goes back to bed. The teen stops answering the cell phone and stops using social networks and computer games. After a couple of weeks in most cases, the episode ends with 1 or 2 days of insomnia. The teen then becomes more talkative and resumes normal sleep, cognition, and behavior for several weeks or months until a new episode appears.

Because the hypersomnia is prominent, KLS has been classified among the hypersomnias of central origin, but many symptoms are suggestive of a wider brain dysfunction, perhaps involving association cortex. These symptoms include exhaustion, severe apathy with loss of spontaneous autoactivation, and a striking derealization that is perceived by others as highly disagreeable. In some but not all patients, the behavioral and psychological changes include various forms of disinhibition similar to frontal lobe dysfunction, including rudeness, ravenous intake of certain foods, or inappropriate overt sexual offers. Blunted or sad mood (with a typically unexpressive face) and possibly hallucinations and paranoiac delusions are other, more psychiatric-related symptoms. Functional brain imaging is now a widely used research tool in KLS, both between and during episodes (which can be challenging), and shows persistent, reduced activity, mostly in the parietotemporal and mesiotemporal associative cortex.[2,3] The cause of KLS is unknown, but genetic factors may contribute, because multiplex families represent 2% to 5% of cases. In addition, metabolic and inflammatory or autoimmune origins are suspected.[4]

HISTORY

Reports suggestive of KLS are found in the nineteenth century literature[5] (Box 100-1). In 1925, Kleine described nine patients with periodic hypersomnia (two with increased food intake and one with menstruation-linked hypersomnia).[6] In 1936, Levin emphasized the association of periodic somnolence with morbid hunger in a single patient.[7] In 1942, Critchley coined the eponym Kleine-Levin syndrome. He claimed that male gender, onset in adolescence, periodic hypersomnia, compulsion to eat, and spontaneous remission were mandatory characteristics of KLS, dismissing as "doubtful" the previously reported female cases and the patients without hyperphagia.[8] His opinion prevailed[9] until a series of cases of KLS in women[10] as well as numerous patients without hyperphagia[11,12] were reported. After 2000, several large case series better characterized the spectrum of KLS,[3,11,13-17] finding that in addition to hypersomnia, slowed cognition, apathy, and

Box 100-1 THE FIRST REPORT OF PROBABLE KLEINE-LEVIN SYNDROME IN 1862*

Dr. Wilson, a physician in the Middlesex Hospital (United Kingdom), observed a very remarkable case of "double-mind" in a child. This patient was timorous, timid, and modest; he ate with moderation; in his usual state, he showed by his actions that he had an honest and scrupulous nature. As soon as the disease recurred, however, he lost all of these qualities. He slept a lot, was difficult to arouse, and as soon as he was awakened, he spontaneously sang, recited, and behaved with great ardor and aplomb. When he was not asleep, he ate ravenously. As soon as he got out of bed, he would approach another patient's sleeping area and then seize, without any hesitation, all the food he could find. Apart from this intriguing disease, he was intelligent and skillful.

*Author's translation.
From Brierre de Boismont A. *Des hallucinations ou histoire raisonnée des apparitions, des visions, des songes, de l'extase, des rêves, du magnétisme et du somnambulisme.* Paris: Germer Baillière; 1862.

Box 100-2 DIAGNOSTIC CRITERIA FOR KLEINE-LEVIN SYNDROME (ICSD-3)

A. The patient experiences at least two recurrent episodes of excessive sleepiness of 2 days to 5 weeks duration.
B. Episodes recur usually more than once a year and at least once every 18 months.
C. The patient exhibits normal alertness, cognitive function, behavior, and mood between episodes.
D. The patient must demonstrate at least one of the following during episodes: cognitive dysfunction, altered perception, eating disorder (anorexia or hyperphagia), disinhibited behavior (such as hypersexuality).
E. The hypersomnia and related symptoms are not better explained by other sleep disorders, other medical, neurologic, or psychiatric disorders (especially bipolar disorder), or use of drugs or medications.

ICSD-3, *International Classification of Sleep Disorders—Third Edition.*

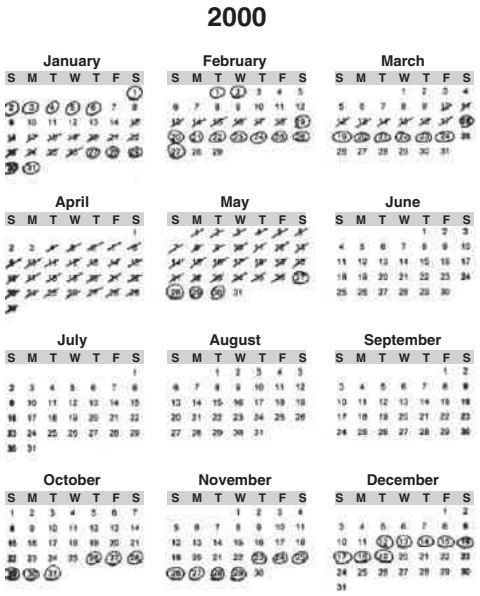


Figure 100-1 Calendar diary recording Kleine-Levin syndrome episodes during the first year of the disease in a 15-year-old girl as reported by her mother. *Circles* indicate the days with complete hypersomnia, whereas *slashes* indicate days marked by abnormal feelings (derealization, apathy, confusion) but no hypersomnia.

derealization were common, and only one third to half of patients had socially disruptive symptoms such as hypersexuality and hyperphagia, and not with each episode.[11,17] Much current KLS research focuses on functional brain imaging, genetics, and possible autoimmunity.

CLINICAL FEATURES

Episodes

The international criteria for diagnosing KLS are presented in Box 100-2.[1] The disease is typically relapsing-remitting (Figure 100-1). KLS episodes start abruptly (i.e., within a few hours) in half of the cases and progressively (within a few days) in the remainder.[11] When the end of an episode is sudden, patients frequently experience some insomnia and mild elation for up to three nights, although this is not the case when the end is gradual. Episodes last a median of 13 days, occurring every 3 months on average.[11,17] However, the clinical picture is highly variable, including short episodes (e.g., 7 days) occurring every month in young patients as well as prolonged episodes of 6 months or more with mostly apathy and altered cognition. Prolonged (lasting more than 1 month) episodes are observed in 28% of patients.[17] By contrast, some patients (mostly when treated) have mini-episodes of exhaustion and derealization lasting 1 day or even a few hours, which disappear after sleeping.[4] In most patients, the episodes are monophasic, with all symptoms always present during the full episode, whereas some rarer patients have biphasic episodes, consisting of a first phase of excitation, disinhibition, and sometimes insomnia for a few days, followed by a longer period of hypersomnia (in this order or in the reverse order). The frequency and duration of episodes are unpredictable. Patients with long episodes at KLS onset usually continue to have long episodes later, with longer disease duration.[17] In a recent series of 108 cases, patients averaged 19 episodes of 13 days duration.[11] The mean interval between episodes was 5.7 months, with a range of 0.5 to 66 months. Over the course of several years, the frequency and duration of episodes often lessen.[12,13]

Disease Duration

In large series, 80% of patients experienced their first episode between 13 and 19 years old, whereas 10% did so before the age of 13 years (childhood forms), and 10% (adult forms) started after the age of 20. The youngest age of disease onset is 4 years,[18] whereas the oldest is 82 years.[19] The episodes tend to be less frequent and then disappear with time. Indeed, most affected teenagers have no more episodes by their thirties.[12,13] The accurate disease duration can be evaluated only over long follow-up periods, especially because patients managed by pediatricians can be lost to follow-up in the transition to adulthood, and because patients may not inform the physician of late relapses (sometimes after 15 years of remission). If the end of the disease is defined as no more episodes after a period that is the double of the longest interval between episodes,

the disease duration was evaluated, using an actuarial curve, as a median of 13.6 ± 4.3 years in a recent series.[11] High episode frequency at the beginning of the disease was associated with a shorter course in a meta-analysis of 168 cases in the literature,[20] but this feature was not apparent in a recent single-center large study.[17] Male sex, onset before the age of 12 or after 20 years, and the presence of hypersexuality during episodes predicted longer disease duration.[11] Approximately 15% of patients (including more than half of those with adult onset) still have recurrent episodes after 25 years of disease.[11]

Triggering Events

Eighty-nine percent of patients remembered an event closely associated with disease onset, most often infections (72%, of whom one fourth had a coldlike syndrome with fever), alcohol use (23%), sleep deprivation (22%), unusual stress (20%), physical exertion (19%), traveling (10%), head trauma (9%), and marijuana use (6%). In infection-triggered KLS, symptoms of KLS occurred shortly (between 3 and 5 days) after the onset of fever. The agents responsible for the first infection included mostly viruses (e.g., Epstein-Barr virus, varicella-zoster virus,[11,21,22] Asian influenza virus,[23] H1N1 influenza virus,[17] enterovirus[24]) and *Salmonella yyphi*[25] and *Streptococcus*.[26,27] Rare postvaccine onset associations include combined typhoid and tetanus vaccine,[27] tuberculosis vaccine,[11] papillomavirus vaccine, and H1N1 vaccine.[17] Factors that trigger relapses are similar, although nothing can be identified in 15% to 20% recurrences. In a cohort of teenagers with KLS in Taiwan carefully followed for at least 2 years, the timing of disease onset and of recurrent episodes correlated significantly (the coefficient of correlation was between 0.45 and 0.55) with community outbreaks of upper respiratory infections such as acute bronchitis, bronchiolitis, and pharyngitis and nasopharyngitis.[15]

SYMPTOMS DURING EPISODES

Some core symptoms are almost always present, at least during the first years of the disease, including hypersomnia, cognitive impairment, derealization, apathy, and personality changes. The frequency of other symptoms such as hypersexuality, hyperphagia, hallucinations, delusions, and headache varies among patients and between episodes.[15]

Sleep Symptoms

All patients with KLS have hypersomnia, and this symptom is required for diagnosis. One of the major characteristics of hypersomnia during a KLS episode is the extremely prolonged duration of sleep (especially in teenagers), with a reported median 18 hours of sleep per day. In a recent series, sleep was enormously increased during episodes, averaging 18 ± 4 hours per 24-hour period. Most patients are difficult to awaken and report intense dreaming and frequent hypnagogic hallucinations during episodes. Sleep paralysis is uncommon, and cataplexy is absent. Patients, however, remain arousable, waking up spontaneously to void and eat, but are irritable or aggressive when awakened or prevented from sleeping. Sleep symptoms evolve from frank hypersomnia during the first episodes to a heavy fatigue accompanied by a feeling "as if in twilight between sleep and waking" during later episodes,[13] especially during episodes of increased duration. Most patients have a main, prolonged sleep episode during the night that ends around midday, followed by a nonrestorative long nap in the afternoon.

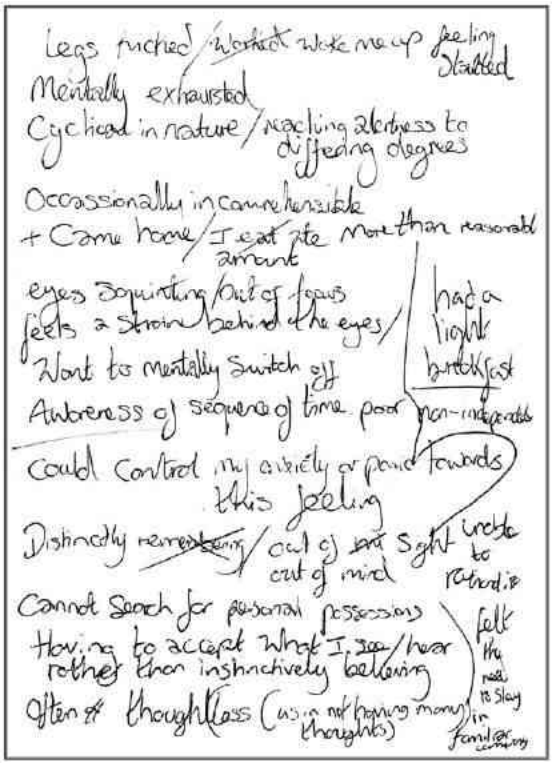


Figure 100-2 Symptom self-report of an 18-year-old boy with Kleine-Levin syndrome during an episode. Both the nature of the reported symptoms and the abnormal, sometimes incoherent writing serve to illustrate his cognitive problems.

Cognitive Symptoms

During an episode, the most obvious cognitive sign is bradyphrenia, or slow thinking. Patients look exhausted; they do not initiate conversation, are slow to answer, and answer with just "yes" or "no" (note examples in Videos 100-1 and 100-2). They frequently are disoriented in time and less often in space. Nevertheless, they can write, read, and calculate when asked and can distinguish left from right. If they write, the short messages they send are frequently confused (Figure 100-2). These confused communications can raise concern regarding alcohol or substance use. Anterograde amnesia is frequent. Some teenagers improperly sent to school during episodes were unable to follow the lessons and to complete examinations and had no memory of what they were taught. A mild apraxia is possible. Some patients report that they do not remember how to put on a T-shirt or how to use tools. The patients report that they do not "understand" the world around them. They can watch television without understanding what happens. Beyond this abnormal mental status, the remainder of the neurologic exam shows nothing out of the ordinary. These findings, together with the frank apathy and derealization, suggest dysfunction of association cortex during episodes. Altered cognition may persist during a few days or weeks after the end of an episode, especially with episodes that end gradually, so some caution is needed before sending a teenager back to school.

Derealization

Almost all (98% to 100%) patients report that the world around them seems unreal, as if they were "in a bubble" or "in a dream,"[28] feeling that their perceptions are unreal or altered in some way.[3] One patient described himself as living in a parallel world without being an actor in his life. Patients scored much higher during episodes compared with between episodes (70 ± 22 versus 13 ± 18) on the Depersonalization/Derealization Inventory.[3] Altered visual, auditory, tactile, temperature, and pain senses are common. Difficulty seeing space in three dimensions and discerning contours,[3] unusual food tastes, perception of abnormal smells, disturbed hearing (hypersensitivity to noise, voices seeming far away), increase in pain threshold, difficulty distinguishing cold from hot water, and dealing with cross-modal sensory stimulations are common. Taking a shower can be disagreeable because a patient may see water flowing on the body without feeling it at exactly the same time, in addition to difficulties evaluating its temperature. Some do not like looking at their face in the mirror, because they can have difficulty recognizing themselves.[3]

Apathy

Striking apathy affects almost all patients with KLS.[17] Patients do not seek novelty and stop their usually enjoyable activities, including reading, smoking, watching TV, playing video games, seeing friends, using social networks, and using cell phones and texting. They have a withdrawn attitude, and 80% of them neglect their hygiene. Typical examples are omission of brushing or styling the hair and not taking a shower for several days until pushed by their parents. The average apathy score on the Starskein Apathy Scale (range 0 to 40) was 30 ± 8 during symptomatic periods versus 9.5 ± 5 during asymptomatic periods.[3] Speaking often seems to require exceptional effort. Some patients remain able to perform their daily routine mechanically. A young patient, who is normally a talented skier, robotically skied after his brother during an episode, without taking any risk or initiative.[4] This complex inability to properly interpret novel environments makes patients unfit to drive during episodes. At the beginning of an episode, one of our patients drove behind his brother's car and followed the wrong vehicle for more than 600 km, crossed a national border, bought gas, and finished in a car crash with complete amnesia of the event.

Disinhibited Behavior

Hyperphagia and hypersexuality often are identified as characteristic or typical symptoms of KLS,[9,29] yet only 57% of patients experience hyperphagia and only 50% to 53% demonstrate hypersexuality, and these are features only in some but not all episodes.[11,15] Single case reports probably have overreported these symptoms,[9,29] because these were mandatory[30] for KLS diagnosis before the 2005 and 2013 international definitions.[1,31] Hyperphagia is different from bulimia in that it does not involve voluntary vomiting or attempts at controlling weight. Patients are disinhibited toward specific food and good manners, to the point of "stealing candies in their friend's bag" or "eating all available food."[5,32,33] Some patients report that they eat a large amount of food once a day and immediately resume sleep. Parents may find candies and cakes hidden under the patient's bed. On the other hand, 34% to 40% of patients actually eat less, sleeping all the time, and when they are called to the family dining table, they eat mechanically.[34,35]

Hypersexuality affects less than one half of patients with KLS during at least one episode, but it can be diagnostically useful because it is uncommon in other psychiatric syndromes. It is one of the most embarrassing symptoms in public and may lead to legal problems if not prevented. This mental, physical, and behavioral loss of sexual inhibition affects boys more often than girls (58% versus 35%), with increased masturbation ("to the point of bleeding"), increased demands on sexual partners, inappropriate sexual behaviors such as exposing or touching the genitals, masturbating in the presence of parents and physicians, swearing, asking overtly sexual questions of a teacher, touching a stranger's breast, and making sexually inappropriate proposals.[25,36] As an example, one of our patients mechanically accompanied his family to the beach during an episode. Once in his swimsuit on the beach, however, he started to masturbate openly in front of his grandmother, aunt, parents, and little sister. A 13-year-old patient entered the bathroom where his stepmother was taking a shower and proposed sexual relations to her. Plasma levels of sex hormones (testosterone, luteinizing hormone, follicle-stimulating hormone) were normal in 14 patients and mildly decreased in 2 patients.[20] Of importance, testosterone is not increased during KLS episodes.

Repetitive, compulsive behaviors also are common. One third of patients sing, pace (Videos 100-3 and 100-4), tap or snap their fingers, or repeatedly listen to the same music or watch the same video in a continuous loop.[11,15,37,38] Some patients exhibit regressive behaviors such as skipping or playing with their fingers,[39] speaking with a childish voice (as in Videos 100-1 and 100-2), using childish words, or asking their mother to sleep nearby.[11,35] Some parents qualify behaviors associated with each episode with a specific (low) mental age. Young patients have tantrums, especially when prevented from rest or sleep, or when brought to medical care.[40-43]

Psychotic, Mood and Anxiety Symptoms

Flattened affect and sad mood (with rare cases of suicidal attempts)[22,44] are observed in approximately half of patients[11,15] and are more prevalent in women than in men.[11] Decreased mood may last only one day (often close to the end of an episode), with teenagers crying and feeling that the disease "will never end." Some patients may ask if they could die from it or express the desire to die if it does not stop.[11] The abrupt termination of an episode commonly is characterized by a feeling of relief and elation lasting 1 to 3 days, with some logorrhea and a feverish desire to reclaim the missing time. When an episode ends slowly, the sad mood may persist. Anxiety can be high during episodes. Patients may fear being left alone at home and can be panicked when left alone in unusual environments such as a hospital, when going outside, or when meeting people.[35] Typically, patients stay in their room, refusing the visits of friends or relatives. During milder episodes, they are able to stay a few hours in the family room, and in rare cases they may accept to go out of the house, but mostly in noncrowded places. They often do not open the door for visitors.

One third of patients report short-lasting hallucinations (such as of a snake near the bed or a dangerous man with a bear in the hospital elevator) and delusions.[35] Most delusions

are mild, characterized by grandiosity, such as patients reporting being able to guess the identity of the caller when the phone rings, being able to pilot the plane they are in, or being able to stop the clock by thought.[45] Reported delusions include being in the movie that the person is watching, the belief that the patient's (still-alive) father is dead, or the feeling of being observed, filmed, or poisoned. Patients also may view themselves as God or hear voices telling them to kill their father. Ideas of reference (the feeling that people look specifically at them, usually with aggressive intent) are typical. Psychotic symptoms usually last only a few hours to a few days and stop spontaneously.

Autonomic Symptoms

Headache, photophobia, and phonophobia are frequent. Most patients remain lying in their room in the dark. In teenagers, during an episode, the face often has an exhausted look, a reduced facial expression, and an empty gaze. Patients may void less often (e.g., only once a day) and have, in rare cases, urinary retention. Orthostatic hypotension is common, especially if the patient has been spending much time in bed.[46,47] Other autonomic signs are rare, including high blood pressure,[34] bradycardia or tachycardia,[48] and ataxic or rapid ventilation.[49,50]

Asymptomatic Periods

During asymptomatic periods, patients generally have normal amounts of sleep and no psychiatric or cognitive symptoms (as evident in Videos 100-2 and 100-4) and normal anxiety, depression, and eating attitude test scores, with a normal[17] or somewhat increased[11] body mass index. The question of whether KLS promotes bipolar disease or psychosis in a subsample of patients cannot be easily answered. When contrasted with the high frequency of siblings affected in patients with bipolar disorder, a familial history of severe depression and bipolar disorder is rare in meta-analyses of cases[20] and was similar in frequency among control subjects in two large controlled studies.[11,17] However, subtle cognitive impairment (mostly affecting verbal memory and executive function) was reported in a few patients formally evaluated with neuropsychological testing.[16,28,51] One long-term follow-up study reported that 25 patients were in good health several years after the cessation of their KLS episodes, suggesting that complete recovery with a good prognosis is the rule for KLS.[13]

CLINICAL SUBTYPES OF THE DISEASE

KLS may be mild, moderate, or severe. In mild forms, teenagers experience 1-week symptomatic periods two or three times a year. In moderate forms (mostly observed in children and teenagers), patients may experience a 7- to 10-day episode every month,[52,53] Episodes lasting more than 1 month are observed in 28% of patients, often since disease onset. A recent study suggests that when KLS starts after the age of 20 years, lasting recovery is less common, with resolution in less than 50% of these patients after 25 years of disease.[11] This last result suggests that in some rare cases, especially those of adult onset, KLS never fully resolves.

Several multiplex KLS families have been reported, representing both horizontal transmission (in three monozygotic twin pairs,[54,55] two brothers,[11] two sisters,[17] and a brother and his sister[33]) and vertical transmission (in a mother and her son,[14] two fathers and their sons,[11] a father and five of his 10 children,[56] an uncle and one nephew,[57] and two uncles and their two nephews.[17,58] The clinical symptoms are similar to those in sporadic cases. Affected siblings do not necessarily experience the episodes at the same periods of the year.

Menstruation-Related Hypersomnia

When episodes are temporally related with menstruation (just before or at time of), the condition is named menstrual/menstruation-related hypersomnia[59] and has been reported in only 18 women worldwide.[9] Hypersomnia episodes in these cases are associated with compulsive eating in 65%, sexual disinhibition in 29%, and depressed mood in 35%.[9] Episodes last 3 to 15 days and recur less than three times a year. Of note, a boy with KLS had a sister affected by menstrual-related hypersomnia.[60] In addition, girls with typical KLS episodes mostly but not exclusively associated with menstruations have been described.[6,27] Because these symptoms are similar to those of KLS, menstrual-related hypersomnia is now considered a variant of KLS in the *International Classification of Sleep Disorders—Third Edition*. Improvement with contraceptive doses of estrogen and progesterone has been reported in rare cases.[61]

DIFFERENTIAL DIAGNOSIS

Because KLS is exceptionally rare, several more common diagnostic possibilities should be considered first. When brought to the emergency department during a first episode, most patients undergo a classic workup for acute confusion and sudden behavioral changes in teenagers: checking for alcohol, prescription drug, and illegal substance intake as well as magnetic resonance imaging (MRI) or other imaging studies to rule out a tumor, trauma, stroke, or inflammation such as that associated with multiple sclerosis. Tumors within the third ventricle (such as colloid cysts, pedunculated astrocytomas, or, in some cases, craniopharyngiomas) may produce intermittent obstructions of ventricular flow, leading to headaches, vomiting, and a paroxysmal impairment of alertness. An electroencephalogram (EEG) will help rule out complex partial seizures. A few, nonspecific sharp waves and localized or generalized slowing of EEG activity may be observed in up to 70% of KLS cases. In addition to basic laboratory tests, clinicians might consider checking serum ammonia to rule out hyperammonemic encephalopathy[62] as well as carnitine, folate, vitamin B_{12}, pyruvate, and lactate level determinations. Evaluation for possible endocrinopathies and autoimmune diseases, as well as tests for intermittent porphyria and Lyme disease, also may be useful. A lumbar puncture (especially in a context of fever) usually is advised to exclude encephalitis. Severe basilar migraine may mimic a KLS episode but typically is of shorter duration.

Recurrent episodes of sleepiness can occur with psychiatric disorders such as depression, bipolar disorder, seasonal affective disorder, and somatoform disorder. Disinhibition also may be prominent in some teenagers with attention deficit–hyperactivity disorder. Hallucinations and delusions in a previously normal teenager are evocative of brief psychosis episodes. As a consequence, some patients with KLS may be admitted to a psychiatric ward before the correct diagnosis is made. KLS with psychotic symptoms differs from psychotic

disorders in the sudden occurrence and disappearance of delusions and hallucinations in KLS, the absence of long-term adherence to delusional beliefs, the presence of associated symptoms (mainly hypersomnia, mental slowness, confusion, and amnesia), and the recurrence of symptoms. The difference between KLS with mood changes and a depressive disorder is in how suddenly the symptoms come and go in a previously happy teenager, and in the association with cognitive and behavioral symptoms. Sleepiness associated with chronic mood disorders typically alternates with periods of insomnia, whereas insomnia is very brief (2 to 3 days) in KLS and occurs only at the onset or cessation of episodes. Excessive sleepiness can occur with drug or substance use, obstructive sleep apnea, narcolepsy, idiopathic hypersomnia, or insufficient sleep. In these disorders, however, sleepiness occurs daily and usually is not recurrent; in idiopathic hypersomnia, the level of sleepiness may fluctuate with some "better" periods but sleep still remains prolonged (e.g., 12 to 14 hours/day). "Idiopathic" stupor is a rare and debated entity, occurring usually in middle-aged subjects, with stuporous episodes lasting no more than 48 hours, associated with benzodiazepine intoxication.[63]

DIAGNOSTIC TESTS

Polysomnography

The polysomnography results depend on whether sleep is monitored only for the night or for 24 hours, at the beginning or the end of episodes, or at onset of the disease or later in its course. Twenty-four hour polysomnography demonstrates prolonged total sleep time (12 to 14 hours),[12-14] up to 18 hours or more in some reports.[64] In a metanalysis, the mean total sleep time was 445 ± 122 minutes (sleep stage N1, 6%; N2, 56%; N3, 19%; REM sleep, 19%) during the night in 40 patients.[20] In 15 patients monitored for 24 hours, sleep time was 740 minutes. In a series of 14 patients in Israel, sleep efficiency was decreased, with frequent awakenings and excess stages N1 and N2.[13] Nocturnal sleep time increased from an average 384 minutes during asymptomatic periods to an average of 568 minutes during symptomatic periods.[13] In 17 children with KLS, nighttime slow wave sleep percentage was decreased during the first half and REM sleep decreased during the second half of episodes.[65]

The Multiple Sleep Latency Test (MSLT) is dependent on the subject's willingness to cooperate, and results may be either normal[65] or abnormal, with short latencies or multiple sleep-onset REM periods and a narcolepsy-like pattern in up to 21% of patients.[20] As the disease develops, patients may not sleep continuously but stay in bed with eyes closed. They report that this behavior is driven by heavy fatigue and reduces the feeling of being unreal.

Other Tests

Spinal fluid studies have shown no abnormalities in cells, protein, or oligoclonal bands. Cerebrospinal levels of hypocretin-1 were within normal range in 16 patients,[33,66-69] although they were slightly decreased during episodes compared with between episodes in some cases.[66,67] Computed tomography scans and MRI studies yield normal results or reveal incidental findings unrelated to the disease. The most interesting tests are brain functional imaging studies, which show abnormalities in most cases (see Pathogenesis section).

EPIDEMIOLOGY

The exact prevalence of KLS is unknown, but it is considered to be an extremely rare disease, with a prevalence of 1 to 2 cases per 1 million population.[11,13,15,17] It has, however, been described in all continents and most countries, including Israel.[70] KLS seems more common in Jewish populations in the United States, but not in France. Reported male-to-female ratios vary, ranging from 2:1 to 3:1. Birth and developmental problems are risk factors for developing the syndrome, with up to one third of patients in two large series having suffered from difficult delivery as an infant or demonstrating childhood developmental problems.[11,17]

PATHOGENESIS

Mechanism of Symptoms

Many KLS symptoms such as derealization, apathy, and disinhibition are suggestive of altered function in association cortex. Possibly, dysfunction of the thalamus, hypothalamus, and basal forebrain contributes to hypersomnia.[2] Functional brain imaging during episodes shows hypometabolism in the thalamus,[15] hypothalamus, mesial temporal lobe, and frontal lobe, persisting during asymptomatic periods in one half of the patients.[2,3,71] Single-photon emission computed tomography (SPECT) imaging in 41 patients during asymptomatic periods showed hypoperfusion in the hypothalamus, thalamus (mainly the right posterior part), the caudate nucleus, and cortical association areas (anterior cingulate, orbitofrontal,[1] and right superior temporal cortices) relative to that in control subjects[3] (Figure 100-3). Two additional hypoperfused areas emerged during symptomatic periods, located in the right dorsomedial prefrontal cortex (possibly underlying apathy) and the right parietotemporal junction (possibly underlying derealization). Patients with KLS performing a working memory task need to recruit a different network (including increased thalamic activity and decreased cingular activity and adjacent prefrontal cortex, as shown by functional MRI) to achieve performance levels similar to or lower than those in control subjects, suggesting a more effortful process and some compensation between episodes in patients.[51,72]

Mechanism of the Disease

The mechanism of KLS is unknown. Neuropathologic examination in four patients with KLS (two primary and two secondary cases) suggests mild localized encephalitis.[49,73-75] Three subjects showed perivascular lymphocytes in the hypothalamus, amygdala, and the gray matter of the temporal lobes, the thalamus, and the diencephalon and the midbrain. In the fourth case, a smaller locus coeruleus and decreased pigmentation in the substantia nigra were reported.

Several mechanisms have been hypothesized to cause KLS, including those with autoimmune, genetic, and metabolic aspects. An autoimmune basis for the disorder is suggested clinically by the onset during adolescence, often in conjunction with an infection, a traumatic head injury, or alcohol intake (which increases blood–brain barrier permeability and may promote the passage of antibodies)[76] and by the relapsing-remitting aspect and by recurrences. However, there is no evidence for systemic immune abnormalities.[76a] An association with HLA DQB1*02 (odds ratio, 2) in 30 European

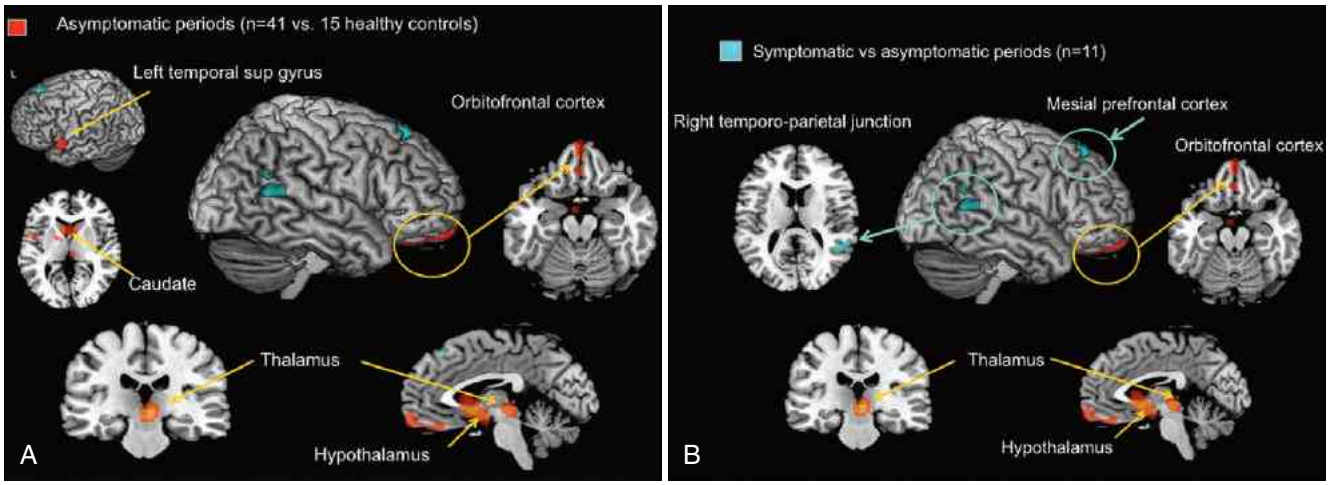


Figure 100-3 Single photon emission computed tomography (SPECT) imaging of brain in Kleine-Levin syndrome (KLS), between episodes (**A**) and during episodes (**B**). This group analysis in 41 patients with KLS shows decreased perfusion (*red*) during asymptomatic periods in the hypothalamus, thalamus, caudate nucleus, left temporal superior gyrus, and orbitofrontal cortex, when compared with healthy control subjects (**A**). During episodes, the right parietotemporal junction and the right dorsomedial prefrontal cortex show hypoperfusion (*green*) relative to the asymptomatic period (**B**). The right parietotemporal junction is more hypoactive in patients with more severe derealization. (From Kas A, Lavault S, Habert MO, Arnulf I. Feeling unreal: a functional imaging study in 41 patients with Kleine-Levin syndrome. *Brain* 2014;137:2077-87, with permission.)

patients[14] was not replicated in 108 American patients[11] or in 28 children in Taiwan.[15] Although familial risk is low (1% per first-degree relative), 5% of cases feature an affected family member, suggesting an 800- to 4000-fold increased risk in first-degree relatives.[11] KLS probably is not a form of epilepsy because epileptiform EEG activity is very rare,[77-79] and the symptoms are not relieved by administration of anticonvulsants during episodes.

TREATMENT

General Management

No formal, established treatment protocol for KLS has been recognized. Drug trials have been disappointing, and none has been performed in a placebo-controlled or blinded manner. Patients and families benefit from reassurance, simple hygiene rules, and home management. During the episodes, it is recommended to let the patient sleep at home in a familiar environment under family supervision, rather than resorting to hospitalization. This approach reduces anxiety related to novelty and the risk of embarrassing public behaviors and is safer for the patient. It is important to explain to the family that attempts to wake up or stimulate the patient are useless and painful. Driving should be firmly forbidden during episodes because sleepiness, automatic behavior, and altered perception increase the risk of a road accident. The family should regularly check during an episode that the patient drinks and eats enough (in case of reduced eating/drinking) or not too much (in case of hyperphagia), urinates at least once a day (in case of urine retention), and has no suicidal ideas. Between episodes, patients should keep a regular sleep-wake schedule (because sleep deprivation can trigger episodes), refrain from alcohol consumption, and avoid contact with others who may be infectious.

Medications during Episodes

Once an episode has started, there is no evidence that medications can stop its development. Still, we recommend trying amantadine (an antiviral and stimulant therapy) at least once, which may help abort episodes, as reported by 50% patients in a cross-sectional interview.[11] During episodes, stimulants (modafinil, methylphenidate, amphetamine) may partially improve alertness, but these medications have no effect on apathy, derealization, and confusion, as reported by 20% of patients[11] and by 40% of treating physicians.[20,22,73] When psychotic symptoms are prolonged and prominent, risperidone seems to be more helpful than other neuroleptics in retrospective and cross-sectional series.[11,20] In cases in which major anxiety occurs, a benzodiazepine can be of some help.

Drugs Preventing New Episodes

When episodes are frequent, disabling, or prolonged, preventive medications can be explored—notably, lithium.[14,20,52] The reported benefit of lithium in reducing the frequency and even stopping the disease varies, ranging from 25% to 80%.[11] In our experience, with serum lithium levels between 0.8 and 1.2 mmol/L,[80] episodes stopped completely in 35% of patients and were less frequent or less severe in another 45% of patients, with immediate relapses within 2 days when lithium was discontinued.[81] The potential risks of lithium therapy are thyroid and kidney insufficiency[82]—hence the importance of adequate hydration and regularly monitoring serum levels of lithium as well as thyroid-stimulating hormone and creatinine.[11,20] Lithium may be tapered after a few years of complete benefit or after the age of 30 years .

Antiepileptic mood stabilizers (e.g., valproate) seem less effective than lithium in our experience.[11,20] In women with menstrual-associated KLS, estrogen-progesterone may be tried,[59] although we have not seen any obvious benefit

(except from preventing undesired pregnancy with severe sexual disinhibition). Antidepressants appear to have no benefit in KLS.[11,20]

CLINICAL PEARLS

- The interview of the patient's family is the most important part of the diagnosis of KLS, because the patient's report often is blurred by amnesia and altered perceptions during episodes. Videos made by the family showing behavioral changes during episodes are very helpful. Sleep monitoring during episodes is too inconsistent to aid in diagnosis.
- It is mandatory to document clear-cut episodes (at least two) with altered sleep, cognition, and behavior, as well as clear-cut asymptomatic periods, at least during the first years of the disease. Parents should keep a calendar tracking episodes and "good" periods.
- The diagnosis is essentially clinical and retrospective—hence the importance of the history, focused on symptoms of apathy (e.g., patient not using the cell phone) and derealization (subjective feeling of being "in a dream" or feeling "unreal"). Functional brain imaging, which is easily feasible during asymptomatic periods, often (70%) shows mild hypoperfusion of the parietotemporal junction and of the mesiotemporal lobe, along with asymmetric perfusion of the thalamus.
- Longer (lasting more than 1 month) episodes affect 28% of patients. The prolonged duration may be a feature at disease onset and is associated with longer disease duration and larger impact on quality of life.

SUMMARY

KLS is a rare, relapsing-remitting disease of unknown origin. The disease usually starts during adolescence and has a male (66%) predominance; one third of patients have a history of birth or developmental problems. KLS is characterized by recurrent episodes of hypersomnia (more than 18 hours of sleep/day), lasting one to several weeks, associated with cognitive impairment, derealization, apathy, and behavioral changes. Less frequently, episodes include disinhibition with hyperphagia and hypersexuality, or depression, hallucinations, and delusions. Between episodes, patients have normal sleep, mood, cognition, and behavior for one to several months. Episodes may be triggered by infection, alcohol intake, or sleep deprivation. Episodes usually become less frequent and less severe by age 30. No diagnostic test for KLS is available, but results of functional brain imaging often are abnormal, both during and between episodes. Lithium therapy probably decreases the frequency of the episodes.

Selected Readings

Arnulf I, Lin L, Gadoth N, et al. Kleine-Levin syndrome: a systematic study of 108 patients. *Ann Neurol* 2008;**63**:482–93.

Arnulf I, Rico T, Mignot E. Diagnosis, disease course, and management of patients with Kleine-Levin syndrome. *Lancet Neurol* 2012;**11**:918–28.

Dauvilliers Y, Mayer G, Lecendreux M, et al. Kleine-Levin syndrome: an autoimmune hypothesis based on clinical and genetic analyses. *Neurology* 2002;**59**:1739–45.

Engstrom M, Landtblom AM, Karlsson T. Brain and effort: brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Front Hum Neurosci* 2013;**7**:140.

Gadoth N, Kesler A, Vainstein G, et al. Clinical and polysomnographic characteristics of 34 patients with Kleine-Levin syndrome. *J Sleep Res* 2001;**10**:337–41.

Huang YS, Guilleminault C, Kao PF, Liu FY. SPECT findings in the Kleine-Levin syndrome. *Sleep* 2005;**28**:955–60.

Huang Y, Guilleminault C, Lin K, et al. Relationship between Kleine-Levin Syndrome and upper respiratory Infection in Taiwan. *Sleep* 2012;**35**:123–9.

Kas A, Lavault S, Habert MO, Arnulf I. Feeling unreal: a functional imaging study in 41 patients with Kleine-Levin syndrome. *Brain* 2014;**137**:2077–87.

Lavault S, Golmard JL, Groos E, et al. Kleine-Levin syndrome in 120 patients: differential diagnosis and long episodes. *Ann Neurol* 2015;**77**:529–40.

Mayer G, Leonhard E, Krieg J, Meier-Ewert K. Endocrinological and polysomnographic findings in Kleine-Levin syndrome: no evidence for hypothalamic and circadian dysfunction. *Sleep* 1998;**21**:278–84.

A complete reference list can be found online at ExpertConsult.com.

Section 13

Parasomnias

Chapter 101

Parasomnias: Overview and Approach

Bradley Vaughn

Chapter Highlights

- Parasomnias are part of a larger group of disorders that produce nocturnal events.
- Parasomnias can be divided into those associated with non–rapid eye movement sleep, those associated with rapid eye movement sleep, and those associated with transitions or no specific sleep stage.
- The evaluation of parasomnias depends on an accurate history and clear description of the events.
- Key features in the history, physical examination, and polysomnography help distinguish the parasomnias.
- Parasomnias offer an opportunity to diagnose other potential sleep, medical, neurologic, or psychiatric disorders.

Parasomnias are part of a larger group of disorders that produce nocturnal events. Most of us envision these nocturnal events as part of entertaining stories from bed partners or roommates. Yet diligent clinicians understand that these events may result in injuries, sleep disruption, and psychosocial impairment and hold the opportunity to diagnose underlying sleep or medical disorders provoking the behavior.[1] Nocturnal events can be typically classified into parasomnias, sleep-related movements, or neurologic, medical, or psychiatric events.

Parasomnias are defined as undesirable physical events or sensory experiences that occur with entry into, during, or arousing from sleep. Many times these events may involve common and usual behaviors, but they also may include bizarre and unusual events like seemingly purposeful movements, perceptions, dreaming, and autonomic output. The *International Classification of Sleep Disorders*, third edition categorizes parasomnias into three major categories: non–rapid eye movement (NREM) sleep-related parasomnias, rapid eye movement (REM) sleep-related parasomnias, and other parasomnias[2] (Table 101-1). Beyond parasomnias some patients may have movement disorders that present in the transition period between wake and sleep or during sleep that can be confused with parasomnias but still represent a risk for harm.[3] Nocturnal events may also include events that go beyond the state of sleep and offer an opportunity to diagnose and potentially treat other medical, neurologic, or psychiatric events.[4] This section (Chapters 101 to 106) addresses the classic parasomnias of NREM and REM sleep–related events and reviews the expansion of other parasomnias and movement issues in sleep.

CLASSIFICATION OF PARASOMNIAS

The categorization of nocturnal events has progressed as our understanding of the mechanisms involved with sleep and wake and of these events has improved. Early classification and nomenclature used the nature of the behavior as the prominent distinguishing feature to categorize the events. Some remnants of this convention still exist in terms such as

Table 101-1 Parasomnia Classification

NREM Sleep–Related Parasomnias

Disorders of arousal (from NREM sleep)
- Confusional arousals
- Sleepwalking
- Sleep terrors

Sleep-related eating disorder

REM Sleep–Related Parasomnias

REM sleep behavior disorder
Recurrent isolated sleep paralysis
Nightmare disorder

Other Parasomnias

Exploding head syndrome
Sleep-related hallucinations
Sleep enuresis
Parasomnia due to a medical disorder
Parasomnia due to a medication or substance
Parasomnia, unspecified

sleepwalking, *sleep-related eating*, and *sexsomnia*. However, as we have developed further understanding of the neurologic drivers of sleep-wake states and the components of consciousness, we have grouped events toward the originating sleep-wake state and to some extent limited common pathology.[5] The brain's three distinct states allow us to understand that the starting physiologic state provides the substrate for which these parasomnias may exist and form the basis of parasomnias associated with NREM sleep, those associated with REM sleep, and those associated with transitions between wake and sleep. This classification scheme also allows us to move toward a classification structure more aligned with physiology and subsequently underlying pathology.

Some parasomnias may represent a mixture of states.[2] This model is best exemplified when considering the NREM sleep–related parasomnias or disorders of arousal (see Chapter 102). The disorders of arousal (sleep terrors, sleepwalking, and confusional arousals) are associated with mixture of NREM sleep and wake (see Table 101-1). These disorders are most likely not distinct but instead represent a continuum of behaviors that share components of NREM sleep associated with minimal cognitive functioning and amnesia for the events, with features of wake such as eyes open (Table 101-2).[1] These events are more likely to be triggered by stimuli, occur after sleep deprivation or psychosocial stressors, and involve a variety of nonstereotyped behaviors. Some patients report a memory of vague visual imagery and auditory impressions. Although no clear neuropathology has been uniformly identified, early studies suggest that these individuals may have compromised ability to inhibit or fully permit arousal from sleep.[6,7] Therefore these parasomnias may be an early indicator of other sleep disruption.

Of the REM sleep–associated parasomnias, recurrent isolated sleep paralysis may represent a mixture of wake and REM sleep (see Table 101-2 and Chapter 103). Although many times associated with narcolepsy, this disorder in isolation is characterized by the intrusion of REM sleep–related paralysis into wakefulness.[7,8] Other REM sleep parasomnias are confined to the state. Nightmare disorder is typically isolated to REM, but the distressing features carry over into wakefulness (see Chapter 104).[9,10] REM sleep behavior disorder is an example of neurologic impairment of the circuitry that produces the REM sleep–associated atonia.[11] This disorder is characterized by dream enactment, typically violent, and lack of the usual paralysis of REM sleep[11] (see Table 101-2). REM sleep behavior disorder has been associated with Parkinson disease, multiple system atrophy, and dementia with Lewy bodies and may predate other symptoms by decades.[12,13] This disorder represents an example of how sleep-dedicated neural circuitry may be uniquely more vulnerable to specific types of degeneration or injury.

Many of the disorders categorized as "other parasomnias" represent events that occur during the transition between wake and sleep (see Table 101-1 and Chapter 105). Some sensory events such as exploding head syndrome and sleep-related hallucinations are events that may occur as the patient enters light sleep but that also may occur on awakening.[14-16] Additionally in this group are parasomnias that occur across the spectrum of sleep states or represent a loss of sleep-wake state distinction, and some movements can be for parasomnias (see Chapter 106). Parasomnias can be initiated by medications or other neurologic, psychiatric, or medical disorders.[4,17-21] Some hypnotics with short half-lives have been implicated in initiating parasomnia behavior.[17,18] Medical issues that provoke arousals such as chronic obstructive pulmonary disease or renal disease have been reported to cause parasomnic behaviors. Parasomnias have been associated with other neurologic degenerative and autoimmune disorders.[19,22] Recent identification of a novel parasomnia disorder associated with antibodies directed toward IgLON5 raise a new pathophysiologic mechanism for nocturnal events.[22] Several researchers have postulated the existence of an overlap disorder in which patients lose the neurologic ability to express discrete sleep stages.[23]

Events can also occur during the night that may not be truly part of sleep. Approximately 20% of patients with known epilepsy have seizures occurring predominantly at night, and the behaviors may frequently overlap with those seen in parasomnias.[24,25] The key element is that these events have stereotypic behavior usually at the beginning of the event. Other neurologic disorders can present with nighttime events, such as confusion in association with encephalopathy or dementia.[26] Psychiatric disorders, such as panic attacks or dissociative events, may be predominantly expressed during the night hours as nocturnal events.[27,28] Other medical disorders can also evoke arousals with disturbed mentation. Hypoglycemia, from nocturnal insulin administration, can cause arousal with cognitive impairment typically in the early morning.[21] Similarly, cognitive depressant medications taken before bedtime may also impair cognition in individuals when waking at night.

APPROACH TO DISTINGUISHING NOCTURNAL EVENTS

The goal of any evaluation of a patient with nocturnal events is to prevent subsequent harm. The initial consultation should focus on the following questions: (1) Is the patient at risk for potential harm or causing harming to someone else? (2) What may be driving the appearance of these events? and (3) Are these events indicating another underlying disorder?

Table 101-2 Distinguishing Features of Nocturnal Events

Feature	Disorders of Arousal	Sleep-Related Eating Disorder	REM Behavior Disorder	Recurrent Isolated Sleep Paralysis	Exploding Head Syndrome	Psychogenic Events	Nocturnal Seizures
Behavior	Confused; semipurposeful movement with eyes open	Eating typically high-calorie foods; eyes open	Sometimes combative with eyes closed	Episodes of inability to move	Painless sensation of explosion inside the head	Variable	Dependent on the portion of brain involved
Age of onset	Childhood and adolescence	Variable	Older adult	Variable	Adult	Adolescence to adulthood	Variable
Time of occurrence	First third of night	First half of night	During REM	Typically on awakening	Usually near sleep onset but can be variable	Anytime	Anytime
Frequency of events	Less than one per night	Variable	Multiple per night	Variable less than weekly	Rare	Variable	Frontal seizures—multiple per night
Duration	Minutes	Minutes	Seconds to minute	Seconds to minutes	Seconds	Variable minutes or longer	Usually under 3 minutes
Memory of event	Usually none	Usually none or limited	Dream recall	Yes	Yes	None	Usually none
Stereotypical movements	No	No	No	No	Similar sensation	No	Yes
Polysomnogram findings	Arousals from slow wave sleep	Arousal from NREM sleep	Excessive electromyogram tone during REM sleep	Arousal from REM sleep	Usually occurs in light sleep	Occur from awake state	Potentially epileptiform activity

Table 101-3 Indications for Polysomnography in Patients with Nocturnal Events

Atypical presentation for a parasomnia (time of night, behavioral description)
Events injurious or with significant potential for injury
Significant disturbance to patient's home life
Unusual age of onset
Events stereotyped or repetitive
Unusual frequency of the events
Patient has excessive daytime sleepiness or complaints of insomnia
Complaints suggestive of sleep apnea, periodic limb movements, or other sleep disorders

In general, one can differentiate parasomnias by looking for key distinguishing features (see Table 101-2). The foundation of any evaluation of nocturnal events is a thorough history and physical examination. Although there are no absolutes, the underpinning of the evaluation is based on a clear description of the events from witnesses who can give an accurate testimony of the behaviors. Historical features such as time of night, duration, frequency of occurrence, behavioral characteristics with each event, eyes open or closed, memory recall, age of onset, and family history of nocturnal events may help differentiate these disorders.[2] The physician should also search for factors that precipitate parasomnias such as poor sleep environment, improper sleep hygiene, sleep deprivation, circadian rhythm abnormalities, other sleep disorders, medical issues, fever or other illnesses, emotional stress, medication use, and ingestion of alcohol or sedatives before sleep onset.[2,17,18,29,30] Additional search for other neurologic symptoms such as decreased sense of smell, constipation, or other autonomic issues may give clues to REM sleep behavior disorder.[31] Similarly, features suggesting cognitive decline in an adult may provide the opening for further investigation of encephalopathic processes or dementia.

Polysomnography can provide important information in determining the etiology of the nocturnal events, with the goals of capturing the physiology of each sleep state and evaluating the possibility of other contributing sleep disorders.[2] Overnight polysomnography is necessary if the history is atypical, sleepiness is significant, other sleep disorders are suspected, or the patient is at risk for self-harm or harming others (Table 101-3).[32] Studies should include complete respiratory monitoring, time-synchronized video monitoring, additional electromyographic recording from all four limbs, a complete set of cephalic electrodes, and ability to extensively review the electroencephalogram.[11,24,30] Incorporation of a full 10- to 20-electrode array and ability to view the tracing in 10-second windows are necessary in evaluating for seizures and the differentiation of the epileptiform discharges from potential normal variants or artifacts.[33]

CLINICAL PEARL

Disorders of arousal are more common in children but can also present in adults. When patients have a recurrence of somnambulism, sleep terrors, or confusional arousals, the clinician should ask about other features that may be causing arousals. Sleep-related events such as sleep apnea or periodic limb movements may provoke parasomnias for which the parasomnia is the sentinel symptom hallmarking the other sleep issue.

SUMMARY

Parasomnias as part of a greater group of nocturnal events allow us to examine the interaction of behaviors we associate with wakefulness during sleep. Although we recognize the three normal states of being as wakefulness, NREM, and REM sleep, we understand that these distinct states may not be as distinct as "a flick of a switch" phenomenon. Disruption of the neuronal processes determining a state of being can cause a mixture of these states. Thus behaviors normally accompanying one state may intrude into another.[2] The mechanisms determining the state distinction also may be impaired and allow the mixture these states, such as in disorders of arousal. This process of state change can be disrupted by disorders causing arousals such as sleep apnea or poor sleeping environment. Similarly, nocturnal events can be the manifestation of a vulnerability of a neuronal circuit to disease process, such as in REM sleep behavior disorder. In addition, sleep-related movements may be easily misinterpreted as part of a parasomnia, as may other medical, neurologic, or psychiatric disorders. A detailed history and clear description of the events may aid in establishing clues about the underlying etiology. The challenge for the clinician is to recognize the prospect of using these nocturnal events as the trigger to diagnose and treat other underlying issues.

Selected Readings

Boeve BF, Silber MH, Ferman TJ, et al. Clinicopathologic correlations in 172 cases of rapid eye movement sleep behavior disorder with or without a coexisting neurologic disorder. *Sleep Med* 2013;**14**:754–62.

Foldvary N, Caruso AC, Mascha E, et al. Identifying montages that best detect electrographic seizure activity during polysomnography. *Sleep* 2000;**23**:221–9.

Howell MJ, Schenck CH. Rapid eye movement sleep behavior disorder and neurodegenerative disease. *JAMA Neurol* 2015;**72**:707–12.

Khachiyants N, Trinkle D, Son SJ, Kim KY. Sundown syndrome in persons with dementia: an update. *Psychiatry Investig* 2011;**8**:275–87.

Nevéus T. Nocturnal enuresis-theoretic background and practical guidelines. *Pediatr Nephrol* 2011;**26**:1207–14.

Pilon M, Montplaisir J, Zadra A. Precipitating factors of somnambulism: impact of sleep deprivation and forced arousals. *Neurology* 2008;**70**:2274–5.

Zucconi M, Ferini-Strambi L. NREM parasomnias: arousal disorders and differentiation from nocturnal frontal lobe epilepsy. *Clin Neurophysiol* 2000;**111**(Suppl. 2):129–35.

A complete reference list can be found online at ExpertConsult.com.

Chapter 102

Non–Rapid Eye Movement Parasomnias: Clinical Spectrum, Diagnostic Features, and Management

Alon Y. Avidan

Chapter Highlights

- Non–rapid eye movement (NREM)–related parasomnias range from mild episodes that often do not disrupt the patient's sleep, such as mild confusion with sleep talking, to aggressive and potentially injurious motor disturbances such as sleepwalking and sleep terrors.
- NREM-related parasomnias generally occur as an incomplete transition from NREM deep sleep to wakefulness, typically with the patient appearing in a state of impaired arousal, lasting for a few seconds to minutes, and is associated with partial or complete amnesia.
- Essential physiologic drives, such as hunger, aggression, and sex, may be manifested by sleep-related eating behaviors, sleep-related violence, or sleep-related sexual behaviors.
- Despite their odd and peculiar clinical presentations, NREM parasomnias are readily explainable, diagnosable, and treatable.
- Other sleep disorders, such as sleep apnea, may cause arousals that provoke a NREM-related parasomnia.

Non–rapid eye movement (NREM)-related parasomnias are defined as *undesirable and often abnormal motor or subjective phenomena that arise during arousals from NREM sleep.*[1-5] The episodes may include abnormal movements, behaviors, emotions, and autonomic activity, most of which are explainable, diagnosable, and treatable.[3,5-10] These parasomnias may occur in response to internal factors, such as apneic episodes and fever, or may be triggered by external stimuli. The latest (third) edition of the *International Classification of Sleep Disorders* (ICSD3) defines NREM parasomnias to include the disorders of arousal (confusional arousals, sleepwalking, and sleep terrors) and sleep-related eating disorder (SRED). Sexsomnia is included as a behavior under confusional arousals and sleep walking, but some sleep specialists consider this as a distinct parasomnia.

CLINICAL FEATURES OF DISORDERS OF AROUSAL

The arousal disorders are classified as such on the basis of the following common characteristics: (1) pathophysiology involving impaired arousal from deep sleep, (2) genetic and familial patterns of inheritance, (3) predisposition secondary to sleep fragmentation, (4) impaired cognitive functioning during the event, and (5) universal amnesia after the event.[6] Although these parasomnias share these common characteristics, the events are phenomenologically distinguished by the clinical presentation and behaviors. Sleep terrors begin with a sudden explosion of sympathetic activity, distress, and expression of fear, which diminish over time, whereas confusional arousals and sleepwalking rarely begin with distress. Confusional arousals usually consist of normal arousal behaviors, with abnormal duration rarely manifesting in explosive distress or motor behavior. Similarly, somnambulistic behaviors consist of normal arousal behaviors at onset, proceeding to nonagitated motor behavior with lack of distress. Figure 102-1 extrapolates in a graphic format the spectrum of behavioral manifestations of the arousal disorders, the duration of the arousal episodes, the magnitude of displacement/ambulation, and the degree and intensity of distress as a function of time.[11] Occasionally, an admixture of two arousal disorders ("hybrid attacks/parasomnias"), comprising waxing and waning of the multiple behavior types, may occur. All events usually start in stage N3 NREM sleep and terminate either in wakefulness or during lighter NREM sleep. The episodes typically are brief (*solid lines* in Figure 102-1) but sometimes can be prolonged (*hatched lines*).

CLASSIFICATION OF THE DISORDERS OF AROUSAL

The ICSD3 diagnostic criteria for the disorders of arousal are outlined in Table 102-1. The behavioral episodes commonly begin as partial arousals from stage N3 slow wave sleep and typically are brief but can be protracted, lasting up to 30 minutes.

Other behavioral manifestations may be based on the following additional features[6]:

Inappropriate speech: Sleeptalking, utterance of random speech, or shouting may occur.

Confused demeanor and stare: During an episode, the eyes often are wide-open, with a confused "glassy" stare.

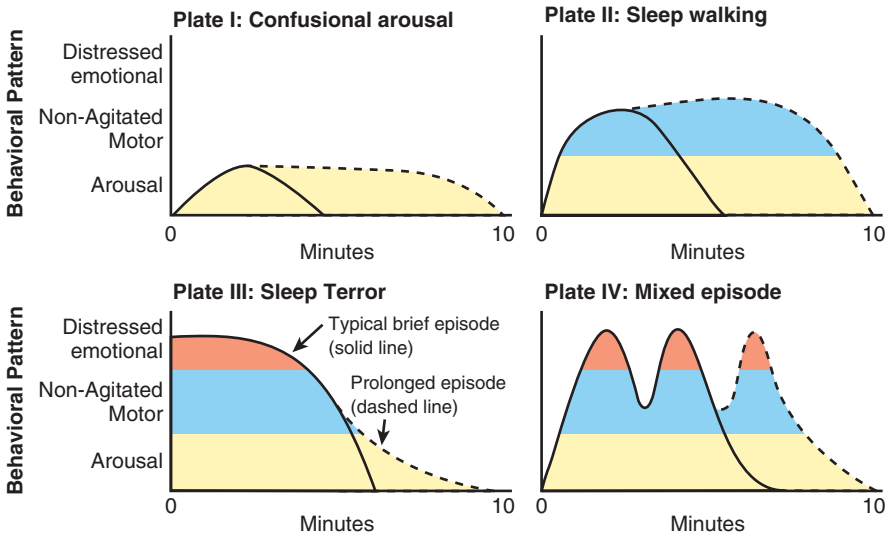


Figure 102-1 Semiology of Disorders of Arousal as a Function of Time. Diagrammatic representation of the common behavioral semiologic pattern as a function of time for the key disorders of arousal (NREM parasomnias). Depicted are hierarchical combinations of the three behavior states on the vertical axis (arousal → nonagitated motor → distressed state) as a function of time (1 to 10 minutes) on the horizontal axis. *Plate I* represents a typical confusional arousal spell. The parasomnia consists of normal arousal behaviors but of abnormal duration only. *Plate II* depicts a classic somnambulistic event comprising normal arousal behaviors at onset, proceeding to nonagitated motor behavior. *Plate III* illustrates a typical episode of sleep terrors, beginning with an intense autonomic discharge, appearance of distress, experience of predominantly negative emotional behavior typically of sudden onset; motor and normal arousal behaviors usually are also seen during these events, either at onset or thereafter. *Plate IV* is a mixed type, comprising two different arousal disorders, referred to hybrid attacks or hybrid parasomnias, with waxing and waning of the multiple behavior types. All events usually start in stage N3 NREM sleep and terminate either in wakefulness or in lighter NREM sleep. The episodes typically are brief (*solid lines*) but sometimes can be prolonged (*hatched lines*). Patients who start out with a confusional arousal who experience locomotion and displacement from the bed would be categorized as experiencing somnambulism. (Modified from Derry CP, Harvey AS, Walker MC, et al. NREM arousal parasomnias and their distinction from nocturnal frontal lobe epilepsy: a video EEG analysis. *Sleep* 2009;32:1637–44.)

Table 102-1 *International Classification of Sleep Disorders* Criteria* for the Disorders of Arousal

1. Repeated episodes of partial or incomplete awakening from NREM sleep
2. Inappropriate response, or failure to respond, to efforts of observers to intervene or redirect the affected person during the episode
3. Diminished to absent cognitive functioning or dream-like mentation or imagery after the event
4. Fragmented or no recollection for the episode
5. Absence of other primary sleep disorder, mental disorder, medical condition, medication, or substance use to help explain the disturbance

*Modified from the *International classification of sleep disorders—third edition* (ICSD-3). Darien (Ill.): American Academy of Sleep Medicine; 2014.

Absence of response to external stimuli: The patient with an arousal disorder may be very difficult to awaken and, even when the efforts succeed, the patient does not return readily to baseline function, remaining in a prolonged state of confusion and disorientation. This is termed *sleep inertia*.

Explosive onset: The patient with spells suggestive of an arousal disorder typically experiences a sudden, abrupt, and *explosive* onset associated with abnormal motor, behavioral, autonomic, or sensory symptoms.[12]

Universal amnesia: Children with disorders of arousal often have complete amnesia for these episodes, whereas adults may retain fragmentary recollection.

Dreaming mentation: Dreamlike mentation or similar altered state of consciousness has been reported occasionally after such arousal episodes.

Diminished cognition: Impairment in or absence of higher cognitive functioning is characteristic of disorders of arousal.

Arousals may be from any NREM sleep stage but primarily arise from stage N3 sleep and have a propensity to occur during the first third of the night. Disorders of arousal are common in children; in this age group, they typically are considered to be normal age-related sleep manifestations, generally not requiring specific interventions.[13,14] Predisposing factors include febrile illness, emotional stress, sleep deprivation, alcohol, a full bladder, and central nervous system (CNS)-acting medications.[15] In both adults and children, primary sleep disorders such sleep apnea and periodic limb movements also may provoke disorders of arousal.[16,17]

PATHOPHYSIOLOGY OF THE DISORDERS OF AROUSAL

Several physiologic mechanisms have been proposed to help explain the disorders of arousal, but a current prevailing theory postulates that they occur as a result of incomplete transition from one sleep state into another. Sleep stage

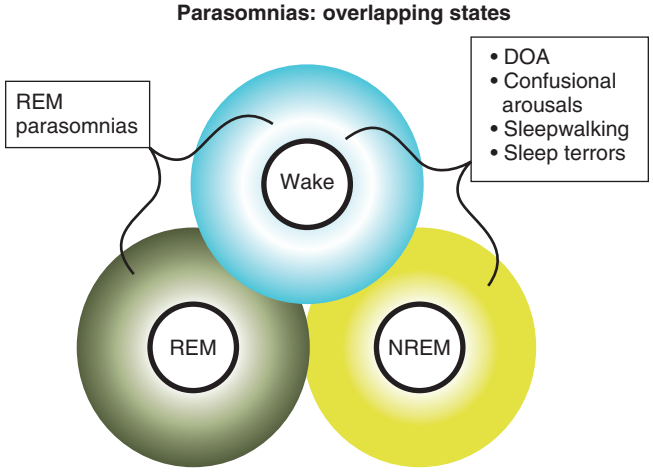


Figure 102-2 Parasomnias as State Dissociation Disorders. Under various conditions, abnormal admixtures of the three sleep-wake states—(1) NREM sleep, (2) REM sleep, and (3) wakefulness—may occur, with consequent overlap, giving rise to parasomnias. Wakefulness and sleep are not mutually exclusive states, and sleep-to-wake dissociation, incomplete transition, or oscillation from one sleep state to another leads to the parasomnias. Parasomnias are hypothesized to be due to changes in brain organization across multiple sleep-wake states: incursion of wakefulness into NREM sleep leads to the disorders of arousal, and intrusion of wakefulness into REM sleep produces REM sleep parasomnias, of which REM sleep behavior disorder (RBD) is the most dramatic and clinically important. (Modified from Mahowald MW, Schenck CH. Non-rapid eye movement sleep parasomnias. *Neurol Clin* 2005;23(4):1077–106, vii; and Avidan AY, Kaplish N. The parasomnias: epidemiology, clinical features, and diagnostic approach. *Clin Chest Med* 2010;31:353–70.)

progression entails coordination of several neuronal centers for an equivocal declaration of stage.[2,5] As illustrated in Figure 102-2, vulnerability to incomplete transition is at a maximum between NREM sleep and the awake state. During this period of sleep-stage admixture, the intrusion of one state into another may result in complex behaviors.[5,6,18,19] Some evidence for this admixture rests on studies showing that affected patients may have impaired awakening mechanisms. Additionally, the partial arousal state may activate central pattern generators in the neuroaxis, producing complex involuntary motor events such as walking.[2]

CLINICAL EVALUATION AND DIFFERENTIAL DIAGNOSIS OF DISORDERS OF AROUSAL

The clinical evaluation of patients suspected to have an arousal disorder should focus on the key items to aid in establishing the diagnosis. These key items rely heavily on historical features as well as a search for other associated features. In general, patients with NREM parasomnias require in-laboratory investigation if they are at risk for hurting themselves or are suspected of having another sleep or medical disorder, or if the events are disruptive to sleep or other functioning. The key elements of the clinical semiology—age at onset, frequency, and duration—are established in the initial evaluation. A clear description of the events, from either the patient or the bed partner, is essential to determine the underlying disorder, and the details of multiple occurrences will further clarify the etiology. In comparison with nocturnal seizures, the parasomnias are characterized by a variety of behaviors with a low frequency, as opposed to the stereotypical behavior and higher frequency of events in a given night with frontal lobe seizures.[20,21] Confusional arousals consist of episodes of abrupt awakening associated with confusion and disorientation. If displacement takes place out of bed, the event is categorized as sleepwalking and may include a range of behaviors from simple automatic non–goal-oriented behavior to more complex violent, inappropriate, agitated behavior. The hallmark of sleep terrors is a piercing scream signaling arousal, associated with increased sympathetic response and aggression. *Age at onset* for arousal disorders commonly is in childhood or the teens, generally after puberty. By contrast, nocturnal seizures, such as nocturnal frontal lobe epilepsy (NFLE), typically emerge between the ages of 10 and 20 years but may persist into young adulthood, and REM sleep behavior disorder (RBD) more typically affects older adults.[20,22] The *frequency* of events in disorders of arousal typically is only a few times per month or week, rarely with more than one event per night, and the arousals usually occur in the first half of the night. NFLE and RBD events may occur several times per night at any time during the night, with the latter more common in the second half of the night. The *duration* of events in disorders of arousal may be more prolonged, especially when sleep inertia is prolonged. NFLE and RBD typically are of brief duration. As a general rule, nocturnal seizures begin with stereotypical, sometimes bizarre semiology (or only minimal behavioral abnormalities), may be followed by postictal confusion, and can occur at any time of night and in any age group, whereas RBD is not characterized by stereotypy, and patients usually retain memory for the coinciding dream mentation[20,23,24] (Table 102-2). A recently described self-administered scale for arousal disorders, *Paris Arousal Disorders Severity Scale* (PADSS), may be useful in clinical practice to assess patients presenting with a history of abnormal or violent nocturnal behaviors. This scale has been found to have reasonably sound psychometric attributes.[25]

Confusional Arousals

Essential and Associated Characteristics

Confusional arousal, also referred to as a state of "sleep drunkenness" or an interval of "excessive sleep inertia," consists of a brief period of confusion and disorientation during and after arousal from slow wave sleep.[26-28] To an observer, patients appear to experience confused mentation, display inappropriate behaviors, and are slow or fail to respond to questioning. The timing of the episodes is most commonly during the first half of the night, in keeping with the higher propensity for arousal out of stage N3 sleep. Confusional arousals typically last a few minutes, are terminated by reinitiation of sleep, and are associated with retrograde and anterograde amnesia. Sexsomnias are considered a subtype of confusional arousals and consist of inappropriate amnestic sexual behaviors, sometimes triggered by primary sleep disorders.[29]

The proposed pathophysiology of confusional arousals is an incomplete awakening from slow wave sleep leading to intensification and prolongation of the normal period of sleep inertia.[30] Predisposing factors may include sleep deprivation and recovery from sleep curtailment, circadian rhythm sleep disorders (shift work disorder), fever, sleep-disordered breathing, CNS depressants (sedative-hypnotics and antihistamines), exposure to stimulants, or any factors that deepen sleep and impair ease of awakening.[31] Confusional arousals may be experimentally induced by attempts at forced arousal from

Table 102-2 Key Features and Treatment for Parasomnias

	Disorders of Arousal			REM Parasomnias		
Feature	**Confusional Arousal**	**Sleepwalking**	**Sleep Terrors**	**RBD**	**Nightmares**	**Nocturnal Seizures**
Timing at night and sleep stage specificity	Usually occurs in the first half of the night and typically out of stage N3 slow wave sleep			Begin out of REM sleep and typically occur in the last third of the night		At any time of night, but usually out of NREM sleep Favoring occurrence when EEG is synchronized
Family history	Frequently positive (in 62% to 96%)			No	May be positive	May be positive ~39% more likely, especially in nocturnal frontal lobe epilepsy (NFLE)
Behavior semiology	Sudden arousals followed by confusion, disorientation, and amnesia for the event	Abrupt arousal, potential for confusion/agitation if interrupted vigorously Amnestic on following day	Sudden arousal with intense screaming, inconsolable crying, agitation, and heightened autonomic discharge	Arousal followed by purposeful dream enactment, including yelling, punching, kicking, or fighting a supposed intruder/animal	Paroxysmal awakenings with anxiety and dream recall	Stereotypical, monomorphic paroxysmal events, often with dystonic limb posturing, vocalizations and confusions; may have partial recall/amnesia
Event duration	Few seconds to minutes	Usually 1–10 minutes	Few seconds to minutes	Usually less than 10 minutes	Few seconds to minutes	Few seconds to few minutes
Age at onset (years)	Typically <10			Typically >60	Typically <10	14 ± 10
Triggering factor(s)	Sleep deprivation, febrile illness, anxiety, stress, sleep apnea apnea, motor disorders of sleep			Medications (antidepressants), caffeine, alcohol	Yes: sleep deprivation, febrile illness	Not typical but may be precipitated by sleep deprivation
Frequency	Few times per month/week; very rarely >1 episode in a single night					Frequent: May occur multiple times in a night
Postspell behavior	Limited to no recall of the events with confusion			Recall is typical at times with vivid details. Patients with RBD often describe needing to protect themselves from an attacker (animal/intruder)		Complete/partial recall to amnesia and confusion
PSG changes	Abrupt arousal from slow wave sleep (stage N3) with expressions of confusion/ambulation/intense fright in all three arousal types, followed by return to sleep; increased cyclic alternating pattern (CAP)			Abnormal increased chin or limb EMG tone (atonia is noted during normal REM sleep)	Dense eye (phasic) movements during REM sleep	Epileptiform activity/muscle artifact/or normal EEG if limited montage
Treatment	Safety interventions/protect patients; remove sharp objects from bedroom, cover windows, barricade furniture, place door alarms Avoidance of precipitating factors, protecting sleep environment, improving sleep hygiene, avoidance of sleep deprivation Hypnosis, anticipatory awakenings are helpful If episodes are frequent, severe, and disruptive to sleep continuity, or result in daytime sleepiness or injury, consider pharmacotherapy			RBD: Level A: promote safety Level B: pharmacotherapy with melatonin or clonazepam Nightmares: Reassurance, avoid injury, treat precipitating factors		Antiepileptic drugs, carbamazepine most frequently used for normal seizures

EEG, Electroencephalogram; EMG, electromyogram; PSG, polysomnography; RBD, REM sleep behavior disorder; SRED, sleep-related eating disorder.

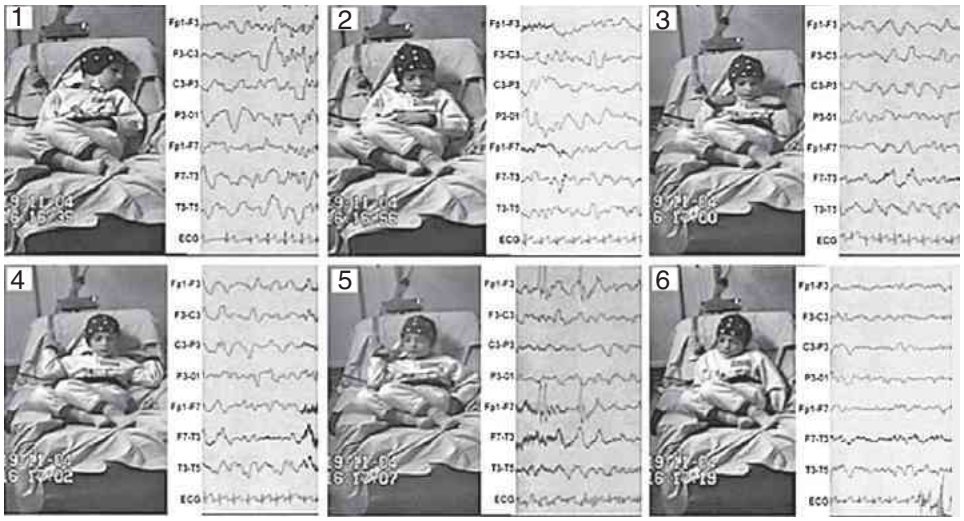


Figure 102-3 Confusional Arousals in a 7-Year-Old Boy. The parasomnia begins with the patient in stage N3 NREM sleep, depicted by delta activity in the right panel of the first fragment of the recorded polysomnogram (1). The episode progresses with the patient's raising the head (2) and placing his hands on the chest (3). The patient then raises his arms to touch the recliner back (4) and repositions his body (5 and 6). The total duration of the episode was longer than 30 seconds, but the electroencephalogram (EEG) remained characteristic of slow wave sleep, as exemplified by delta-to-theta range activity concurrent with motor behavior. Compared with patients with sleepwalking and sleep terrors, patients with confusional arousals present with less complex ambulatory events in bed but never with walking or terror-specific behavior. (From Tinuper P, Bisulli F, Provini F. The parasomnias: mechanisms and treatment. *Epilepsia* 2012;53[Suppl. 7]:12–9.)

slow wave sleep and, in adults, during recovery from sleep deprivation.

Demographic Features and Epidemiology

Confusional arousals are almost universal among children younger than 5 years of age and are less common in older childhood. The prevalence in the adult population is approximately 4%.[32] In a European study involving patients aged 15 or older, confusional arousals were reported by 2.9% of subjects in the sample.[33] In one recent cross-sectional study in the United States surveying 19,136 healthy adults older than 18 years of age, 15.2% of participants reported episodes of confusional arousals in the previous year. Of this sample, 8.6% described full or partial amnesia for the episodes and 14.8% had confusional arousals and nocturnal wandering episodes.[34] In adults, demographic assessment determined that populations at risk include those with an underlying psychiatric disorder. The overlap between confusional arousal and psychiatric comorbidity is significant: 51% of adult patients with confusional arousal reported concomitant symptoms of anxiety, 60% reported depressed mood, and 22% were diagnosed with bipolar disorder.[32,33] A similar epidemiologic study found that confusional arousals were associated with sleep and mental disorders (bipolar and panic disorders) or psychotropic drugs (especially antidepressants) in 84% of the cases, and underlying sleep disorders accounted for 70.8% of the confusional arousals.[34] Additional risk factors include circadian rhythm sleep disorder, sleep deprivation, and prolonged sleep duration of 9 hours or longer.[34]

Objective and Polysomnographic Features

Confusional arousals represent an admixture of wakefulness into non-REM sleep.[6-8,35] Patients awaken partially exhibiting marked confusion, slow mentation, disorientation, perceptual impairment, and errors of logic. Polysomnographic recordings during the episodes demonstrate arousals from deep slow wave associated with brief episodes of delta activity, stage N1 theta patterns, with repeated micro-sleeps, or a diffuse and poorly reactive alpha rhythm (Figure 102-3). The duration of the episode typically is between 30 seconds to several minutes, and the electroencephalogram (EEG) during the recordings depicts characteristics of stage N3 sleep, with delta-to-theta range activity.[15] Unlike in somnambulism and sleep terrors, motor events in confusional arousals are less complex and typically do not include any form of ambulation or sympathetic activation.[15]

Differential Diagnosis

Distinguishing confusional arousals from other paroxysmal disorders and normal arousals can present a diagnostic challenge. Confusional arousals are differentiated from other parasomnias, such as somnambulism, by the lack of displacement out of bed or walking. Confusional arousal episodes lack an acute fear component, intense screaming/crying, and increased autonomic hyperarousal, seen with sleep terrors, nor are there the complex dream enactment behaviors of RBD.[36] Unlike in RBD, confusional arousal episodes tend to occur in children, during delta sleep, and in the first half of the night. Amnesia for the event preceding the arousal is typical in confusional arousals. Complex partial nocturnal seizures of frontal lobe or inferomesial temporal origin may manifest with confusional amnestic semiology. Unlike confusional arousals, however, nocturnal seizures with ictal confusion episodes begin with more stereotypical behavior, have a variable frequency, and may be associated with EEG changes.[23,24]

Kleine-Levin syndrome (KLS), one of the periodic hypersomnolence disorders, is associated with sleep inertia and occasionally may be difficult to differentiate from confusional arousals. During the periods of prolonged sleep, patients with KLS experience the mental confusion and incoherent speech common to sleep inertia. KLS, however, has the additional distinctive features of prolonged periods of sleep and abnormal behaviors of irritability, hypersexuality, and hyperphagia during wakefulness.[37,38] KLS also may be associated with episodes of sleep eating and sleep sexual activity, which further blurs the clinical distinction.[39]

Epidemiologic data showed that 13.2% of patients with confusional arousals had obstructive sleep apnea, compared with 2% of those without such arousals.[33] Patients with sleep-related breathing disorders may engage in a variety of peculiar behaviors after an apneic episode. Figure 102-4 shows a diagnostic polysomnogram of a 54-year-old man who presented with a history of disruptive nighttime confusion and singing behaviors. During the recorded event, an arousal from slow wave sleep, the patient is seen with the arms abducted; he was then observed to "flap" his arms and was described by the technicians to be "quacking like a duck."[18] In the setting of sleep apnea, confusional arousals may result from associated hypoxemia, triggering an arousal during sleep, which in some cases provokes waking neurobehavioral deficits.[33]

Behavioral and Pharmacologic Treatment Options

Treatment is rarely indicated, because in most cases the disorder remits with age and episodes can be prevented or limited by avoidance of the facilitating factors (e.g., sleep deprivation, stimulants). Confusional arousals may be managed conservatively by avoiding sleep deprivation, preventing irregular sleep-wake schedule patterns, limiting exposure to CNS depressants, and managing coexisting sleep disorders. Pharmacologic management often is not necessary because the arousal episodes are self-limiting. In refractory cases, some patients respond to tricyclic antidepressants such as clomipramine (Table 102-3).

Confusional Arousals Subtype: Sexsomnias

Clinical Features and Typical Characteristics

One of the more bizarre and intriguing arousal disorders is a variant of confusional arousals known as *sexsomnia*, somnambulistic sexual behavior, or "sleepsex."[29,40,41] Sexsomnias consist of abrupt nocturnal episodes of often-inappropriate sexual

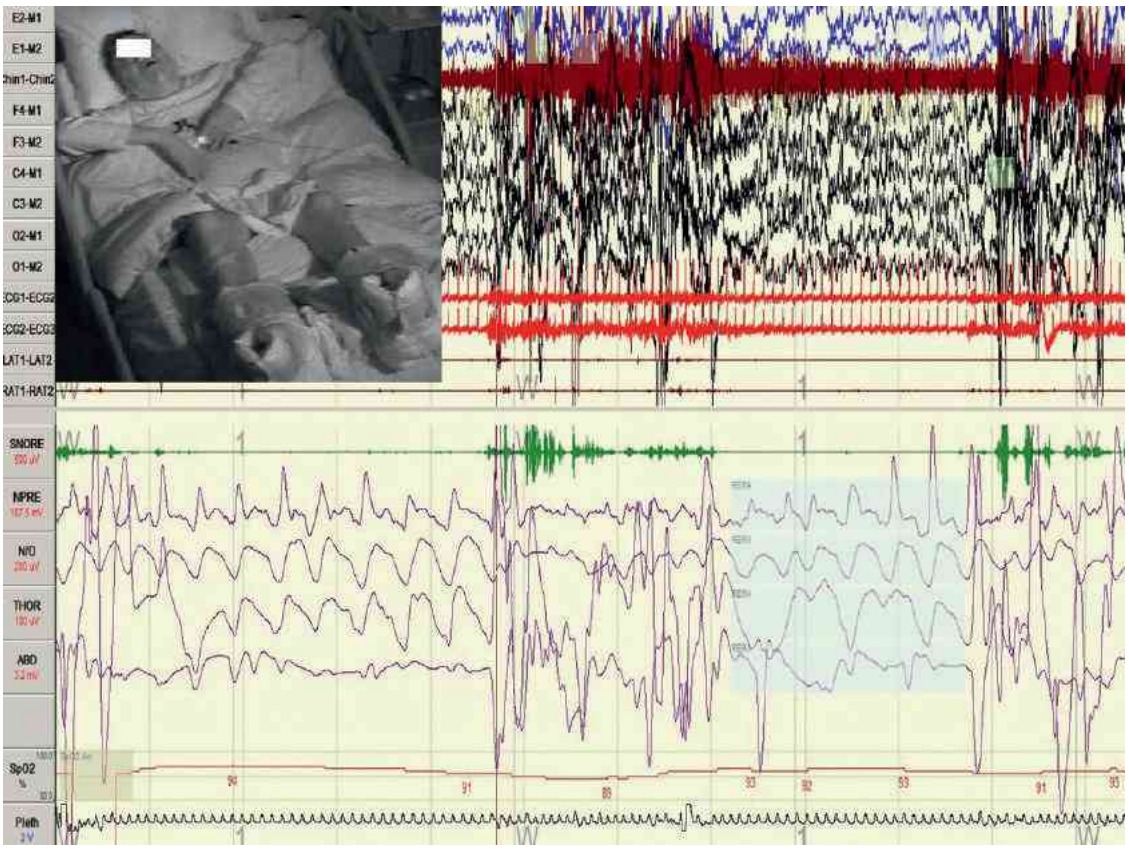


Figure 102-4 Confusional Arousal Secondary to Sleep-Disordered Breathing. A 120second epoch of a diagnostic polysomnogram (PSG) conducted to evaluate arousals with confusion and singing behavior in a 54-year-old man. A representative event for the patient is illustrated: an arousal from slow wave sleep, as demarcated by the *star,* with the patient's arms abducted (in "flapping" his arms; he also was described by the technicians to be "quacking like a duck"). Channels are as follows: electrooculogram (*left*: E1-M2; *right*: E2-M1), chin EMG (Chin1-chin2), EEG (*left*: frontal-F3, central-C3, occipital-O1, left mastoid-M1; *right*: frontal-F4, central-C4, occipital-O2, right mastoid-M2); two ECG, two-limb EMG (LAT, RAT), snore, nasal-oral airflow-N/O, nasal pressure signal-NPRE, respiratory effort (thoracic, abdominal), and oxygen saturation (Sa_{O_2}). ECG, Electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; LAT, left anterior tibialis; RAT, right anterior tibialis. (Modified from Avidan AY,. Kaplish N. The parasomnias: epidemiology, clinical features, and diagnostic approach. *Clin Chest Med* 2010;31:353–70.)

Table 102-3 Treatment Options for Disorders of Arousal

Component	Confusional Arousal	Sexomnias	Somnambulism (Sleepwalking)	SRED	Sleep Terrors
Environmental safety	X	X	X	X	X
Scheduled anticipatory awakening	X		X		X
Behavioral management	Reassurance of benign nature Avoid precipitants: Sleep deprivation Alcohol CNS depressants	Enhancing sleep hygiene, ensuring optimal sleep duration Psychotherapy and stress management Comorbid mood disorder or anxiety[1]	Avoid precipitants: Sleep deprivation Lithium Nonbenzodiazepines receptor agonists		Reassurance of benign nature Relaxation therapy Hypnosis/autogenic training* Psychotherapy
Pharmacologic management	Imipramine Clomipramine Clonazepam	Benzodiazepines Clonazepam Antidepressants Sertraline GABAergic agents Lamotrigine Valproic acid Serotonin reuptake inhibitors	Benzodiazepines Clonazepam (0.5–1 mg) Diazepam (10 mg) Triazolam (0.25 mg) Imipramine (50–300 mg) Melatonin Paroxetine	Dopamine agonists, selective serotonin reuptake inhibitors, and topiramate[†]	Paroxetine (20 to 40 mg) Clonazepam (0.5–1 mg) Trazodone Hydroxytryptophan Imipramine/clomipramine

*Autogenic training is a special relaxation technique similar to meditation.
[†]Side effects include weight loss, cognitive impairment, paresthesias, visual symptoms, and less frequently, renal calculi.
CBT, Cognitive-behavioral therapy; CNS, central nervous system; GABAergic, gamma-aminobutyric acid–ergic; SRED, sleep-related eating disorder.

behaviors occurring with limited awareness during the act, relative unresponsiveness to the external environment, and amnesia for the event.[41]

Patients with sexsomnia engage in often inappropriate amnestic sexual behavior ranging from masturbation, attempting sexual activity with a partner cosleeping in the same bed, and may include attempted sex with a nonpartner with whom the patient does not share a bed or a room.[29] The consequences of sexsomnia can be serious, leading to marital distress and even forensic repercussions in aggressive, inappropriate cases or those involving minor children.[31] Although the underlying cause of sexsomnias is unknown, factors such as fatigue, stress, alcohol use, and substance abuse and physical contact with another person in the bed can precipitate the episodes.[42] The prevalence of sexsomnia is unknown, and the disorder probably is underreported. A recent review from Spain, behavioral semiology in the sexsomnias ranges from gentle and affectionate with the bed partners to violent explosive sexual acts, typically out of character for the afected person.[43] As with any behaviors, sexual behaviors can occur in a variety of disorders including as part of disorders of arousal, RBD, epileptic seizures, and psychiatric events.[42] These events can have significant forensic implications (Chapter 65). Anecdotal evidence suggests that successful amelioration of sexsomnia events entails the following approach: optimizing safety, enhancing sleep hygiene measures, ensuring optimal sleep duration, and, especially in patients with anxiety, psychotherapy and stress management techniques.[40,44,45] Benzodiazepines, particularly clonazepam, may be considered for first-line pharmacotherapy, with a recommended dosage ranging from 0.25 mg to 2 mg at bedtime (Table 102-3).[46]

Somnambulism

Essential Features and Associated Characteristics

Somnambulism, or sleepwalking, consists of complex behaviors that are initiated during slow wave sleep and result in walking during sleep. The episodes generally last from 1 to 5 minutes and may consist of a wide range of activities. Typical spells consist of complex behaviors ranging from simple and calmly sitting up in bed, simple walking to agitated walking, and rarely, in extreme cases, frantic efforts to escape a perceived threatening situation, sometimes accompanied by inappropriate behavior such as urinating.[47] Typical frequency ranges from several times a week to only when precipitating factors are present.[5,19] Somnambulism occasionally may result in falls and injuries incurred during attempts to "escape" or while walking.

Although most who experience sleepwalking are children, this usually benign and self-limiting condition can be associated with violent behaviors, injury to the sleepwalker or the bed partner, sleep disruption, hypersomnia, anxiety, psychological distress, and altered quality of life.[48] Precipitating factors include the use of sedatives, acute sleep deprivation, and specific causes of arousals such extrinsic stimuli (noise).[49] In addition, drugs with effects recognized as possible contributing factors include selective serotonin reuptake

inhibitors,[50,51] bupropion,[52] mirtazapine,[53] paroxetine,[54] and norepinephrine reuptake inhibitors.[49,55,56]

Other medical events and conditions such as fever, untreated sleep apnea, stress, and distended bladder may exacerbate the frequency of sleepwalking episodes. These occurrences demonstrate that sleep fragmentation and deprivation increase sleepwalking episodes and support the hypothesis that an impairment of either the sleep homeostatic mechanism or the arousal mechanism is at the root of these events.[49,50]

Demographic Features and Epidemiology

In a recent review of data for a large population of adults, the lifetime prevalence of nocturnal wandering with abnormal state of consciousness was 29.2%, whereas 3.6% of the sample subjects experienced more than one episode of nocturnal wandering in the previous year. The prevalence of sleepwalking in the general pediatric population is between 1% and 17% and approaches 4% in adults.[32,57] A large study involving more than 1000 children from Los Angeles reported a prevalence of 2.5% of sleepwalking,[58] whereas a study of Swedish schoolchildren revealed a prevalence of 7%,[59] peaking between the ages 4 and 8 years.[60,61] Male-to-female ratio is 1, and familial patterns are common.[3] Recently, a specific human leukocyte antigen gene (*DQB1*) was found to increase susceptibility to sleepwalking.[62]

Objective and Polysomnographic Features

The utility of formal polysomnography in the evaluation of somnambulism is mainly in the exclusion of potential mimics such as nocturnal seizures and RBD, and of other primary sleep disorders that may potentially contribute to this parasomnia. However, two fundamental difficulties impede the ability to accurately describe them: (1) Somnambulistic behaviors do not occur nightly, and (2) when the episodes do occur, they often are less dramatic or complex than previously described by the patient or family members. Polysomnographic findings in patients with sleepwalking episodes usually show frequent arousals from slow wave sleep, the emergence of hypersynchronous slow wave EEG just before and during arousals, and diminished delta activity. These findings may represent instability of slow wave sleep in the early portion of the sleep period[31,63,64] (Figure 102-5). However, these frequent polysomnographic variables are neither diagnostic of nor confirmatory for sleepwalking. The lack of specificity and sensitivity for these biomarkers degrades their usefulness as tools in the clinical or forensic confirmation of parasomnias.[63,65]

Differential Diagnosis

Considerations in the differential diagnosis for sleepwalking include other disorders of arousal, such as sleep terrors and confusional arousals, and RBD. Whereas in RBD, patients are in REM sleep and appear to be acting out a dream sequence, sleep terror and confusion arousals are more abrupt in onset, with the typical association of hypersynchronous and high-amplitude delta activity on polysomnography.[66] The first reports of sleep-related partial complex seizures with ambulatory automatisms appeared almost four decades ago, published by Pedley and Guilleminault,[73] who termed these events "episodic nocturnal wanderings," characterized by paroxysmal ambulation and bizarre behavioral manifestations during sleep.[49,67] Episodic nocturnal wandering episodes may represent complex partial seizures but also can be seen in alcohol intoxication (differentiated by drunken behavior while awake), dementia, and CNS drug–related effects.[49,68]

Behavioral and Pharmacologic Treatment Options

The cornerstone to managing sleepwalking and the other disorders of arousal is maximizing patient safety, achieved by avoiding the precipitating factors and ensuring a safe living environment by removal of sharp-edged furniture, covering windows, eliminating obstacles that may lead to injury during a sleepwalking episode, locking doors, using door alarms, and providing the safest possible sleeping area, on the ground floor rather than upstairs, or on the lower rather than an upper bunk bed.[44] Behavioral intervention of scheduled awakenings is successful in ameliorating sleepwalking episodes but, compared with pharmacologic treatment modalities, has the added advantage of avoidance of CNS-acting medications.[47] The technique calls for the parents to gently wake the patient approximately 15 to 30 minutes before the child typically experiences a sleepwalking episode.[69] Once the child responds, parents are instructed to allow the child to fall back asleep. Parents are instructed to continue the scheduled awakenings for 4 weeks and to log the frequency of the episodes during this period.[69] Although robust efficacy data currently are lacking, the theory behind the scheduled awakening tecchnique suggests that altering the patient's sleep patterns decreases the disruption in slow wave sleep.[70] Alternatively, the technique may condition the patient for self-arousal just before the arousal event, thereby avoiding it altogether, or may normalize the patient's total sleep time and improve sleep efficiency.[69] In adults with sleepwalking, sleep disorder–focused

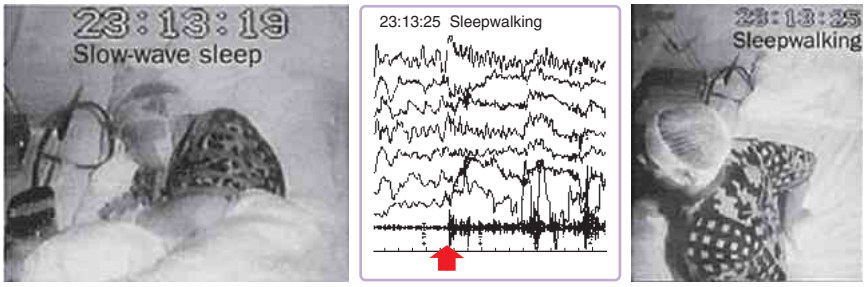


Figure 102-5 Polysomnography in Sleepwalking: Slow Wave Sleep Instability. Polysomnographic fragment highlights instability during stage N3 during sleepwalking episode (*arrow*), demarcated by sudden onset of a diffuse, high voltage rhythmic delta activity. The patient is ambulating and appears confused; the electroencephalogram (EEG) recorded concurrently shows a diffuse theta activity with intermixed alpha, beta, and delta activities. (Modified from Bassetti C, Vella S, Donati F, et al. SPECT during sleepwalking. *Lancet* 2000;356:484–5.)

psychotherapy and hypnosis constitute an alternative to pharmacotherapy.[71] Definitive pharmacotherapy with tricyclic antidepressants or benzodiazepines is necessary when the sleepwalking spells are frequent or severe or may possibly lead to injury.[3,72-84] Treatment of sleepwalking is summarized in Table 102-3.

Sleep Terrors

Essential Features and Associated Characteristics

Sleep terrors, also known as night terrors (pavor nocturnus), consist of a sudden arousal from deep sleep manifested by a piercing scream accompanied by significant autonomic arousal and behavioral manifestations of intense fear. This is the most dramatic of the arousal disorders. In addition to the characteristic sudden arousal plus the piercing scream, the patient often exhibits signs of intense fear, extreme panic, and confusion, associated with sympathetic activity manifested by tachycardia, tachypnea, reduced galvanic skin resistance reflecting diaphoresis, flushing of the skin, and mydriasis. The episodes typically are followed by amnesia and disorientation and occasionally prominent motor activity and displacement resulting in bodily injury.[5,35,85,86] Sleep terror episodes may become violent, potentially resulting in injury to the patient and bed partner, at times with forensic implications.[87-89] Sleep terrors reflect a state of cerebral hyperresponsiveness consisting of incoherent vocalizations, extreme agitation, escape behavior, and marked confusion.[3,90] The duration of the event usually is between 30 seconds to a few minutes. A universal feature is inconsolability, and attempts by the observer to interrupt the episode or soothe the patient may exacerbate, intensify, or prolong the episodes. The patient typically appears to be awake and may sometimes misperceives the nature of the environment or engages in automatic activity such as running for a door or window.[12]

Demographic Features and Epidemiology

Prevalence is approximately 1% to 6% among prepubertal children and 1% among adults. Males are more commonly affected than females, with a peak incidence between 5 and 7 years of age.[91] Episodes tend to decrease in frequency or cease during early adolescence. Psychopathology often is rare in affected children but may have a more significant role in adult sufferers.[92] As with sleepwalking, sleep terrors are much more prevalent in adults than has been generally acknowledged.[71]

Objective and Polysomnographic Features

Although sleep terrors usually are diagnosed on the basis of clinical criteria alone, video polysomnography, with multiple EEG channels included in the recording montage, may be performed for atypical episodes (i.e., repetitive episodes, occurrence several times per night, or a stereotypical behavior pattern) or in patients with possible underlying sleep disorders or neurologic or psychiatric issues.[63,93]

The onset of sleep terror episodes typically is within the first few hours of the night, during stage N3 sleep. Before a sleep terror spell, the EEG typically reveals high-voltage, symmetric, hypersynchronous slow wave activity. The characteristic polysomnogram during the sleep terror episode may reveal a sudden and incomplete arousal from slow wave sleep associated with a regular, rhythmic slow wave activity pattern, accompanied by a marked increase in muscle tone and change in respiratory and heart rates and sympathetic hyperactivation.[94-96] Although these changes may be difficult to appreciate consistently, the arousal from N3 sleep (as noted by the star in the figure), with inconsolable crying along with screaming, often is a unique feature that can be revealed with video monitoring (Figure 102-6).

In children with sleep-disordered breathing, periodic limb movements in sleep may precipitate sleepwalking or sleep terrors; improvement may be obtained after treatment of these primary sleep disorders.[16,97] When patients with arousal disorders undergo formal polysomnography, detailed examination of the nasal cannula/pressure transducer system and/or esophageal manometry is key to fully evaluating the possibility of respiratory events provoking the parasomnia.[16]

Differential Diagnosis

The differential diagnosis should include REM nightmares, nocturnal anxiety attack related to obstructive sleep apnea, nocturnal cardiac ischemia, and sleep-related epileptic seizures. Differentiation from nightmares is most important. Nightmares are distinguishable from sleep terrors in that the former often are associated with a vivid recall of the dream event during REM sleep, the clinical presentation is less dramatic, and many patients do not experience the typical autonomic hyperarousal that characterizes the latter[2,3,35] (Table 102-2). Sleep terrors preferentially occur during stage N3 slow wave sleep[95] and are universally characterized by amnesia, but fragmentary indistinct recollections of threats (spiders, monsters, snakes) from which patients have to defend themselves are sometimes reported.[16] Differentiating between sleep terrors and sleep-related epilepsy occasionally is difficult, and the use of electroencephalography is helpful in patients in whom the episodes are atypical, frequent, and less responsive to management. Nocturnal seizures, especially complex partial seizures, may have a major fear component and manifest with many of the characteristics found in sleep terrors including screaming, panic, fear, tachycardia, and vague frightening perceptions. For this reason, history and clinical semiology alone are not sufficient to conclusively differentiate sleep terrors from seizures. In these challenging cases, video polysomnographic recording with a full set of EEG electrodes is essential when the spells are frequent, repetitive, or refractory to conventional therapy or have atypical features.

Behavioral and Pharmacologic Treatment Options

Treatment often is unnecessary when episodes are rare but is essential when events are frequent, intense, or disruptive to the patient's sleep. The most essential first step is to facilitate safety measures, because these are paramount in protecting the patient from injury. As with sleepwalking, security and protection of the patient from harm in the external environment are prudent. Typical measures include (1) ensuring that the patient's sleeping area is on the ground floor, (2) avoiding the upper deck of a bunk bed, (3) removing sharp furniture close to the bed area, (4) barricading windows, (5) providing special bolts for windows and doors, and (6) using door alarms and bells to alert the family members should the child leave the room.[98] The parents and the patient need to be educated and reassured that these episodes are transient and self-limiting and typically will be outgrown with time. Additional measures should focus on maintaining a regular sleep-wake schedule, reducing or entirely eliminating caffeinated beverages. As with the other arousal disorders considered here, it is

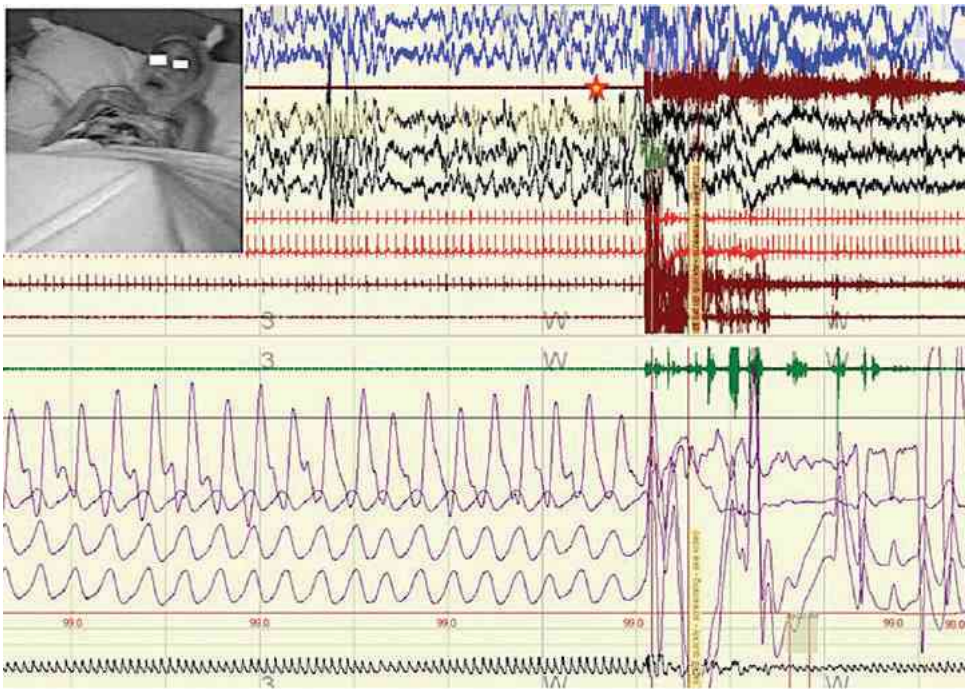


Figure 102-6 Two-minute epoch of a diagnostic polysomnogram performed to evaluate arousals associated with screaming and inconsolable crying in a 9-year-old boy. The recording was obtained during one of the patient's representative spells: an arousal out of slow wave sleep demarcated by the *star*, accompanied by screaming, with the patient's arms flexed and held close to the chest (as if afraid and protecting himself). Channels are as follows: electrooculogram (*left*: E1-M2, *right*: E2-M1), chin EMG (Chin1-chin2), EEG (*right*: frontal-F4, central-C4, occipital-O2, right mastoid-M2), two ECG, two-limb EMG (LAT, RAT), snore, nasal-oral airflow-N/O, nasal pressure signal-NPRE, respiratory effort (thoracic, abdominal), and oxygen saturation (SaO_2). ECG, Electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; LAT, left anterior tibialis; RAT, right anterior tibialis. (Modified from Avidan AY, Kaplish N. The parasomnias: epidemiology, clinical features, and diagnostic approach. *Clin Chest Med* 2010;31:353–70. Polysomnogram slide courtesy Timothy Hoban, MD, Professor of Pediatrics and Neurology, University of Michigan, Ann Arbor, Michigan.)

best not to confront, restrain, or awaken the patient during the episodes, because such interventions may prolong and worsen the behavior.[98]

Scheduled or anticipatory awakenings several minutes before the time of the regular occurrence of episodes may be of considerable benefit. The technique is especially helpful when the sleep terror episodes occur at a consistent time. Hypnosis is another practical, inexpensive, and effective treatment, especially when posthypnotic suggestions are used to help decrease awareness of the unpleasant nocturnal sensory experience.[75,99-103] Adult sufferers with a history of a psychiatric disorder may benefit from psychotherapy and stress reduction as well as reassurance.[6,104-106] Pharmacotherapy should be reserved for patients with severe, frequent, and refractory sleep terror episodes. Treatment options consist of clonazepam and tricyclic antidepressants. Low-dose benzodiazepines (clonazepam, diazepam) may be effective when administered close to bedtime[107] (Table 102-3).

Sleep-Related Eating Disorder

Essential Features and Associated Characteristics

Sleep-related eating disorder (SRED) consists of episodes of amnestic nocturnal sleepwalking associated with compulsive eating behavior, occurring with fluctuating levels of impaired consciousness and sometimes associated exposure to psychotropic agents (e.g., zolpidem, olanzapine).[73-76] These events of eating occur after a partial arousal from sleep, and the patient may consume unusual food items, such as inedible substances (e.g., bar of soap) or cat food on a sandwich, or may display atypical eating (eating mayonnaise out of a jar, biting into a frozen pizza). The main concerns with SRED are related to the safety and welfare of the patient during the preparation of food (i.e., cutting and cooking) and the potential metabolic consequences (obesity, poor glucose control) of the compulsive eating.

A number of terms may be used to describe patients that experience abnormal eating behaviors at night.[77,78] SRED should not be confused with *evening hyperphagia* or *night eating syndrome* (NES). Evening hyperphagia describes excessive and unrestrained eating after the last meal and before falling asleep, while the patient is fully awake and cognizant. NES similarly indicates consciously eating greater than 50% of the daily calories in late evening and before bed or during full awakenings during the sleep period. This syndrome is associated with mood disturbances, obesity, and substance abuse but also may have links to a circadian delay in meal timing.

Demographic Features and Epidemiology

The precise prevalence of SRED is somewhat difficult to ascertain on account of misunderstanding and confusion regarding diagnostic criteria and a low awareness of SRED and NES.[77] Nevertheless, the current data indicate a prevalence of approximately 1.5% of NES in the adult

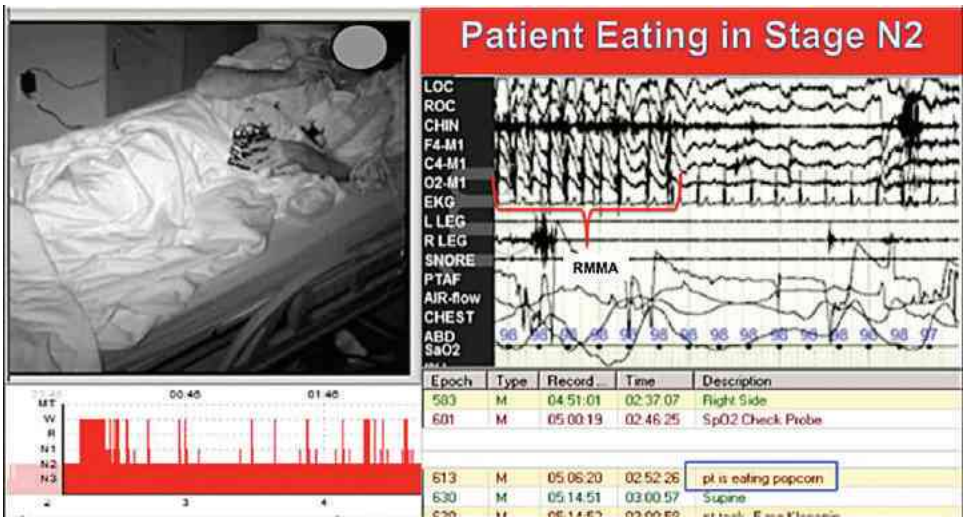


Figure 102-7 Characteristic behavior in a patient with known amnestic sleep-related eating disorder (SRED). A recording of arousal activity, obtained during NREM sleep, shows the patient eating popcorn, his favorite sleep-related eating food item, which he was instructed to bring with him to the sleep laboratory. The patient had no recollection of this activity on awakening in the morning. The repetitive chewing movements are evident as rhythmic masticatory muscle activity (RMMA) on the electromyogram tracings, as noted in the *bracketed* segment, associated with arousals during NREM sleep.

population.[80,92] The prevalence of SRED in a student population was 4.6%; an inpatient eating disorder sample reported more than triple (16.7%) and an outpatient eating disorder sample reported nearly double (8.7%) the prevalence for the healthy sample.[82]

Objective and Polysomnographic Features

Formal polysomnography may be indicated for the workup of SRED as it may be related to other comorbid sleep disorders, which predispose to arousal such as restless legs syndrome, periodic limb movements of sleep, or obstructive sleep apnea.[83,84,108-110] Patients should be invited to bring their habitually eaten nocturnal food items to the sleep laboratory for the study.[77] If the patient is observed to eat during the sleep study recording, the accompanying sleep-wake state is delineated and the sleep technologist is enlisted to assess and document the level of awareness of the event and ability to recall the eating episode in the morning.[77] Most of the SRED events occur out of NREM sleep, and patients are noted to have increased periodic limb movements.[83] Electromyographic findings may show repetitive chewing movements—rhythmic masticatory muscle activity (RMMA)—during the event (Figure 102-7). The specific pathophysiologic implication of the RMMA in SRED is unknown at present but may be related to dopaminergic mechanism.[83]

Recently, it has been shown that SRED is more common in patients with restless legs syndrome, potentially induced by hypnotics to help address the sleep-onset insomnia in this condition. This new finding broadens the clinical manifestations of restless legs syndrome to include nocturnal eating and validates that both NES and amnestic SRED may be relieved by dopaminergic therapy.[111]

Differential Diagnosis

SRED is distinct from NES, which is characterized by morning anorexia, evening hyperphagia (while awake), and insomnia and is associated with hypothalamic-pituitary axis abnormalities.[112-115] Both SRED and NES need to be differentiated from other conditions associated with nocturnal eating. In contrast with SRED, in KLS, Prader-Willi syndrome, and compulsive eating disorder, the compulsive eating occurs during waking periods.[116]

Behavioral and Pharmacologic Treatment Options

Treatment includes management of any underlying sleep disorder (such as obstructive sleep apnea). As with other disorders of arousal, safety measures are of upmost importance and include padding or removing sharp furniture, barricading windows, and using alarms to alert the household to safely guide the patient back to bed. If the patient has a history of SRED spells that involve food preparation, cabinets and drawers housing sharp utensils should be locked, and stove/oven knobs should be removed until the episodes resolve.

If the patient suffers from other disorders of arousal, recommended treatments for these conditions, such as benzodiazepines or tricyclic antidepressants, may help (Table 102-3). Dopamine agonists, selective serotonin reuptake inhibitors, and topiramate have all been reported to help improve symptoms, although lack the support from large trials is a limitation as in the other parasomnia categories.[117,118] Patients with restless legs syndrome who also have SRED often responds to therapy with a combination of a dopaminergic agonist as well as an opiate or topiramate.[108,119] A typical starting dose of topiramate is 25 mg near bedtime, and titrated to clinical efficacy at an increasing dose of 25 mg every week.[120]

CLINICAL PEARLS

Disorders of arousal may place patients at high risk for sleep-related injuries and can adversely affect quality of life. Accordingly, health care providers need to adequately screen for, correctly classify, and appropriately manage these disorders. Investigation of nocturnal events can be challenging to even

Continued

CLINICAL PEARLS—cont'd

the most ardent clinician and requires careful attention to detail. Essential to any degree of therapeutic success are identification and elimination/treatment of predisposing factors including environmental issues, provoking medications and other substances, and concomitant sleep disorders. Minimization of stimuli in the bedroom including extraneous sounds and lights may limit occurrence of events. Alcohol remains a known precipitant for some arousal disorders, but its role in provoking somnambulism is controversial. Formal evaluation and treatment of other sleep disorders are key to successful management.[121,122-126]

SUMMARY

The disorders of arousal are a unique group of sleep disorders that share similar characteristics and have a common underlying pathophysiology. Successful amelioration of disturbing arousals will depend on accurate diagnosis, focusing on key distinguishing features, and appropriate treatment strategies. Often reassurance, addressing safety issues, and education are important for the less dramatic parasomnias. Safety precautions and good general sleep hygiene measures are vital recommendations, because such disorders can be exacerbated by sleep deprivation and various other factors. When the nocturnal episodes are frequent or involve aggressive or dramatic behaviors, pharmacotherapy with a benzodiazepine, particularly clonazepam, at night is the preferred approach with most of the disorders of arousal. Other therapy relies on the use of short-acting benzodiazepines, such as diazepam and alprazolam, or tricyclic antidepressantss in the case of sleep terrors. Relaxation training and guided imagery may be helpful strategies for some patients, especially those with disorders of arousal.

Suggested Readings

Arino H, Iranzo A, Gaig C, Santamaria J. Sexsomnia: parasomnia associated with sexual behaviour during sleep. *Neurologia* 2014;**29**:146–52.

Attarian H. Treatment options for parasomnias. *Neurol Clin* 2010;**28**:1089–106.

Derry CP, Harvey AS, Walker MC, et al. NREM arousal parasomnias and their distinction from nocturnal frontal lobe epilepsy: a video EEG analysis. *Sleep* 2009;**32**:1637–44.

Fois C, Wright MAS, Sechi G, et al. The utility of polysomnography for the diagnosis of NREM parasomnias: an observational study over 4 years of clinical practice. *J Neurol* 2015;**262**(2):385–93.

Foldvary-Schaefer N, Alsheikhtaha Z. Complex nocturnal behaviors: nocturnal seizures and parasomnias. *Continuum* 2013;**19**:104–31.

Galbiati A, Rinaldi F, Giora E, et al. Behavioural and cognitive-behavioural treatments of parasomnias. *Behav Neurol* 2015;**2015**:786928.

Howell MJ, Khawaja IS, Schenck CH. Parasomnias: an update. *Psychiat Ann* 2015;**45**(1):30–4.

Howell MJ, Schenck CH, Crow SJ. A review of nighttime eating disorders. *Sleep Med Rev* 2009;**13**:23–34.

Kotagal S. Treatment of dyssomnias and parasomnias in childhood. *Curr Treat Options Neurol* 2012;**14**:630–49.

Kothare SV, Ivanenko A. *Parasomnias: clinical characteristics and treatment*. New York: Springer; 2013.

Pressman MR. Factors that predispose, prime and precipitate NREM parasomnias in adults: clinical and forensic implications. *Sleep Med Rev* 2007;**11**:5–30.

Tinuper P, Bisulli F, Provini F. The parasomnias: mechanisms and treatment. *Epilepsia* 2012;**53**(Suppl. 7):12–9.

Zadra A, Desautels A, Petit D, Montplaisir J. Somnambulism—clinical aspects and pathophysiological hypotheses. *Lancet Neurol* 2013;**12**:285–94.

A complete reference list can be found online at ExpertConsult.com.

Rapid Eye Movement Sleep Parasomnias

Michael H. Silber; Erik K. St. Louis; Bradley F. Boeve

Chapter 103

Chapter Highlights

- Rapid eye movement (REM) sleep behavior disorder (RBD) is a unique parasomnia characterized by loss of REM sleep atonia and dream enactment behavior. This chapter explores the epidemiology, clinical features, pathophysiology, diagnosis, and management of the condition.
- RBD is associated with the group of neurodegenerative disorders known as the synucleinopathies, including Parkinson disease, dementia with Lewy bodies, and multiple system atrophy. Robust evidence confirms that a majority of patients with idiopathic RBD also harbor synucleinopathy pathology. Other etiologic factors include the use of antidepressants and certain autoimmune disorders.
- RBD is diagnosed through clinical history and polysomnography. Quantitative assessments of muscle tone in REM sleep can result in greater diagnostic accuracy. Management includes addressing bedroom safety and the use of medications such as melatonin and clonazepam.
- Other REM sleep parasomnias include nightmare disorder, recurrent isolated sleep paralysis, and sleep-related painful erections. The epidemiology, clinical features, and management of the latter two conditions are discussed in this chapter.

RAPID EYE MOVEMENT SLEEP BEHAVIOR DISORDER

Overview

Rapid eye movement (REM) sleep comprises a complex combination of phasic and tonic phenomena, including desynchronized electroencephalographic activity, rapid eye movements, dreaming, and skeletal muscle atonia (Figure 103-1). In disorders such as narcolepsy, these phenomena can dissociate, resulting, for example, in intrusion of muscle paralysis during wakefulness (cataplexy and sleep paralysis). Conversely, muscle tone can be retained in REM sleep, a finding known as REM sleep without atonia (RSWA) (Figure 103-2). This allows dream enactment motor activity to occur in the condition known as REM sleep behavior disorder (RBD).

The first description of such a dissociated state was in experimental cats in which introduction of pontine lesions resulted in motor behaviors during REM sleep.[1] Retained muscle tone in REM sleep caused by clomipramine was described in human subjects in 1972.[2] Similar findings were reported in sleep during alcohol withdrawal, a state given the name "stage1-REM sleep."[3] The definitive descriptions of the disorder, as well as its name, are attributed to Carlos Schenck and Mark Mahowald, who published reports of 10 cases in 1986 and 1987, 3 associated with neurodegenerative disorders.[4,5] Extensive subsequent studies of the clinical features, pathophysiology, etiology, and management have led to a deeper understanding of the complexity and broad implications of this unique parasomnia.

Epidemiology

The prevalence of RBD is uncertain. In a Hong Kong community-based study of 1034 subjects 70 years of age or older, prevalence of RBD was estimated at 0.38%.[6] Telephone interviews of 19,961 European subjects suggested a prevalence of sleep-related violent behavior of 1.6%; the study authors estimated that less than 0.5% had RBD.[7] RBD becomes more frequent with increasing age. In population-based RBD prevalence studies, the age- and sex-adjusted estimate for polysomnography-confirmed RBD in a Korean population aged 60 years or older was 2.01%, and the idiopathic RBD (iRBD) prevalence estimate was 1.34%.[8] In Olmsted County, Minnesota, the prevalence of probable RBD symptoms reported by bed partners of persons aged 70 to 99 years was 8.9%.[9]

RBD occurs more frequently in men. Pooled data for 514 patients in six series from the United States, Europe, and China indicate a male predominance of 78%.[10-15] This difference is less marked in patients with RBD younger than 50 years of age (52% to 59% of affected men).[12,13,16] RBD commences most frequently in middle-aged to older persons; the mean age at onset of symptoms ranged between 45 and 61 years in several large series,[10-13] and the mean age at diagnosis, between 52 and 65 years.[10-13,15]

In a questionnaire series of 316 patients with iRBD, a proxy-reported family history of presumed RBD was found in 13.8% of cases, compared with 4.8% of control subjects.[17] An environmental risk factor study of 347 patients with iRBD

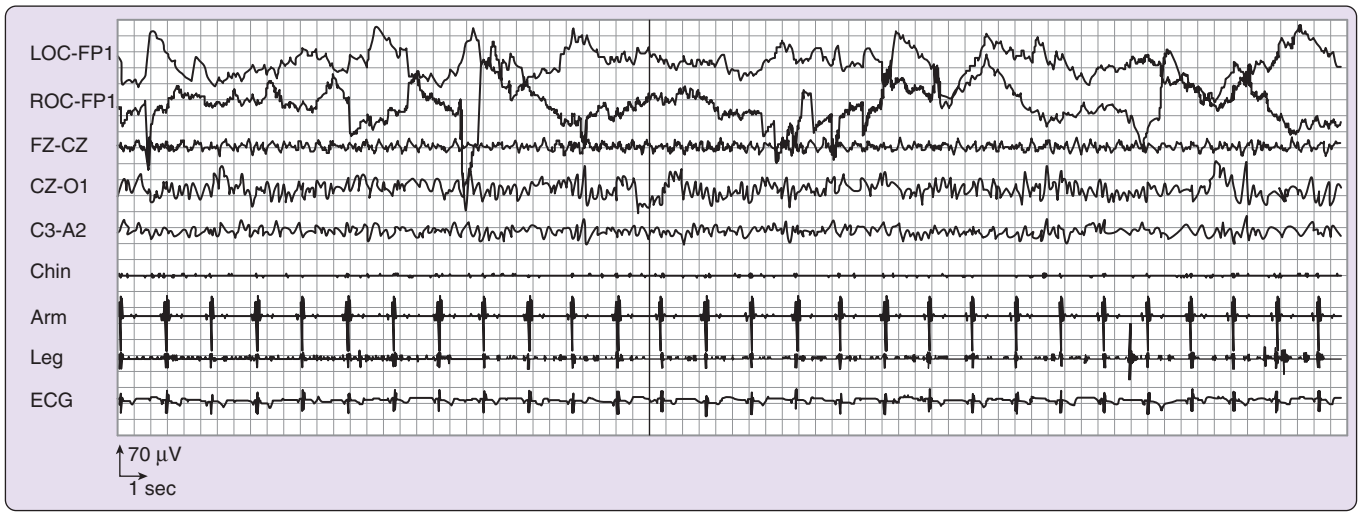


Figure 103-1 Normal REM Sleep Atonia. A 30-second polysomnogram epoch demonstrates normal REM sleep levels of atonia in submentalis, arm, and anterior tibialis leg leads on the electromyogram (EMG).

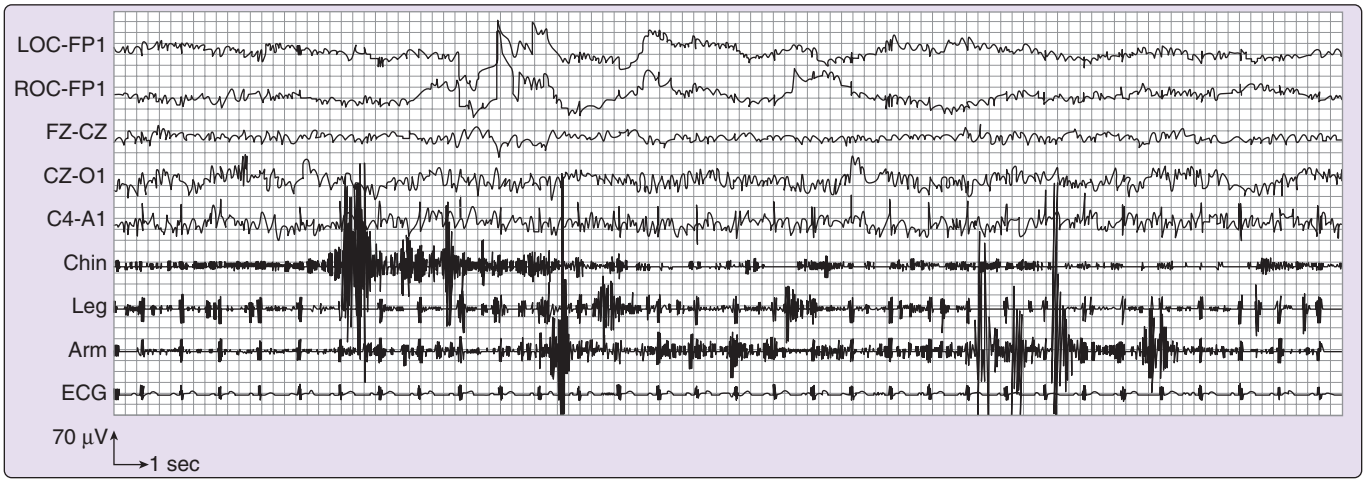


Figure 103-2 Abnormal REM Sleep without Atonia (RSWA). Increased phasic/transient muscle activity is shown in the submentalis, anterior tibialis, and arm electromyographic leads in this 30-second polysomnogram epoch. A more sustained lower-grade elevation of muscle tone lasting for longer than one half of the epoch represents abnormal tonic muscle activity, seen in the submentalis and arm channels of the electromyogram (EMG).

found cigarette smoking, previous head injuries, farming occupation, and pesticide exposure to be associated with RBD more frequently than in control subjects.[18] Although most of these factors also have been linked to Parkinson disease, smoking, by contrast, is associated with lower risk for Parkinson disease.[19]

Clinical Features

Patients with RBD exhibit abnormal motor behaviors during REM sleep, predominantly while in bed. These behaviors include talking, screaming, swearing, gesturing, arm flailing, punching, kicking, and leaping or falling off the bed.[11,20] Walking or running away from the bed occurs in 11% of patients, in addition to the more typical in-bed activities.[11] Although violent behaviors are most common, nonviolent activities such as laughing, whistling, singing, and masturbation occur at times in 18% of patients.[21] Self-injuries are reported in 32% to 76% of patients,[11,13,20,22] including lacerations, ecchymoses, limb fractures, and subdural hematomas[11,22,23] requiring medical attention in approximately 11% of the cases. Injuries occur when patients fall off the bed or strike the limbs against the wall, a headboard, or bedside furniture. Occasional patients have attempted to leap through a window.[11,24] Sixty-four percent of bed partners report being assaulted, and many have been injured.[11,22] Injurious behaviors include punching, slapping, kicking, pulling of hair, and attempted strangulation.[6,11,24,22] Periorbital hematomas and dental injuries have been reported.[5,11] Not understanding the organic nature of the disorder, patients and their partners often are mortified by the behaviors; some patients have built elaborate barriers in their beds or even slept in restraints to avoid injuring their spouse.[11,25] The frequency and severity of

behaviors are highly variable. In some patients with concomitant neurodegenerative disorders, the frequency of events appears to diminish over time, for uncertain reasons.[11,26]

Dream content changes in RBD and becomes more violent in nature in most cases. In a series of 93 cases, dreams involved defense of the sleeper against attack in 80% (65% human assailants and 35% animals), defense of relatives against attack in 7%, adventures in 9%, sporting activities in 2%, but aggression by the dreamer in only 2%.[11] In a study of 98 dreams reported by patients with RBD compared with 69 control dreamers, the RBD dreams featured a higher level of aggression and a higher frequency of animal characters.[27] Despite the aggressive nature of the dreams, daytime aggressiveness questionnaire scales have yielded findings in the normal range[28] or have even shown lower values on the physical aggressiveness subscale than for control data.[27]

Etiology and Associated Disorders

Association with Neurodegenerative Disorders

Synucleinopathies are a group of neurodegenerative disorders in which the protein alpha-synuclein accumulates abnormally to form inclusions in the cell bodies or axons of neurons or oligodendrocytes. The synucleinopathies include Lewy body disease—manifested as phenotypes of Parkinson disease, dementia with Lewy bodies, or pure autonomic failure—and multiple system atrophy. Associated RBD is very common in these disorders, and increasing evidence suggests that most cases of apparent iRBD are due to otherwise asymptomatic synucleinopathies.

Parkinson disease, characterized clinically by often asymmetric rest tremor, rigidity, bradykinesia, and postural instability responsive to levodopa, is associated on histopathologic examination with α-synuclein–containing intraneuronal Lewy bodies in substantia nigra and other structures, with resultant degeneration of the dopaminergic nigrostriatal pathway.[29] The reported frequency of RBD in Parkinson disease ranges between 15% and 65%.[30-34] Clinical features of Lewy body dementia differ from those of Alzheimer disease and may include impaired attention and visuospatial organization, fluctuating course, visual hallucinations, parkinsonism, delusions, and depression.[35] Lewy bodies are found in cortical, limbic, and substantia nigra neurons. When dementia develops later in the course of Parkinson disease, the designation Parkinson disease with dementia is used. Patients with dementia and RBD have clinical and psychometric features much more suggestive of Lewy body dementia than of Alzheimer disease.[36-38] The frequency of RBD in dementia with Lewy bodies is reported to be 68% to 80%.[39]

Multiple system atrophy is a neurodegenerative disorder with dysautonomia and variable combinations of parkinsonism (poorly responsive to levodopa) and dysfunction of the cerebellar and corticospinal systems.[40] Sleep-disordered breathing, including obstructive sleep apnea (OSA), central sleep apnea, and nocturnal stridor, is common. In contrast with Parkinson disease and dementia with Lewy bodies, alpha-synuclein is found in non–Lewy body inclusions in oligodendrocytes. RBD occurs in 60% to 90% of patients with multiple system atrophy, both the cerebellar and the parkinsonian phenotypes.[41,42]

Many cross-sectional studies have demonstrated that patients with the apparently idiopathic form of RBD have physiologic and imaging abnormalities suggestive of the prodromal phase of a synucleinopathy[43] (Table 103-1). Loss of olfactory function has been noted in 58% to 93% of patients with iRBD, significantly more than in control subjects[44-46] and similar to that in patients with Parkinson disease.[47] Similarly, abnormalities in color vision have been detected both in patients with iRBD and in those with Parkinson disease.[48] The presence of abnormalities of olfaction and color vision in patients with iRBD at baseline increases the risk of phenoconversion to parkinsonism or dementia over 5 years.[49] Autonomic symptoms[50,51] and abnormalities on tests of cardiovascular autonomic function[51-54] are more frequent in patients with iRBD than in control subjects. Subtle abnormalities in motor function, including altered gait and decreased hand dexterity, have been detected in patients with iRND.[51,55] Abnormalities in visuospatial abnormalities are evident on cognitive testing,[56] and electroencephalogram (EEG) spectral analysis has demonstrated increased theta power during wakefulness, decreased beta power during REM sleep,[57] and increased delta power in NREM sleep[56] compared with control data.

A range of advanced imaging studies have shown abnormalities in patients with iRBD, particularly in the nigrostriatal system implicated in parkinsonism (Table 103-1). Transcranial sonography has shown midbrain hyperechogenicity in 36% to 63% of patients with iRBD[45,47,57-59]—a finding also noted in patients with Parkinson disease. Single photon emission tomography (SPECT) scans with use of the tracer ^{123}I-fluoropropyl (FP)-CIT (2β-carbomethoxy-3β-[4-iodophenyl]-N-[3-fluoropropyl]-nortropane) or ^{123}I-IPT ([N]-[3-iodopropene-2-yl]-2beta-carbomethoxy-3beta-[4-chlorophenyl] tropane) have shown reduction of striatal dopamine transporters in 36% to 40% of patients with iRBD.[45,59,60] Reduced activity has been associated with progression to a

Table 103-1 Prodromal Features of Fully Expressed Synucleinopathies in Patients with Idiopathic Rapid Eye Movement Sleep Behavior Disorder (iRBD)

Physiologic Abnormalities
Reduced olfaction
Reduced color vision
Autonomic dysfunction (symptoms, cardiovascular tests, ^{123}I-MIBG myocardial scintigraphy)
Motor dysfunction
Cognitive dysfunction
EEG power abnormalities
Imaging Abnormalities
Midbrain—transcranial sonography
Striatal dopamine transporters—SPECT scans
Putaminal volume—MRI scans
Parkinson disease–related covariance pattern—PET and SPECT scans
Hyper- and hypoperfusion of various brain regions—SPECT scans
Pons and midbrain abnormalities—MRI diffusion tensor imaging
Hippocampal gray matter—voxel-based morphometry
Cerebellum and pontine tegmentum—voxel-based morphometry

EEG, Electroencephalogram; MIBG, metaiodobenzylguanidine; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single photon emission computed tomography.

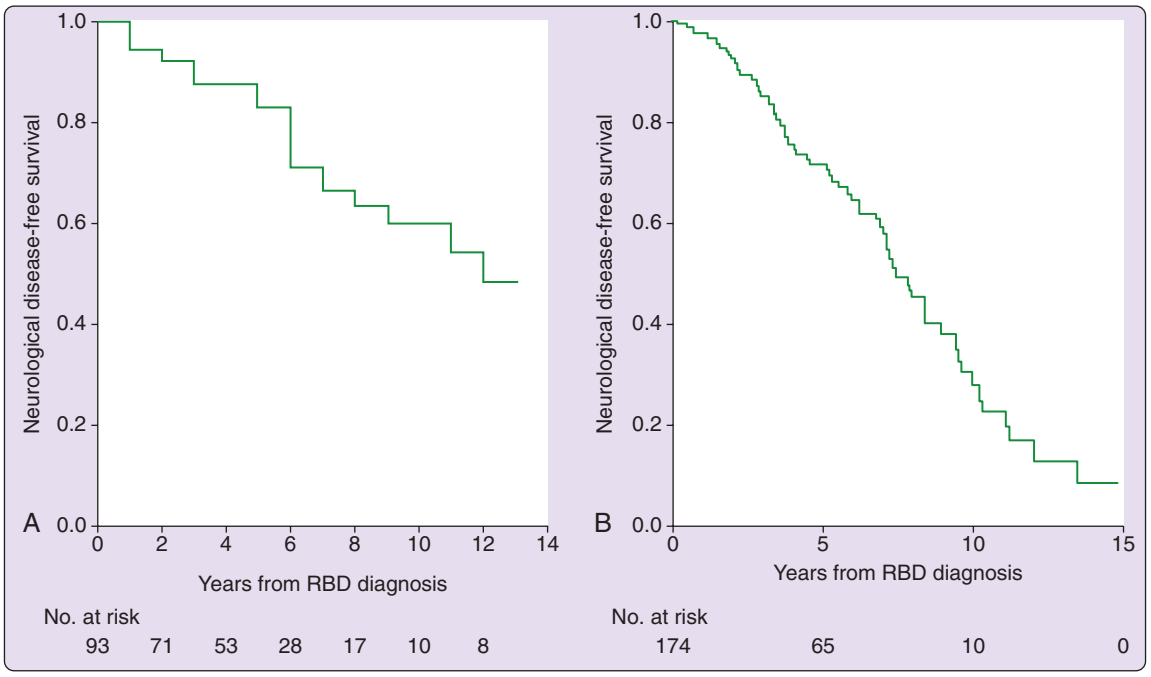


Figure 103-3 Progression from Idiopathic REM Sleep Behavior Disorder (RBD) to Fully Developed Synucleinopathies. Kaplan-Meier curves showing the rate of progression from idiopathic RBD to fully developed synucleinopathies in two studies. (**A,** Modified from Postuma RB, Gagnon JF, Vendette M, et al. Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology* 2009;72: 1296–1300, with permission.[72] **B,** Modified from Iranzo A, Fernandez-Arcos A, Tolosa E, et al. Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: study in 174 patients. *PLoS One* 2014;9:e89741, with permission.[74])

fully expressed synucleinopathy.[59,61] In volumetric measurements on 3T magnetic resonance imaging (MRI) scans, putaminal volumes were smaller in patients with iRBD than in control subjects.[62] ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) and ethyl cysteinate dimer (ECD) SPECT scans assessing a brain network known to be abnormal in Parkinson disease (i.e., Parkinson disease–related covariance pattern) showed similar increased activity in patients with iRBD, and those exhibiting such abnormalities had a greater likelihood of subsequent phenoconversion to Parkinson disease or dementia with Lewy bodies.[63] SPECT scans with ^{99mc}Tc ECD in patients with iRBD have shown variable changes in multiple brain regions compared with control subjects.[64-66] MRI diffusion tensor imaging has shown abnormalities in the pons and midbrain of patients with iRBD,[67,68] and two studies of voxel-based morphometry have shown inconsistent changes in various brain regions.[67,69]

These data conclusively indicate that many patients with iRBD have abnormalities found in fully expressed synucleinopathies and that the presence of some of these changes predicts phenoconversion to parkinsonism or dementia. The rate of such conversion has been assessed in several prospective studies. A parkinsonian disorder emerged over a mean of 3.7 years in 38% of 29 men with iRBD diagnosed after the age of 50 years.[70] When the same cohort was reassessed 16 years later, parkinsonism or dementia had developed in 80.8%, with a mean of 14.2 years from iRBD onset to phenoconversion.[71] In a study of 93 patients (80% men) with iRBD, parkinsonism or dementia developed in 28% over a mean of 11.5 years from iRBD onset.[72] A survival analysis estimated the risk of phenoconversion at 10 years from polysomnographic diagnosis of RBD to be 40.6% and at 12 years from diagnosis, 52.4% (Figure 103-3, *A*). A similar survival analysis of a cohort of 44 patients with iRBD followed for a mean of 10.5 years since diagnosis estimated a 34.8% rate of conversion at 5 years, 73.4% at 10 years, and 92.5% at 14 years.[73] In a larger series of 174 patients with iRBD from the same center followed for a mean of 4 years from diagnosis, survival analysis predicted phenoconversion rates of 33.1% at 5 years, 75.7% at 10 years, and 90.9% at 14 years (Figure 103-3, *B*).[74] Thus it appears that parkinsonism or dementia will develop in most patients with iRBD if they live long enough. A significant conversion rate has been confirmed in the only community-based study of 44 neurologically normal subjects with iRBD aged 70 to 89 years. After a median of 3.8 years, mild cognitive impairment, generally a precursor of dementia, had developed in 14, and Parkinson disease in 1. Compared with that in control subjects, the risk of phenoconversion estimated by hazard ratio was 2.2 (range, 1.3 to 3.9).[75] The latency between the onset of iRBD and phenoconversion can be extremely prolonged. A retrospective study of 27 patients diagnosed with iRBD at least 15 years before development of a fully-expressed neurodegenerative syndrome found a median interval of 25 years and a maximum interval of 50 years.[76]

RBD occasionally has been reported in other neurodegenerative disorders that are not synucleinopathies. In a series of 45 patients with clinically diagnosed progressive supranuclear palsy (PSP), a disorder associated with tau protein inclusions, 13% were reported to have polysomnography-confirmed RBD, and another 14%, RSWA.[77] By contrast, none of the bed partners of nine patients with PSP, one with the diagnosis subsequently confirmed pathologically, reported the presence of dream enactment behavior.[78] In a series of 15 patients with clinically diagnosed Alzheimer disease, 1 patient (7%) had

polysomnography-confirmed RBD, and 27%, RSWA.[79] However, the pathologic changes of Lewy body dementia and Alzheimer disease often occur concurrently in patients with suspected Alzheimer disease, and this may not be diagnosable ante mortem. RBD also has been described in association with Huntington disease[80] and with Guadeloupean parkinsonism, another tauopathy.[81]

The association of RBD with synucleinopathies has been confirmed in an international study of 172 patients with RBD, diagnosed either clinically or by polysomnography, who subsequently underwent autopsy.[82] Synucleinopathies were found in 93% (95% of 82 patients with polysomnography-confirmed RBD). These included 82% with Lewy body pathology (34% with concomitant Alzheimer-type changes) and 11% with multiple system atrophy. In the group with polysomnographic confirmation, Alzheimer disease alone was found in only one patient, and PSP in another. In one patient with iRBD who died without the development of another overt neurologic disease, pathologic changes of Lewy body disease were found at autopsy, confirming an earlier reported similar case.[83] Whether a specific neurodegenerative disorder is associated with RBD appears to depend on the propensity for the disorder to involve pontomedullary neurons, and not on the chemistry of the accumulating protein. This issue is discussed further in the subsequent Pathophysiology section.

Association with Other Neurologic Disorders

REM sleep motor dyscontrol is a common feature of narcolepsy.[xx] On the basis of data obtained in a questionnaire and telephone interview study, 35% of 55 narcoleptic patients were suspected of having RBD.[84] Polysomnography studies in narcoleptic patients have suggested frequencies of RBD from 36% to 43%.[85] RBD starts at a younger age than in patients without narcolepsy[86] and may be present at the start of the illness.[87]

RBD has been described in several rare paraneoplastic and autoimmune encephalopathies. Voltage-gated potassium channel complex antibodies are associated with limbic encephalitis and Morvan syndrome, a disorder characterized by profound insomnia, dysautonomia, and peripheral neuromuscular irritability, sometimes associated with tumors such as malignant thymoma. RBD has been described in patients with both phenotypes.[88,89] Ma1 and Ma2 autoantibody–related encephalopathies are associated with testicular and other cancers. The clinical phenotype may include narcolepsy and RBD.[90,91] A newly described autoantibody directed against IgLON5, a neuronal cell adhesion molecule, is associated with a distinctive neurologic syndrome characterized by progressive gait disturbances, bulbar symptoms, stridor, central hypoventilation, dysautonomia, and abnormal ocular movements.[92] All eight affected patients exhibited abnormal movements during REM and NREM sleep, with RBD being diagnosed in the four patients in whom REM sleep was recorded during polysomnography. In one series of patients with iRBD, systemic autoimmune disease seemed to be common in women with RBD.[93] However, a controlled study of 318 patients with iRBD did not reveal a higher frequency of autoimmune diseases in the RBD group.[94]

RBD also has been described in patients with a range of other neurologic disorders, including Machado-Joseph disease (spinocerebellar atrophy type 3),[95] adult-onset autosomal dominant leukodystrophy,[96] myotonic dystrophy type 2,[97] autism,[98] Tourette syndrome,[99] Möbius syndrome,[100,101] and Smith-Magenis syndrome.[101] Structural brainstem lesions occasionally may cause RBD, including multiple sclerosis,[102] astrocytomas,[101] acoustic neuroma,[103] vascular malformations,[104,105] central nervous system vasculitis,[106] and cerebral infarcts.[107,108]

Association with Antidepressants

RSWA[109,110] and RBD[12,111,112] have been related to antidepressant use. In comparison with patients with iRBD not taking antidepressants, RBD commenced earlier in patients in the antidepressant group, which also had a higher percentage of women.[113] In a study of early- versus late-onset RBD, the frequency of psychiatric diagnoses and antidepressant use was higher in the early-onset group.[16] A multicenter, controlled study that included 318 iRBD cases showed a higher percentage of depression and antidepressant use for the RBD group.[94] Various hypotheses have been suggested to explain this apparent association.[16,114] Antidepressants might cause RBD idiosyncratically in a minority of treated patients through a mechanism unrelated to synucleinopathies. Alternatively, antidepressants might unmask RBD in patients with otherwise subclinical synucleinopathies, resulting in an earlier presentation of RBD. Finally, antidepressant use may be a surrogate marker for depression, which is known to be an independent prodromal feature of synucleinopathies.[115] To address these possibilities, 100 patients with iRBD, 27 of whom were taking antidepressants, were followed for a mean of 4.5 years.[114] The groups with and without antidepressants showed the same frequency of abnormalities in olfactory, color vision, autonomic, and motor tests, suggesting that the antidepressant group patients were no more predisposed to development of a fully expressed synucleinopathy. However, the rate of phenoconversion to a diagnosis of a fully expressed synucleinopathy was slower in the antidepressant group, implying that RBD had emerged earlier than it would have if the patients were not taking antidepressants. By contrast, if depression were the initial manifestation of a synucleinopathy, one might have predicted phenoconversion to occur at least at the same rate as for the group not taking antidepressants, and perhaps faster.

Association with Other Drugs

An acute, transient form of RBD induced by withdrawal from barbiturates[116] and ethanol[117] has been described. Beta blockers[119] and caffeine abuse[120] also may possibly induce RBD.

Pathophysiology

The primary generator of REM sleep lies in the lateral pontine tegmentum in a small area ventral and slightly rostral to the locus coeruleus. Depending on researchers and species studied, the region has been termed the subcoeruleus nucleus, perilocus coeruleus alpha, sublateral dorsal tegmental nucleus, or nucleus reticularis pontis oralis.[118,119] Predominantly glutamatergic neurons in the vicinity of this nucleus are responsible for the atonia of REM sleep. These neurons project both directly to inhibitory interneurons in the ventral horn of the spinal cord and also to neurons in the ventromedial medulla, including the gigantocellular and magnocellular neuronal groups.[118,120] The spinal interneurons and ventromedial medullary neurons inhibit anterior horn cells in the ventral horn of the spinal cord by release of the inhibitory

neurotransmitters glycine and gamma-aminobutyric acid (GABA), resulting in skeletal muscle atonia.[118,119] Theoretically, then, RBD could arise from interruption of these pathways at pontine, medullary, or spinal cord levels.

Various animal models confirm these hypotheses, illustrated by the following examples.[121] Bilateral pontine tegmental lesions in cats cause loss of REM atonia and dream enactment behavior, depending on the exact site and extent of the lesions.[122] Loss of REM atonia and complex motor behavior in REM sleep are seen in rats with lesions in the sublateral dorsal nuclei in the pons.[123] Lesions of the medullary gigantocellular nuclei result in loss of REM atonia in cats.[124] Transgenic mice expressing reduced glycine and GABA inhibitory activity exhibit motor behaviors during REM sleep with loss of REM atonia.[125] In humans, rare cases of focal lesions in the pontine tegmentum have been reported to cause RBD.[102,104,106,126,127] In Parkinson disease, Lewy bodies are found in neurons in the subcoeruleus nuclei in the pons and magnocellular reticular formation in the medulla.[128] In an MRI study of 24 patients with Parkinson disease, 12 with RBD and 12 without, reduced signal intensity was seen in the locus coeruleus–subcoeruleus complex, more marked in the subgroup with RBD.[129] The degree of signal intensity reduction correlated with the percentage of abnormal muscle tone in REM sleep.

This model fits with the Braak staging scheme for the pathologic features of Parkinson disease.[130] In this scheme, pathologic changes of Lewy body disease develop first in the dorsal nucleus of the vagus in the medulla and the olfactory bulb—stage 1—and subsequently ascend to affect pontomedullary structures, including the magnocellular and sublateral dorsal nuclei—stages 2 and 3. Only in stage 4 is substantia nigra involved, resulting in overt parkinsonism, whereas mild cognitive impairment and dementia develop with cortical involvement in stages 5 and 6. This anatomic progression can explain the phenomenon of iRBD being in many cases a prodrome for the later development of Parkinson disease. Experimental evidence suggests that this relatively stereotypical progression of pathology may be explained by a prionlike mechanism whereby alpha-synuclein aggregates can be induced transneuronally and serve as templates for further misfolding of the protein and subsequent neurodegeneration.[131] Clearly the Braak staging scheme does not fit all cases of Lewy body disease. In some patients with Parkinson disease, RBD develops simultaneously with the onset of parkinsonism or thereafter,[11] whereas in Lewy body dementia, cognitive impairment precedes definite parkinsonism.[128] Further animal and human studies are needed to define more precisely the nature of progression in RBD-associated neurodegeneration.

Diagnosis

Diagnostic criteria for RBD include the presence of RSWA (Figure 103-2) on polysomnography, sleep-related injurious or potentially injurious disruptive behaviors by history, and/or abnormal REM sleep behaviors during polysomnography, absence of epileptiform activity during REM sleep (unless RBD can be clearly distinguished from any concurrent REM sleep–related seizure disorder), and no better alternative explanation for the sleep disturbance.[132] In view of the resource-intensive nature of confirmatory polysomnography, however, a designation of "probable RBD" can be applied to patients presenting with a clear-cut clinical history of dream enactment behaviors but lacking polysomnographic evidence for RSWA, either owing to unavailability of the test or on account of the failure of polysommnography to record REM sleep. Also, because dream enactment is only rarely captured during in-laboratory polysomnography, RBD diagnosis usually instead relies on the history of dream enactment behaviors, together with recording of RSWA.[132,133]

For use in epidemiologic research studies or clinical practices without readily accessible polysomnography, well-validated questionnaire screening instruments for probable RBD are available.[134-141] The REM Sleep Behavior Disorder Single-Question Screen (RBD1Q) poses a single "yes/no" question about dream enactment, providing 93.8% sensitivity and 87.2% specificity for RBD diagnosis in Parkinson disease in a large, multicenter validation study.[137] The Mayo Sleep Questionnaire (MSQ), a screening questionnaire that can be administered to either the patient or the bed partner, is another well-validated tool, with an Informant (Bedpartner) Version being particularly useful in older patients.[141,142] The MSQ-Informant (Bedpartner) Version is 100% sensitive and 95% specific[141] and the MSQ-Patient Version is 100% sensitive and 73% specific for RBD diagnosis.[143] The RBD-HK poses 13 questions about the frequency and severity of dream enactment behavior, with a cutoff score of 18 of 100 possible points yielding positive and negative predictive values greate than 80%.[135] The REM Sleep Behavior Disorder Screening Questionnaire (RBDSQ) is a 10-item questionnaire self-rated by the patient with scores ranging from 0 to 13 points, yielding average scores of 9.5 in patients with RBD and 4.6 in control subjects without RBD.[135] Despite the usefulness of these measures for screening purposes, excluding mimickers of RBD (e.g., OSA with atypical arousals during REM sleep, NREM sleep parasomnias, and nocturnal epilepsy) is difficult and usually requires confirmation of RSWA by polysomnography.[14,28,132,144-147]

Polysomnographic RSWA manifests in submentalis, anterior tibialis, and arm leads of the electromyogram (EMG), especially flexor digitorum superficialis (FDS) and biceps brachii.[132,148-150] An expanded EMG montage, including the FDS, biceps, and abductor pollicis brevis muscles, has been shown to be most sensitive for identifying RSWA.[149,150] The prevailing "gold standard" for RSWA determination is the use of visual scoring methods as first proposed in 1992,[151] with subsequent modifications.[132,149,150,152,153] RSWA is classified as either excessive phasic (transient) or tonic muscle activity (Figure 103-2). Phasic RSWA is scored when a 3-second "mini-epoch" (with each 30-second epoch containing 10 mini-epochs) contains phasic (transient) muscle activity bursts exceeding the REM background EMG amplitude by a factor of at least four, and lasting at least 0.1 second.[132,148,151,152] The American Academy of Sleep Medicine has defined abnormal RSWA as the presence of five or more 3-second mini-epochs containing excessive measured phasic (transient) muscle activity within a single 30-second REM epoch.[132] Tonic muscle activity is scored when the REM EMG background exceeds twice the background voltage amplitude and lasts longer than 15 seconds within a 30-second epoch.[132,148,152] The designation "any" referring to muscle activity percentage, which includes either phasic, tonic, or both activities present within a 3-second mini-epoch, is a more inclusive and more easily applied metric for clinical use.[152,153] Percentage muscle activity for phasic, tonic, and "any" muscle activity is then determined by dividing the total number of 3-second mini-epochs (for phasic and

"any") EMG activity and the number of 30-second tonic epochs by total REM time.[148,152,153] Many studies have shown clear elevations of phasic and tonic muscle activity in patients with RBD over that in control subjects without RBD, with diagnostic cutoffs determined for both individual and combined muscles.[148,152-155]

Various phasic, tonic, and "any" density cutoff values have been established for RBD diagnosis with 100% specificity. The SINBAR ("Sleep Innsbruck Barcelona") group of investigators has proposed a 31.9% "any" muscle activity cutoff for RBD diagnosis using combined submentalis and bilateral FDS EMG derivations.[152] A 43.4% "any" muscle activity diagnostic cutoff has been suggested for the commonly used combined submentalis and anterior tibialis EMG derivations from either split- or full-night polysomnograms in patients with RBD plus concomitant OSA.[153] Measuring the phasic muscle burst duration was found to further help distinguish patients with RBD from control subjects.[153]

Abnormal levels of RSWA also may be found incidentally during polysomnogram recordings in patients without dream enactment behavior, especially in patients taking antidepressants.[156,157] The significance of incidental RSWA for future development of dream enactment or neurodegeneration is unknown although anecdotal cases of progression to synucleinopathy have been reported.[136] RSWA without clinical dream enactment occurs frequently in Parkinson disease and other synucleinopathies, suggesting that additional lesions in other structures may be required to mediate full RBD.[31]

Considerable progress has been made in the development of computerized automated methods for RSWA analysis, including the REM Atonia Index,[154,155] an enveloping technique for mentalis short- and long-duration muscle activity measurement,[158] and an automated SINBAR method analyzing combined submentalis and FDS muscle activity.[159] Diagnostic yield comparable to that for visual RSWA scoring methods has been demonstrated for the REM Atonia Index[153,154,160] and automated SINBAR methods.[159] A disadvantage of the visual methods is their labor-intensiveness, but validation of automated RSWA methods for limb, as opposed to mentalis, muscles is limited.[159]

REM Sleep Behavior Disorder Variants

RBD has been described in association with three variant syndromes in which dream enactment and RSWA are associated with other parasomnias or disturbances of sleep architecture. Overlap parasomnia disorder is characterized by overlapping clinical features of both NREM disorder of arousal parasomnias (such as sleepwalking, night terrors, confusional arousals, or related events including sleep eating or sexsomnia behaviors) and RBD.[161,162] A recent review of 144 cases suggested that overlap parasomnia disorder is a distinct clinical entity. In 69% of patients with this disorder, the RBD plus parasomnia was idiopathic, with an earlier age at onset than for patients with typical RBD alone. Patients with overlap parasomnia exhibited more prominent NREM than REM parasomnia characteristics, with RBD either being mild or discovered incidentally.[163] The association of overlap parasomnia with synucleinopathies also appears to be less clear than for typical RBD, although overlap parasomnia has been described in Parkinson disease. Further research is necessary to clarify whether overlap parasomnia has a natural history different from that for typical RBD.[164]

Status dissociatus is a rare condition involving completely eroded wake and sleep state boundaries, with resultant disturbances in vigilance, cognitive functioning, and sleep motor control. Conventional sleep stage scoring during polysomnography is almost impossible owing to the characteristic admixture of neurophysiologic features of wake, NREM, and REM sleep states.[165,166] Agrypnia excitata is another rare, closely related condition with the distinctive features of inability to fall asleep coupled with excessive motor and autonomic hyperactivity. Polysomnography in agrypnia excitata shows absence of N3 sleep, spindles, and K complexes, with persisting N1 architecture and brief periods of REM sleep with loss of normal atonia.[167] Oneiric behaviors (complex motor acts resembling simple daytime activities and gestures, such as chewing, hair combing, or pointing) also are characteristic of agrypnia excitata.[168] Status dissociatus may be seen in certain pontomesencephalic brain lesions and late-stage neurodegenerative disorders, including multiple system atrophy and other synucleinopathies. Agrypnia excitata is thought to arise from thalamic-limbic dysfunction associated with delirium tremens, fatal familial insomnia, and certain autoimmune encephalopathies including voltage-gated potassium channel antibody (Morvan) syndrome.[169] Clonazepam has been reported to be effective in the treatment of status dissociatus[165]; treatment approaches for agrypnia excitata beyond supportive care have not been described.

Differential Diagnosis

Mimickers of RBD include nightmares, NREM parasomnias, nocturnal epilepsy, OSA with "pseudo-RBD"–type confusional arousals from REM sleep, and psychiatric disorders such as posttraumatic stress disorder (PTSD) or nocturnal panic disorder.[144,170,171] Patients with nightmare disorder may vocalize or move minimally during their dreams but dream enactment involving complex motor behaviors paralleling dream content usually is lacking. These patients do not exhibit RSWA during polysomnography. Nocturnal epilepsy, particularly frontal lobe or extratemporal seizures, may very closely mimic RBD in some cases, although nocturnal epilepsy usually is distinguished by relative stereotypy across events and lack of paralleling dream mentation.[170] Video EEG polysomnography, including a full EEG montage, is necessary in some cases to rule out nocturnal epilepsy with confidence. OSA may manifest with dream enactment behavior that subsides with successful nasal continuous positive airway pressure therapy, and in patients with OSA alone, without comorbid RBD, RSWA is not observed during polysomnography.[171] Psychiatric disorders such as PTSD and panic disorder are sometimes difficult to distinguish from RBD, although PTSD and panic disorder are not associated with RSWA. In keeping with the frequent association of RBD with comorbid depression, some patients may have both conditions. Confident RBD diagnosis requires a thorough clinical history and examination, as well as confirmation of RSWA by the "gold standard" of polysomnography.[14,132,144]

Management

The goals of RBD treatment are to reduce dream enactment behavior frequency and severity and to prevent injury. Bedroom safety principles should be advised for all patients, including removal or padding of bedside furniture with sharp corners, minimizing fall-related injury potential by

Table 103-2 Principal Treatments for Rapid Eye Movement Sleep Behavior Disorder (RBD)

Drug	Possible Mechanism of Action	Dose	Adverse Effect(s)
Clonazepam	$GABA_A$ receptor agonist	0.25–2 mg	Sedation, dizziness, sexual dysfunction, worsened sleep-disordered breathing
Melatonin	Unknown	3–15 mg	Sedation

$GABA_A$, Gamma-aminobutyric acid type A (receptor).

adding bed rails, lowering the mattress to floor level or placing cushions near the bed, erecting pillow barriers between the patient and the bed partner, and removing firearms from the bedroom.[136,144] A bed alarm system also may be useful to alert the patient or caregiver if the patient leaves the bed.[172]

Melatonin and clonazepam are the two mainstays of RBD pharmacologic treatment[11,173-176] and appear to be similarly effective in reducing dream enactment behaviors. However, although clonazepam and melatonin both effectively reduce RBD symptoms, these drugs only rarely actually stop the behaviors.[174] Consequently, the development of other novel agents to treat RBD is needed, including prospective comparative treatment trials.[177]

Clonazepam has been the traditional drug of choice for RBD, with a reported median effective dose between 0.25 and 2.0 mg at bedtime. Approximately 90% of subjects in open label studies reported a partial or complete response to the drug, but no controlled trials have been performed.[11,174-176,178] However, clonazepam has several adverse effects that may limit its application in most elderly patients, including worsened OSA, cognitive dysfunction, and dose-related adverse effects of sleepiness, dizziness, unsteadiness, and sexual dysfunction.[144,174] Both melatonin and clonazepam decrease polysomnographic motor activity during REM sleep.[151,175,176,179-182]

Several retrospective and small prospective studies support the use of melatonin treatment for RBD.[173,174,178,180,181,183-185] Melatonin may have a slightly more robust effect in reducing injuries, with fewer adverse effects,[174] although a prospective comparison with clonazepam has not been performed.[177] Melatonin may be more effective and tolerable in elderly patients, especially those with cognitive impairment or parkinsonism.[174] Adverse effects of melatonin include daytime sleepiness and dizziness, with headache and hallucinations occasionally reported.[173,174] Use of melatonin doses between 3 and 15 mg at bedtime (with occasional patients using doses up to 25 mg) has been reportedly effective, with a median dose of 6 mg[174] (Table 103-2).

Other drugs also have been reported to reduce RBD symptoms, including pramipexole,[186,187] zopiclone,[179] zonisamide,[188] donepezil,[189,190] ramelteon,[191] agomelatine,[192] memantine,[193] cannabidiol,[194] and the herbal supplement Yi-Gan San.[195]

OTHER RAPID EYE MOVEMENT SLEEP PARASOMNIAS

Nightmare Disorder

Nightmare disorder is discussed elsewhere in this book (see Chapters 53 and 104).

Recurrent Isolated Sleep Paralysis

Sleep paralysis has long been recognized in different cultures, often with postulation of a supernatural explanation.[196] It consists of an inability to move or speak at sleep onset, or on awakening from sleep, with at least partial preservation of consciousness. The trunk and all four limbs are affected.[197,198] The paralysis resolves after sensory stimulation, such as with someone touching the patient. At least initially, events usually are associated with severe anxiety.[197] Hypnagogic or hypnopompic hallucinations are experienced by 21% to 24% of patients with sleep paralysis.[199] Events occur more commonly on awakening than at sleep onset.[199] A large population-based European study suggested that sleep starts, sleep-related cramps, and sleeptalking were more common in patients with hypnagogic hallucinations than in control subjects.[199] Polysomnographic studies have suggested that sleep paralysis may occur during dissociated REM sleep, with persistence of consciousness and alpha activity intruding into the otherwise desynchronized REM sleep EEG.[196]

Sleep paralysis is common in narcolepsy but occurs frequently in isolation in the general population. Reported prevalence of sleep paralysis varies widely between studies, with higher rates reported in samples of students, patients with psychiatric illnesses, and patients with panic disorder.[200] The prevalence within a large population-based study of 8085 subjects was 6.2%.[199] Predictive variables for sleep paralysis included bipolar disorder and the use of anxiolytic medications. Sleep paralysis can start at any age, but onset may be most common during young adulthood and middle age.[199] A population study from Hong Kong suggested a bimodal age at onset, with peaks in adolescence and after age 60 years.[201] Familial sleep paralysis has been described, [197] with recognition of three- to four-generation families and a possible maternal inheritance pattern.[198]

Sleep paralysis should be considered a disorder only if it is recurrent and results in clinically significant distress including bedtime anxiety or difficulty initiating sleep.[132] Explanation and reassurance usually are sufficient to address the patient's concerns. Antidepressants occasionally have been tried, presumably to suppress REM sleep phenomena, but no adequately reported studies of pharmacologic interventions are available.

Sleep-Related Painful Erections

Sleep-related painful erections is a rare disorder of uncertain etiology. The mean age at diagnosis is 39.8 years, with symptoms present for a mean of 5.4 years.[202] Painful erections occur during REM sleep, usually nightly and sometimes several times a night.[203] Erections are painless during intercourse or

masturbation.[204] Structural abnormalities of the penis are not present. Polysomnography shows reduced sleep efficiency, increased wake time after sleep onset, and decreased percentage REM sleep compared with control data.[203] One study showed reduced vagal activity during sleep and a trend toward heart rate acceleration during spontaneous body movements.[203] Various medications have been reported to be effective for weeks to months in usually single-patient reports, including propranolol,[202] clozapine,[205] clonazepam,[206] baclofen,[206] and various antidepressants (amitriptyline,[206] paroxetine,[202] and venlafaxine[206]), but no controlled or systematic long-term study of management has been performed.

CLINICAL PEARLS

- RBD occurs predominantly in men (78% of patients), but this difference becomes less marked when the disorder commences before the age of 50 years.
- RBD can result in serious injuries to both patients and their bed partners, including fractures, subdural hematomas, ecchymoses, lacerations, and dental injuries.
- RBD is strongly associated with fully expressed synucleinopathies (Parkinson disease, dementia with Lewy bodies, and multiple system atrophy), and in most patients with idiopathic RBD, a fully expressed synucleinopathy will develop within 15 years of RBD diagnosis.
- Although both clonazepam and melatonin are effective in treating RBD, clonazepam has many side effects, especially in elderly and neurologically disabled persons, and melatonin should be the first drug tried in these patients.
- Recurrent isolated sleep paralysis may occur in 6% of the population but should be regarded as a disorder only if it results in significant distress including bedtime anxiety or difficulty initiating sleep.

SUMMARY

Dissociation of the phenomena characterizing the REM sleep state results in a number of disorders, with broad implications for elucidation of sleep and neurologic disease. Clinical and basic science research over the past 50 years has resulted in an understanding of RBD that exceeds that of any other parasomnia. It is an example of a sleep disorder in which the anatomic, physiologic, and pathologic substrate correlates closely with the clinical and polysomnographic manifestations. Increasingly sophisticated diagnostic tools are becoming available, and effective forms of management have emerged. RBD serves as a biomarker for the development of fully expressed synucleinopathies, and an important goal of future research will be finding therapeutic agents that may delay or prevent phenoconversion to overt neurodegenerative disorders. The contribution of antidepressants to the pathogenesis of RBD remains unsettled, and further studies are needed to determine definitively whether these drugs are primary etiologic agents or merely unmask an evolving synucleinopathy. Other REM parasomnias range from the very common (nightmares and sleep paralysis) to the rare (sleep-related painful erections). Further investigations of their pathophysiology and optimal management are needed.

Selected Readings

Boeve BF, Silber MH, Saper CB, et al. Pathophysiology of REM sleep behavior disorder and relevance to neurodegenerative disease. *Brain* 2007;**130**:2770–80.

Boeve B, Silber MH, Ferman TJ, et al. Clinicopathologic correlations in 172 cases of rapid eye movement sleep behavior disorder with or without a coexisting neurologic disorder. *Sleep Med* 2013;**14**:754–62.

Frauscher B, Iranzo A, Hogl B, et al. Quantification of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder. *Sleep* 2008;**31**:724–31.

Iranzo A, Fernandez-Arcos A, Tolosa E, et al. Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: study in 174 patients. *PLoS ONE* 2014;**9**:e89741.

McCarter SJ, Boswell CL, St Louis EK, et al. Treatment outcomes in REM sleep behavior disorder. *Sleep Med* 2013;**14**:237–42.

McCarter SJ, St Louis EK, Duwell E, et al. Diagnostic thresholds for quantitative REM sleep muscle densities, phasic burst duration, and REM atonia index in REM sleep behavior disorder with and without co-morbid obstructive sleep apnea. *Sleep* 2014;**37**:1–14.

Olson EJ, Boeve BF, Silber MH. Rapid eye movement sleep behavior disorder: demographic, clinical and laboratory findings in 93 cases. *Brain* 2000;**123**:331–9.

Postuma RB, Gagnon JF, Bertrand JA, et al. Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. *Neurology* 2015;**84**:1104–13.

Postuma RB, Gagnon J, Montplaisir J. Rapid eye movement sleep behavior disorder as a biomarker for neurodegeneration: the past 10 years. *Sleep Med* 2013;**14**:763–7.

Postuma RB, Gagnon JF, Tuineaig M. Antidepressants and REM sleep behavior disorder: isolated side effect or neurodegenerative signal? *Sleep* 2013;**36**:1579–85.

Postuma RB, Adler CH, Dugger BN, et al. REM sleep behavior disorder and neuropathology in Parkinson's disease. *Mov Disord* 2015;**30**(10):1413–17.

Schenck CH, Mahowald MW. REM sleep behavior disorder: clinical, developmental, and neuroscience perspectives16 years after its formal identification in SLEEP. *Sleep* 2002;**25**:120–38.

A complete reference list can be found online at ExpertConsult.com.

Chapter
104

Nightmares and Dream Disturbances

Isabelle Arnulf

Chapter Highlights

- Nightmares and nightmarish dreamlike content are not limited to rapid eye movement (REM) sleep.
- Most adults with arousal disorders occasionally remember dreamlike and nightmarish mental content associated and congruent with the motor episodes of sleepwalking or sleep terrors.
- Patients with REM sleep behavior disorders report enacted dreams containing more elements of aggression and animals, with behavior isomorphic to the dream content.
- In arousal disorders, patients often run out of the bed to avoid an imminent, life-threatening danger, whereas patients with REM sleep behavior disorders counterattack their dreamed attacker while staying recumbent in the bed.
- Patients with narcolepsy have long and intense dreams during sleep-onset REM periods (with pleasurable flying expeditions), with a high number of lucid dreams and nightmares. The multimodal hypnagogic hallucinations in narcolepsy suggest awake dreaming.

A 62-year-old patient experienced recurrent nightmares in which he was obliged to swallow several big, yellow snakes every night. As a consequence, he was scared of going to bed and delayed this moment every night. The patient interview revealed that he did not use any drug capable of producing these dreams and did not have posttraumatic stress disorder (PTSD), previous experience with snakes, sleep terrors, rapid eye movement (REM) sleep behavior disorder (RBD), or general nightmare disorder. However, with information provided by video polysomnography (PSG) and Multiple Sleep Latency Tests, the patient was discovered to suffer from hypnagogic hallucinations linked with a long-lasting form of narcolepsy. This case illustrates that patients complaining of nightmares or disturbed dreaming may not use the word "nightmare" as a sleep specialist would (i.e., restricting this word to REM sleep-associated nightmares). Patients use this word in various situations, including the classic REM sleep nightmare disorders, poststress disorders, distressing dreams linked with psychiatric conditions, and disturbed dreaming associated with drugs and substances, neurologic diseases or conditions (including hypnagogic hallucinations), sleep terrors, RBD, status dissociatus, and epic dreaming. The sensorial characteristics of their mental experience; its construction, timing, duration, and context; their associated behaviors, autonomic signs, ability to recognize the unreal nature of the experience, and associated disorders; and aspects and timing of PSG help to determine a diagnosis (Table 104-1). In terms of mechanisms, recent insights have shown that most of these disturbed mental experiences associated with neurologic diseases result from a dissociated sleep-wake state.

NIGHTMARE DISORDER

Diagnosis

Nightmare disorder is characterized by recurrent episodes of intense dysphoric dreaming (involving feelings of threat, anxiety, fear or terror, anger, rage, embarrassment, and disgust) that arise primarily during REM sleep and that often result in awakening. The episodes tend to occur during the second half of the major sleep episode when the REM pressure is most pronounced. Nightmare content most often focuses on imminent physical danger to the individual but may also involve other distressing themes. Most patients are able to detail the nightmare's contents on awakening. Nightmares are distinct from anxiety dreams, which are frightening dream experiences remembered only after waking in the morning. Nightmare disorder can lead to sleep avoidance and deprivation and thereby to more intense nightmares, which can produce insomnia. The disorder is distinguished from parasomnia-associated nightmares by the absence of enacted dreaming.

Epidemiology

Occasional nightmares are frequent in children (60% to 75%, beginning as young as 2.5 years old and peaking between 6 and 10 years old). Nightmares are repeated more than once a week in only 2% to 8% of children, but they often persist in adulthood. Girls and boys are equally affected until late adolescence, when girls are more affected than boys. About 4% of adults suffer from a nightmare disorder. The nightmare frequency is higher in psychiatric disorders, including PTSD, substance abuse, stress and anxiety, borderline personality, and other psychiatric illnesses, including schizophrenia spectrum disorders. Nightmares particularly occur immediately after a traumatic stress. Nightmares beginning within 3 months of a trauma are present in up to 80% of patients with PTSD. Approximately 50% of PTSD cases resolve within 3 months, but posttraumatic nightmares may persist throughout life. Recurrent nightmares may also be a deleterious effect of various drugs,[1] including antidepressants, antihypertensives (beta blockers, α-adrenergic receptor agonists, enalapril, losartan, verapamil), dopamine receptor agonists, cholinesterase blockers (donepezil, rivastigmine, tacrine), varenicline (a

Table 104-1 Causes and Semiology of the Complaints of "Nightmares" and Disturbed Dreaming in Neurologic Diseases

	Aspects	Associated Disorders
Vivid dreaming	Dreams of flight, nightmares Fights with animals, aggression	Narcolepsy Parkinson disease Dementia with Lewy bodies
	Associated with autonomic dysfunction and status dissociatus	Guillain-Barré syndrome
Hypnagogic hallucinations	Multimodal, frightening Often associated with sleep paralysis and REM sleep behavior disorder	Narcolepsy
	Presence, passage, visual hallucinations of humans and animals Visual, spatial tilt, auditory Associated with autonomic dysfunction and status dissociatus	Parkinson disease Dementia with Lewy bodies Guillain-Barré syndrome
NREM parasomnia	Associated with screams, tachycardia, sudden arousal, escaping from the bed; mostly misfortunes (buried alive, collapsing ceiling, life-threatening danger)	Sleep terrors Confusional arousals
REM parasomnia	Enacted dreams with aggression by humans or animals, kicking, boxing, shouting, swearing	REM sleep behavior disorder (idiopathic, parkinsonism, dementia, narcolepsy)
Status dissociatus	Continuously enacted dreams with visual hallucinations (patients may seem awake)	Guillain-Barré syndrome Delirium tremens Fatal familial insomnia Morvan chorea Neurodegenerative diseases

nicotinic acetylcholine blocker), and ganciclovir. Also, the withdrawal of REM sleep–suppressive agents (antidepressants, benzodiazepine, barbiturate, ethanol) or the end-of-night REM sleep rebound after using short-acting hypnotics such as zolpidem can promote nightmares.

Criteria

Patients with nightmare disorders should meet the following criteria: (1) repeated occurrences of extended, extremely dysphoric, and well-remembered dreams that usually involve threats to survival, security, or physical integrity; (2) rapidly becoming oriented and alert on awakening from the dysphoric dreams; and (3) clinically significant distress or impairment in social, occupational, or other important areas of functioning that is caused by the dream experience or the sleep disturbance produced by awakening from it.[2] The impairment may be demonstrated in a variety of ways, including mood disturbance, resistance or anxiety about sleep, cognitive impairments, behavioral problems, negative impact on family functioning, daytime sleepiness or fatigue, and impaired occupational, education, and social function.

Evaluation

Overnight PSG is not routinely used to assess nightmare disorder but may be appropriately performed to exclude other parasomnias or sleep-disordered breathing. Sleep recordings during actual nightmares occasionally show abrupt awakenings from REM sleep preceded by accelerated heart and respiratory rates. Of note, posttraumatic nightmares emerge both from REM (sometimes just after less than 1 minute of REM sleep) and non–rapid eye movement (NREM) sleep (including sleep onset). Highly disturbing dream content frequently contrasts strikingly with relatively minor autonomic changes (e.g., no visible tears) and no acting out of the nightmare (no shouting, no attempt to suddenly escape the bed), which constitute the main differences with sleep terrors (which happen only on emerging from N3 sleep). RBD appears in middle age and is characterized by enacted dreams of defense against aggressions (often with injuries). Hallucinations and sleep paralysis may be described as "nightmares," but they specifically occur at sleep onset and offset, and the paralysis affects the whole body and induces a dyspnea. Nocturnal panic attacks are not associated with a detailed mental imagery. Severe sleep apnea may be associated with disagreeable sleep-associated perceptions or images that resolve with the treatment of apnea.

Treatment

The best-established treatment of idiopathic nightmare disorder is image rehearsal therapy, whereas systematic desensitization and progressive deep muscle relaxation training are suggested.[3] There is lower grade evidence for using lucid dreaming therapy and self-exposure therapy. Among the treatment of PTSD-associated nightmares, prazosin carries a level A recommendation, whereas clonidine benefit is less clear. Treatments with lower grade evidence include several drugs (trazodone, atypical antipsychotic medications, topiramate, low-dose cortisol, fluvoxamine, triazolam and nitrazepam, phenelzine, gabapentin, cyproheptadine, and tricyclic antidepressants) and behavioral therapies (exposure, relaxation, rescripting therapy, sleep dynamic therapy, hypnosis, eye-movement desensitization and reprocessing, and the testimony method).

DREAMLIKE AND NIGHTMARISH MENTATIONS DURING NREM PARASOMNIAS

General Context

Sleepwalking and sleep terrors consist of a series of abnormal mental experiences and complex behaviors associated with

sudden adrenergic discharges that occur during partial awakening from slow wave sleep—hence their name *arousal disorders*.[2] Most motor episodes begin with raising the head, opening the eyes, and looking about in a confused manner, sometimes with verbal utterances, a condition referred to as *confusional arousals*.[4] Following this common pattern of behavior, sleepwalkers sit up in bed, stand up, and walk, usually in a quiet but confused manner, whereas patients with sleep terrors scream or display the behavioral and autonomic signs of intense fear, escape the bed as if to avoid imminent danger, or defenestrate. Sleepwalking and sleep terrors show considerable overlaps, including difficult arousal, mental confusion when awakened from an episode, complete or partial amnesia for the episode, and dangerous behaviors during the episode. Both disorders arise from slow wave sleep, frequently co-occur in the same family or patient, and have a strong genetic background.[5] An inventory of behaviors assessing their severity has been recently developed for sleepwalking and sleep terrors with good psychometric properties.[6]

Sleepwalking– and Sleep Terror–Associated Mentations

Contrary to common belief, most adults with sleepwalking or sleep terrors occasionally remember the mental content (which they frequently identify as a dream or a nightmare) associated with their abnormal motor behavior.[7] Some isolated case reports first mentioned this association. A young patient rushed into the room where his parents were sitting and threw the butter dish out of the window, believing it to be a bomb.[8] A sleepwalker threw his wife on the floor, ran to his two children, took them into his arms, and ran outside. He afterward said he believed that the house was on fire.[9] Three patient series confirmed these clinical impressions. In 12 patients with sleep terrors who underwent electroencephalography (EEG) and heart rate monitoring for several nights, 58% recalled their mental content after spontaneous arousal from slow wave sleep associated with heart rate acceleration (but not necessarily with a scream).[10] In a series of 43 adults with sleepwalking or sleep terrors, 71% reported at least one dream associated with a nocturnal motor episode.[7] The dreams were mostly short, visual, and unpleasant. There was no major difference regarding the frequency and nature of the dream content in sleepwalking compared with sleep terrors. In another series of 73 adults with sleepwalking or sleep terrors, 53% reported nightmarish mental content combined with the sensation of a vital threat and the need to escape danger during the motor episodes, including the feeling of a dangerous intruder (sometimes a murderer) trespassing the room ($n = 6$); collapsing walls, ceilings, or houses or being buried alive ($n = 5$); crocodiles, bugs, and spiders ($n = 3$); choking ($n = 5$); the urgent need to flee or being chased ($n = 3$); a baby falling out of the bed or suffocating ($n = 1$); or being in a running car or a falling lift without brakes ($n = 1$).[6]

Two case series addressed the characteristics of these dreamlike mentations through content analysis. Using the categorization method developed by Hall and Van Castle, 39% of patients with sleepwalking or sleep terrors described the presence of at least one person in the enacted dream, mostly unknown to the dreamer, whereas only 33% were relatives of the dreamer. Eleven percent of patients saw animals, which were generally aggressive or frightening, during the episode. Most (80%) mentations were negative, associated with aggression (26% of the total dreamlike mentations) and misfortune (54%, vs. no case of good fortune), and 84% were apprehensive. In the case of aggressive mentations, the patient was not the primary aggressor. In addition, 12% of the dreamlike mentation reports contained at least one act of friendliness. In all cases, the patient befriended someone and attempted to protect them (generally a relative) from danger.[7] In another series of 32 patients with sleepwalking or sleep terrors, as many as 70% of the 74 enacted dreams contained threats, which were more commonly misfortunes and disasters than aggressions. As a response to these threats, the sleepwalkers mostly fled from the disasters (25% fought back when attacked). The scenario included their bedroom in nearly half of the sleepwalkers' dreams.[11] The projection of the mental images on the room background (suggesting activation of mental images and real images seen through the eyes of the sleepwalker) is illustrated by Figure 104-1. During this episode, a sleepwalker showed some abnormal images in the sleeping room to her husband.[12] A few case reports confirmed the mind-body isomorphism using a video or video PSG of the event, followed by a detailed report of a dream coherent with the previously observed behavior.[13] The following video clips, obtained in our patients, illustrate this isomorphism: a woman protected her head from a moving, heavy boom as she dreamed of sailing with the wing astern (Video 104-1); a patient ran out of bed and cried for help as he saw a truck swooping down on him (Video 104-2); and a patient dreamed of trying to remove an insect trapped in her nose (Video 104-3).

Mechanisms

There have been several hypotheses regarding the mechanisms of these abnormal dream mentations. The brief scenes may be hypnagogic hallucinations emerging from slow wave sleep rather than more classic dreams; alternatively, they could be the terminal part of a longer dream partially forgotten at the time of arousal,[7] or a phasic, short mental creation elicited before or just at the time of arousal by ambient noises or physical contacts.[12,14] The negative emotions associated with the mentations parallel the activation of the amygdala-temporoinsular areas disengaged from the control of the prefrontal cortex (motor and emotional activation, such as fear and wandering) observed in functional imaging and deep EEG during sleepwalking.[15,16] The emotional and motor features of parasomnias could be interpreted as a release of inhibition of the subcortical central pattern generators, which regulate innate behavioral automatisms and survival behaviors.[17] The relatively consistent mental contents during sleep terrors across individuals (ceiling collapse, being buried alive, and escaping a life-threatening event) followed by a common flight response supports this concept,[11] as predicted by the Threat Simulation Theory of dreams.[18] These mental contents may also be biased toward unpleasant experiences in clinical series because these patients would more frequently seek medical advice.

The source of the abnormal mental experiences is sometimes obvious to the patients. In the series of 73 adults with sleepwalking or sleep terrors, four patients replayed a recent event that had happened during daytime, in real life or on TV, during the motor episode.[6] The history of sleepwalkers being able to replay (at least partially) a recent daytime event was

Figure 104-1 A 33-year-old woman with sleepwalking dreamed that her baby was about to fall from her cradle. **A,** She suddenly stood up, caught the baby, and brought her cautiously (with slightly folded knees) onto her bed, illustrating the isomorphism between dream content during sleepwalking and real behavior (from infrared video clip). **B** and **C,** During several episodes, the patient stared and pointed with her finger toward an invisible person, trying to convince her husband of the apparition, as if she were hallucinating, illustrating how the dreamed images are projected on the real bedroom scene.

further substantiated by a cognitive experiment performed on 19 sleepwalkers.[19] In the evening, the sleepwalkers were intensively trained on a sequence of large movements (learned during a variant of the serial reaction time task), resulting in a hands' choreography. During the nighttime, one sleepwalker reenacted a fragment of the recently trained motor behavior during a sleepwalking episode (with a concomitant delta and theta EEG background). This study provided the first evidence for the temporally structured replay of a learned behavior during sleep in humans.

Distressing mentations and nightmares during sleepwalking or sleep terrors are suggestive of abnormal emotion processing during sleep. Psychological trauma has been reported to influence the content of the sleepwalking or sleep terrors episode.[20] Additionally, recent stressful events may trigger abnormal nocturnal episodes and have even been replayed during episodes. Compared with sleepwalkers, higher levels of anxiety, obsessive-compulsive traits, phobias, and depression have been found in adult patients with sleep terrors in one study.[21] However, most studies showed no psychiatric disorders in these patients. Patients with sleepwalking or sleep terrors scored slightly higher than healthy controls on depression and anxiety scales,[22] but their scores did not differ from those of patients with RBD.[11] In addition, they had normal daytime aggression scores.[11] Another study of 105 sleepwalkers showed they scored on depression and anxiety scales similar to those of the general adult population.[23] The presence of psychopathology in sleepwalkers was associated with a higher frequency of nightmares and with potentially injurious behaviors. The successful treatment of a comorbid depressive disorder in 100 adults with sleepwalking or sleep terrors had no effects on the course of parasomnias, suggesting that the concurrent psychopathology did not play an essential role.[24]

Treatment

If sleepwalking or sleep terrors are causing distress or danger despite safety measures (avoiding sleep deprivation and other priming factors, such as alcohol, heavy meals, hyperthermia or daytime stress; quietly guiding back the patient to bed; closing windows and doors), behavioral (relaxation training, hypnosis) or medical (benzodiazepines, antidepressants, carbamazepine, melatonin; these drugs must be given at least 1 hour before going to bed to guarantee an effect as soon as 30 minutes after sleep onset) therapies are indicated. None of these therapies, however, has been tested with randomized trials. In our experience, the treatment of patients with sleepwalking or sleep terrors typically improves the dreams as well.

DISTURBED DREAMING IN REM SLEEP BEHAVIOR DISORDER

Dreams and Nightmares

Patients with RBD enact violent dreams during REM sleep in the absence of normal muscle atonia.[2] This disorder is frequent in patients with synucleinopathies (including dementia with Lewy bodies, Parkinson disease, and multiple sleep atrophy) and rare in patients with other neurodegenerative disorders (see Chapter 103). The dreams associated with RBD are usually different from those experienced by patients before RBD onset. The patients report enacted dreams containing more elements of aggression and animals than control subjects when they are asked about the dreams they remember in the last month.[25] This is also true for the dreams of patients with Parkinson disease, with or without RBD.[26] When dream reports are collected daily over several weeks in patients with treated RBD and controls, there are no differences in the content of the dreams, suggesting either a bias of

recall shifted toward selectively remembering the enacted violent dreams or a benefit of clonazepam on the abnormal dreaming process itself.[27] One may note, however, that the dream content of patients with Parkinson disease (whether they have RBD or not) includes heightened aggressiveness and an increased presence of animals compared with those of normal subjects.[28]

RBDs constitute a unique window to study the dreaming process from a point of view external to the dreamer. Behaviors, facial expressions, and verbal utterances are concordant with the dream reports obtained on awakening. In addition, some dream-enacted behaviors can be prolonged and scenic and may be more prolonged than the sleepwalking or sleep terrors behaviors. They include gesturing, reaching, grabbing, arm flailing, slapping, punching, kicking, sitting up, and leaping from bed. A demonstrative example of a patient with Parkinson disease using his pillow to defend himself against pterodactyls in his dream is shown in Video 104-4. In contrast to sleepwalking, only a minority of patients with RBD (3%) occasionally stand, and most have their eyes closed. Most descriptions emphasize the forceful and violent aspect of these motor behaviors, which are usually associated with vivid, unpleasant, and active dreams.[29] In 58 patients with Parkinson disease and RBD, the most commonly associated dream is fighting in response to danger (91%), whereas pleasant activity is reported in 20% of patients and daily activity in 22% of patients.[30] Nonviolent elaborate behaviors, however, occur in 18% of patients with Parkinson disease (coexisting in this case with violent behaviors within the same or other nights), as well as in patients with idiopathic RBD and RBD associated with other diseases.[31] They include eating and smoking (in the absence of real food or cigarettes); picking apples; dancing; teaching; gesturing thumbs up; kissing; giving a lecture; selling textiles; clapping at a show; sorting buttons; displaying pelvic, coitus-like thrusting; masturbating; urinating (while dreaming of urinating in a river as a child); scoring a goal; bicycling; greeting; flying; building a staircase; getting dressed and inspecting the army; and searching for treasure. Most behaviors are learned behaviors in accordance with the cultural and social context of the patient. Patients display various types of vocalizations, such as mumbling, talking, shouting, swearing profanities, laughing, and crying.[29] However, most patients mumble or speak during RBD, sometimes quite easily, and they speak with appropriate prosody, gestures, fluency, and syntax.[31] Singing and whistling are possible with correct musicality, and the local dialect is maintained (Video 104-5).[31]

Mechanisms of Abnormal Dreaming

RBDs are strongly linked to brainstem focal or neurodegenerative lesions within the system causing muscle atonia during REM sleep.[32,33] In this case, the motor correlates of dreams should be rendered visible for observers, whether these dreams are agreeable or not. In animal models with focal brainstem lesions that cause RBD, the RBD behaviors also contain a majority of violent behaviors (chasing or fighting) and a minority of nonviolent behaviors (licking or grooming).[34] In human RBDs, 80% of RBD-associated dreams are violent or unpleasant. The aggressive dream content correlates with more severe frontal dysfunction in Parkinson disease, regardless of whether RBD is present.[28] In this case, disturbed dreaming in RBD may be a consequence of the frontal dysfunction rather than of a brainstem lesion. In addition, the intense apparent emotions observed in RBD patients suggest an increased activation of the amygdala, as shown in normal REM sleep. Another mechanism could be related to the general function of dreams, as suggested by the Threat Simulation Theory.[18] This theory suggests that the function of dreaming is to simulate threatening events in a virtual environment and to rehearse threat perception and threat avoidance for the evolutionary purpose of increased survival. If true, this threat simulation would be exacerbated in the dreams of patients with Parkinson disease (and even more in cases of associated RBD), possibly linked with the frontal dysfunction. These dreams of fighting wild animals and aggressors are not a consequence of personality changes because they contrast with the placid personality and absence of aggressiveness during the daytime in RBD patients.[11,25]

Treatment

Several large series show the benefit of melatonin (3 to 12 mg before sleeping; one study was double blinded and placebo controlled[35]) and clonazepam (0.5 to 2 mg before sleeping)[29] on nightmares and the corresponding behaviors during RBD.[36] A few recent observations suggest that zopiclone and cholinesterase inhibitors (rivastigmine, donepezil) may also help.[37] In contrast, the effects of pramipexole are debated. Two open series reported alleviation of events in RBD, but a double-blind trial failed to show benefit.[38] Whether these drugs attenuate the negative emotions in dreams or just reduce the motor expression of dreams is yet incompletely determined.

DISTURBED DREAMING, NIGHTMARES, AND HALLUCINATIONS IN NARCOLEPSY

Dreams

Patients with narcolepsy often report frequent and intense dreaming (Figure 104-2, *left*). In a controlled series of 53 patients, narcoleptics remembered, on average, 49 dreams/month, compared with 15 dreams/month in controls, including more frequent dreams of false awakening.[39] The intense characteristics of dreaming activity are more obvious during sleep onset in REM periods,[40] with reports containing longer and a more complex organization.[41] Narcoleptic patients experienced more negative emotional tone and stronger positive emotions of dreams than healthy controls.[42] Narcoleptic patients also experienced more recurrent dreams than controls[39] or insomniacs.[43] The level of consciousness during dreams was also increased,[40] with narcoleptic patients having much more lucid dreams than controls.[39] As many as 85% of narcoleptics also experienced "dream delusion," or difficulty distinguishing between dream and reality, and mistook the memory of a vivid dream for a real experience.[44] One man, after dreaming that a young girl had drowned in a nearby lake, asked his wife to turn on the local news in full expectation that the event would be covered. Another patient experienced sexual dreams of being unfaithful to her husband. She believed this had actually happened and felt guilty about it.

Nightmares

Between 33% and 83% of narcoleptic patients report nightmares (Figure 104-2, *right*),[39,45,46] more often[39] or as frequently[46] in narcolepsy with than without cataplexy and more frequently than in controls[39] and in insomniacs.[43] Whether

Figure 104-2 Vivid dreams painted by their author, a 19-year-old patient with narcolepsy; "We were on field trip with my class, traveling in a giant ball silently floating over a desert. Suddenly, we saw a large building with a moving surface. As we approached, we discovered that the moving surface was composed of thousands of Spidermen. We leant with interest" (*left*). "In this nightmare, I was obliged to kill my whole family in a tube placed in the backyard. I had to kill them with a ballroom shoe, which was a long, tedious and horrible experience" (*right*).

the nightmares result from an overactive REM system, sleep onset in REM periods with more perceived sleep paralysis, fragmented REM sleep, abnormal amygdala activation in narcolepsy, overall higher daytime stress, or side effects of antidepressants and stimulants is yet undetermined. In a series of 54 narcoleptic patients, the occurrence of nightmares was higher in patients with RBD.[46] Narcoleptic patients with nightmares had longer wakefulness time and higher percentages of N1 stage sleep, suggesting either that nightmares disrupted sleep or that superficial sleep promoted their onset.[46] Lucid dreaming, which is easily achieved by three fourths of narcoleptic patients, may be used to turn recurrent nightmares into agreeable dreams,[39] as illustrated in the following examples: "I was chased by an aggressor and had to kill myself by throwing myself onto an electrified fence to avoid the aggressor and wake up"; "I was chased by soldiers and chose to fly to escape them"; "I built a panic room in myself where I mentally went during my nightmare to escape dangerous people"; "I was seeing my family tortured to death and decided that it was not true, because we were rehearsing a play or a movie. I was behind the camera"; and "I built myself a suit of armor made of lightning that protected me from my enemies."[39] A small (11%, 6 of 54 patients) proportion of patients reported that nightmares had been eliminated by narcolepsy medications, mostly anticataplectic drugs.[46]

Hallucinations

Hallucinations are experienced by 45% to 80% of narcoleptic patients with cataplexy[47-49] and by 29% of narcoleptic patients without cataplexy.[49] The formed hallucinations are multimodal, with visual, auditory, olfactory, sexual, tactile, and motor perceptions. The visual hallucinations are the most frequent, including human or fantastic characters that are complete or incomplete (subjects standing still or moving, transparent bodies, isolated head moving, Picasso-like faces, vampires, decapitated heads, or a rare guardian angel; Figure 104-3, *A*) and animals (Figure 104-3, *B*).[49] Auditory hallucinations come second in terms of frequency, with nonverbal (footsteps, doors opening, alarm clock, music box, singing sparrows, bells, gun detonation, and broken china) and verbal (hearing one's own first name is the most frequent, bits of conversations, but with no imperative voice or replacement of will as in schizophrenia). Motor (or kinetic) hallucinations include the sensation of laterally moving the body, being sucked into the bed, flying, levitating, and having out-of-body experiences (Figure 104-3, *C*). Tactile hallucinations include feelings of blowing fresh air, a smooth touch, drops of water, burning body sensations, and being bitten, trampled, strangled, or sexually abused.[48,49] Multimodal hallucinations may be experienced together as a holistic, real phenomenon, often referred to by patients as "waking dreams." Hallucinations are frightening in 80% patients,[50] and their major impact is anxiety.[48] As a consequence, some patients close their door or sleep with weapons on their bed table. They can be associated with sleep paralysis, which is, with RBD, their major determinant (Figure 104-3, *D*).[49] Hallucinations typically occur at sleep onset and offset, but they also occur under fully awake conditions in 40% of patients.[48,49] They begin at about 20 years of age. Because hallucinations can be pervasive, some narcolepsy patients have even been misdiagnosed as having schizophrenia.[51] This misdiagnosis is rare, however,[52] because narcoleptic hallucinations differ from schizophrenic hallucinations by their major association with sleep, more frequent visual and multimodal hallucinations, and exceptional delusional thinking.[48,49,53] Hallucinations may be phasic elements of sleep (possibly abnormal ponto-geniculo-occipital wave

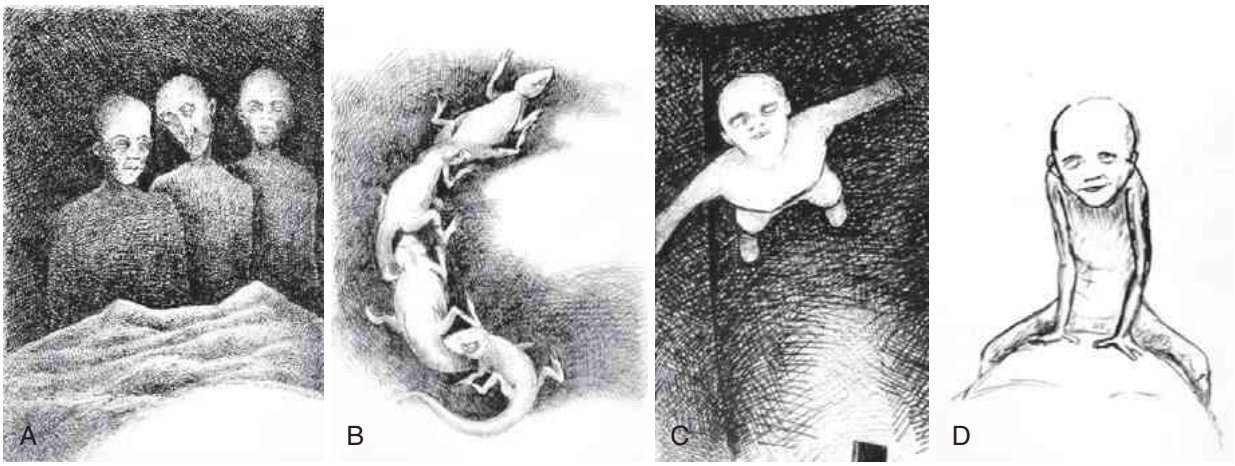


Figure 104-3 Hypnagogic hallucinations drawn by a 45-year-old man with narcolepsy. **A,** The patient experienced recurrent hallucinations of "friends" around his bed at sleep onset or offset. They were not identifiable but perceived as friendly. **B,** Another recurrent hallucination was a salamander crawling on the bed. **C,** Additionally, the patient had an out-of-body experience (but no heautoscopy), then floated in the corner or just below the ceiling of the room. **D,** Eventually, he felt a small, malicious child sitting on his chest and pressing it, inducing dyspnea and the feeling that it was impossible to move (sleep paralysis and hallucination).

activities) occurring during the wake state. They are usually reduced by anticataplectic drugs.

DREAM DISTURBANCES AND HYPNAGOGIC HALLUCINATIONS IN OTHER NEUROLOGIC DISEASES

Parkinson Disease

Compared with normal controls, the dreaming activity changes in Parkinson disease. As many as 46% of patients report altered dream phenomena,[54,55] with a high frequency of nightmares and violent or unpleasant dreams in Parkinson disease patients, especially when levodopa therapy is introduced.[54-56] In early Parkinson disease stages, patients' dreams differ from those of the control group in features related to aggressive actions (in which they frequently had a passive role), the presence of animals, a relatively higher frequency of friendly acts toward other characters, and a lower frequency of bodily misfortunes.[28] Because the altered dreaming is related to frontal cognitive impairment and not to RBD, the authors speculate that the higher level of aggression reflects intensification of the limbic preponderance during sleep owing to a loss of the prefrontal regulatory influence. In contrast, Borek and colleagues found a relatively higher frequency of aggressive features in PD patients with than without RBD (with no further difference in men vs. women with RBD), although dreams were less aggressive in women with Parkinson disease than in men.[26] The altered dreaming activity was associated with more frequent awakenings and illusions or hallucinations, but not with specific (levodopa, dopamine agonist) medications.[56] A "kindling" phenomenon starts from altered dreaming and evolves toward illusions, then hallucinations of minor then major severity; eventually, psychosis is suspected.[54] The presence of vivid dreams and nightmares correlated with concurrent hallucinations, but did not predict the future development of hallucinations when they occurred in nonhallucinators in a 10-year prospective study.[57] The interest in vivid dreams and nightmares as a step toward psychosis, however, did not include RBD, which was not identified as a disorder at the time of the study.[54] When RBD was examined at cohorts' entry into the study, it proved to be a major determinant for concurrent and incident hallucinations, as well as later development of psychosis and dementia.[57,58]

In addition to altered dreaming, visual hallucinations are experienced by an average of one third of patients with Parkinson disease. They consist of visions of human figures or faces, or animals, sometimes in miniature, or scenery of outstanding beauty.[49,59,60] In addition to older age and longer disease course, several risk factors have been associated with the occurrence of hallucinations, including dopaminergic and anticholinergic therapy, cognitive impairment, visual deficits, excessive daytime sleepiness, and sleep disturbances.[59-63] It is likely that hallucinations in Parkinson disease are better explained by state dissociation and intrusion of dreamlike phenomena into wakefulness (as in primary narcolepsy) than by "psychosis" of cortical origin. Evidence for this mechanism consists of the frequent association of hallucinations with RBDs[64] and with excessive daytime sleepiness,[59] as well as their timely occurrence at the end of nighttime and daytime REM sleep episodes (Figure 104-4).[65] This narcolepsy-like mechanism was also postulated in patients with excessive daytime sleepiness.[66] Eventually, a brain examination showed numerous Lewy bodies within the subcoeruleus nucleus but not in the cortex in a patient with visual hallucinations plus RBD.[65] Eventually, a loss of hypocretin neurons, which can contribute to the sleepiness and hallucinations in patients with Parkinson disease, was demonstrated.[67,68] These results supported a brainstem-hypothalamus origin for visual hallucination in Parkinson disease, rather than a cortical origin (and onset of dementia with Lewy bodies), as thought before. The treatment of nightmares and hallucinations in Parkinson disease and other neurodegenerative diseases usually requires decreasing or stopping the provocative drugs (usually dopamine agonists and antidepressants) to assess for RBD (and treat adequately with melatonin or clonazepam) and, if severe, use of atypical neuroleptics such as quetiapine or clozapine.

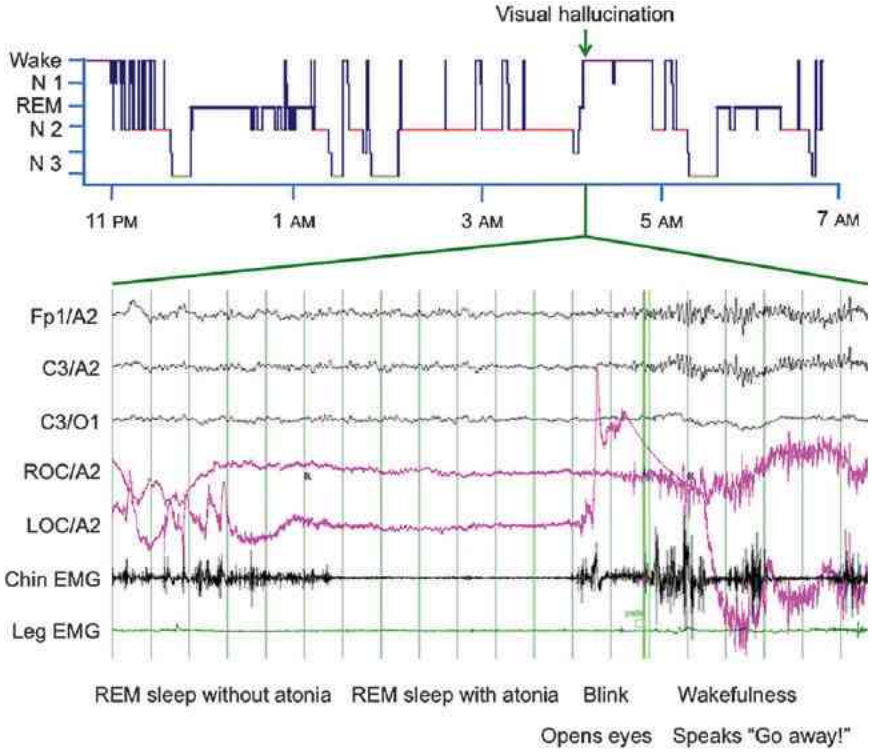


Figure 104-4 Visual hallucination experienced by a patient with Parkinson disease on emergence from REM sleep. On the hypnogram (*upper panel*), the visual hallucination emerges from a short REM sleep episode and triggers a prolonged awakening. The electroencephalogram shows REM sleep without atonia followed by REM sleep with atonia, then a blink, followed by a sudden awakening. On the video clip, the patient opens the eyes (blink) and speaks, saying "Go away!" to her hallucination. She later reported to the nurse that she saw her grandchildren around her in the bedroom.

Guillain-Barré Syndrome

Guillain-Barré syndrome is a rare, acute, and severe polyradiculopathy, probably of autoimmune origin. Although it mostly affects the peripheral nervous system, signs of central nervous dysfunction, such as RBD,[69] sleepiness, hallucinations, abnormal antidiuretic hormone secretion, and low cerebrospinal fluid levels of hypocretin are noted.[70] In a consecutive series of 139 patients with Guillain-Barré syndrome without any psychotropics and opiates, we found that patients frequently reported vivid dreams (19%), illusions (30%, including an illusory body tilt), hallucinations (60%, mainly visual), and delusions (70%, mostly paranoid).[71] Examples include a dream of floating over the streets of New York early in the morning and seeing people taking out their garbage (a vivid dream that the patient remembered several years afterward); an intravenous bag mistaken for a banana tree (visual illusions); the bed feeling as if it were vertical, with the nurses walking on the walls (spatial illusions); and animals or painted deceased relatives raising their arms through the frame (visual hallucinations). A patient believed that he was a prisoner of the Germans during World War II trying to cross a border and was continually asking the nurses for his passport (a brief psychosis). In patients monitored during the hallucinatory period, the PSG showed a status dissociatus characterized by major insomnia; REM sleep without atonia during 92% of the REM sleep time in all hallucinators (RBD could not be assessed because patients were totally paralyzed by the Guillain-Barré syndrome process); bursts of rapid eye movements during stage N2 sleep; sleep onset in REM periods; and continuous, rapid switching between waking, N1 sleep, and REM sleep stages across day and night. In addition, there were signs of dysautonomia. The status progressively disappeared when hallucinations subsided. In this case, status dissociatus (rather than "isolated" RBD) was associated with disturbed dreaming and hallucinations.

Delirium Tremens, Fatal Familial Insomnia, and Morvan Chorea: Agrypnia Excitata

Notably, some forms of status dissociatus include continuous dreaming and hallucinations, motor activity, and a complete loss of wake-sleep boundaries, with patients fluctuating continuously between N1/REM sleep and wakefulness.[72] Such status dissociatus is observed in Parkinson disease, dementia with Lewy bodies, Guillain-Barré syndrome, alcohol withdrawal syndrome (delirium tremens),[73] fatal familial insomnia, and Morvan chorea. This phenomenon has been named *agrypnia excitata* by the Bologna group and is sometimes called *oneirism* in French groups, but these names refer to the same observation of severe visual hallucinations and enacted dreams (resembling a continuous RBD, but with open eyes). Two specific variants of status dissociatus included a case of frontal dementia of unknown origin with continuous dreamlike activity across all sleep stages, bursts of rapid eye movements during NREM sleep, and RBD,[74] as well as eight patients with a tauopathy and abnormal antibodies plus obstructive sleep apnea, stridor, abnormal sleep architecture with undifferentiated NREM sleep or poorly structured stage N2 sleep with

simple movements and finalistic behaviors, normalization of NREM sleep by the end of the night, and RBD.[75]

CLINICAL PEARLS

The clinician should realize that some patients use the words "disturbed dreaming" and "nightmares" not only for indicating the classic REM sleep nightmare disorder, PTSD, and distressing dreams associated with psychiatric conditions, drugs, and substances but also for reporting the mental experiences associated with hypnagogic hallucinations, sleep terrors, RBDs, and status dissociatus. The characteristics (type, construction, timing, duration, and context) of their mental experience, along with associated behaviors, autonomic signs, and associated disorders, and aspects and timing on PSG are crucial for making a diagnosis.

SUMMARY

Most adult patients with arousal disorders remember a mental content timely associated with sleepwalking and sleep terrors, mostly a brief visual scene including misfortune, apprehension, and life-threatening, imminent dangers that they must flee. The observed behaviors and speeches are congruent with the mental content. During RBDs, patients mostly enact violent dreams in which they are aggressed by humans and animals and counterattack, despite their placid personality when awake. About 20% of the behaviors and dreaming are elaborated, nonviolent content, including laughing, performing their job, or speaking with people. Narcolepsy patients have longer, more intense and lucid dreams than controls, with more frequent nightmares that are associated with time of sleep onset in REM periods. Vivid dreams, illusions, hallucinations, and psychosis in several neurologic diseases caused by neurodegeneration (mostly synucleinopathies, including Parkinson disease, dementia with Lewy bodies), autoimmunity (primary narcolepsy, Guillain-Barré syndrome, Morvan chorea), and alcohol withdrawal (delirium tremens) seem linked to abnormal REM sleep and status dissociatus.

Selected Readings

Bugalho P, Paiva T. Dream features in the early stages of Parkinson's disease. *J Neural Transm* 2011;**118**:1613–19.

Cochen V, Arnulf I, Demeret S, et al. Vivid dreams, hallucinations, psychosis and REM sleep in Guillain-Barre syndrome. *Brain* 2005;**128**:2535–45.

Fantini ML, Corona A, Clerici S, Ferini-Strambi L. Aggressive dream content without daytime aggressiveness in REM sleep behavior disorder. *Neurology* 2005;**65**:1010–15.

Fénelon G, Mahieux F, Huon R, Ziegler M. Hallucinations in Parkinson's disease: prevalence, phenomenology and risk factors. *Brain* 2000;**123**:733–45.

Fortuyn HA, Lappenschaar GA, Nienhuis FJ, et al. Psychotic symptoms in narcolepsy: phenomenology and a comparison with schizophrenia. *Gen Hosp Psychiatry* 2009;**31**:146–54.

Leclair-Visonneau L, Oudiette D, et al. Do the eyes scan dream images during rapid eye movement sleep? Evidence from the rapid eye movement sleep behaviour disorder model. *Brain* 2010;**133**:1737–46.

Mahowald MW, Schenck CH. Status dissociatus: a perspective on states of being. *Sleep* 1991;**14**:69–79.

Montagna P, Lugaresi E. Agrypnia excitata: a generalized overactivity syndrome and a useful concept in the neurophysiopathology of sleep. *Clin Neurophysiol* 2002;**113**:552–60.

Oudiette D, Leu S, Pottier M, et al. Dreamlike mentations during sleepwalking and sleep terrors in adults. *Sleep* 2009;**32**:1621–7.

Uguccioni G, Golmard JL, de Fontreaux A, et al. Fight or flight? Dream content during sleepwalking/sleep terrors vs rapid eye movement sleep behavior disorder. *Sleep Med* 2013;**14**:391–8.

Valli K, Frauscher B, Peltomaa T, et al. Dreaming furiously? A sleep laboratory study on the dream content of people with Parkinson's disease and with or without rapid eye movement sleep behavior disorder. *Sleep Med* 2015;**16**(3):419–27.

van Schagen AM, Lancee J, de Groot IW, et al. Imagery rehearsal therapy in addition to treatment as usual for patients with diverse psychiatric diagnoses suffering from nightmares: a randomized controlled trial. *J Clin Psychiatry* 2015;**76**(9):e1105–13.

A complete reference list can be found online at ExpertConsult.com.

Chapter 105

Other Parasomnias

Alex Iranzo

Chapter Highlights

- The heterogeneous group of other parasomnias includes sleep-related hallucinations, exploding head syndrome, sleep enuresis, and sleeptalking.
- Other parasomnias are very frequent, transient, and usually benign in nature, but some rare disorders can be progressive and devastating.
- Patients' education and reassurance that they are harmless conditions is necessary. Therapy is rarely warranted but could be considered when experiences are bothersome, unpleasant, terrifying, or very frequent and causing sleep difficulties.
- Although most cases are benign and idiopathic, some are associated with serious conditions such as narcolepsy in patients with hypnagogic hallucinations, brainstem stroke in subjects with complex nocturnal visual hallucinations, and obstructive sleep apnea and recurrent urinary tract infections in children with sleep enuresis.

Parasomnias are undesirable physical events or experiences that occur during entry into sleep, within sleep, or during arousal from sleep.[1] The *International Classification of Sleep Disorders,* third edition (ICSD3) classifies "other parasomnias" to include a miscellaneous group of phenomena not linked to a specific sleep stage such as sleep-related hallucinations, exploding head syndrome (EHS), sleep enuresis, the normal variant sleeptalking, and parasomnias secondary to a medical disorder and to a medication or substance.[1] This chapter describes the conditions listed in the other parasomnias category and will discuss the possible epidemiology, pathophysiologic consequences, and clinical approach.[1]

SLEEP-RELATED HALLUCINATIONS

Definition, Diagnostic Criteria, and Classification

Hallucinations are sensory experiences that are perceived in the absence of environmental stimuli. Consciousness is preserved, and subjects can describe in detail their perceptions with or without insight into their unreality. Hallucinations should be distinguished from illusions, which correspond to misperceptions or distorted perceptions of real external stimuli. They also should not be mistaken for dreams, which are experiences that occur when the individual is asleep and are recalled on awakening.[2] Hallucinations are the result of an abnormal neuronal sensory processing in the setting of the wake-sleep transition (e.g., hypnagogic hallucinations, hypnopompic hallucinations), sensory deprivation (e.g., severe vision loss in the Charles Bonnet syndrome), structural brain lesions (e.g., stroke), neurologic diseases (e.g., Parkinson disease), psychiatric diseases (e.g., schizophrenia), metabolic diseases (e.g., hypoxic encephalopathy), and the use or withdrawal of a medication (e.g., dopaminergic agents).

Although hallucinations can occur during daytime wakefulness, at sleep onset, and on awakening from sleep, daytime hallucinations are more frequent than sleep-related hallucinations.[2] According to the ICSD3, diagnostic criteria of sleep-related hallucinations include (1) a complaint of recurrent hallucinations that are experienced just before sleep or on awakening during the night or in the morning, (2) hallucinations that are predominantly visual, and (3) disturbance that is not better explained by another sleep disorder (especially narcolepsy), mental disorder, medical disorder, medication, or substance use.[1]

Sleep-related hallucinations may be classified into two different forms according their time of occurrence, clinical characteristics, and underlying substrate: (1) hypnagogic (those that occur at sleep onset) and hypnopompic (those that occur upon awakening) hallucinations, which are experiences that are either idiopathic or related to other conditions, particularly narcolepsy; and (2) complex nocturnal visual hallucinations (CNVHs), which occur after sudden awakening during the night, are not simple experiences, and are almost always linked to an underlying pathology, usually neurologic, psychiatric, metabolic, and ophthalmic in nature.[2]

Hypnagogic and Hypnopompic Hallucinations

Clinical Findings. Hypnagogic hallucinations (HHs) and hypnopompic hallucinations (HPHs) are visual (e.g., feeling someone or something present in the room; simple elementary forms such as sparks, lines, flashes, confetti, and shadows; complex forms like waterfalls, cucumbers, animals, known or unknown people or faces, dwarfs, thieves, firemen, and lifelike scenes), auditory (e.g., footsteps, explosions, shots, a beep from a cell phone, voices of known or unknown people, familiar or unfamiliar songs), tactile (e.g., someone grabbing the subject, bugs crawling on the skin, tingling, pain), gustatory (e.g., metallic taste), olfactory (e.g., perfume, cologne, feces, smoke), and kinetic (e.g., floating, flying, jumping, falling, out-of-body experiences, levitation). Most of the people who experience HH and HPH know that the perceptions are not true and do not exist. Despite this, some hallucinations, such as seeing and hearing persons, are very vivid and real and may be experienced as unpleasant, frightening, and hard to

distinguish from true events.[3] Kinetic hallucinations are so bizarre they may lead to paranormal beliefs or be mistaken for delusional psychosis.

Epidemiology. Epidemiologic studies are scarce and give different estimations of prevalence because of methodologic differences in the definition of hallucination and the accuracy of how data were obtained (telephone interviews, questionnaires). HHs are more common than HPHs. In a community sample of 49,772 people from the United Kingdom aged 15 to 100 years, who were interviewed by telephone, 37% reported HH and 12.5% HPH.[4] In an observational study involving 134 healthy medical students from Spain with a mean age of 22 years, 13% reported sleep-related hallucinations.[5]

Associations. There is no known familial or genetic predisposition. The most important precipitant in predisposed individuals is sleep deprivation leading to HH or HPH during daytime naps or during the night. Many drugs and psychoactive substances have been related to HH and HPH, such as hashish, opiates, amphetamines, cocaine, hypnotics, and zopiclone.[3] Simultaneous occurrence of HH or HPH with sleep paralysis is common in healthy people and in narcoleptics.

In *narcolepsy*, HHs are more frequent than HPHs. In one study hallucinations occurred in 59% of patients with cataplexy and in 28% without cataplexy.[6] They were experienced in the context of sleep onset (55%), sleep offset (3%), and onset-offset (42%). Hallucinations were visual in 95% of the patients, auditory in 75%, kinetic in 55%, and tactile in 33%, and 10% experienced passage or presence of a real person. Olfactory and gustatory hallucinations were not reported. When narcoleptic patients experience sleep paralysis, the most common simultaneous hallucinations are tactile (e.g., a frightening pressure on the chest) and visual (e.g., feeling the presence of someone or a shadow). Events can be so vivid that some undiagnosed narcoleptic patients may be convinced that they are experiencing paranormal events.[7] In narcolepsy, HH and HPH occur predominantly in the supine position.[8]

Etiology. HH and HPH are thought to be intrusions of the characteristic dreamlike phenomena of rapid eye movement (REM) sleep into wakefulness.

Management. When HH and HPH are idiopathic, patients need to be assured that they correspond to a very frequent and almost normal phenomenon that occurs in healthy people and not a sign of psychosis, narcolepsy, or a paranormal experience. Patients with narcolepsy need to be informed that hallucinations are part of the disease. When treatment is needed because hallucinations are frequent or bothersome, clomipramine 10 to 75 mg at bed time may be effective, particularly in those associated with sleep paralysis. Sodium oxybate may not help diminish events in patients with narcolepsy.[9,10]

Complex Nocturnal Visual Hallucinations

Clinical Findings. CNVHs occur following a sudden awakening. They may be primary or much more frequently secondary to an underlying condition. Despite the variety and large number of conditions that may cause CNVH, the type of visual hallucinations are very similar among them. CNVHs are complex, vivid, detailed, relatively stereotyped, static or mobile, colorful images of people, animals, and elaborated scenes resembling a dream. They can be brief, lasting a few seconds, or prolonged, lasting several hours. They usually disappear if the eyes are opened or if the lights of the room are switched on. Sometimes the images are distorted in shape or size, such as those seen in the Alice in Wonderland syndrome. In rare cases, visual hallucinations can be associated with auditory hallucinations (e.g., a song, a crowd of people talking or chanting) or tactile (e.g., tingling, pain). Insight regarding the hallucinations is more reduced than in HH and HPH, and subjects may believe they are true and leave the bed to investigate whether the images are real or not. In other patients, insight is preserved but hallucinations are perceived as bothersome and distressing. CNVH may co-occur with HH and HPH. The secondary, but not the primary, form of CNVH may coexist with hallucinations during wakefulness.[1,11,12]

Associations. Precipitants of CNVH depend on the nature of the underlying condition and include sleep deprivation, fever, trauma, electrolyte disturbances, changes in medication, poor vision, and low ambient illumination. In contrast to HH and HPH, CNVHs are secondary to a large number of conditions of different origin (Table 105-1).[11-27]

In *Parkinson disease* hallucinations occur in about 25% of patients.[13] Risk factors are older age, long disease duration,

Table 105-1 Conditions Associated with Sleep-Related Hallucinations

Hypnagogic and Hypnopompic Hallucinations
Sleep deprivation[1]
Narcolepsy[6,7]
Complex Nocturnal Visual Hallucinations
Parkinson disease[6,11,13-16]
Dementia with Lewy bodies[17,18]
Peduncular hallucinosis[11,19,21,22]
Charles Bonnet syndrome[11,23]
Schizophrenia[11]
Metabolic encephalopathy[11]
Posterior cerebral artery infarction[11]
Delirium tremens[11]
Migraine[11]
Focal epilepsy[11]
Guillain-Barré syndrome[26]
Sleepwalking[1,2,27]
Night terrors[1,2]
Idiopathic hypersomnia[12]
Anxiety disorder[12]
Acute alcohol withdrawal[12]
Acute barbiturate withdrawal[12]
Acute benzodiazepine withdrawal[12]
Lipophilic beta blockers[1,2]
Dopaminergic agents[1,2]
Substances with hallucinogenic properties like mescaline, LSD, amphetamine, and cocaine[1,2]

increased motor disability, depression, cognitive impairment, hypersomnia, REM sleep behavior disorder (RBD), and the use of dopaminergic agonists and anticholinergics.[6,13] Hallucinations are mostly visual, recurrent, static or mobile, stereotyped, vivid, involving shadows, silent moving animals and people, and the presence or passage of a quiet known person or pet. They appear mostly in dim surroundings or at night. Tactile, auditory, and kinetic hallucinations are rare. Although the etiology of these hallucinations in Parkinson disease is unclear, the connection to REM sleep intrusions into wakefulness or perceptual disturbance is debated.[6-16]

Dementia with Lewy bodies is dependent on visual hallucinations as one of the major diagnostic criteria.[17] In one study, a patient with dementia with Lewy bodies experienced an episode of CNVH, and polysomnography showed that it resulted from an awakening arising from N2 sleep stage.[18] Patients with dementia with Lewy bodies or other forms of dementia may find it difficult to distinguish CNVH from dreams.

Peduncular hallucinosis is characterized by CNVH in the setting of pontine, midbrain, or thalamic damage.[11,19] The most common etiology is vascular, but tumors and inflammatory processes have also been described.[20] These stereotypic events are usually self-limited, starting a few days after the stroke and commonly associated with hypersomnia. CNVHs occur in the evening, last several minutes or hours, and disappear when the eyes are open.[11] Patients may lack insight and may interact with the hallucinations, performing pseudopurposeful activities (e.g., using a screwdriver, putting on trousers, or talking). Polysomnographic studies have shown that the hallucinations occur during wake-sleep transitions with preserved occipital alpha activity.[21,22]

Charles Bonnet syndrome is characterized by CNVH in elderly subjects with severe visual loss secondary to bilateral ocular pathology (e.g., macular degeneration, cataracts, glaucoma). The syndrome may be caused by any lesion affecting the visual pathway, including the optic nerve, optic chiasm, and temporal and occipital lobes. For these patients CNVH occurs in the evening and night when the patient is drowsy but the eyes are still open. Patients usually have insight into the nature of their perceptions. The course and treatment depends on the underlying pathology. Symptomatic therapy with atypical antipsychotics such as olanzapine may be effective.[11,23]

Polysomnography. The few studies that have described the polysomnographic pattern of CNVH have shown that CNVHs arise from N2 and N3 sleep and not from REM sleep. During the events, electroencephalography shows occipital alpha rhythm without epileptic activity, indicating that the subject is awake.[12,24] CNVHs experienced by Parkinson disease patients may arise from four distinct circumstances: (1) during REM sleep without atonia, (2) on arousal from non–rapid eye movement (NREM) sleep, (3) in the daytime during drowsiness within the wake-NREM sleep transition, and (4) during wakeful state.[25]

Etiology. The anatomic substrate of CNHV shares a final common pathway where the occipital visual cortex generates false images from reduced sensory inputs. Input reduction may arise from the occipital cortex itself (e.g., in dementia with Lewy bodies) or its afferents from the thalamus (e.g., peduncular hallucinosis), the brainstem (e.g., Parkinson disease), and the retina (e.g., Charles Bonnet syndrome).[11]

Management. In most cases reassurance is sufficient. Adequate treatment of the underlying condition may eliminate the CNVH. Withdrawal of the offensive agent (e.g., β-adrenergic antagonist, dopaminergic agents, anticholinergics) often results in resolution of CNVH.[12] In cases in which therapy is warranted, antidopaminergic psychotics (e.g., risperidone, quetiapine, olanzapine, and clozapine) or anticholinesterase inhibitors (e.g., rivastigmine) may be effective.

EXPLODING HEAD SYNDROME

Definition and Diagnostic Criteria

EHS is characterized by a sense of a loud explosion in the head that awakens the individual.[28] Diagnostic criteria of the ICSD3 include (1) a complaint of a sudden noise or sense of explosion in the head either at the wake-sleep transition or on awakening in the night; (2) abrupt arousal following the event, often with a sense of fright; and (3) experience not associated with significant complaints of pain.[1]

Clinical Findings

Patients are abruptly woken by an imagined, sudden, violent, and frightening sensation of explosion deep in center or back of the head or near them.[29] Patients have a clear recollection of the events without any postictal confusion. Episodes last between a split second and a few seconds and disappear completely the moment the subject is awakened. The noise is perceived over both ears,[30] although a few subjects describe hearing the sound in only one ear.[31,32] Patients report noise, not pain, and describe the experience in many ways, such as "sudden bang in the head,"[29] "thunderclap,"[29] "shouts,"[30] "shotgun,"[33] "loud metallic noise,"[33] "clash of cymbals,"[33] "door slamming,"[33] "the earth moved,"[33] "electric shocks,"[33] "bomb like explosion,"[33] "electrical current running,"[34] "a beep,"[35] and "a buzzing noise."[36] Some patients are so alarmed by its violence that think that it is a symptom of a stroke or a brain tumor.[30] The explosive noise may occasionally be accompanied by a simultaneous flash of light in front of the eyes,[30,33] hypnic jerks of the whole body[30] or one limb,[36,37] and the sensations of fear, difficult breathing, palpitations, and sweating on awakening.[33] In one case, EHS was followed by episodes of sleep paralysis.[35] Attacks occur exclusively during sleep, during the transition from wakefulness to sleep, or less often on awakening from sleep, particularly during the nighttime. In a few cases the attacks only appear on awakening from nighttime sleep[31,37] or during daytime naps.[37]

Epidemiology

There are no epidemiologic studies focused on EHS, and less than 100 cases are reported in the medical literature, but it is suspected to be much more common.[29] In fact, a recent study found that recurrent EHS occurred in 16% of college students.[37a] Onset is usually after the age of 50 years, but it may start in childhood and adolescence. There is a slight predominance in females.

Natural Course

The frequency and course of episodes are highly variable, ranging from two or three spells in a lifetime up to several per

night for several weeks. Some patients have clusters lasting weeks to a few months separated by prolonged periods of remission. The condition improves and resolves spontaneously.[28]

Laboratory Investigations

Neurologic examination, brain magnetic resonance imaging, angiography, electroencephalography, and interictal polysomnography are normal. Sleep studies reveal that the attacks may either evolve at the transition from wakefulness to N1,[30,38] and from N2,[32,38] or in the transition from N1 to wakefulness.[36] No epileptic activity, apneic events, or electrocardiographic abnormalities are seen in the episodes.

Associations

Family history of EHS may occur, but it is very uncommon.[36] There are no identifiable precipitants, although a few patients state that attacks tend to occur when they are under stress or very tired. Events are not associated with any type of headache disorder or neurologic, psychiatric, or auditory diseases.

Etiology

The etiology of the EHS is unknown. It is an abnormal sensory phenomenon occurring at the transition from wakefulness to sleep. The events may represent an auditory sleep-related hallucination or a sensory variant of hypnic jerks.[2] Some investigators suggest events may be related to abnormalities in the brainstem reticular formation,[37] an acoustic migraine aura,[38] or middle-ear eustachian tube dysfunction.[31]

Management

Treatment is generally not required because EHS is a benign disorder that remits with time. Education and reassurance that EHS is a harmless condition can be necessary to reduce fear and anxiety. A medication can be tried when episodes interfere with sleep or cause stress. Anecdotal reports suggest benefit with clomipramine,[30] topiramate,[36] flunarizine,[37] nifedipine,[39] or clonazepam.[40]

SLEEP ENURESIS

Definition

Sleep enuresis (SE), also called bedwetting and nocturnal enuresis, is defined as involuntary recurrent micturition during sleep in subjects older than 5 years at least twice a week for more than 3 months.[1]

Classification

SE is classified into a primary and a secondary form depending on whether the patient has been dry during sleep for at least 6 months. This distinction is made because primary and secondary forms have different etiologies despite the same symptoms. Primary SE is more frequent than secondary SE.[1]

Primary Sleep Enuresis

Diagnostic Criteria. According to ICSD3, diagnostic criteria include the following: (1) the patient is older than 5 years; (2) the patient exhibits recurrent involuntary voiding during sleep, occurring at least twice a week; (3) the condition has been present for at least 3 months; and (4) the patient has never been consistently dry during sleep for at least 6 months.[1]

Epidemiology. SE affects approximately 4% to 16% of school-aged children.[41-43] A study on the development of parasomnias from childhood to early adolescence showed that enuresis occurred in 16% of 1353 children and persisted in adolescence in 12% of these cases.[43] The prevalence of SE decreases with time, being present in about 10% of 6-year-olds and 3% of 12-years-olds.[43] Adult enuresis is rare, occurring in about 0.5% to 3% of the population, and is usually linked to an underlying condition. In children, primary SE is more common in boys than in girls. In adults it is more frequent in women.[1]

Polysomnography. Sleep studies show that the episodes of SE occur from wakefulness and any sleep stage. During sleep, 39% of episodes occur in N2, 20% in N3, and 9% in REM sleep. Most of the events occur in the first third of the night. Polysomnography studies show an increased number of arousals.[44] Obstructive sleep apnea occurs in 8% to 47% children with SE.[45]

Associations. About two thirds of cases of primary SE are familial, and one third are sporadic. SE occurs in 77% of children when both parents had SE, and in 44% when one of the parents had a history of SE.[46] There is a higher concordance rate for monozygotic twins than in dizygotic twins.[47] Models of heritance can be either autosomal dominant or autosomal recessive.[47] Linkage studies have demonstrated associations with chromosomes 4, 8, 12, 13, and 22.[47] Primary SE is usually not associated with involuntary voiding during wakefulness. SE may cause embarrassment and low self-esteem, causing psychosocial stress for the patient.

Etiology. The etiology of primary SE is unknown, but several mechanisms are implicated in its pathophysiology: polyuria leading to nocturia, reduced bladder functional capacity (the volume at which the bladder empties itself is reduced), increased nocturnal bladder activity, decreased nocturnal secretion of vasopressin (the antidiuretic hormone), and central difficulty of arousing from sleep. The latter may be the most important factor in primary SE due to a delayed development of the central nervous system leading to a failure in the arousal system in response to the sensation of a full bladder. Enuretic boys are found to be more difficult to arouse from sleep than age-matched controls,[48,49] and a small proportion of children with primary SE have reduced vasopressin levels during sleep, resulting into high urinary volume.[50]

Management. Parent and child involvement, motivation, and cooperation are crucial. To avoid embarrassment and low self-esteem causing the patient psychosocial stress, parents must be supportive of the child and use positive reinforcement for desired behaviors. Fluid restriction in the evenings and forced awakenings in the middle of the night to urinate are not usually effective. Treatment includes behavioral therapy with daytime bladder control training,[51] alarm systems as a conditioning therapy,[52,53] and drugs such as imipramine, oxybutynin, and desmopressin.[54]

Secondary Enuresis

Diagnostic Criteria. According to the ICSD3, diagnostic criteria include the following: (1) the patient is older than 5 years; (2) the patient exhibits recurrent involuntary

voiding during sleep, occurring at least twice a week; (3) the condition has been present for at least 3 months; and (4) the patient has been consistently dry during sleep for at least 6 months.[1]

Associations. Secondary SE is the result of an acquired factor such as excessive liquid intake, diabetes mellitus, diabetes insipidus, malformations of the genitourinary tract, recurrent urinary tract infections, chronic constipation and encopresis producing an extrinsic pressure on the bladder, significant psychosocial stress, attention-deficit hyperactivity disorder, developmental disorders in children with intellectual disability, obstructive sleep apnea, congestive sleep heart failure, nocturnal epilepsy, stroke, Parkinson disease, dementia of any type, neurogenic bladder in spinal cord lesions and multiple sclerosis, and the use of diuretics.[1]

Etiology. Etiology depends on the underlying condition. Overall, pathophysiology of secondary SE includes an inability to concentrate urine, increased urinary production, hyperactivity of the bladder, and genitourinary tract malformations. Failure of arousal from sleep when the bladder is full of urine may also occur in a few cases with secondary SE, but this is a factor mostly associated with primary SE.

Management. The underlying cause should be treated first (urinary infection, urinary malformation, sleep apnea, diabetes mellitus, diabetes insipidus) before symptomatic therapy. In children, specific therapy for primary SE should be tried in secondary SE.

SLEEPTALKING

Definition and Classification

Sleeptalking, also known as somniloquy, consists of talking while the individual is asleep. This should be distinguished from periods of talking during nocturnal awakenings. Sleeptalking may not be mistaken for other sounds occurring in sleep, such as catathrenia, snoring, stridor, choking, and coughing. There are two forms of sleeptalking: a primary form in which sleeptalking is an isolated manifestation, and a secondary form in which it is one the clinical features of a different underlying condition, such as confusional arousals and RBD.

Isolated Sleeptalking

Clinical Findings. Sleeptalking is a normal sleep variant, a benign phenomenon in which the individual remains immobile (except for the act of talking) and usually does not remember the event the next morning on awakening. In most instances, the spells do not cause an awakening and are noticed by the bed partner or household members who report them.[55] In most cases, the episodes are sporadic, self-limited, and brief, usually lasting less than 1 minute. Sometimes the events may occur nightly in clusters, particularly when the subject is under emotional stress. Episodes may range from speaking a single word or a sentence to a long elaborated dialogue. Content may be either meaningful or nonsense. Sleeptalkers seem to be either talking to themselves or having a dialogue. The voice can be the same or different as in the waking state. Sometimes sleeptalking may be accompanied by other sounds and vocalizations, such as whispering, muttering, humming, moaning, weeping, giggling, and shouting. It is rarely accompanied by crying, laughing, and singing.[55] In a study performed in the Basque country, a region in northern Spain in which two completely different official languages are spoken (Euskera and Spanish), most children used their dominant (native) language during sleeptalking, and less than 4% used their nondominant language.[56]

Epidemiology. Sleeptalking is very common in the general population. About half of children present somniloquy at least once a year, but less than 10% present it every day.[56,57] Sleeptalking in children decreases with time and is slightly more common in boys (53%) than in girls (47%).[43]

Natural Course. Onset is usually between the ages of 3 and 10 years, but it may start in adolescence and even in early adulthood. The course is variable because it may be present for a few days only, recur in clusters, or last for several months or many years. In children, sleeptalking usually resolves spontaneously during adolescence or adulthood.

Polysomnography. There are few formal studies assessing isolated sleeptalking with polysomnography. These have shown that sleeptalking occurs in all sleep stages; 50% to 60% of cases are associated with N1 and N2 sleep, 20% to 25% with N3 sleep, and 20% to 25% with REM sleep.[58] Some subjects only experience sleeptalking during NREM sleep, others exclusively in REM sleep, and the majority in both NREM and REM sleep.[58] When a person is experiencing sleeptalking and is experimentally awakened, dream recall occurs in 79% of the episodes arising from REM sleep, in 46% of the events arising from N2 sleep, and in 20% of the spells arising from N3 sleep. Sleeptalking episodes related to REM sleep are longer and clearer than those occurring in N2 and N3 sleep.[58] In REM sleep, episodes of sleeptalking are characterized by sustained alpha electroencephalographic trains.[59]

Associations. Most cases of sleeptalking are not associated with a medical illness or psychopathologic conditions. Predisposing factors include anxiety, sleep deprivation, and fever. Children are more likely to experience recurrent sleeptalking if parents had sleep talking during childhood.[60]

Etiology. The etiology is unknown. Some cases may have a clear familial predisposition suggesting a genetic background.[61]

Management. Isolated sleeptalking is benign and an uncommon reason for consultation in a sleep center. However, it may become disturbing for both the sleeptalker and the bed partner if it is frequent and excessively long or loud and if the content includes obscenities and discusses intimate experiences. Isolated sleeptalking has no specific medical treatment, but adequate sleep hygiene and reduction of emotional stress may help.

Secondary Sleeptalking

Sleeptalking may occur as one of the clinical manifestations of several disorders (Table 105-2) that are covered in other chapters of this book.

Table 105-2 Conditions Associated with Secondary Sleeptalking

Confusional arousals[88]
Night terrors[88]
Sleepwalking[88]
Sleep-related eating syndrome[66]
Sexsomnia[63]
REM sleep behavior disorder[74,89-91]
Status dissociatus[68]
Parasomnia overlap disorder[69]
Agrypnia excitata[92]
IgLON5 parasomnia[73]
Nocturnal frontal lobe epilepsy[1]
Periodic limb movement disorder[93]
Obstructive sleep apnea[94]
Nocturnal panic attacks[1]
Sleep-related dissociative disorder[1]

PARASOMNIA DUE TO A MEDICAL DISORDER

Parasomnia due to a medical disorder consists on the presence of a parasomnia that is attributable to an underlying neurologic disorder or medical condition.[1] Secondary NREM sleep and REM sleep parasomnias are covered in Chapters 102 and 103, respectively.

NREM Sleep Parasomnias

Confusional arousals, night terrors, and sleepwalking are idiopathic disorders of arousal that are not linked to any specific psychiatric, psychological, neurologic, or medical disorder. Patients with disorders of arousal do not develop a neurodegenerative disease with time. Episodes of disorders of arousal can be precipitated by sleep deprivation, night shifts, fever, alcohol, noise, touch, psychological stress, and anxiety.[62] Coexistent obstructive sleep apnea and periodic leg movements in sleep may provoke partial arousals, resulting in episodes of disorders of arousal and sleepwalking variants (sleep-related eating disorder, sleep driving, and sexsomnia) in predisposed individuals.

Sleep-related eating disorder is associated with sleepwalking, sleep smoking, sexsomnia,[63] narcolepsy with cataplexy,[64] restless legs syndrome,[65] nocturnal eating syndrome,[65] and depression.[66,67]

Combined NREM and REM Sleep Parasomnias

Parasomnia Overlap Disorder

Parasomnia overlap disorder is the coexistence of a disorder of arousal (NREM parasomnia) and RBD (REM parasomnia) in the same patient. This entity mainly affects young subjects in whom NREM sleep clinical and polysomnographic features of the parasomnia predominate over RBD. Two thirds of these patients have an idiopathic form,[68-70] and the remaining third have a form linked to several heterogeneous disorders, such as Mobius syndrome,[68] narcolepsy,[68] multiple sclerosis,[68] removal of fourth ventricle astrocytoma,[68] head trauma,[68] atrial fibrillation during sleep,[68] posttraumatic stress disorder,[68] ethanol abuse,[68] harlequin syndrome,[71] and brainstem structural lesions.[72] On the other hand, the link between NREM sleep parasomnia and RBD can be by chance in an older subject who has had a disorder of arousal since childhood that persists and who decades later develops RBD. In some cases, this may explain the fact that a few subjects with idiopathic RBD who are older than 60 years report that their abnormal behaviors started 30 to 50 years earlier. In more severe cases of overlap disorder, patients have degradation of sleep and wake features and behaviors to point of the being indistinguishable in status dissociates.

IgLON5 Parasomnia

The recently described IgLON5 parasomnia[73] is a novel neurologic disorder initially described in eight unrelated adults in whom subacute (two cases) and chronic (six cases) presentations were noted. Patients presented clinically with witnessed apneic events, stridor, abnormal sleep behaviors, and additional waking neurologic symptoms such as gait instability, dysarthria, dysphagia, chorea, and mild dysautonomia. Patients had no previous history of disorders of arousal, they denied insomnia, and they were unaware of their sleep behaviors that were only noted by the bed partner. Video polysomnography showed a complex, distinct pattern characterized by (1) normal occipital alpha rhythm during wakefulness; (2) mild to moderate reduction of total sleep time and sleep efficiency; (3) a distinctive temporal sequence of sleep stages and behaviors taking place, from most abnormal at the beginning of the night to more normal at the end; (4) initiation of sleep and reentering of sleep after awakenings characterized by prolonged periods of theta activity with motor activation and rapid repetitive leg movements that do not fit criteria periodic leg movements in sleep; (5) reduced amount of normal N2 sleep; (6) periods of diffuse delta activity, typical of normal N3 sleep, mixed with spindles; (7) poorly structured stage N2 sleep characterized by clear spindles and K-complexes with frequent vocalizations (e.g., talking, laughing, crying) and simple (e.g., raising the arm, punching) and finalistic (e.g., goal-directed behaviors like sucking the thumbs while apparently eating, manipulating wires) behaviors; (8) RBD characterized by limb and body jerks but no violent or finalistic behaviors; and (9) obstructive sleep apnea and inspiratory stridor secondary to vocal cord palsy, particularly intense during normal N3 stage (Figures 105-1 and 105-2). Longitudinal follow-up by polysomnography showed no deterioration of these features with time.

Autoantibodies against IgLON5, a neuronal cell adhesion protein, were identified in all patients. The haplotypes DQB1*0501 and DRB1*1001 were detected in all four patients tested. Most of the patients died suddenly during wakefulness or in sleep. Neuropathology performed in two patients showed a tauopathy mainly involving the tegmentum of the brainstem and the hypothalamus, making the underlying pathophysiologic process unclear. To date (September 2015) two additional cases with the same HLA haplotype have been published where insomnia, hypersomnia, chorea, and dementia were part of the clinical spectrum.[73a,73b]

REM Sleep Parasomnias

As pointed out in the ICSD3, when diagnostic criteria for RBD are met, the more specific diagnosis of RBD should be

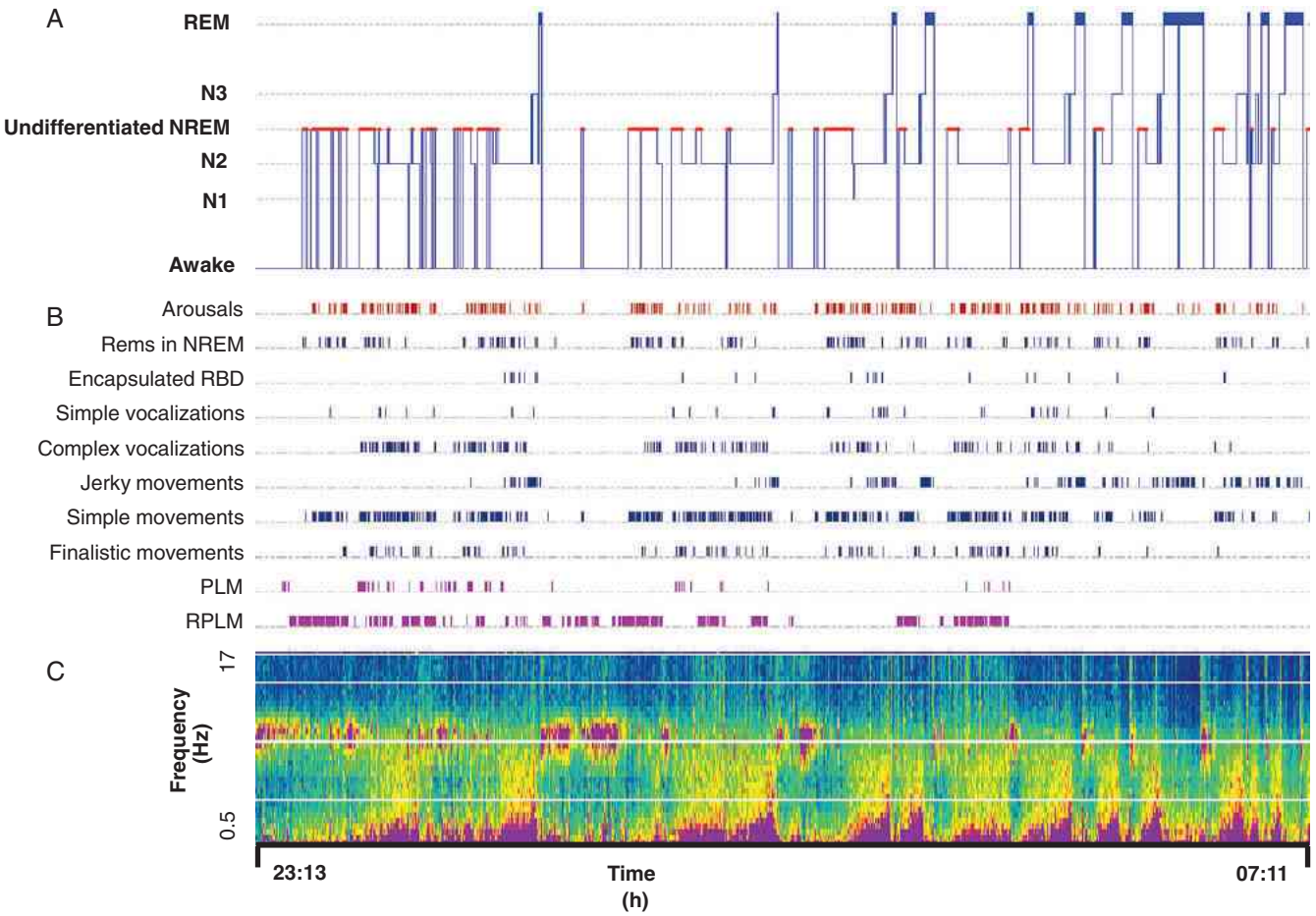


Figure 105-1 Sleep recording with continuous positive airway pressure in a patient with IgLON5 parasomnia. **A,** Hypnogram. **B,** Arousals, dissociations, and abnormal movements. **C,** Density spectral array showing the power spectrum of electroencephalographic frequencies (0–17 Hz) in electrode C3 referenced to electrode O2. *Warmer colors* indicate more dominant frequencies. PLM, Periodic limb movements; REMs, rapid eye movements; RBD, REM sleep behavior disorder; RPLM, rapid periodic leg movements.

made, as with the other parasomnias.[1] In some instances, RBD may be an important clinical feature, and in others it is not clinically significant and is overlooked by other symptoms and signs (e.g., dementia, confusion, seizures). RBD is frequent among 25% to 58% patients with Parkinson disease,[74-78] 70% to 80% with dementia with Lewy bodies,[74,75,79] and 90% to 100% with multiple-system atrophy[74,75,80] as well as other neurodegenerative diseases, autoimmune diseases like narcolepsy, and structural brainstem lesions (see Chapter 103).[74]

PARASOMNIA DUE TO A MEDICATION OR SUBSTANCE

The assumption of drug-induced parasomnia is based on the temporal association between the introduction of the drug and the onset of the abnormal sleep behaviors and their cessation after the drug is stopped. The emergent parasomnia can be de novo parasomnia, the aggravation of a chronic intermittent parasomnia, or the reactivation of a previous parasomnia.[1]

NREM Sleep Parasomnias

The most common drugs inducing NREM sleep parasomnias are zolpidem (used at therapeutic doses for insomnia)[81] and less frequently sodium oxybate (at high doses within the normal range for narcolepsy).[82] Both medications can cause sleepwalking, sleep-related eating disorder, sleep driving, and sleep sex.[81,82] There is no strong evidence that alcohol causes sleepwalking or other NREM parasomnias, but alcohol could increase sleep apnea evoking a parasomnia event.[1]

REM Sleep Parasomnias

Antidepressants, including tricyclics, selective serotonin reuptake inhibitors, and selective noradrenaline reuptake inhibitors, have been described to trigger or aggravate RBD. They include clomipramine, imipramine, nortriptyline, mirtazapine, fluoxetine, venlafaxine, paroxetine, sertraline, citalopram, and escitalopram.[74,83,84]

The introduction of lipophilic beta blockers such as bisoprolol can also induce RBD parasomnias.[85] There are also reports of RBD induced by withdrawal from meprobamate and alcohol.[86] Suvorexant may induce sleep paralysis.[87]

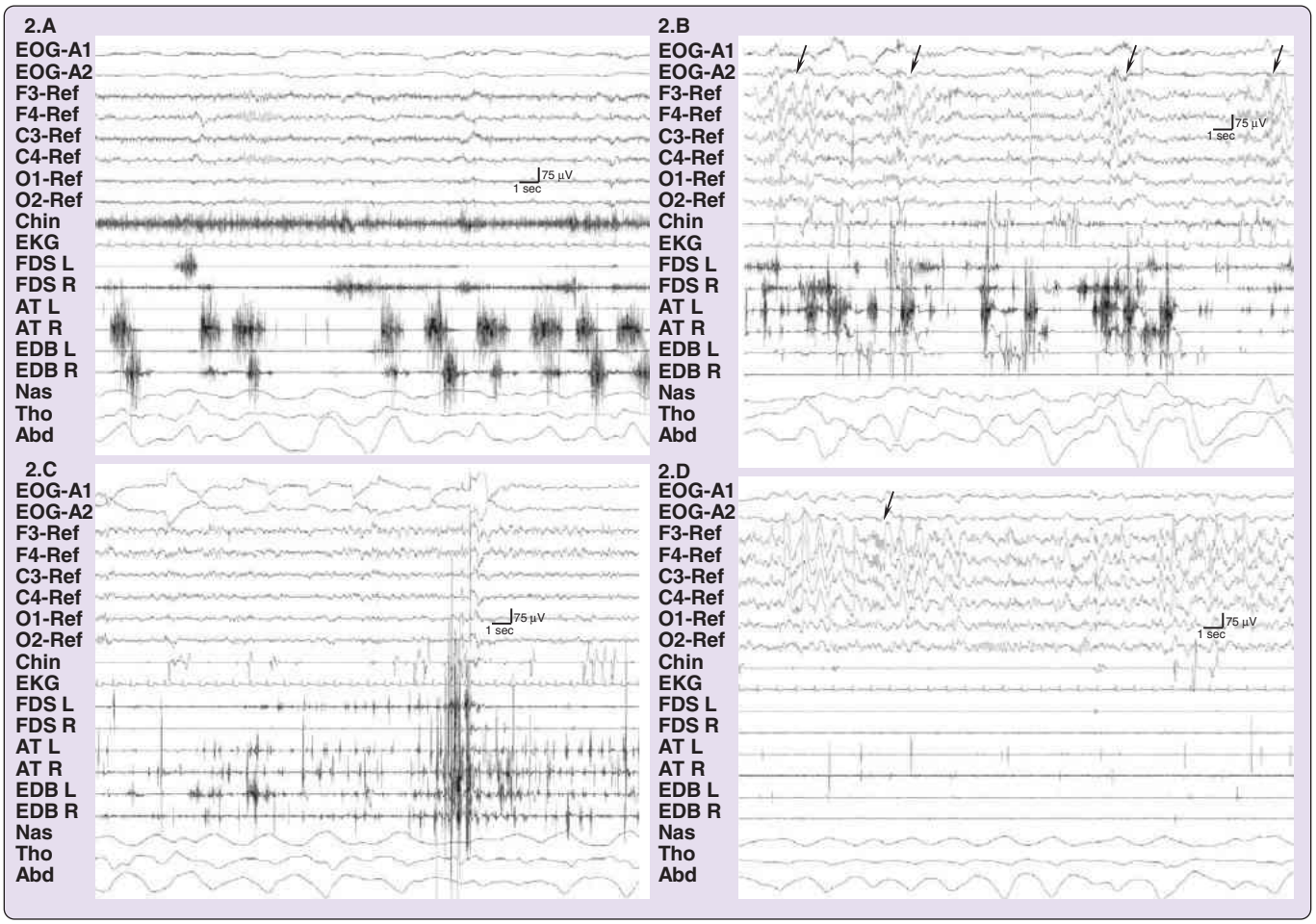


Figure 105-2 Polysomnographic epochs illustrative of each sleep state in a patient with IgLON5 parasomnia. **A,** Sleep onset characterized by undifferentiated NREM sleep with diffuse theta activity and rapid periodic leg movements particularly prominent at the right anterior tibialis electromyographic channel. **B,** N2 sleep with chains of two or three consecutive K-complexes (*arrows*) with frequent muscular phasic activity in electromyogram (EMG) surface of the limbs that correlate with finalistic movements. **C,** REM sleep with typical rapid eye movements and electroencephalogram (EEG) features with excessive phasic muscular activity and limb jerks indicative of REM sleep behavior disorder. **D,** N3 with diffuse delta activity mixed with well-defined sleep spindles at 13 Hz (*arrows*) without vocalizations nor movements. ABD, Abdominal respiratory movement; AT, anterior tibialis left (L) and right (R); Chin, electromyography of mentalis muscle; EDB, extensor digitorum brevis muscle left (L) and right (R); EKG, electrocardiogram; EOG, electrooculogram; FDS, flexor digitorum superficialis muscle left (L) and right (R); NAS, nasal air flow; THO, thoracic respiratory movement. Note the calibration mark for time/EEG voltage.

CLINICAL PEARL

When facing a patient who seeks medical advice for abnormal behaviors or experiences during sleep, clinicians should be aware that, in addition to NREM sleep parasomnias, REM sleep parasomnias, and sleep-related movement disorders, there still remains a group of other parasomnias that include sleep-related hallucinations, exploding head syndrome, sleep enuresis, and sleeptalking.

SUMMARY

This chapter describes a miscellaneous group of parasomnias that are included in the ICSD3 under the category "other parasomnias." Sleep-related hallucinations are experiences perceived in the wake-sleep transition or on awakening during the night or in the morning. Hypnagogic and hypnopompic hallucinations are common in the general population and in narcolepsy, whereas complex nocturnal visual hallucinations are usually linked to an underlying pathologic condition such as Parkinson disease, brainstem structural lesions, and severe vision loss. EHS is possibly a type of auditory hallucination characterized by a benign sense of a loud explosion in the head that awakens the individual. SE is characterized by recurrent involuntary voiding during sleep in subjects older than 5 years. SE is usually idiopathic and transient, but other conditions such as obstructive sleep apnea and genitourinary tract malformations may be associated. Isolated sleeptalking is a very common normal sleep variant that consists of talking while the individual is asleep. However, sleeptalking also occurs in the setting of the NREM- and REM-related parasomnias. RBD is associated with neurodegenerative diseases, autoimmune diseases, structural brainstem lesions, and medications such as antidepressants. All types of NREM-related

parasomnias may be induced by medications, particularly by zolpidem. The recently described IgLON5 parasomnia is characterized by obstructive sleep apnea, abnormal sleep architecture, and abnormal behaviors during both NREM and REM sleep in subjects with a brainstem and hypothalamic tauopathy and antibodies against the protein IgLON5.

Selected Readings

American Academy of Sleep Medicine. *International classification of sleep disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.

Arkin AM, Toth MF, Baker J, Hastey MS. The frequency of sleep talking in the laboratory among chronic sleep talkers and good dream recallers. *J Nerv Ment Dis* 1970;**151**:369–74.

Högl B, Heidbreder A, Santamaria J, et al. IgLON5 autoimmunity and abnormal behaviours during sleep. *Lancet* 2015;**385**(9977):1590.

Leu-Semenescu S, De Cock VC, Le Masson VD, et al. Hallucinations in narcolepsy with and without cataplexy: contrasts with Parkinson disease. *Sleep Med* 2011;**12**:497–504.

Manford M, Andermann F. Complex visual hallucinations. *Brain* 1998; **121**:1819–40.

Sabater L, Gaig C, Gelpi E, et al. A novel non-rapid-eye movement and rapid-eye-movement parasomnia with sleep breathing disorder associated with antibodies to IgLON5: an observational study. A case series, characterisation of the antigen, and postmortem study. *Lancet Neurol* 2014;**13**:575–86.

Sharpless BA. Exploding head syndrome. *Sleep Med Rev* 2014;**18**(6): 489–93.

Sharpless BA. Exploding head syndrome is common in college students. *J Sleep Res* 2015;**24**(4):447–9.

Simabukuro MM, Sabater L, Adoni T, et al. Sleep disorder, chorea, and dementia associated with IgLON5 antibodies. *Neurol Neuroimmunol Neuroinflamm* 2015;**2**(4):e136. [eCollection 2015 Aug].

Wen JG, Wang QW, Chen Y, et al. An epidemiological study of primary nocturnal enuresis in Chinese children and adolescents. *Eur Urol* 2006; **49**:1107–13.

A complete reference list can be found online at ExpertConsult.com.

Chapter
106

Sleep-Related Movement Disorders and Their Unique Motor Manifestations

Rachel E. Salas; Seema Gulyani; Anthony B. Kwan; Charlene E. Gamaldo

Chapter Highlights

- The *International Classification of Sleep Disorders*, 3rd edition, currently recognizes 10 sleep-related movement disorders and 4 isolated sleep-related movements that are generally characterized as simple and stereotyped movements primarily affecting sleep initiation, maintenance, or quality.
- To be classified as a disorder, these movements must be associated with a functional complaint (e.g., daytime sleepiness).
- Sleep clinicians should have a familiarity with all sleep-related movement disorders to distinguish them from other more clinically significant medical and neurodegenerative disorders with which they may share some overlapping clinical features.

A variety of simple and complex movements can occur during sleep. Many movements are considered part of the normal sleep-wake state or may be normal variants, whereas others may be more disturbing. The third edition of the *International Classification of Sleep Disorders* (ICSD3)[1] currently lists 10 sleep-related movement disorders (SRMDs) and 4 isolated sleep-related movements that are relatively simple, stereotyped movements disturbing to sleep or sleep onset.[1] SRMDs are important to recognize and to differentiate from normal physiologic movements during sleep and other more serious movements that occur during sleep in medical disorders, such as nocturnal seizures and parasomnias. The presence of significant clinical consequences, such as interference with sleep quality, daytime dysfunction, or body injury requiring medical treatment, is what differentiates sleep-related rhythmic movement disorder from normal sleep-related movements.[2] In the case of sleep-related rhythmic movement disorder, timely recognition is necessary because it often warrants further evaluation or treatment. This chapter provides a discussion of the movements commonly observed during sleep (Figure 106-1), focusing primarily on the sleep-related movements (both physiologic and pathologic) that are not specifically covered in designated chapters elsewhere within this textbook. These movements of sleep are divided into three groups: (1) normal physiologic movements typically observed in sleep, (2) isolated sleep-related movements and normal variants, and (3) SRMDs. Normal physiologic movements typically observed in sleep, movements classified as isolated sleep-related movements, and normal sleep-related movement variants are important events to recognize as benign and not indicators of other possible pathologic processes. SRMDs include a range of simple and repetitive movements that may be associated with downstream clinical symptoms. Some features are useful clues to distinguish normal movements from more clinically significant events during sleep. Key elements in the history, such as a clear description of the events, specific region of the body, time of occurrence, age at onset, and findings on polysomnography (PSG), are helpful for determining the etiology.

NORMAL PHYSIOLOGIC MOVEMENTS OF SLEEP

Phasic Twitches During REM Sleep

Twitches during rapid eye movement (REM) sleep are sudden, brief physiologic movements. These twitches typically occur during the phasic periods of REM sleep and can usually be observed in the context of visible eye movements. The movements are seen in the leg electromyography leads and take on a short (usually <0.10 second) repetitive burst pattern that is superimposed on the suppressed "atonic" background electromyographic (EMG) baseline (Figure 106-2). These twitches usually involve small distal muscle groups and as a result are not usually manifested as an overt segmental body movement.

Major Body Movements

The American Academy of Sleep Medicine (AASM) scoring manual currently recognizes the phenomenon of major body movements as a normal manifestation of body movements due to individuals' moving or shifting position during a sleep period.[2] These movements can be scored as contiguous with the sleep stage that occurred before and after the major body movement epoch if the preceding epoch does not demonstrate overwhelming evidence of prolonged arousal or awakening (e.g., slow eye movements). If evidence suggests prolonged arousal or awakening within the epoch following the major body movement, the major body movement epoch is considered a wake state; if the body movement occurs contiguous

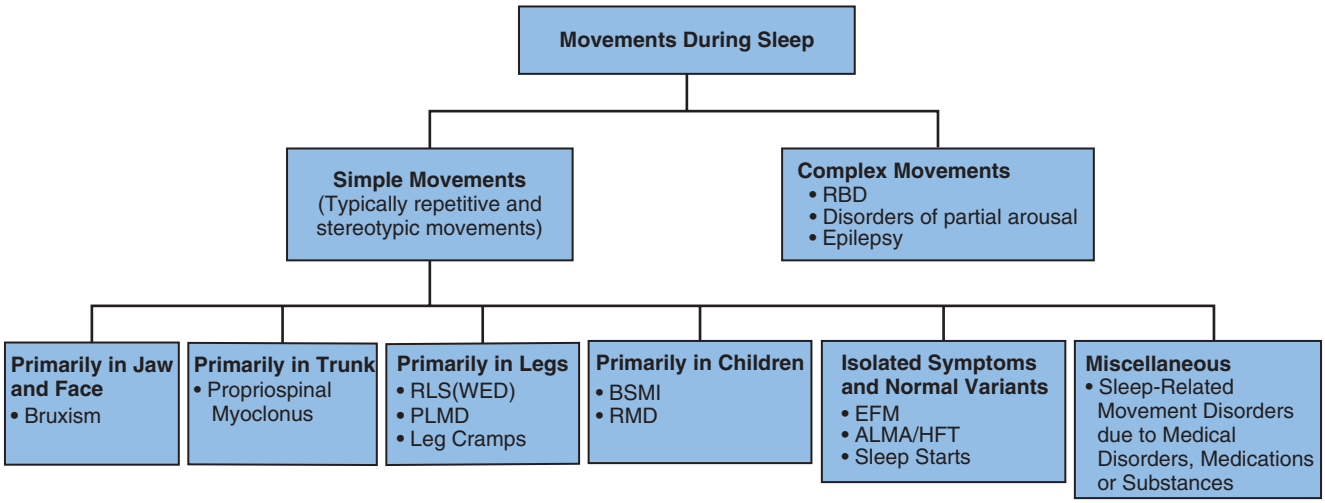


Figure 106-1 Flow chart for the approach to the differential diagnosis of sleep-related movement disorders. ALMA, Alternating leg muscle activation; BSMI, benign sleep myoclonus of infancy; EFM, excessive fragmentary myoclonus; HFT, hypnagogic foot tremor; PLMD, periodic limb movement disorder; RBD, rapid eye movement (REM) sleep behavior disorder; RLS (WED), restless legs syndrome (Willis-Ekbom disease); RMD, rhythmic movement disorder.

with two definitive wake epochs of wakefulness, it is also considered a wake epoch.

ISOLATED SLEEP-RELATED MOVEMENTS AND NORMAL VARIANTS

Periodic Limb Movements in Sleep

Polysomnography Scoring Criteria and Motor Features

Periodic limb movements in sleep (PLMS) in isolation (i.e., not associated with another sleep or medical disorder) are benign, sporadic episodes of involuntary leg movements that occur during sleep. PLMS typically involve extension of the big toe in combination with partial flexion of the ankle, the knee, and sometimes even the hip.[1] Although these movements predominantly occur in the legs, they may occur in the upper extremities as well. According to the AASM scoring criteria, PLMS must consist of a series of four or more consecutive movements, each lasting 0.5 to 10 seconds with intermovement intervals of 5 to 90 seconds and an amplitude greater than 8 mV above the resting baseline EMG signal (Figure 106-2). The PLMS index is defined as the number of PLMS divided by the observation time in hours. In general, a PLMS index is considered abnormal if it is more than 5 per hour in children and more than 15 per hour in adults.[1] Five percent to 6% of all asymptomatic adults have been reported to demonstrate PLMS on PSG, and in these cases they are considered benign epiphenomena. For this reason, PLMS is included in this chapter as a clinical consideration under the category of a normal variant. Figure 106-2 provides unique physiologic features to distinguish PLMS from other isolated sleep movements and normal variants that can be mistaken for PLMS on PSG, and Table 106-1 offers some of the pathological associations with periodic limb movements. In the context of coexisting clinical or functional symptoms, further clinical evaluation is recommended. The reader is referred to Chapter 95 for discussion of PLMS as it relates to periodic limb movement disorder, restless legs syndrome, and other sleep disorders.

Excessive Fragmentary Myoclonus

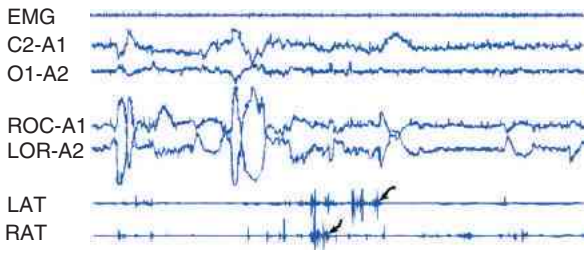
Excessive fragmentary myoclonus (EFM) is characterized by very small (often nonvisible to the blind eye) movements of the corners of the mouth, fingers, or toes. EFM is usually noted as an incidental EMG finding and to date has not been associated with any negative clinical sequelae. EFM is usually seen during non-REM (NREM) sleep. Morphologically, the EMG bursts seen in EFM occur for at least 20 minutes during NREM sleep with at least 5 bursts per minute, each with a duration of approximately 150 milliseconds (see Figure 106-2).[2] EFM activity typically occurs without evidence of alterations in the electroencephalogram (suggesting an associated arousal). However, in the case of relatively high altitude EFM bursts, associated K-complexes or shifts to faster electroencephalographic (EEG) frequencies may be observed.[3] EFM predominantly occurs in adult men. On rare occasions, EFM may result in sleep fragmentation and may be associated with obstructive sleep apnea (OSA), central sleep apnea, sleep-related hypoxemic/hypoventilation syndromes, narcolepsy, periodic limb movement disorder, and various causes of insomnia.[1]

Alternating Leg Muscle Activation

Alternating leg muscle activation (ALMA) was first coined by Chervin et al[4] to describe alternating EMG bursts in the legs during sleep. Much shorter in duration than PLMS, ALMA is defined by a duration of 100 to 500 milliseconds with a frequency range for the EMG bursts that alternate between the legs of 0.5 to 3.0 Hz (see Figure 106-2). A single ALMA episode represents at least four movements that alternate between the legs. Although the pathophysiologic mechanism of ALMA is currently unknown, researchers suspect a relation between ALMA and serotoninergic and dopaminergic pathways in the spinal network. Although ALMA does not require treatment in most cases, a sleep specialist may consider treatment in select patients who also exhibit sleep disruption. In fact, treatment of ALMA with pramipexole, a

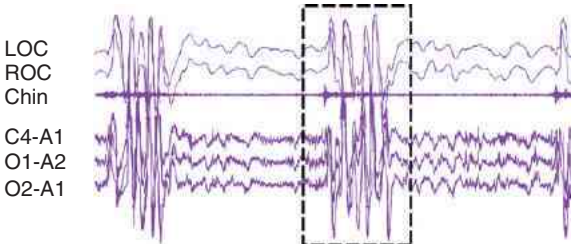
Disorders with their Distinct Features

REM twitches



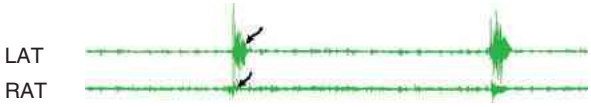
- Occur during phasic periods of REM sleep and often seen along with visible eye movements
- Leg movements are typically short (usually <.10 sec) repetitive burst pattern superimposed on the suppressed "atonic" background EMG baseline
- Twitches usually involve small distal muscle groups and as a result do not usually manifest as an overt segmental body movement

RMD



- Repetitive, stereotypes, and rhythmic motor behaviors involving large muscle groups (e.g. body rocking or head banging)

PLMS

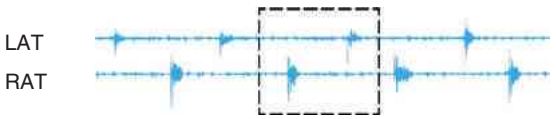


- Typically involve extension of the big toe, along with partial flexion of the ankle, the knee and rarely hip
- Consist of a series of > 4 movements, 0.5 - 10 seconds and 5 to 90 seconds between movements and an amplitude > 8 mV above the resting baseline EMG signal

EFM

- Very small to no visible movement (e.g. minor movement or twitching of the fingers, toes or corners of the mouth)
- Persists in all stages of sleep (at least 20 mins of NREM sleep with EFM must be recorded)

ALMA

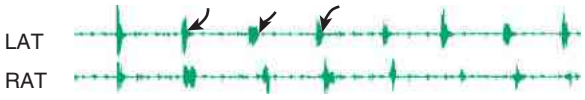


- Alternating contractions between legs
- Usual range of duration is 100-500 msec

HFT

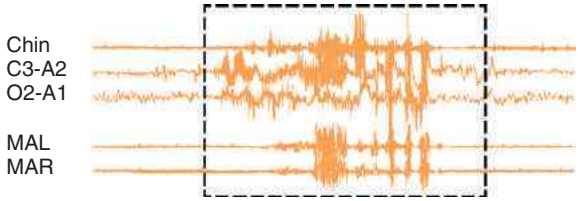
- Rhythmic foot movements in one or both feet
- Occurs at the transition between wake and sleep or during light sleep (N1 and N2)
- Usual range of duration is 250-1000 msec

Leg Cramps



- Non-period bursts of gastrocnemius EMG activity rarely associated with any other physiological changes during sleep (i.e. arousal heart rate increase etc.)
- Can be observed in all sleep stages and wake

Bruxism



- May consist of either brief/phasic or sustained/tonic elevations of EMG activity

Denotes an individual movement Highlights a movement cluster

Figure 106-2 Common Movements in Sleep. Distinct clinical and polysomnographic features of rapid eye movement (REM) twitches, rhythmic movement disorder (RMD), periodic limb movements in sleep (PLMS), excessive fragmentary myoclonus (EFM), alternating leg muscle activation (ALMA), hypnagogic foot tremor (HFT), leg cramps, and bruxism. A, Auricle (as in C4-A1); C, central (as in C4-A1); EMG, Electromyography; LAT, left anterior tibialis; LOC, left outer canthus; MAL, masseter muscle left; MAR, masseter muscle right; NREM, non–rapid eye movement; O, occipital (as in O1-A2); RAT, right anterior tibialis; ROC, right outer canthus.

Table 106-1 Occurrence of Periodic Limb Movements in Sleep due to Medical Disorders or Medication or Substance

Sleep Disorders	Medical Disorders Not Primarily Affecting Sleep	Medications
Restless legs syndrome (Willis Ekbom Disease)	Severe congestive heart failure	Clomipramine
Periodic limb movement disorder	Hypertension	Lithium
Narcolepsy	Tourette syndrome	Fluoxetine
Obstructive sleep apnea	End-stage renal disease	Venlafaxine
REM sleep behavior disorder	Syringomyelia Juvenile fibromyalgia Systemic sclerosis (scleroderma) Spinal cord injury Morbus Parkinson Multiple system atrophy Posttraumatic stress disorder	Selective serotonin reuptake inhibitors Bupropion Doxepin

dopamine agonist, may be considered because it has been shown to significantly reduce the occurrence of ALMA, insomnia, and daytime sleepiness.[5]

Hypnagogic Foot Tremor

Hypnagogic foot tremor (HFT) is categorized as a rhythmic movement of the feet or toes that occurs at the transition between wake and sleep or during light sleep.[6] To be designated an HFT, at least four EMG bursts, 250 to 1000 milliseconds in duration, must occur in a frequency range series between 0.3 and 4.0 Hz (see Figure 106-2). HFT appears as trains of recurrent EMG potentials or movements in either one or both legs, whereas ALMA appears as a brief activation in the anterior tibialis of one leg alternating with the other leg. PLMS, in contrast, have a longer interval between the movements. Whereas PLMS have a minimum interval of 5 seconds, the interval is 0.25 second for HFT and 0.3 second for ALMA. Currently, no clinical consequences of HFT have been reported; thus, treatment outside of patient education on the diagnosis and the benign nature of this movement disorder during sleep is typically all that is warranted. HFT appears to be a fairly common disorder. In 2001, Wichniak et al[6] looked at the prevalence of this disorder in 375 consecutive patients with sleep disorders and found a prevalence of 7.5%.

Sleep Starts (Hypnic Jerks)

Sleep starts, also known as hypnic jerks, are sudden, brief, simultaneous contractions of the body or one or more body segments occurring at sleep onset. Sleep starts occur at the wake-sleep transition and are usually not repetitive.[1] Sleep starts (hypnic jerks) usually consist of a single contraction that often affects the body asymmetrically. The jerks may be either spontaneous or induced by stimuli. PSG recordings show that hypnic jerks generally occur during transitions from wakefulness to sleep, mainly at the beginning of the sleep episode. Superficial EMG recordings of the involved muscles show brief (generally 75- to 250-milliseconds) high-amplitude potentials, either singly or in succession. The electroencephalogram typically shows drowsiness or stage N1 sleep during these movements. The motor activity is often associated with sensory (e.g., a feeling of falling, pain, tingling), auditory (e.g., banging, snapping, or crackling noises), or visual (e.g., flashing lights, hypnagogic hallucination) sensations. Interestingly, a sharp cry may even occur with the movement. The patient may not recall these often "dramatic" movements if they do not result in an awakening. When sleep starts are frequent, intense, or repetitive, they could potentially result in insomnia. The prevalence for sleep starts may be as high as 60% to 70%.[1] Sleep starts affect all ages and both sexes. Patient education on the diagnosis and the benign nature of these phenomena should be pursued. Excessive caffeine or other stimulant intake, prior intense physical work or exercise, sleep deprivation, and emotional stress all can increase the frequency and severity of sleep starts and should be a point of discussion with patients presenting for evaluation.

SLEEP-RELATED MOVEMENT DISORDERS

SRMDs encompass a broad range of simple and typically stereotyped movements that are rarely associated with clinical dysfunction (e.g., poor sleep quality, nonrestorative sleep, fatigue). These phenomenologic features are helpful clues in distinguishing SRMDs from the more complex and clinically significant movements observed during sleep (e.g., in the case of parasomnias, which can involve behaviors and movements that may result in injuries, sleep disruption, adverse health effects, and unpleasant psychosocial effects). In this section, we discuss the following disorders: sleep-related leg cramps, bruxism, sleep-related rhythmic movement disorder, benign sleep myoclonus of infancy, and propriospinal myoclonus at sleep onset. These are all conditions currently classified in the ICSD3 under the SRMD category. The diagnoses of SRMD due to medical disorder, SRMD due to medication or substance, and SRMD unspecified are not covered in detail in this chapter. Differentiating between the SRMDs can be accomplished by recognizing features that are unique to their respective presentations, including the specific body regions affected, age at onset, and unique polysomnographic features. Detailed discussion of these characteristic features is presented in the following section.

Sleep-Related Leg Cramps

Clinical Features

Sleep-related leg cramps are painful sensations usually in the calf or small muscles of the foot and are caused by sudden and intense involuntary contractions during which there is muscle

spasm and hardness for several seconds.[1] In addition, the diagnosis requires that the painful muscle contractions occur during the sleep period and that the pain is relieved by forcefully stretching the affected muscle. Although the sleep-related leg cramps normally begin suddenly, they may also begin slowly with less painful warning signs.[7] These cramps can last for a few seconds up to several minutes. They may also occur as rarely as once per year in some individuals, whereas others can have many cramps every night, potentially resulting in insomnia. Individuals with sleep-related leg cramps report, in general, more sleep disturbances, snoring, less adequate sleep, and excessive daytime sleepiness.[8] In addition, individuals who experience leg cramps tend to have a lower quality of life.[8] Although generally considered to be an idiopathic condition, sleep-related leg cramps have been reported to be associated with vascular disease, lumbar canal stenosis, cirrhosis, hemodialysis, pregnancy, neuromuscular disorders, and other medical conditions (e.g., metabolic disorders). Moreover, certain medications may also result in sleep-related leg cramps.

Differential Diagnosis

Sleep-related leg cramps are often cited as a condition easily confused with the leg discomfort experienced in those individuals with restless legs syndrome (RLS; Willis Ekbom Disease [WED]). The leg discomfort of RLS, however, does not typically involve cramps that can be resolved by stretching. Because both disorders are characterized by leg discomfort and RLS patients may report cramping sensations, careful clinical history and examination are needed to distinguish between these disorders. The critical differentiating feature of sleep-related leg cramps is the actual spasm or hardening of the muscle that occurs during the leg pain, which is not seen in RLS. Furthermore, leg cramp events tend to be much briefer compared with the typical symptoms of RLS, which can persist for hours. Nocturnal leg cramps can also be confused with epilepsy, particularly in infants. Features to help distinguish between the two include the fact that sleep-related cramps last less than a second, whereas epileptic movement is more rhythmic and lasts longer and is typically associated with other features suggestive of seizure. Sleep-related leg cramps can also be confused with various other disorders, such as chronic myelopathy, peripheral neuropathy, and muscular pain fasciculation syndrome, so it is important to distinguish sleep-related leg cramps by clinical history and physical examination. Unlike nocturnal cramps, peripheral neuropathy tends to be associated with sensory or motor findings.[9] Focal dystonias of the feet can be distinguished electrophysiologically from leg cramps by demonstration of ongoing co-contraction of agonist and antagonist muscles. Whereas the history and physical examination are necessary in distinguishing sleep-related leg cramps from other medical conditions, laboratory studies and other more specialized testing (e.g., electromyography, nerve conduction studies) are typically unnecessary to make the diagnosis.

Polysomnography Scoring Criteria and Motor Features

Individuals who experience chronic sleep-related leg cramps often also report insomnia. Sleep-related leg cramps reveal nonperiodic bursts of gastrocnemius EMG activity that arise without any specific preceding physiologic changes during sleep.[1] Sleep-related leg cramps may also be observed throughout all sleep stages, which further demonstrates that they cannot be attributed to physiologic changes during particular stages.

Pathophysiology

The pathophysiologic mechanism of sleep-related leg cramps is still uncertain. These muscle cramps occur when sustained recruitment of motor units causes the leg muscles to contract. Based on electrophysiologic recordings, leg muscle contractions are associated with high-frequency and high-voltage discharges after spontaneous firing of anterior horn cells, which suggests abnormal excitability in the spinal cord.[9] Other mechanisms that may contribute to leg cramps include spinal disinhibition, abnormal terminal motor nerve excitability, and enhanced muscle contraction propagation through cross-activation of adjacent neurons.[10] Pain experienced in leg cramps may result from local ischemia or from local metabolite accumulations.

Prevalence and At-Risk Groups

Although these nocturnal leg cramps generally occur in older adults, leg cramp disturbances may occur at any age. Sleep-related leg cramps occur in about 7% of children and adolescents (typically not occurring before the age of 8 years), 33% of adults older than 60 years, and 50% of adults older than 80 years, with both older groups reporting a symptom frequency of at least once every 2 months. Pregnancy has also been associated with sleep-related leg cramps; about 33% to 50% of pregnant women experience leg cramps that also tend to become worse as pregnancy progresses but tend to go away after delivery. Vigorous exercise, use of certain medications (e.g., naproxen, intravenous iron sucrose, conjugated estrogens, and teriparatide), dehydration, fluid and electrolyte disturbances, and disorders that reduce mobilization in individuals can trigger nocturnal cramps. In addition, individuals with certain medical disorders, such as diabetes, sleep apnea,[11] blood vessel disease, metabolic disorders, and nerve or muscle diseases, may be more likely to have sleep-related cramps.

Treatment

Treatment of leg cramps includes frequent stretching or massaging of the affected muscle, application of heat, and movement of the affected limbs. In one report, quinine was considered to be an effective option, but the benefits may be modest (about 20% to 25% reduction in the number of cramps).[12] In another study, quinine (300 to 500 mg) did show clinically significant improvement of leg cramps.[13] Limitations of quinine also relate to severe side effects (e.g., thrombocytopenia and cardiac arrhythmias) that, for most clinicians, are thought to outweigh the potential benefits.[12] Although diltiazem has been shown to be effective, supporting data are limited.[12] Anticonvulsants, such as gabapentin and levetiracetam, have not been adequately assessed. Studies have also suggested that vitamin E may be an effective treatment. Recently published case series reported almost complete resolution of leg cramps with the use of continuous positive airway pressure in patients with moderate to severe sleep apnea.[11] Magnesium supplements are marketed for the prophylaxis of cramps, but systematic review reported that it is unlikely that magnesium supplementation provides clinically meaningful cramp prophylaxis to older adults experiencing skeletal muscle cramps.[14]

Sleep-Related Bruxism

Clinical Features

The repetitive jaw muscle activity characterized by clenching or grinding of the teeth and by bracing or thrusting of the mandible is referred to as bruxism, the Greek word for "gnashing of teeth." The clinical diagnosis of sleep-related bruxism, which is described in much more detail in Chapter 144, can be made when an individual demonstrates or reports regular or frequent tooth grinding sounds during sleep and one or more of the following: abnormal tooth wear consistent with the reports of tooth grinding during sleep, transient morning jaw muscle pain or fatigue, temporal headache, and jaw locking on awakening consistent with reports of tooth grinding during sleep.[1] The intensity and duration of sleep-related bruxism are variable, but it often occurs hundreds of times during a sleep period. Although there is likely to be a relationship between bruxism and stress or anxiety, this has not been adequately supported.

Prevalence and At-Risk Groups

Sleep-related bruxism tends to occur in families; approximately 20% to 50% of affected individuals have at least one direct family member with a history of tooth grinding, and childhood sleep-related bruxism appears to persist into adulthood in two thirds of reported cases. Despite this, no genetic variants or genetic inheritance patterns have been identified with sleep-related bruxism. Bruxism is more common during childhood. Prevalence is approximately 8% in middle-aged adults and 3% in older adults, which may be explained by edentulism, use of dentures, and changes in sleeping behaviors (i.e., no bed partner) likely influencing reporting.[1] There is no gender difference for sleep-related bruxism. Curiously, individuals who are highly motivated or characteristically maintain high vigilance may have an increased prevalence of sleep-related bruxism.[1] Precipitating factors can include anxiety related to current life events, tasks requiring high levels of performance, and repetitive tasks with short deadlines.[1] The use of cigarettes or caffeine in the hours before sleep also can contribute to the occurrence of sleep-related bruxism (probably because of the increased arousals and sleep instability).

Differential Diagnosis

Sleep-related bruxism needs to be differentiated from other faciomandibular activities occurring during sleep, such as faciomandibular myoclonus, REM sleep behavior disorder, abnormal swallowing, gastroesophageal reflux, night terrors, confusional arousals, dyskinetic jaw movements persisting in sleep (dystonia, tremor, chorea, dyskinesia), and, rarely, sleep-related epilepsy.

Unpleasant muscle and tooth sensations, limitation of jaw movements, orofacial pain, temporal or tension headaches, tooth wear, fractured teeth, and buccal lacerations may result from excessive bruxism.[1] Headaches, frequently reported by both adults and children, typically involve the temporal regions and tend to be reported either in the morning (with sleep bruxism) or during the day (with wake bruxism). Sleep-related bruxism can be so loud that it can be disruptive to the affected person's bed partner and to those sleeping nearby.

Polysomnography Scoring Criteria and Motor Features

PSG is not required to make the diagnosis, but bruxism can be recorded with EMG activity on the PSG study.[2] PSG may be indicated to evaluate for associated respiratory disturbances, gastroesophageal reflux, REM sleep behavior disorder, night terrors, faciomandibular myoclonus, or epilepsy.[1] The muscle potential increase, as recorded on the polysomnogram, must be differentiated from simple body or head movements; myoclonus, which is periodic short muscle contractions during sleep; head banging, rhythmic movements of the head occasionally occurring at a frequency of about 1 Hz, which is within the frequency of bruxism; and other rhythmic oromandibular activities (e.g., sleep chewing automatism). To detect and to study sleep bruxism, standard PSG has to include additional EMG derivations; surface electrodes are placed over bilateral masseter and temporal muscles, sometimes even on the frontal muscles; audio-video recordings help confirm the nature of the sounds (e.g., grinding, snoring) and the type of movements (e.g., sigh, swallowing, coughing, myoclonus, body rocking). The electrographic picture of a bruxism episode can vary from a sustained tonic contraction to a phasic burst, with an increase in muscle potential (more than 20% of the maximal voluntary activity awake) lasting 0.5 second or more with an intermovement interval of more than 3 seconds.[15] These increases in muscle potential, to be clearly identified as bruxism, have to be correlated to the loud "grinding sounds" as noted online by a vigilant technician or recorded for later analysis. There are no standard criteria for reports of bruxism episodes; usually indices are used to give the total number of episodes per hour of sleep.

There are distinct morphologic features of bruxism on PSG that include brief (phasic) or sustained (tonic) elevations of chin EMG activity that are at least twice the amplitude of background EMG activity. Phasic elevations occur in a regular sequence at least three times with each elevation and about 0.25 to 2 seconds in duration (Figure 106-2). For tonic elevations, EMG activity must occur for more than 2 seconds to be scored as bruxism. New episodes of bruxism are scored after a period of at least 3 seconds of stable background chin EMG activity has occurred. The sensitivity of PSG in detecting sleep-related bruxism in severe cases is moderate to high, whereas it is low in mild cases because of the night-to-night variability. Interestingly, bruxism has been found to be associated with cyclic alternating pattern[16-18] preceding EEG activity in the alpha range as well as tachycardia.[19,20] Ambulatory home monitoring may be used for screening, diagnosis, and treatment outcome assessment by studying the individual in his or her native environment, but it is characterized by lower diagnostic specificity.[1]

Pathophysiology

Sleep-related bruxism can be primary or idiopathic when there is no identified cause, whereas secondary sleep-related bruxism may result from the use of medications or recreational drugs or a variety of medical disorders (e.g., Parkinson disease, REM sleep behavior disorder, Down syndrome).[1] Interestingly, catecholamines (e.g., norepinephrine and dopamine) have been potentially implicated as having a role in the pathophysiologic process of bruxism.[21-23] Sleep-related bruxism may occur in those with cerebral palsy and mental retardation and

in adult patients with abnormal movements, such as oromandibular myoclonus/faciomandibular myoclonus, or with sleep-related breathing disorders, such as sleep apnea.[1]

Treatment

Sleep specialists should evaluate for secondary causes of bruxism, such as undiagnosed and untreated sleep apnea, epilepsy, tics, and medication-induced dyskinesias or dystonias. Bruxism typically involves a constellation of stereotypical somatic complaints, such as dental attrition, tooth pain, temporomandibular joint dysfunction, and headaches, whereas other disorders within the differential, such as sleep apnea, epilepsy, tics, and medication-induced dyskinesias, typically do not.[24] In addition, sleep apnea is often associated with snoring and witnessed apneas and excessive daytime sleepiness, whereas tics tend to occur during the day and not during sleep. Medication-induced dyskinesias and dystonias (e.g., oromandibular dystonia and orofacial dyskinesia) are temporally related to an onset associated with the initiation or use of particular medications. The utility of dental splints and mouth guards may be helpful in many cases in the appropriate patient. In severe cases, medications such as anxiolytics and muscle relaxants may be considered. Behavioral modifications can also be helpful. Consider referral to a dentist or sleep behavioral psychologist for a team-based approach in patients who appear to have severe bruxism or are experiencing reduced quality of life.

Sleep-Related Rhythmic Movement Disorder

Clinical Features

Sleep-related rhythmic movement disorder (RMD) is characterized by repetitive, stereotyped, and rhythmic motor behaviors. It is a clinical diagnosis based entirely on the patient's history and neurologic examination. Subtypes of RMD include body rocking, head banging, head rolling (side-to-side movements of the head), and a combined type involving two or more of the individual types. Whereas other movements, such as body rolling, leg banging, and leg rolling, may also be noted, movements such as hand banging are less common. Interestingly, different types of RMD may sometimes occur in the same individual during the same night.[25-27]

The episodes usually last less than 15 minutes, but the duration may vary from a few minutes to an hour, with a frequency of 0.5 to 2 per second. Rhythmic humming may be heard along with the movements.[1] The movements and the associated sounds may be quite loud. Environmental noise or distractions during the movements may lead to cessation of the event. Whereas RMD was thought to occur mostly during transitions between sleep and wakefulness, these movements may occur during any stage of sleep[28,29] and during quiet wakeful activities, such as listening to music or riding in a vehicle.[1] Children with RMD are otherwise normal but may experience insomnia, parasomnia (e.g., sleepwalking), and daytime sleepiness.[30] Significant daytime sleepiness may also be observed in adults.[31] RMD in infants and toddlers is typically not associated with significant risk of injury. Some patients may develop a soft tissue swelling over the forehead known as "head banger tumor." On rare occasions, enlargement of the diploic space in the parietal and occipital bones with loss of adjacent gray matter on magnetic resonance imaging in head bangers has been reported.[32]

In adults, RMD is more likely to be reported in association with other primary sleep disorders, such as OSA, narcolepsy, REM sleep behavior disorder, and attention-deficit/hyperactivity disorder, which may be of potential therapeutic benefit.[33] For instance, RMD such as body rocking, head banging, and head rolling may be noted in severe RLS.[34] Furthermore, RMD during sleep onset for an individual with RLS may even be used as a strategy to relieve internal dysesthesia symptoms of RLS. Interestingly, OSA-associated RMD often improves with positive airway pressure when patients are treated for OSA. Curiously, individuals with narcolepsy may demonstrate rhythmic movements that result in the termination of episodes of sleep paralysis.[35] These movements are usually more concerning and bothersome for the bed partner than for the patient. Patients with RMD may experience insomnia, but it is not clear whether sleep difficulties are evoked by the movements or simply accompany them.[12]

Differential Diagnosis

The diagnosis of RMD is one of exclusion. Diagnosis of RMD can typically be established by clinical history or video recordings provided by the patient. PSG is indicated when the clinical history alone is insufficient to provide diagnostic certainty or when the movements are atypical or particularly violent. Seizure is in the differential if movements are associated with other clinical evidence suggestive of a seizure (e.g., bowel or bladder incontinence, tongue biting, foaming at the mouth) or supported by the patient's history (i.e., seizure risk). In cases in which there is a concern for seizure, full electroencephalography should be considered. Other conditions to consider in the differential diagnosis for RMD include parasomnias such as REM sleep behavior disorder (which tend to be complex and "goal directed," but in the context of inappropriate behavior, whereas RMD is more monophasic and not goal directed). Other conditions in the differential diagnosis include sleep myoclonus (which occurs in all stages of NREM and REM sleep) and bruxism. RMD is specifically distinguished from tremor, another common neurologic disorder in the differential, as it is defined as a rhythmic and involuntary movement of any body part. The neurologic and medical history often elucidates the underlying cause of tremor, particularly when it is temporally associated with a specific event (e.g., stroke or head injury) or disease manifestation or progression (e.g., multiple sclerosis). On the other hand, RMD is classified as a disorder only if the behaviors interfere with normal sleep and cause impairment in daytime function.

Polysomnography Scoring Criteria and Motor Features

The AASM scoring manual currently defines RMD as a minimum of four stereotypical movements occurring in a cluster pattern at a frequency of 0.5 to 2.0 Hz (Figure 106-2). RMD can be witnessed during quiet wakeful activities or during sleep initiation, presumably as a potential self-soothing ritual. Video-PSG studies have shown that nearly half of all rhythmic movements occur during stages N1 or N2 exclusively, with approximately one third occurring throughout both NREM and REM sleep and one quarter occurring exclusively during REM sleep. Interestingly, the rhythmic movements exclusive to REM occur more frequently in adults.

Prevalence and At-Risk Groups

Although it is seen in adults, RMD is an SRMD primarily observed in children. At 9 months of age, 59% of all infants have been reported to exhibit one or more sleep-related rhythmic movements; the overall prevalence has been reported to decline to 33% by 18 months and to only 5% by 5 years.[1] Temper tantrums and attention-deficit disorder[36-39] and high levels of general distress and generalized anxiety disorder[40] may be associated with RMD. Attention-deficit/hyperactivity disorder, RLS, and sleep apnea[33,34,41,42] have also been associated with RMD. REM sleep behavior disorder has also been reported to be associated with RMD.[43,44]

Pathophysiology

The neurobiologic mechanisms underlying RMD remain uncertain. Because these movements are common in infants and young children, some believe that the cause is related to the soothing effect of vestibular stimulation by the rhythmic movements. The absence of organic causes in the majority of cases has led to behavioral and psychological theories[45] proposing that perhaps RMD is linked to fluctuations in arousal that are mediated through central motor pattern generators of the brainstem.

Treatment

There are no proven therapies for RMD. On the occasion when treatment is required (i.e., concerns about self-injury), safety precautions are typically emphasized (i.e., protective head gear), followed by the use of clonazepam and imipramine based on a limited amount of efficacy data.[46] Other treatments that have been used include antidepressants,[47,48] behavioral interventions,[49] hypnosis,[50] and sleep restriction.[51] RMD requires specific treatment only if risk of body injury or severe disruption to the sleep of a bed partner is present.[12] Subsequently, induced sleep disorders such as insomnia can be independently evaluated and addressed.

Benign Sleep Myoclonus of Infancy

Clinical Features

Benign sleep myoclonus of infancy (BSMI) is the myoclonic jerking movement of the elbow, fingers, toes, and face during sleep in infants. The diagnosis can be made when there is observation of repetitive myoclonic jerks involving the limbs, trunk, or whole body in an infant (e.g., birth to 6 months of age).[1] The movements tend to be bilateral and massive, involving large muscle groups, and can occur in the whole body or only in the limbs, the trunk, or rarely the face. Symptoms may be present for only a few days or may last for several months. These movements occur during sleep only and stop when the infant is aroused. Diagnosis of BSMI can be made if the infant has all the aforementioned criteria and symptoms cannot be better explained by the use of a medication or by another sleep, medical, or neurologic disorder.[1]

Differential Diagnosis

BSMI is considered benign and rare, but it is often confused with epilepsy, particularly myoclonic seizures and myoclonic encephalopathy. An EEG recording during episodes of myoclonus will not show parallel seizure activity in the brain. Neurologic examination findings are usually normal for patients with BSMI, although some may exhibit nonspecific neurologic signs or expressive language delays. Workup may include electroencephalography, a metabolic panel, and screening for toxins in the case of equivocal clinical presentations.[52] Provocative maneuvers, such as gently rocking the child's crib, have been shown to induce a BSMI event and may be useful with EEG monitoring to help differentiate BSMI from seizures. Other key features to distinguish BSMI from other neurologic disorders are listed in Table 106-2.

Table 106-2 Common Mimics of Benign Sleep Myoclonus of Infancy (BSMI)

Mimics	Distinguishing Features
Myoclonic seizures	Can occur while infants are awake, whereas BSMI will stop abruptly and consistently when infants are aroused Often associated with perinatal disorders (i.e., hypoxic-ischemic encephalopathy, infection, or metabolic abnormalities), whereas BSMI is typically manifested in neurologically normal infants
Infantile spasms (West syndrome)	Often seen after the first month of life but may occur earlier Manifested by sudden head flexion with arm extension and lower extremity flexion Usually associated with hypsarrhythmia electroencephalographic pattern
Pyridoxine-dependency seizures	Responsive to vitamin B_6 treatment
Myoclonic encephalopathies	Can occur while infants are awake, whereas BSMI will stop abruptly and consistently when infants are aroused
Hyperekplexia (startle disease)	Can occur while infants are awake, whereas BSMI will stop abruptly and consistently when infants are aroused Generalized stiffness while awake Exaggerated startle reflex
Jitteriness	Typical movements occur as an excessive response to stimulation, such as touch or loud noise

Polysomnography Scoring Criteria and Motor Features

On PSG, a BSMI event generally demonstrates paroxysmal body jerks that typically occur in clusters of four or five in number that are usually 40 to 300 milliseconds in duration. Once the BSMI episode has ensued, the paroxysmal body jerks can be repeated for several minutes in most cases but rarely continue past an hour. The video monitoring and EEG channels, during an episode of BSMI, will not demonstrate concurrent arousals, awakening, or (as noted before) ictal or interictal activity, which helps differentiate BSMI from seizures.

Pathophysiology

The pathophysiologic mechanism of BSMI remains unknown. It has been proposed that BSMI may be due to the inability of an immature (and unmyelinated) central nervous system's descending pathway to sufficiently inhibit the underlying movements generated from the cervical spinal cord.[1]

Prevalence and At-Risk Groups

The prevalence for BSMI is unknown, and the incidence has been estimated at 0.8 to 3 per 1000 live births.[54] Interestingly, boys are affected more than girls. Whereas BSMI has been reported in a case of opioid withdrawal syndrome, the majority of infants are neurologically normal and born to mothers with no history of illicit drug use.[53]

Treatment

No specific therapeutic intervention is necessary, except parental education on the disorder and reassurance. Recognition of this condition is important, especially to avoid potentially harmful and certainly unwarranted antiepileptic treatment. Workup may include electroencephalography, a metabolic panel, and screening for toxins.[52]

Propriospinal Myoclonus at Sleep Onset

Clinical Features

Propriospinal myoclonus at sleep onset (i.e., spinal myoclonus, plurisegmental myoclonus, intersegmental myoclonus, axial myoclonus) is a rare disorder characterized by jerks involving the abdominal and truncal muscles occurring at the transition from wakefulness to sleep.[1] The diagnosis can be made when a patient reports sudden jerks, mainly of the abdomen, trunk, and neck, during relaxed wakefulness and drowsiness as the patient attempts to fall asleep. The jerks disappear with sleep onset and are not observed during sleep.

These movements directly result in complaints of sleep disruption and insomnia not otherwise explained by another cause. Movements are typically spontaneous but can be evoked by external stimulations. The movements disappear with mental activation and with the onset of a stable sleep stage. Propriospinal myoclonus can have a negative impact on quality of life and is described as a chronic, unremitting condition that may result in injury to the patient or the bed partner.

Differential Diagnosis

Because of the stereotyped and rhythmic pattern of the often "full body jerking" appearance of these movements, they can easily be confused with seizures. In the case of propriospinal myoclonus at sleep onset, EEG signal will be normal during the event. Because of its temporal relation to sleep onset, observers may also confuse it with movements triggered by the dysesthesias of RLS since these dysesthesias can also be triggered with relaxed wakefulness in the recumbent position. A simple clinical history will help distinguish the voluntary movements to relieve RLS dysesthesias compared with the involuntary movements seen with propriospinal myoclonus of sleep onset.

Polysomnography Scoring Criteria and Motor Features

PSG confirms that the myoclonic activity is restricted to the wakefulness period preceding sleep onset or falling to sleep and is manifested as trunk flexion or, less frequently, extension.

Table 106-3 Common Mimics of Propriospinal Myoclonus

Sleep starts
Phasic REM twitches
Fragmentary myoclonus
Epileptic myoclonus
Periodic limb movements in sleep
Psychogenic myoclonus

During relaxed wakefulness, the myoclonic activity appears when the EEG alpha rhythm spreads from the occipital to the frontal leads or as the alpha rhythm drops out. Isolated or repetitive jerks may occur. The EMG discharges last typically 100 to 300 milliseconds but sometimes longer, with both reciprocal and co-contracting agonist-antagonist activity.

During relaxed wakefulness, the myoclonic activity appears when the EEG alpha rhythm spreads from the posterior to the anterior cortical areas or when it drops out. The jerks disappear with sleep onset and are not observed during sleep. Jerks may recur at quasi-periodic intervals (every 5 to 40 seconds), and this pattern may repeat itself more than a hundred times with consequent delayed sleep onset and sleep fragmentation.[45] No EEG abnormalities in the routine recording have been reported. Similar to RLS, these movements appear to be related to the recumbent position and a state of relaxed wakefulness. Table 106-3 highlights some of the mimics for propriospinal myoclonus.

Pathophysiology

Although the pathophysiologic mechanism for propriospinal myoclonus remains unknown, it is believed to involve the spinal cord serving as the underlying generator for the movement. Interestingly, persistent propriospinal myoclonus occurring during the day has been linked to structural spinal cord disease, whereas propriospinal myoclonus during sleep onset at this point in time has not.[1]

Prevalence and At-Risk Groups

It is likely that propriospinal myoclonus is a rare condition. The disorder has been reported to affect adults and to be more common in men, but it has not been reported to affect children.[1]

Treatment

As with most SRMDs, education is recommended for treatment after a thorough evaluation to rule out more serious disorders. Some patients may respond to clonazepam.

CLINICAL PEARL

Key features to help distinguish SRMDs from other movements observed during the night typically include the following: normal physical and neurologic examination findings, absence of comorbid neurologic conditions, absence of abnormal neurophysiologic findings (e.g., normal sleep and wake EEG findings), and presence of simple and stereotyped movements (as

CLINICAL PEARL—cont'd

opposed to the more complex movements and behaviors seen with parasomnias or seizures). Because of the simple and stereotypical appearance of SRMDs, each individual type can be differentiated by the characteristic age at onset, characteristic timing at which it occurs during the night, body region affected, and unique polysomnographic characteristics.

SUMMARY

The ICSD3 currently recognizes 10 different SRMDs and 4 isolated sleep-related movements. Although most SRMDs are considered benign, they can often result in sleep disruption and reduced quality of life. Thus, because of the relatively high prevalence of SRMDs and the phenomenologic overlap with other neurodegenerative or medical conditions, sleep health care providers and scientists should be familiar with this class of sleep conditions.

ACKNOWLEDGMENT

The authors wish to acknowledge Mr. Anthony Kwan for his original work on Figure 106-1 and Ms. Seulah Choi for her original work on Figure 106-2.

Selected Readings

Anderson KN, Smith IE, Shneerson JM. Rhythmic movement disorder (head banging) in an adult during rapid eye movement sleep. *Mov Disord* 2006;**21**:866–7.

Berry RB, Brooks R, Gamaldo CE, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.2*. Darien, Illinois: American Academy of Sleep Medicine; 2015.

Carra MC, Rompre PH, Kato T, et al. Sleep bruxism and sleep arousal: an experimental challenge to assess the role of cyclic alternating pattern. *J Oral Rehabil* 2011;**38**:635–42.

Durmer JS, Quraishi GH. Restless legs syndrome, periodic leg movements, and periodic limb movement disorder in children. *Pediatr Clin North Am* 2011;**58**:591–620.

Guertin PA. Central pattern generator for locomotion: anatomical, physiological, and pathophysiological considerations. *Front Neurol* 2012;**3**:183.

Koo BB, Sillau S, Dean DA 2nd, et al. Periodic limb movements during sleep and prevalent hypertension in the multi-ethnic study of atherosclerosis. *Hypertension* 2015;**65**(1):70–7.

Li J, Moore H 4th, Lin L, et al. Association of low ferritin with PLM in the Wisconsin Sleep Cohort. *Sleep Med* 2015;**16**(11):1413–8.

Marx C, Masruha MR, Garzon E, Vilanova LC. Benign neonatal sleep myoclonus. *Epileptic Disord* 2008;**10**:177–80.

Rana AQ, Khan F, Mosabbir A, Ondo W. Differentiating nocturnal leg cramps and restless legs syndrome. *Expert Rev Neurother* 2014;**14**:813–18.

Vetrugno R, Provini F, Plazzi G, et al. Propriospinal myoclonus: a motor phenomenon found in restless legs syndrome different from periodic limb movements during sleep. *Mov Disord* 2005;**20**:1323–9.

A complete reference list can be found online at ExpertConsult.com.

Section
14 Sleep Breathing Disorders

Chapter
107 Sleep Related Breathing Disorders: Classification

Richard B. Berry

Chapter Highlights

- This chapter presents a classification and clinical overview of sleep breathing disorders based on the recently published *International Classification of Sleep Disorders*, third edition (ICSD3).
- Diagnostic criteria for sleep related breathing disorders in adults, including the obstructive sleep apnea disorder, central sleep apnea disorders, and sleep-related hypoventilation disorders, are provided based on the ICSD3.
- New additions to the ICSD3 classification of sleep related breathing disorders include diagnostic criteria for the obesity hypoventilation syndrome and treatment-emergent central sleep apnea.
- Out-of-center sleep testing (home sleep apnea testing) is now included in the diagnostic criteria for obstructive sleep apnea disorder in adults.
- Central sleep apnea with Cheyne-Stokes breathing is uniquely defined in the ICSD3 based on symptoms and comorbid conditions as well as polysomnographic findings.
- Areas of controversy and unresolved issues relating to the ICSD3 diagnostic classification of sleep related breathing disorders are discussed.

Box 107-1 ADULT SLEEP RELATED BREATHING DISORDERS IN THE ICSD3 (ICD-9) [ICD-10]

Obstructive Sleep Apnea Disorders

Obstructive Sleep Apnea, Adult (327.23) [G47.33]

Central Sleep Apnea Disorders*

Central Sleep Apnea with Cheyne-Stokes Breathing (786.04) [R06.3]
Central Sleep Apnea Due to a Medical Disorder without Cheyne-Stokes Breathing (327.27) [G47.37]
Central Sleep Apnea Due to High Altitude Periodic Breathing (327.22) [G47.32]
Central Sleep Apnea Due to a Medication or Substance (327.29) [G47.39]
Primary Central Sleep Apnea (327.21) [G47.31]
Treatment-Emergent Central Sleep Apnea (327.29) [G47.39]

Sleep-Related Hypoventilation Disorders

Obesity-Hypoventilation Syndrome (278.03) [E66.2]
Idiopathic Central Alveolar Hypoventilation (327.24) [G47.34]
Sleep Related Hypoventilation Due to Medication or Substance (327.26) [G47.36]
Sleep Related Hypoventilation Due to a Medical Disorder (327.26) [G47.36]

Sleep Related Hypoxemia Disorders

Sleep-Related Hypoxemia (327.26) [G47.36]

*In ICSD3, Central Sleep Apnea Syndromes (here, "disorders" is used for consistency).

ICD-9, *International Classification of Diseases*, 9th revision; ICD-10, *International Classification of Diseases*, 10th revision; ICSD3, *International Classification of Sleep Disorders*, 3rd edition.

This chapter presents a classification and clinical overview of sleep breathing disorders based on the recently published *International Classification of Sleep Disorders*, third edition (ICSD3).[1] The term *Sleep Related Breathing Disorders* is used throughout the text and tables of this chapter to conform to the ICSD3 terminology. The goal is to provide a summary of diagnostic criteria for the major disorders, with brief illustrative clinical cases for selected disorders. Box 107-1 lists the ICSD3 sleep related breathing disorder classification in adults; the *International Classification of Diseases*, 9th (ICD-9) and 10th (ICD-10) revisions' codes for each diagnosis are also listed. The ICSD3 provides a timely update of diagnostic criteria for sleep related breathing disorders because the ICSD2 (second edition) was published in 2005. Where appropriate, areas of controversy and unresolved issues are discussed.

Of note, the ICSD3 does not itself define sleep-related respiratory events, but instead refers to the most recent version of the *American Academy of Sleep Medicine (AASM) Manual for the Scoring of Sleep and Associated Events* (hereafter referred to as the AASM scoring manual).[2] Where relevant for understanding diagnostic criteria, the definition of selected respiratory events is reviewed. The fact that patients may fit into more than one diagnostic category of sleep-related breathing disorders is emphasized in the ICSD3.

The ICSD3 often uses the term *disorder* rather than *syndrome*. A disorder usually refers to a disease process for which a cause is known. In contrast, a syndrome is a constellation of symptoms. In general, use of disorder versus syndrome is somewhat arbitrary and often variable in the medical literature. For example, the term *congenital central hypoventilation syndrome* continues to be used in the literature although the genetic cause has been identified (mutation in *PHOX2B* gene).

OBSTRUCTIVE SLEEP APNEA DISORDERS

In the ICSD3, obstructive sleep apnea (OSA) disorders are classified separately as adult and pediatric. However, this chapter refers solely to adult disorders. OSA is classified as a disorder in the ICSD3. One could argue that OSA is better described as a syndrome because the pathophysiology may be structural in one patient and due to unstable respiratory control in another patient.[3] It is to be noted that the term *obstructive sleep apnea syndrome* is frequently used in the literature.

The adult OSA diagnostic criteria currently include a provision for use of both in-center polysomnography (PSG) and out-of-center sleep testing (OCST), in which sleep is usually not recorded. OCST has also been termed *portable monitoring*, *home sleep testing*, and *home sleep apnea testing*. Obstructive respiratory events in sleep that account for the diagnosis of adult OSA include obstructive and mixed apneas, hypopneas, and respiratory effort–related arousals (RERAs). The ICSD3 does not provide a definition for these events. A summary of such respiratory event definitions as defined by the AASM scoring manual is provided in Box 107-2. There are currently two hypopnea definitions in the manual ("recommended" and "acceptable"). Of note, if sleep is not recorded, hypopneas must be scored based on airflow attenuation and arterial oxygen desaturation criteria alone because there is an inability to score arousals from sleep without documentation of sleep.

The most recent version of the AASM scoring manual defines the apnea-hypopnea index (AHI) as the number of apneas and hypopneas per hour of sleep; the respiratory disturbance index is defined as the AHI plus the number of RERAs per hour of sleep. If OCST is performed, sleep is usually not recorded, and the diagnostic metric is the number of apneas and hypopneas per hour of monitoring time (sometimes called the *respiratory event index*). The ICSD3 does not define the terms AHI, RDI, or respiratory event index.

Diagnostic criteria for OSA (Box 107-3) require either 15 or more obstructive respiratory events per hour of sleep using PSG (or per hour of monitoring using OCST) *or* a combination of symptoms, manifestations, and comorbidities (see later) and at least 5 but less than 15 obstructive respiratory events per hour of sleep (PSG) or per hour of monitoring if OCST is used. The rationale is that patients with 15 or more obstructive respiratory events per hour are thought to have a clinically significant disorder even if symptoms are absent. The best evidence for increased risk for cardiovascular morbidity is in patients with more than 30 obstructive respiratory events per hours of sleep.[4] In the case of at least 5 but less than 15 obstructive respiratory events per hour, the clinician may be concerned about overdiagnosis in an asymptomatic patient without a significant clinical disorder. The requirement of symptoms in the milder range of an increased AHI (or RDI) will exclude asymptomatic patients. The clinical significance of an AHI (or respiratory disturbance index [RDI]) in the mild range must be evaluated by the clinician,

Box 107-2 ADULT RESPIRATORY EVENT DEFINITIONS

Apnea

1. Score a respiratory event as an apnea when *both* of the following criteria are met:
   a. There is a drop in the peak signal excursion by ≥90% of pre-event baseline using an oronasal thermal sensor (diagnostic study), positive airway pressure (PAP) device flow (titration study), or alternative apnea sensor (diagnostic study).
   b. The duration of the ≥90% drop in sensor signal is ≥10 seconds.
2. Score an apnea as obstructive if it meets apnea criteria and is associated with continued or increased inspiratory effort throughout the entire period of absent airflow.
3. Score an apnea as central if it meets apnea criteria and is associated with absent inspiratory effort throughout the entire period of absent airflow.
4. Score an apnea as mixed if it meets apnea criteria and is associated with absent inspiratory effort in the initial portion of the event, followed by resumption of inspiratory effort in the second portion of the event.

Hypopnea

1A. Score a respiratory event as a hypopnea if *all* of the following criteria are met (**RECOMMENDED**):
   a. The peak signal excursions drop by ≥30% of preevent baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative hypopnea sensor (diagnostic study).
   b. The duration of the ≥30% drop in signal excursion is ≥10 seconds.
   c. There is ≥3% oxygen desaturation from preevent baseline, or the event is associated with an arousal.

1B. Score a respiratory event as a hypopnea if *all* of the following criteria are met (**ACCEPTABLE**):
   a. The peak signal excursions drop by ≥30% of preevent baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative hypopnea sensor (diagnostic study).
   b. The duration of the ≥30% drop in signal excursion is ≥10 seconds.
   c. There is a ≥4% oxygen desaturation from preevent baseline.

Respiratory Effort–Related Arousal

Score a respiratory event as a respiratory effort-related arousal (RERA) if there is a sequence of breaths lasting ≥10 seconds characterized by increasing respiratory effort or by flattening of the inspiratory portion of the nasal pressure (diagnostic study) or PAP device flow (titration study) waveform leading to arousal from sleep when the sequence of breaths *does not meet criteria for an apnea or hypopnea.*

Adapted from Berry RB, Brooks R, Gamaldo CE, et al, for the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Version 2.2. www.aasmnet.org. Darien, Illinois: American Academy of Sleep Medicine, 2015.

taking into account symptoms as well as comorbid cardiovascular conditions. Certainly, symptomatic patients with an AHI (RDI) in the mild range may experience benefit from treatment of OSA.[4]

Identification of RERAs is adjudicated based on associated arousal rather than oxygen desaturation and therefore requires the recording of sleep (i.e., PSG). Hypopneas are included as OSA respiratory events without specific determination as to whether they are obstructive or nonobstructive (i.e., central). It is emphasized that symptoms and comorbid conditions thought to be associated with or adversely affected by the presence of OSA are required for the diagnosis if the index of obstructive events is equal to or greater than 5 per hour but less than 15 per hour (see Box 107-3). Such comorbid conditions qualifying a patient for a diagnosis of OSA include systemic hypertension, a mood disorder, cognitive dysfunction, coronary artery disease, stroke, congestive heart failure, atrial fibrillation, type 2 diabetes mellitus, nocturia, and erectile dysfunction.

Although a major study published in 1993 found OSA prevalence (defined as symptoms *and* AHI ≥5 per hour) to be 2% in women and 4% in adult men,[5] the current prevalence of OSA is believed to be much higher, likely because of increased prevalence of obesity.[6] The prevalence of OSA in fact varies with the characteristics of the population studied and respiratory event definitions.[4] Note that although patients with obesity-hypoventilation syndrome are currently classified under the sleep-related hypoventilation disorders (see later), 80% to 90% of patients so classified also meet diagnostic criteria for OSA.

There is continuing controversy concerning the most appropriate definition of hypopnea.[4] The current AASM scoring manual *acceptable* definition of hypopnea, a 30% or greater drop in airflow plus 4% or greater arterial oxygen desaturation (see Box 107-2), is consistent with that accepted by the Centers for Medicare and Medicaid Services. The recommended definition of hypopnea is based on a lesser degree of oxygen desaturation (≥3%) or the presence of an associated arousal, or both. The AHI-acceptable and AHI-recommended terminology will be used here to denote the AHI determined using either the acceptable hypopnea (H-acceptable) or recommended hypopnea (H-recommended) definition. Note that, because RERAs by definition do not meet diagnostic criteria for hypopnea,[2] classification of a given event as RERA versus hypopnea often depends on the definition of hypopnea that is used to score events. For example, an event characterized by flattening of the airflow profile with a 30% reduction in flow for 15 seconds followed by a 2% arterial oxygen desaturation and an arousal would meet diagnostic criteria for H-recommended but not H-acceptable. If hypopneas are scored based on criteria for H-acceptable, the event would meet criteria for RERA. If one uses the ICSD3 criteria, which include the RERA index as part of the diagnostic metric (e.g., AHI-acceptable + RERA index), a wider spectrum of patients will be diagnosed as having OSA than using AHI-acceptable. The addition of the RERA index will diagnose patients often labeled as having the upper airway resistance syndrome as having OSA. These patients report symptoms due to respiratory arousals but do not have significant arterial oxygen desaturation.[4] Of note, upper airway resistance syndrome is not included in the ICSD3 diagnostic criteria; such patients are thought to have a variant of OSA (see Chapter 112, Snoring and Pathologic Upper Airway Resistance Syndromes). If the H-recommended definition is used, most events scored as

Box 107-3 ICSD3 DIAGNOSTIC CRITERIA: OBSTRUCTIVE SLEEP APNEA, ADULT

(A and B) or C satisfy the criteria.

A. The presence of one or more of the following:
   1. The patient complains of sleepiness, nonrestorative sleep, fatigue, or insomnia symptoms.
   2. The patient wakes with breath holding, gasping, or choking.
   3. The bed partner or other observer reports habitual snoring, breathing interruptions, or both during the patient's sleep.
   4. The patient has been diagnosed with hypertension, a mood disorder, cognitive dysfunction, coronary heart disease, stroke, congestive heart failure, atrial fibrillation, or type 2 diabetes mellitus.

B. Polysomnography (PSG) or out-of-center sleep testing (OCST*) demonstrates:
   1. Five or more obstructive respiratory events† (i.e., obstructive or mixed apneas, hypopneas, or respiratory effort–related arousals [RERAs])‡ per hour of sleep during PSG or per hour of monitoring (OCST*)

   or

C. PSG or OCST* demonstrates:
   1. Fifteen or more obstructive respiratory events† (i.e., obstructive or mixed apneas, hypopneas, or RERAs)‡ per hour of sleep during PSG or per hour of monitoring (OCST*)

*OCST may underestimate the true number of obstructive respiratory events per hour because actual sleep is not usually recorded.

†Respiratory events defined according the latest version of the *AASM Manual for the Scoring of Sleep and Associated Events*.

‡RERAs cannot be scored using OCST because arousals cannot be identified.

RERAs when using H-acceptable would meet criteria for H-recommended such that AHI-acceptable + RERA index ≈ AHI-recommended.[4] Data indicate that using a diagnostic metric including RERAs or an AHI-recommended will identify an increased percentage of individuals studied as having OSA.

The following clinical case illustrates the hypopnea definition issue. A thin 30-year-old man reported loud snoring and daytime sleepiness. A diagnostic PSG recorded 420 minutes of sleep, during which there were 10 obstructive apneas, 10 hypopneas (H-acceptable definition) and 40 RERAs. The AHI was therefore 2.8/hour, and the RDI was 8.6 per hour. Many of the events scored as RERAs were associated with a 1% or 2% drop in the oxygen saturation. Using the AHI-acceptable as a diagnostic metric, the patient does not have the OSA disorder. If the same sleep study were scored using the recommended rather than the acceptable hypopnea definition, 10 obstructive apneas, 45 hypopneas, and 5 RERAs would be scored, and the AHI would be 7.8 per hour (the RDI would have continued to be 8.6 per hour). A diagnosis of OSA would then be possible (if associated with symptoms/comorbidity). This assumes that 35 of the 40 RERA events were associated with a 30% or greater drop in flow as well as inspiratory flattening for at least 10 seconds.

CENTRAL SLEEP APNEA DISORDERS

Patients with the central sleep apnea (CSA) disorders are a diverse group with a wide spectrum of etiologies.[7,8] Many of the diagnostic entities associated with CSA are associated with either unknown or multiple etiologies. The common finding is absence of airflow and respiratory effort (central apnea) or reduced airflow and respiratory effort without clear evidence of partially obstructed breathing (e.g., inspiratory airflow signal flattening or paradoxic thoracoabdominal respiratory effort [central hypopnea], or both). In the AASM scoring manual,[2] scoring hypopneas as central or obstructive is noted as an option, with the following specifications: hypopnea is scored as obstructive if any of the following are present: there is snoring during the event; there is increased inspiratory flattening of the nasal pressure or positive airway pressure (PAP) device flow signal compared with baseline breathing; or there is an associated thoracoabdominal paradox that occurs during the event but not during preevent breathing. A hypopnea is scored as central if none of these are present. The option for scoring hypopneas as central was added because some patients with CSA have a significant proportion of events that are central hypopneas.[4,9,10] Further, a number of large clinical trials of patients with CSA have included central hypopneas in the inclusion criteria.[9,10]

Some clinicians have proposed that the CSA disorders can be categorized on the basis of the disorder being associated with awake- or sleep-associated hypocapnia (i.e., high ventilatory drive) and normocapnia or hypercapnia (i.e., normal or low ventilatory drive).[7] Primary CSA, CSA with Cheyne-Stokes breathing (CSB), CSA due to high-altitude periodic breathing, and treatment emergent CSA are thought to occur because the sleeping arterial partial pressure of carbon dioxide ($Paco_2$) is below the apneic threshold (AT)—the $Paco_2$ value below which there is no ventilatory effort.[7] High hypercapnic ventilatory drive (both awake and during sleep) characterizes such patients and is associated with a small difference between the sleeping $Paco_2$ and the AT (delta $Paco_2$). While awake, such patients have a normal or low $Paco_2$. The low awake or sleeping $Paco_2$ does not cause the small delta $Paco_2$ but rather is a marker of high ventilatory drive. CSA due to a medical disorder without CSB and CSA due to a medication or substance are thought to be associated with normal or low ventilatory drive.

Central Sleep Apnea with Cheyne-Stokes Breathing

In the ICSD2,[11] this entity was listed as CSB pattern without a requirement for symptoms or associated conditions. One could argue that diagnostic criteria for a breathing *pattern* would be more appropriately defined in the scoring manual rather than in the classification of sleep disorders. As defined in the ICSD3, central sleep apnea with Cheyne-Stokes breathing (CSA-CSB) is a clinical disorder rather than simply a PSG finding (Box 107-4),[1] with the main rationale being that the identification of patients with CSA-CSB, which is a very common disorder, has clinical implications, including etiology (most have congestive heart failure) and prognosis.[12] Thus symptoms (sleepiness, insomnia, awakening short of breath, snoring, witnessed apneas) or the presence of comorbid conditions (atrial fibrillation, congestive heart failure, or a neurologic disorder), or both, are required in the ICSD3 for a diagnosis of CSA-CSB. CSA-CSB is common in patients with both stable and decompensated congestive heart failure[13] and can occur following a cerebrovascular accident.[14-16] Renal failure has also been associated with CSA-CSB, although such documentation in the literature is very

Box 107-4 ICSD3 DIAGNOSTIC CRITERIA: CENTRAL SLEEP APNEA WITH CHEYNE-STOKES BREATHING

(A or B) + C + D satisfy the criteria.

A. The presence of one or more of the following:
   1. Sleepiness
   2. Difficulty initiating or maintaining sleep, frequent awakenings, or nonrestorative sleep
   3. Awakening short of breath
   4. Snoring
   5. Witnessed apneas

B. The presence of atrial fibrillation/flutter, congestive heart failure, or a neurologic disorder

C. Polysomnography (PSG) (during diagnostic or positive airway pressure titration) shows all of the following:
   1. Five or more central apneas or central hypopneas per hour of sleep
   2. The total number of central apneas and/or central hypopneas is >50% of the total number of apneas and hypopneas
   3. The pattern of ventilation meets criteria for Cheyne-Stokes breathing

D. The disorder is not better explained by another current sleep disorder, medication use (e.g. narcotics), or substance use disorder

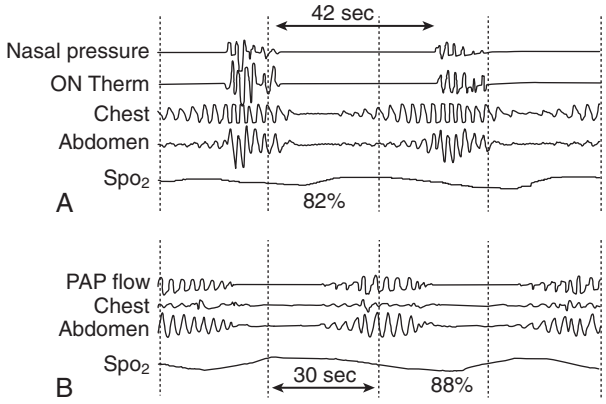


Figure 107-1 Respiratory events during a diagnostic **(A)** and positive airway pressure sleep study **(B).** The middle event in the top tracing is a "mixed" apnea; the flanking events are "obstructive" apneas. Central apneas with Cheyne-Stokes breathing are shown in the lower tracing. The patient was on continuous positive airway pressure of 12 cm H_2O. ON Therm, Oronasal thermal flow sensor; PAP, positive airway pressure.

limited.[17] Patients with idiopathic CSA-CSB have also been reported.[18]

In the current AASM scoring manual,[2] CSA-CSB (the respiratory event, not the clinical disorder) is scored when both of the following criteria are met:

1. Three or more consecutive central apneas or central hypopneas, or both, separated by a crescendo and decrescendo change in breathing amplitude with a cycle length of at least 40 seconds
2. Five or more central apneas or central hypopneas, or both, per hour of sleep associated with the crescendo and decrescendo breathing pattern recorded over at least 2 hours of monitoring.

The ICSD3 criteria for CSA-CSB are listed in Box 107-4. If central apneas of CSB morphology as defined previously are *not* frequent enough to be more than 50% of the total respiratory events, CSB is simply listed as a PSG finding. Note that CSA-CSB and OSA diagnoses may coexist in a patient who meets PSG and clinical criteria for both. Further, as per the ICSD3, diagnostic criteria for CSA-CSB may be met during *either a diagnostic PSG or PAP titration*. Many patients with CSA-CSB associated with heart failure have a significant number of both obstructive and central apneas,[13] and the predominance of obstructive versus central events may vary over time or during the night[19] in a given patient. If a patient with a mixture of obstructive and central events is placed on continuous positive airway pressure (CPAP), with titration specifically for the OSA, central events, including outright CSA-CSB, may persist after obstructive events have resolved.

A typical clinical example is the following. A 50-year-old man with a history of witnessed apnea, snoring, atrial fibrillation, and congestive heart failure underwent diagnostic and therapeutic PSG. During the diagnostic portion of the study, 120 minutes of sleep were recorded. The AHI was 30/hour with 40 obstructive apneas, 10 mixed apneas, and 10 central apneas. Typical obstructive and mixed events in this patient are shown in Figure 107-1, *A*. During the CPAP titration portion of the PSG, 300 minutes of sleep were recorded, with an AHI of 15/hour, 15 obstructive apneas, 50 central apneas, and 10 hypopneas (adjudicated as obstructive). The central AHI was 10/hour, and central events composed more than 50% of the respiratory events on CPAP. A typical period of central apnea, in fact central apnea with CSB, in this patient is shown in Figure 107-1, *B*. This patient meets diagnostic criteria for OSA (diagnostic portion) and CSA-CSB (the CPAP titration portion). The clinical implications of such results remain to be defined; presumably, the patient continues to carry the cardiovascular morbidity and mortality associations of CSA-CSB in this setting.

Central Sleep Apnea Due to a Medical Disorder without Cheyne-Stokes Breathing

CSA that is attributed to a medical disorder, but does not have the characteristic periodic breathing pattern of CSB, is classified as CSA due to a medical disorder without Cheyne-Stokes breathing (Box 107-5). Here the term *medical disorder* is used as an all-inclusive nomenclature that includes cardiovascular, respiratory, and neurologic conditions; however, most of these patients have brainstem lesions of developmental, vascular, neoplastic, degenerative, demyelinating, or traumatic origin. Examples of such neurologic conditions that are typically associated with CSA without CSB include prior cerebrovascular accident (CVA),[14-16] Chiari malformation,[20] brainstem neoplasms, and multiple system atrophy.[21] A predominance of OSA versus CSA is more common after a CVA.[15] However, many post-CVA patients have a mixture of OSA and CSA; CSA following CVA can be present with and without a pattern of CSB.

Patients classified in the diagnostic category of CSA without CSB may have awake or sleep-related hypoventilation, or both. Further, in such patients, if diagnostic criteria for CSA and sleep-related hypoventilation are met (see later), both diagnoses are made (i.e., CSA *and* sleep–related hypoventilation).

Box 107-5 ICSD3 DIAGNOSTIC CRITERIA: CENTRAL APNEA DUE TO MEDICAL DISORDER WITHOUT CHEYNE-STOKES BREATHING

Criteria A to C must be met:
A. The presence of one or more of the following:
   1. Sleepiness
   2. Difficulty initiating or maintaining sleep, frequent awakenings, or nonrestorative sleep
   3. Awakening short of breath
   4. Snoring
   5. Witnessed apneas
B. Polysomnography (PSG) shows all of the following:
   1. Five or more central apneas and/or or central hypopneas per hour of sleep
   2. The number of central apneas and/or central hypopneas is >50% of the total number of apneas and hypopneas
   3. Absence of Cheyne-Stokes breathing
C. The disorder occurs as a consequence of a medical or neurologic disorder but is not due to medication use or substance use.

Box 107-6 ICSD3 DIAGNOSTIC CRITERIA: CENTRAL SLEEP APNEA DUE TO HIGH-ALTITUDE PERIODIC BREATHING

Criteria A to D must be met:
A. Recent ascent to high altitude*
B. The presence of one or more of the following:
   1. Sleepiness
   2. Difficulty initiating or maintaining sleep, frequent awakenings, or nonrestorative sleep
   3. Awakening with shortness of breath or morning headache
   4. Witnessed apnea
C. The symptoms are clinically attributable to high-altitude periodic breathing, or polysomnography, if performed, demonstrates recurrent central apneas or hypopneas primarily during non–rapid eye movement sleep at a frequency of ≥5/hour.
D. The disorder is not better explained by another current sleep disorder, medical or neurologic disorder, medication use (e.g., narcotics), or substance use disorders.

*Typically at least 2500 meters (8202 feet), although some individuals may exhibit the disorder at altitudes as low as 1500 meters.

Box 107-7 ICSD3 DIAGNOSTIC CRITERIA: CENTRAL SLEEP APNEA DUE TO A MEDICATION OR SUBSTANCE

Criteria A to E must be met:
A. The patient is taking an opioid or other respiratory depressant.
B. The presence of one or more of the following:
   1. Sleepiness
   2. Difficulty initiating or maintaining sleep, frequent awakenings, or nonrestorative sleep
   3. Awakening short of breath
   4. Snoring
   5. Witnessed apneas
C. Polysomnography (PSG; diagnostic or on positive airway pressure) shows all of the following:
   1. Five or more central apneas and/or central hypopneas per hour of sleep (PSG)
   2. The number of central apneas and/or central hypopneas is >50% of the total number of apneas and hypopneas
   3. Absence of Cheyne-Stokes breathing
D. The disorder occurs as a consequence of an opioid or other respiratory depressant.
E. The disorder is not better explained by another current sleep disorder.

Central Sleep Apnea Due to High-Altitude Periodic Breathing

Periodic breathing during sleep is a common response to altitude. Typically, an altitude of at least 2500 meters (8202 feet) is required for such periodic breathing to be present. However, some individuals may exhibit the disorder at altitudes as low as 1500 meters. There are no specified levels of central AHI (central apneas or central hypopneas per hour) nor frequency of episodes of periodicity of breathing that clinically separate a normal from an abnormal response to altitude. The ICSD3 diagnostic criteria require 5 or more central events per hour of sleep *and* associated symptoms to make the diagnosis of CSA due to high-altitude periodic breathing (Box 107-6). The cycle length of this periodic breathing is commonly less than 40 seconds and often as short as 12 to 20 seconds.[22,23] The periodic breathing present at altitude does *not* specifically meet the criteria for CSB as defined earlier. A diagnosis of CSA due to high-altitude period breathing may be made concomitantly with a diagnosis of OSA if criteria for each of these are clearly present. A study by Pagel and colleagues[24] of sleep study results at three sleep centers at progressively higher altitudes found that central apnea becomes significantly more common at increasing altitude in both the diagnostic and treatment portions of PSG in patients with OSA, with an apparent exponential increase in the percentage of OSA patients with a central apnea index greater than 5 occurring with increasing altitude.

Central Sleep Apnea Due to a Medication or Substance

Patients with this diagnosis have CSA adjudicated as secondary to an opioid medication or other respiratory depressant (Box 107-7). The condition has been described in patients taking methadone and long-acting forms of morphine or oxycodone as well as individuals being treated with fentanyl patches or continuous narcotic infusions.[25] Suboxone (a combination of buprenorphine and naloxone) is often used for treatment of patients with narcotic dependence and pain and can also be associated with central apnea. Patients taking these medications can also manifest OSA and ataxic breathing. In addition, some patients may exhibit mostly obstructive respiratory events during a diagnostic study but mainly central events on positive airway pressure, and the required frequency of central events to meet diagnostic criteria for this entity may be present on either a diagnostic or PAP PSG. Although most patients with this disorder have normal or only mildly elevated awake values of $Paco_2$, some manifest sleep-related

Box 107-8 ICSD3 DIAGNOSTIC CRITERIA: PRIMARY CENTRAL SLEEP APNEA

Criteria A to D must be met:

A. The presence of at least one of the following:
   1. Sleepiness
   2. Difficulty initiating or maintaining sleep, frequent awakenings, or nonrestorative sleep
   3. Awakening short of breath
   4. Snoring
   5. Witnessed apneas

B. Polysomnography demonstrates all of the following:
   1. Five or more central apneas or central hypopneas per hour of sleep
   2. The total number of central apneas and/or central hypopneas is >50% of the total number of apneas and hypopneas
   3. Absence of Cheyne-Stokes breathing

C. There is no evidence of daytime or nocturnal hypoventilation.

D. The disorder is not better explained by another current sleep disorder, medical or neurologic disorder, medication use, or substance use disorder.

hypoventilation (see later) as well as central apneas. If sleep-related hypoventilation is present, a diagnosis of sleep-related hypoventilation due to a medication or substance is made *in addition to* a diagnosis of CSA due to a medication or substance.

Primary Central Sleep Apnea

Primary CSA, which has also been referred to as idiopathic CSA, is a disorder of unknown etiology and is characterized by recurrent central apneas that do not have CSB morphology (Box 107-8) or a specific known etiology. Primary CSA is believed to occur because the sleeping arterial $Paco_2$ is below the AT,[26,27] as discussed previously. High hypercapnic ventilatory drive (both awake and during sleep) characterizes such patients; while awake, they therefore have a normal or low $Paco_2$. The reason for the high ventilatory drive in patients with primary CSA is unknown.

Clinical manifestations of primary CSA typically include daytime sleepiness, insomnia, awakening short of breath, and witnessed apneas. PSG diagnostic criteria include the requirement that there must be 5 or more central apneas or central hypopneas per hour of sleep and that central, rather than obstructive, events make up greater than 50% of the total number of scored respiratory events. Further, the central apneas or hypopneas cannot, by definition, occur with the periodic breathing morphology of CSB. If the central apnea is believed to be due to a medical or neurologic condition or to a medication, then a diagnosis of primary CSA is not made. Primary CSA is a diagnosis based on *exclusion of identifiable causes of this pattern of breathing*. Clinical implications regarding etiology and sequelae of this disorder are not clear.

A typical clinical example is the following. A 30-year-old man underwent a PSG to evaluate complaints of snoring, witnessed breathing pauses, and mild daytime sleepiness. The PSG showed a total sleep time of 420 minutes and an AHI of 15/hour. There were 20 obstructive apneas and 85 central apneas. The central apneas did not have a Cheyne-Stokes morphology, although runs of central apneas did occur, most commonly associated with stages 1 and 2 non–rapid eye movement sleep (Figure 107-2). The patient was taking no medications and was otherwise healthy. Magnetic resonance imaging revealed no structural lesion of the brainstem (note that such a lesion would define the disorder as CSA due to a medical disorder without CSB).

Treatment-Emergent Central Sleep Apnea

Treatment-emergent CSA as defined in the ICSD3 is similar to a phenomenon described in the medical literature as complex sleep apnea. In this situation a patient with predominantly obstructive respiratory events (obstructive and mixed apneas or hypopneas) during a diagnostic PSG manifests predominantly central respiratory events (persistent or emergent) on PAP treatment (CPAP, or bilevel PAP without a backup rate) after the obstructive events have resolved.[28] This remains an area of controversy[29] because some clinicians consider all patients manifesting the pattern described previously as having complex sleep apnea (or TE-CSA), whereas others reserve the diagnosis of TE-CSA for patients in whom a clear diathesis for CSA is not present. A diagnosis of treatment emergent CSA as defined in ICSD3 (Box 107-9) is more specific than the typically applied terminology of complex sleep apnea because it requires that the CSA present during a PAP sleep study *not be* better explained by another CSA disorder or diathesis (e.g., CSA-CSB or CSA due to a medication or substance). Perhaps a more accurate term would be *idiopathic treatment emergent CSA*, when a reason or even risk for the phenomenon is unclear; in any event, the current term has clinical importance and specificity in that CSA has been reported to be present in 2% to 20% of patients in whom PAP is initiated,[28-29] whereas the TE-CSA as defined by the ICSD3 is believed to resolve with chronic PAP treatment in most patients.

In contrast, patients with complex sleep apnea associated with opioids and patients with CSB-CSA and heart failure each are less likely to experience resolution of CSA with chronic CPAP.[30,31,32] Although about 50% of patients with CSA-CSB due to heart failure will respond to the first night of exposure to CPAP, after 3 months about 43% still have an AHI greater than 15/hour.[30] However, the long-term morbidity associated with the diagnosis of treatment emergent CSA (ICSD3 criteria) and the pathophysiology have not been defined. Patients with persistent CSA on chronic CPAP treatment may not experience benefit from PAP treatment and may fail to be adherent to CPAP.

An illustrative clinical example of treatment emergent CSA is follows. A 50-year-old man underwent a PSG to evaluate complaints of snoring, daytime sleepiness, and witnessed apnea. He was being treated for systemic hypertension, had no history of heart failure, and was not taking opioid medications. During the diagnostic portion of the PSG, 120 minutes of sleep were recorded, with an AHI of 40/hour, including 55 obstructive apneas, 5 central apneas, and 20 hypopneas. During the CPAP titration, 300 minutes of sleep were recorded, with a final overall AHI of 20/hour. Respiratory events consisted of 10 obstructive apneas, 10 hypopneas (uncharacterized as central or obstructive), and 80 central apneas. The central apneas did not have Cheyne-Stokes morphology. A diagnosis of OSA and treatment emergent CSA was made.

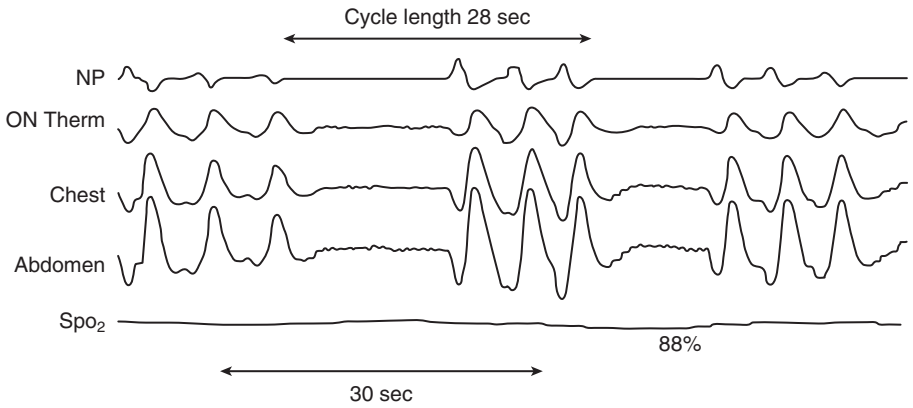


Figure 107-2 Central apneas in a patient without an apparent cause of central apnea. Note that the cycle length of the central events is less than 30 seconds. NP, Nasal pressure; ON Therm, oronasal thermal flow sensor.

Box 107-9 ICSD3 DIAGNOSTIC CRITERIA: TREATMENT EMERGENT CENTRAL SLEEP APNEA

Criteria A to C must be met:

A. Diagnostic polysomnography (PSG) shows five or more predominantly obstructive respiratory events (obstructive or mixed apneas, hypopneas, or respiration effort–related arousals [RERAs]) per hour of sleep.
B. PSG during use of positive airway pressure without a backup rate shows significant resolution of obstructive events and emergence or persistence of central apnea or central hypopnea with all of the following:
   i. Central apnea–central hypopnea index [CAHI] ≥5/hour
   ii. Number of central apneas and central hypopneas is ≥50% of total number of apneas and hypopneas
C. The central sleep apnea (CSA) is not better explained by another CSA disorder (e.g., CSA with Cheyne-Stokes breathing or CSA due to a medication or substance).

Box 107-10 ICSD3 DIAGNOSTIC CRITERIA: OBESITY-HYPOVENTILATION SYNDROME

Criteria A to C must be met:

A. Presence of hypoventilation during wakefulness (Pa_{CO_2} >45 mm Hg) as measured by arterial P_{CO_2}, end-tidal P_{CO_2}, or transcutaneous P_{CO_2}.
B. Presence of obesity (body mass index >30 kg/m^2; >95th percentile for age and sex for children).
C. Hypoventilation is not primarily due to lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder (other than mass loading from obesity), medication use, neurologic disorder, muscle weakness, or a known congenital or idiopathic central alveolar hypoventilation syndrome.

SLEEP-RELATED HYPOVENTILATON DISORDERS

The sleep-related hypoventilation disorders are listed in Box 107-1. The current criteria for scoring hypoventilation during sleep for adults are addressed by the most recent version of the *AASM Manual for the Scoring of Sleep and Associated Events*,[2] as follows:

1. Score hypoventilation during sleep if *either* of the following occur:
   a. Increase in the arterial P_{CO_2} (or surrogate) to a value >55 mm Hg for ≥10 minutes
   b. Increase of ≥10 mmHg in arterial P_{CO_2} (or surrogate) during sleep (in comparison to an awake supine value) to a value exceeding 50 mm Hg for ≥10 minutes

Because monitoring of arterial P_{CO_2} during sleep is not practical and is not widely done during PSG, acceptable surrogates include end-tidal P_{CO_2} and transcutaneous P_{CO_2}. Arterial oxygen desaturation is often present but is not required for the diagnosis.

The rationale for the previous definition of sleep-related hypoventilation is discussed in a review paper[4] and is based on limited "consensus" evidence. A normal increase in Pa_{CO_2} during sleep compared with awake has been described, with a range of about 2 to 8 mm Hg. The criteria attempt to define both an abnormal increase in Pa_{CO_2} and an absolute Pa_{CO_2} level considered to represent hypoventilation. Note that criterion 1a, although stated as an increase, in fact represents an absolute value during sleep rather than a comparison to awake and therefore not necessarily an effect of sleep in a person with awake hypoventilation. On the other hand the arterial P_{CO_2} invariably increases during sleep, irrespective of the presence or absence of daytime hypoventilation. The ICSD3 diagnostic criteria for sleep-related hypoventilation are simply those as defined in the AASM scoring manual.

Obesity-Hypoventilation Syndrome

The obesity-hypoventilation syndrome (OHS) is the only sleep-related hypoventilation disorder that requires documentation of awake hypoventilation (Pa_{CO_2} >45 mm Hg) (Box 107-10). In the other ICSD3 defined sleep-related hypoventilation disorders, awake hypoventilation may or may not be present.

During PSG, if a noninvasive estimate of the Pa_{CO_2} is measured in OHS patients, sleep-related hypoventilation is invariably present (i.e., criterion 1a for scoring hypoventilation during sleep), with increases in the Pa_{CO_2} from awake to sleep typical (i.e., criterion 1b for scoring hypoventilation during sleep). Arterial oxygen desaturation in usually present during sleep and may be present during wakefulness, but neither is required for a diagnosis of OHS. In 80% to 90% of

Box 107-11 ICSD3 DIAGNOSTIC CRITERIA: IDIOPATHIC CENTRAL ALVEOLAR HYPOVENTILATION

Criteria A and B must be met:
A. Sleep-related hypoventilation is present.
B. Hypoventilation is not primarily due to lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder, medication use, neurologic disorder, muscle weakness, or obesity or congenital hypoventilation syndromes.

Box 107-12 ICSD3 DIAGNOSTIC CRITERIA: SLEEP-RELATED HYPOVENTILATION DUE TO A MEDICATION OR SUBSTANCE

Criteria A to C must be met:
A. Sleep-related hypoventilation is present.
B. A medication or substance known to inhibit respiration and/or ventilatory drive is believed to be the primary cause of sleep-related hypoventilation.
C. Hypoventilation is not primarily due to lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder, neurologic disorder, muscle weakness, obesity-hypoventilation syndrome, or a known congenital central alveolar hypoventilation syndrome.

Box 107-13 ICSD3 DIAGNOSTIC CRITERIA: SLEEP-RELATED HYPOVENTILATION DUE TO A MEDICAL DISORDER

Criteria A to C must be met:
A. Sleep-related hypoventilation is present.
B. A lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder, neurologic disorder, or muscle weakness is believed to be the primary cause of hypoventilation.
C. Hypoventilation is not primarily due to obesity-hypoventilation syndrome, medication use, or a known congenital central alveolar hypoventilation syndrome.

OHS patients, a diagnosis of OSA can also be made.[33] The other 10% to 20% manifest awake hypoventilation that worsens during sleep in association with reduced tidal volume. In OHS, by definition, the etiology of the hypoventilation is not primarily due to lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder (other than mass loading from obesity), medication use, neurologic disorder, muscle weakness, or a documented congenital or idiopathic central alveolar hypoventilation syndrome.

The etiology of OHS also likely varies among individuals. In clinical practice there is often uncertainty concerning the most appropriate diagnosis in an obese individual with OSA, obstructive lung disease, and awake hypoventilation; for example, does the patient have OHS or the "overlap syndrome" (OSA + chronic obstructive pulmonary disease [COPD])? Such clinical considerations affect the optimal treatment of the awake- and sleep-related hypoventilation disorder. If a patient diagnosed with OHS undergoes CPAP treatment for the OSA component, normalization of the awake $Paco_2$ can occur; similarly, treatment of the OSA in the overlap syndrome can favorably affect both OSA and COPD outcomes. An unresolved issue concerning the nosology is the recommended exclusion of a diagnosis of OHS in the presence of "a known idiopathic central hypoventilation syndrome" (see later); such a diagnosis in an obese patient would require that the clinician affirm that the hypoventilation developed before obesity or persisted after resolution of obesity, an unlikely clinical possibility.

Idiopathic Central Alveolar Hypoventilation

In this disorder, the cause of sleep-related alveolar hypoventilation is unknown (Box 107-11). It is a diagnosis of exclusion, and the disorder appears to be rare. By definition, the hypoventilation is *not* primarily due to an identifiable cause of hypoventilation such as lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder, medication use, neurologic disorder, muscle weakness, obesity, or congenital hypoventilation syndromes. The congenital central hypoventilation syndrome usually manifests at birth, but late-onset cases have been described. The definitive diagnosis of congenital central hypoventilation syndrome requires demonstration of a mutation in the *PHOX2B* gene.[34]

Sleep-Related Hypoventilation Due to a Medication or Substance

In this disorder the sleep- related hypoventilation is due to a medication or substance that inhibits ventilatory drive (Box 107-12). Hypoventilation may be present during wakefulness but is not required for the diagnosis. The medications associated with the closely related diagnosis of CSA due to medication or substance (see earlier) may also be associated with sleep-related hypoventilation. Hypoventilation is not primarily due to another identifiable cause of hypoventilation, such as lung parenchymal or airway disease, pulmonary vascular pathology, chest wall disorder, neurologic disorder, muscle weakness, OHS, or a known congenital central alveolar hypoventilation syndrome. Although obstructive and central apneas may be present, the predominant respiratory pattern is usually one of reduced tidal volume or ataxic breathing and associated arterial oxygen desaturation. However, some patients may manifest frequent central apneas and also meet the criteria for sleep-related hypoventilation. When relevant criteria are met, as defined above, a diagnosis of OSA or CSA, or both, due to a medication or substance, as well as sleep-related hypoventilation due to a medication or substance, may be made.

Sleep-Related Hypoventilation Due to a Medical Disorder

In this disorder of breathing during sleep, sleep-related hypoventilation is believed to be due to a defined medical disorder (Box 107-13). Here the term "medical disorder" is inclusive and includes disorders of the lung parenchyma, airways, or pulmonary vasculature; chest wall disorders (other than mass loading for obesity); neurologic disorders such as neuromuscular disorders (disorders of the brain, spinal cord, or phrenic nerve); and myopathies.[35,36] By definition, hypoventilation is *not* associated with OHS or a central alveolar

Box 107-14 ICSD3 DIAGNOSTIC CRITERIA: SLEEP-RELATED HYPOXEMIA

Criteria A and B must be met:

A. Polysomnography, out-of-center sleep testing, or nocturnal oximetry shows the arterial oxygen saturation (Spo_2) during sleep of ≤88% in adults or ≤90% in children for ≥5 minutes.
B. Sleep-related hypoventilation is not documented.

hypoventilation syndrome. Although awake hypoventilation may be present, this is not required for this diagnosis. The predominant respiratory pattern is one of reduced tidal volume or ataxic breathing and associated arterial oxygen desaturation, rather than a predominant pattern of central or obstructive apnea. However, if appropriate criteria are met, as defined previously, a diagnosis of OSA or CSA, or both, due to a medical or neurologic condition, as well as sleep-related hypoventilation due to a medical disorder, may be made.

SLEEP-RELATED HYPOXEMIA DISORDER

This disorder has also been referred to as *nocturnal oxygen desaturation* and is based on either PSG, OCST, or oximetry findings (Box 107-14). Note that a diagnosis of sleep-related hypoventilation with such hypoxemia cannot be made if Pco_2 is not measured during sleep or, if measured, remains within normal limits. Conversely, if sleep-related hypoventilation is documented (as measured by arterial blood gas, transcutaneous Pco_2, or end-tidal CO_2 sensors), the disorder is classified as sleep-related hypoventilation rather than sleep-related hypoxemia disorder. OSA or CSA may be present, but such events are not directly associated with the majority of the sleep time spent with hypoxemia. Physiologic causes, if known, should be indicated (e.g., shunt, ventilation-perfusion [V/Q] mismatch, low mixed venous oxygen, and low partial pressure of inspired O_2, as with high altitude). In the ICSD3, sleep-related hypoxemic disorder is listed separately from the sleep-related hypoventilation disorders. Note, however, that most disorders associated with a diagnosis of sleep-related hypoventilation due to a medical disorder will be classified as sleep-related hypoxemia disorder if the sleeping $Paco_2$ or a surrogate has not been measured, or if such a measure does not meet criteria for sleep-related hypoventilation. For example, a patient with COPD (whether or not awake $Paco_2$ is elevated) undergoing nocturnal oximetry without a CO_2 measure may show significant nocturnal arterial oxygen desaturation; such a patient would fit criteria for sleep-related hypoxemia disorder. One could argue that this diagnostic category should be combined with the sleep-related hypoventilation disorders. In fact, the ICD-9 (International Classification of Diseases) diagnostic code 327.26 denotes sleep-related hypoxemia or hypercapnia associated with a condition classified elsewhere. Further, such hypoxemia during sleep may also be associated with sleep-related hypoventilation not clearly documented by a CO_2 sensor because of limitations in the technology of CO_2 sensors, which typically substitute for direct arterial blood gas measurement.

OTHER SLEEP RELATED BREATHING DISORDERS

As may be expected, a number of sleep related breathing disorders do not fit easily into the previous diagnostic categories. Many types of lung disease, including COPD, cystic fibrosis, asthma, interstitial lung disease, and pulmonary vascular disease (often with pulmonary arterial hypertension), may result in hypoxemia during sleep with or without documented awake or sleep-related hypoventilation. If sleep-related hypoventilation is not documented but nocturnal hypoxemia is present, a diagnosis of sleep-related hypoxemic disorder is made. Further, many of these clinical entities may "overlap" with either CSA or OSA. The most classically described overlap syndrome,[37] consisting of the combination of OSA and COPD, is a common condition that clearly does not fit easily into a diagnostic category. These patients may have episodes of obstructive events (apnea, hypopnea, or both) superimposed on a low baseline sleeping arterial oxygen saturation. Patients with the overlap syndrome would by definition meet diagnostic criteria for OSA, but the recognition of the role of COPD is important because patients benefit from treatment of both components. It has also been increasingly shown that treatment of the OSA component when present in combination with the previously noted pulmonary and pulmonary vascular disorders may also be necessary for optimal treatment of each entity.

CLINICAL PEARL

Sleep related breathing disorders encompass a wide and clinically important spectrum of disordered breathing and related medical morbidities, which often overlap. The ICSD3 has published new diagnostic criteria for sleep related breathing disorders. Separate diagnostic criteria are now available for CSA with CSB, OHS, and treatment emergent CSA. Definitions of specific respiratory events are not provided in the ICSD3, and the clinician must use the *AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* for such definitions. A diagnosis of treatment emergent CSA requires that the CSA during positive-pressure initiation is *not* better explained by another CSA disorder. In this sense, *treatment emergent CSA* is more specific than the widely used term *complex sleep apnea*. A diagnosis of sleep-related hypoventilation disorders requires measurement of $Paco_2$ or a surrogate of $Paco_2$ during sleep.

SUMMARY

A diagnosis of OSA in adults requires either 15 or more obstructive events (obstructive or mixed apneas, hypopneas, and RERAs) per hour or the combination of symptoms or comorbid conditions and 5 or more obstructive events per hour. The diagnoses of CSA with CSB, CSA due to a medical disorder without CSB, CSA due to a medication or substance, and primary CSA all require 5 or more central apneas or central hypopneas per hour of sleep, with at least 50% of the total number of respiratory events present during sleep being central apneas or central hypopneas. A diagnosis of sleep-related hypoventilation disorders requires demonstration of hypoventilation during sleep (in practice, usually assessed by a surrogate measure of arterial Pco_2 such as end-tidal Pco_2 or

transcutaneous Pco_2 monitoring). OHS requires demonstration of awake hypoventilation; most of these patients also have OSA. In the other sleep-related hypoventilation disorders, daytime hypoventilation may or may not be present. A diagnosis of sleep-related hypoventilation due to a medical disorder requires that hypoventilation is primarily due to a disorder of the lung parenchyma, airways, or pulmonary vasculature; chest wall disorders (other than mass loading for obesity); neurologic disorders (disorders of the brain, spinal cord, or phrenic nerve); or myopathy.

Selected Readings

Aboussouan LS. Sleep-disordered breathing in neuromuscular disease. *Am J Respir Crit Care Med* 2015;**191**(9):979–89.

American Academy of Sleep Medicine. *International classification of sleep disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.

Berry RB, Brooks R, Gamaldo CE, et al, for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications,* Version 2.2. Darien, Illinois: American Academy of Sleep Medicine; 2015 <www.aasmnet.org>.

Eckert DJ, Jordan AS, Merchia P, Malhotra A. Central sleep apnea: pathophysiology and treatment. *Chest* 2007;**131**:595–607.

Escourrou P, Grote L, Penzel T, et al; ESADA Study Group. The diagnostic method has a strong influence on classification of obstructive sleep apnea. *J Sleep Res* 2015;**24**(6):730–8.

Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;**3**:141–63.

Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. *Lancet* 2014;**383**(9918):736–47.

A complete reference list can be found online at ExpertConsult.com.

Sleep Breathing Disorders: Clinical Overview

Reena Mehra; Douglas E. Moul; Kingman P. Strohl

Chapter 108

Chapter Highlights

- Sleep breathing disorders (SBDs) is an umbrella term used to describe a range of breathing disorders of sleep that include obstructive sleep apnea, central sleep apnea, periodic breathing, and sleep-related hypoventilation.
- Overall, SBDs are highly prevalent in adults, with each type of disturbance having distinct pathophysiologic and clinical features. The widespread prevalence and scope of SBDs, afflicting more than 50 million Americans alone, along with the compounding adverse effects on medical, mental health, quality-of-life, and productivity outcomes, provides compelling motivation to design and engage in strategies to optimize their diagnosis and management.
- SBDs uniquely intersect with a host of medical, surgical, and neurologic comorbid conditions and directly affect health outcomes. Optimal management strategies for SBDs therefore must involve not only multidisciplinary expertise but also a team-based approach incorporating efforts from primary care physicians, medical and surgical subspecialists, and sleep specialists.
- Novel therapeutics will likely be forthcoming based on addressing the specific pathophysiologic attributes of SBDs, including critical closing pressure of the upper airway, upper airway muscle recruitment, arousal threshold, and loop gain.
- From a holistic standpoint, it is imperative to take into account the health status and life course of the patient when tailoring treatment strategies for an SBD such that these are effectively incorporated in the patient's personal life narrative.

The broad impact of disorders of respiration in sleep on human health and disease has spurred the development of major new conceptual frameworks and treatments over the past 50 years.[1] We now know in much greater detail how sleep is a critical biologic function for brain development and maintenance, behavioral vigilance, memory, and general well-being. Respiratory disturbances during sleep dramatically interfere with not only brain but also other organ functions, resulting in accompanying cardiovascular, endocrine, neurologic, and psychiatric disorders. The functional benefits of sleep reside in the amount and continuity of sleep and its various stages, effects that are inevitably disrupted in those with sleep breathing disorders (SBDs).

The purpose of this overview to the section on SBDs is to orient and represent a broad view of pulmonary-related sleep medicine and to provide a framework for consideration of the SBDs in a clinical preventative paradigm. More detailed and annotated work on each aspect of what we will cover will be found in the subsequent chapters of this section.

HISTORICAL PERSPECTIVE

SBD is an umbrella term used to describe a range of sleep-related breathing disorders that includes obstructive sleep apnea (OSA), central sleep apnea (CSA), periodic breathing (including Cheyne-Stokes breathing), and sleep-related hypoventilation (see Chapter 107). Historically, recurrent central or nonobstructive events (airflow and effort absent) were clinically identified as part of the pathobiology of congestive heart failure and central nervous system disorders, including stroke. Now OSA (airflow absent but efforts persist) is considered more prevalent, with each apnea type exhibiting somewhat different clinical presentations

OSA as a clinical entity began to be recognized in the 1960s when several groups in Europe reported observations on sleep in patients with excessive daytime sleepiness.[2] Suspecting upper airway obstruction as the cause for disturbed sleep, several groups took a radical step by performing tracheostomy as treatment. This regularized breathing during sleep, with the signs and symptoms of awake-time sleepiness[3] and cardiopulmonary disease resolving.[4] Other centers soon used the same diagnostic and therapeutic approach, confirming the disease ("upper airway apnea") in adults and children. In the 1970s, case series of OSA were reported, and the term *polysomnography* became widely used to describe the process of monitoring sleep and its associated movements and cardiopulmonary functions. Motta and colleagues in 1978 reported success with tracheostomy for OSA in a large case series.[5] In the 1980s there was an emphasis on the neurophysiology of respiratory control and the anatomic features of a vulnerable upper airway, the interactions of which created recurrent nonobstructive and obstructive apneas and hypopneas,

respectively.[6] OSA was still considered a rare and curious disease, to be considered when other causes for polycythemia, pulmonary hypertension, or extreme inability to stay awake were excluded. OSA treatments first described in 1981 included positive airway pressure (PAP)[7] and uvulopalatopharyngoplasty surgery[8] directed at abnormalities of the size and compliance of the nasopharynx and oropharynx.

In 1993, the first organized epidemiologic investigation published the figure of a 2% to 4% prevalence of OSA syndrome (elevated apnea-hypopnea index [AHI] with symptoms) in adults, placing OSA in the United States among common chronic medical conditions such as asthma and diabetes.[9] Since then, established clinical pathways for management have been developed in addition to improved methods of positive-pressure therapy delivery and to the alternative of oral appliance therapy. Outcomes of both the untreated and the treated state of OSA have been described in enough detail to articulate general thresholds for OSA severity: mild (AHI, 5 to 15), moderate (AHI, 15 to 30), and severe (AHI, >30).

Central sleep apnea and Hunter-Cheyne-Stokes breathing are among the SBDs that received initial attention in the early nineteenth century. Although these disorders are less prevalent than OSA, the symptom-based characteristics and physiologic perturbations are perhaps more readily apparent. In 1818, John Cheyne, a physician in Dublin, was one of the first to describe the disorder in a 60-year-old patient with breathing oscillations that he observed during sleep in the latter phases of the patient's illness. On postmortem examination, Cheyne discovered that the patient's heart was enlarged with fatty tissue invasion of the cardiac muscle.[10] In 1854 another physician from Dublin, William Stokes, went further and attributed the abnormal breathing pattern to a "weakened state of the heart."[11] For the purposes of historical accuracy, it should be noted that 37 years before Cheyne's report, John Hunter, an English surgeon, was the first to describe the classic crescendo-decrescendo breathing pattern: "The patient's respiration was extraordinary. He ceased to breathe for twenty or thirty seconds and then started breathing again. First faintly, and then with increasing force until the climax, after which his breathing again weakened until it disappeared completely."[12] Treatment options for CSA associated with Hunter-Cheyne-Stokes breathing initially included stabilization of the underlying contributing condition (e.g., optimizing heart failure management) and some drugs. More advanced technologies are currently available, including novel PAP approaches.

OVERVIEW OF THE PATHOPHYSIOLOGIC BASIS OF SLEEP BREATHING DISORDERS INFORMING THE TREATMENT APPROACH

Our review in 1987 described the clinical understanding of sleep apnea and the intent of different therapies to address suspected anatomic and neural factors in SBDs and the elements of clinical risk gleaned in the examination of the patient.[6,13] Although long-term prognostic generalizability has leapt forward in the past 25 years regarding SBDs more generally, the fine-grained understanding about findings on sleep studies still has areas of substantial ambiguity, particularly when patient-specific findings and preferences need to be incorporated into treatment planning. For example, two clinically salient domains of ambiguity concern treatment emergent CSA and residual sleepiness in the OSA patient who is fully compliant with continuous positive airway pressure (CPAP). Accordingly, the field has moved to more fundamental issues,[14-16] including the development of a process of separating and measuring features during sleep that correlate with the expression of recurrent apneas in an individual patient.[17,18] These features include (1) the pharyngeal critical closing pressure (Pcrit), (2) the recruitment (or "gain") function of muscles that keep the airway patent, (3) the threshold for arousal from sleep, and (4) the tendency for a disturbance to "set up" the person for a subsequent apnea ("loop gain") (see Chapters 109 and 122).

The pressure at which the compliant upper airway closes can be measured during sleep and calculated as the Pcrit. A positive pressure needed to keep the airway open is called a positive Pcrit and is the principle for using CPAP; if closing requires a negative pressure (as in a healthy person or in those with nonobstructive apneas or hypopneas), it is a negative Pcrit. Moving Pcrit to a more negative level is therefore beneficial in terms of treating OSA. Upper airway muscles can keep an airway open through brainstem activation using reflex mechanisms and direct actions on size or stiffness of the airway wall. Sleep reduces both activation and reflex recruitment, and in severe OSA there are additional reductions in reflex activations.[19-21] Another factor is the inherent gain of recruitment with the increasing chemoreceptor stimulation.[22] Furthermore, muscle efferent output must be translated into mechanical changes in the airway, which may be more difficult in the presence of edema.[23]

Arousal thresholds operate in SBDs in two ways.[24,25] Arousals are a mechanism that not only shorten events but also rapidly increase ventilation for any given carbon dioxide level. However, a longer time to arousal is thought to increase the chance for a person to increase drive to upper airway muscles, improve ventilation, maintain oxygenation, and reduce arousals during sleep,[24,26] permitting muscle activation to rise up to open the airway.

The fourth mechanism to discuss is loop gain, which is the propensity of the respiratory feedback control system to oscillate in response to a perturbation like a nonobstructive or obstructive apnea. A high loop gain indicates a relative inability for a person's control system to return to steady breathing after a disturbance such as an apnea, and a number of studies indicate that high loop gain is a factor in OSA and in the recurrent nonobstructive apneas of Cheyne-Stokes breathing. One major aspect of this control system relates to the effectors for ventilation (nerves, muscles, chest wall); another is related to the response organized by the brainstem to the disturbance. Oxygen administration can decrease loop gain and reduce OSA in some patients who have a high loop gain,[27] and indomethacin can increase loop gain and increase the tendency for apneas to reoccur.[28] CPAP itself will lower loop gain,[29] as the ventilatory response to CO_2 decreases (i.e., ventilation becomes less oscillatory) following treatment with CPAP.[22,30] Thus an individuality of and change in loop gain could be a mechanism of the appearance of emergent nonobstructive apnea, the variability among individuals in drug trials, the success of oxygen therapy, and obstructive and nonobstructive tendencies in disease presentation.

In summary, the anatomy of an apnea in sleep is determined in the following manner. There is a reduction in global respiratory muscle drive to create opportunity for an apnea. The nonobstructive apnea occurs when the upper airway has

a negative Pcrit, whereas the obstructive apnea occurs when there is a mechanical collapse with a reduction in drive to upper airway dilator muscles. In the middle of the apnea, drive increases to one degree or another, with both negative pressure and chemoreflex stimulus, until there occurs either an arousal or sufficient muscle activation to reopen the airway. The degree of overshoot in ventilation sets up the opportunity for a reduction in drive to set up the next apneas. Less well known are the factors that lead to the length of an apnea, the interapneic interval, and the cycle length between the onset of one apnea and the onset of the next.

All of these four factors contribute alone and in combination to the development and level of severity of both nonobstructive and obstructive apneas and will affect the response to therapy. There is some prospect that these elements will be ultimately measurable using a CPAP interface to clamp the airway pressure and initiate interventions to collect values related to muscle recruitment, arousals, and loop gain.[26] Currently, longitudinal studies of these factors and how they influence the development of sleep apneas over time are absent, and the collected traits at presentation are more epidemiologically oriented (e.g., obesity, hypertension) than based on data on causes (e.g., Pcrit, loop gain). In addition to the functional factors just discussed, some supportive evidence for risks associated with molecular markers associated with sympathetic activity, oxidative stress, and inflammation are present in some but not all cross-sectional studies in the presence of hypoxia-reoxygenation with recurrent obstructive and nonobstructive apneas.[31] Genetic and epigenetic factors will undoubtedly emerge, but currently the genetic studies are underpowered for common alleles and inconclusive for rare alleles sought in family studies.[32] Although some causal generalizability has arisen epidemiologically about the effects of SBDs, and etiologic frameworks have undergone robust clarifications, the current level of knowledge leaves the sleep clinician with limited guidance, relying mainly on modification of an abnormal Pcrit, and clinical management is based on algorithms rather than measures of the four pathophysiologic causes.

CLINICAL ASSESSMENTS AND DECISION MAKING

Detection of SBDs is at a historical inflection point. Rather than "rare, curious, or untreatable," SBDs are inevitably present in any clinical cohort. Treating SBDs improves quality of life and reduces cardiovascular risks. SBDs are common, treatable, and likely to some extent preventable. Because OSA is itself as common as asthma or diabetes, the sleep medicine specialist cannot take care of all SBD patients. The role of the specialist, like that of the diabetes or asthma specialist, is to assist in diagnosis and treatment of more complex cases and, in the future, to institute care paths to serve as a guide for the primary care physician or subspecialist for informed management of SBDs.

The prevalence and scope of SBDs, afflicting more 50 million Americans alone, along with its compounding effects on adverse medical, mental health, behavioral, and quality-of-life outcomes, as well as loss of productivity,[33] provides compelling motivation to engage in strategies to optimize the diagnosis and management of these conditions. SBDs have been notably underrecognized, with 85% to 90% of cases estimated to be undiagnosed.[34] This is partly attributable to limited awareness and understanding of the relevance of SBDs and to suboptimal guides and tools available for practitioners to efficiently diagnose and treat SBDs. A challenge facing many primary care physicians, our first-line "gatekeepers," is to choose which health problem to address for each patient, given clinical time constraints and patients' own health care preferences. In contrast, subspecialists face the challenges of addressing focused domains of care in a targeted and comprehensive manner, but not necessarily the composite of a patient's comorbidities. In the middle zone, treatment of SBDs encompasses somewhat specialized aspects of respiratory care but, owing to the prevalence and protean effects of untreated SBDs, necessitates the involvement of a wide range of medical specialties.

Hence there exist multiple opportunities to identify SBDs not only in everyday settings of health maintenance and disease prevention but also along the patient's trajectory through acute inpatient settings and subsequent outpatient follow-up. For example, while addressing a variety of common medical issues such as blood pressure and diabetes control, the patient and clinician may not prioritize or view SBD symptoms (e.g., snoring, daytime sleepiness) as clinically relevant or important. Nonetheless, a systematic approach for identifying and managing respiratory and nonrespiratory sleep disorders can have strategic importance for the effective longitudinal control of cardiovascular and cardiometabolic health consequences. Treatment of SBDs is part of a holistic approach of medical management and optimization of health care.

The clinical evaluation of a person suspected of having an SBD or being assessed after initiation of treatment requires a multisystem assessment. Because sleep patterns and preferences are determined by embodiments of personal, occupational, and social goals, effective care of SBDs can take on a complexity and richness not often needed for other organ-specific specialty assessments. Certain instruments, such as the Berlin Questionnaire and the STOP-BANG questionnaire, are designed to provide risk factor analysis as well as a psychometrically valid, generalizable suspicion of OSA. The Berlin Questionnaire is geared toward recognition of OSA in the primary care office, and it best considered for its negative predictive value, rather than as a predictor of severity of disease.[35] The STOP-BANG was developed for preoperative screening for general anesthesia in a population particularly at risk for OSA because of many chronic illnesses,[36] and it is gaining increasing use primarily because of ease of implementation; note, however, that the original STOP-BANG includes neck size measurement, which is often overlooked. These instruments highlight that the clinical skill set for the sleep specialist has come to include understanding of self-report scales and their utility, appropriate context, and limitations, especially in regard to determining the need for clinical management. Similarly, assessments of diagnostic tests such as polysomnography require a skill set in evaluating the plausibility of the test compared with pretest probability, the risk and difficulty of testing, the ability to initiate a treatment decision based on its result, and patient collaboration. Interpreting information, scales, and tests related to SBDs requires topic-related skills in identifying a study question, interpreting study design, and data analysis. Insurance providers and patients are asking tough questions about process and value. Sleep medicine specialists have to acquire these skills in communication, but translating some of these issues to other specialties remains a challenge for sleep clinicians.

Growing attention has been paid to research in health services and quality improvement in the treatment of SBDs. This trend has engaged funding agencies, health care delivery systems, and the government, and each has increasingly emphasized this as an area of focus as population-level health impacts become better known. Such research will increasingly shape the financing and administration of SBD-associated health care. As the volume of systematic research studies increases, assessing the nature and quality of the studies, particularly when they suggest contrasting findings, is essential for the development of wise clinical practices. Meta-analyses and systematic reviews can be helpful to integrate findings from several studies and appraise biases. Cost-effectiveness and comparative effectiveness analyses attempt to scale the real-world benefit of new interventions to current standards of care while simultaneously evaluating associated costs.

Incorporating proven state-of-the-art innovations at the bedside will continue to be an ethical priority in evaluating and treating SBDs in a clinical system that embraces continuous quality improvement. Important outcomes commonly evaluated include reduction of cost, reduction of morbidity and mortality, and improvement in patient satisfaction. A vital part of quality improvement is sustainability—creating systems to ensure the desired outcomes are regularly achieved. Accordingly, SBD specialists will need to oversee quality improvement projects to better understand how the process works, how to improve individual practice, and how to sustain successful practice in the changing health care environment.

SLEEP BREATHING DISORDERS IN DISEASE-SPECIFIC POPULATIONS

SBDs are uniquely comorbid with a whole host of major, common clinical conditions and medical settings, spanning, but not limited to, cardiac disease, obstructive and restrictive pulmonary disease, neurologic conditions, kidney disease, and the perioperative and anesthesia setting. The common clinical scenarios discussed in this section and highlighted in Table 108-1 provide a sense of the need for high-level awareness of SBDs and a low threshold for testing for, and managing, associated SBDs in such conditions and settings because an SBD often represents a modifiable risk factor for adverse health outcomes related to these.

Cardiovascular Disease

SBDs and cardiac disease have clear bidirectional relationships. As an example, a family medicine practitioner evaluates a patient with recalcitrant hypertension who meets the criteria for resistant hypertension. The practitioner notes that the patient snores and has a history of witnessed apneas. In this clinical scenario, it must be considered that concomitant and sustained increases in blood pressure are observed in OSA, likely through pathways involving hypoxia as a potent stimulus for sympathetic activation, hypercapnia, reductions in baroreflex sensitivity, and direct increases in sympathetic nervous system activation.[37,38] OSA-induced increases in activation of the renin-angiotensin-aldosterone system may in particular play a role in resistant hypertension. Results pooled across many randomized controlled trials designed to examine the effect of OSA reversal with CPAP on blood pressure outcomes consistently demonstrate statistically significant and clinically meaningful reductions in office systolic and diastolic blood pressure levels.[39] In fact, the Joint National Committee has recognized OSA as a secondary contributor to hypertension.[40] OSA has also been associated with nocturnal nondipping patterns[41,42] (i.e., does not undergo a standard dipping pattern of blood pressure reduction of at least 10% of the wake value), which is predictive of future cardiovascular risk.[43-45] Furthermore, resistant hypertension (approximately 15% of the hypertensive population[46]), which involves blood pressure above goal despite at least three antihypertensive medications,[47] is an entity in which OSA represents an important promulgating contributor. In fact, a large randomized clinical trial has demonstrated 3-month improvement in mean and diastolic blood pressure in those with resistant hypertension and OSA treated with CPAP versus no therapy.[48]

As another example, an internist admits a patient for management of acute non-ST elevation myocardial infarction; the patient undergoes coronary catheterization. Phenotypic characterization in this case most likely involves a background risk for dyslipidemia, hypertension, and diabetes mellitus. In terms of how undiagnosed SBD fits into this clinical paradigm, it should be recognized that SBD prevalence is greater than 70% in acute coronary syndrome and appears to be characterized by mainly obstructive respiratory physiology.[49] Furthermore, compared with patients without OSA, those with OSA and coronary artery disease have a higher degree of late lumen loss,[50] a marker for coronary restenosis after percutaneous coronary intervention, and there is evidence that OSA inflicts direct cardiac injury.[51] The prevalence of OSA exceeds that of other known risk factors for coronary artery disease on those who are revascularized and therefore represents a key modifiable therapeutic target to consider after coronary intervention.[52] Untreated severe OSA also predicts increased mortality and subsequent myocardial infarction in patients who have experienced an ST elevation myocardial infarction.[53] Similarly, CPAP treatment in those with cardiovascular risk appears to reduce cardiac event risk[54] and minimizes need for revascularization.[55] The treating physician also needs to consider the benefits of SBD treatment on cardiovascular risk factors (e.g., hypertension) to mitigate future adverse cardiovascular events.

Down the hall from the catheterization laboratory, in the heart failure unit, a cardiologist is managing a patient with acutely decompensated reduced ejection fraction heart failure (left ventricular ejection fraction of 30%) with diuresis and afterload reduction pharmacologic therapy. A high index of suspicion for SBD in systolic heart failure should be in place given the known high prevalence in this setting (ranging from 50% to 75%[56,57]), and a combination of OSA and CSA with Cheyne-Stokes breathing is typically represented. Symptom-based screening for SBDs in heart failure patients poses a challenge because standard symptoms of SBD such as excessive daytime sleepiness are often not present. This is likely because of the overall enhanced sympathetic nervous system state inherent in heart failure. Moreover, symptoms of nocturnal awakenings due to shortness of breath and nocturia may be attributed to heart failure rather than an SBD, potentially leading to underdiagnosis of an SBD. Hypoxia-induced impairments in myocardial oxygen delivery and increased cardiac strain from sympathetic activation are likely factors in the initiation and progression of SBD-related heart failure. The dire importance of the detection and treatment of an SBD in those admitted with acute decompensated heart

Table 108-1 Key Considerations in Specific Clinical Conditions and Settings

Clinical Condition	Key Considerations
Hypertension	• Sleep breathing disorder prevalence of 50% to 60% is driven predominantly by obstructive sleep apnea.[40] • Elevation in blood pressure in obstructive sleep apnea occurs both awake and in the sleep period. • Nondipping blood pressure patterns are observed in obstructive sleep apnea, which is predictive of adverse cardiovascular sequelae.[43-45] • More than 15 randomized controlled trials have demonstrated improvement in blood pressure profiles with continuous positive airway pressure.[39] • The Joint National Committee recognizes obstructive sleep apnea as a secondary contributor to systemic hypertension.[40]
Resistant hypertension	• Obstructive sleep apnea is associated with resistant hypertension (prevalence of 60% to 90%)[86] and biologic plausibility for modulation by a hyperaldosteronism state. • Interventional trial data support blood pressure improvement with treatment of obstructive sleep apnea with continuous positive airway pressure in resistant hypertension.[48]
Coronary artery disease	• The prevalence of sleep breathing disorders is approximately 70%, and they primarily involve obstructive physiology.[49] • Obstructive sleep apnea is associated with late lumen loss (marker for restenosis) after coronary intervention[50] and appears to inflict direct myocardial injury (i.e., increased troponin-I levels).[51] • The prevalence of obstructive sleep apnea as a risk factor in coronary artery disease exceeds that of other traditional risk factors such as hypertension and diabetes mellitus.[52] • Untreated severe obstructive sleep apnea is associated with increased cardiovascular-specific mortality in several large scale epidemiologic studies.[87,88]
Heart failure	• Sleep breathing disorder prevalence of 50% to 75% is characteristic; a combination of obstructive sleep apnea, central sleep apnea, and Cheyne-Stokes breathing is represented.[56,57] • Clinical presentation is a challenge because typical symptoms of excessive daytime sleepiness are not present (presumably owing to sympathetic nervous system activation) and symptoms of sleep-related nocturnal dyspnea overlap with symptoms characteristic of fluid overload and paroxysmal nocturnal dyspnea characteristic of heart failure. • Sleep breathing disorders in acute decompensated heart failure are associated with a 50% to 60% increase in mortality after discharge and increased hospital readmissions.[58] • The effect of improvement of sleep breathing disorders with adaptive servoventilation in heart failure on mortality and other outcomes is biologically tenable and is currently the focus of study in large multicenter interventional trials.
Atrial fibrillation	• The prevalence of sleep breathing disorders in atrial fibrillation is estimated to be approximately 50%, with a stronger magnitude of association with central than with obstructive sleep apnea.[63] • Autonomic nervous system fluctuations are strongly implicated in atrial arrhythmogenesis based on experimental data, as are the mechanisms of intermittent hypoxia, the resolution phase of hypercarbia, and intrathoracic pressure alterations resulting in direct mechanical effects on the thin-walled atria. • Several retrospective studies have demonstrated a significant decrease in atrial fibrillation recurrence after ablation or cardioversion with the treatment of sleep breathing disorders compared with no treatment.[60,61]
Chronic obstructive pulmonary disease	• The presence of concomitant chronic obstructive pulmonary disease and obstructive sleep apnea has been termed overlap syndrome. • Patients with overlap syndrome experience more profound hypoxia and pulmonary vasoconstriction compared with either disorder in isolation. • REM sleep likely represents a particular state of vulnerability for obstructive lung disease given enhanced cholinergic mediated bronchoconstriction and further blunting of the hypoxic and hypercapnic ventilatory drives. • Continuous positive airway pressure and bronchodilators should be used along with supplemental oxygen in overlap syndrome because this strategy appears to confer a survival advantage.[67]
Obesity-hypoventilation syndrome	• This is defined by obesity (body mass index >30 kg/m^2), resting hypercapnia, and is characteristically associated with sleep-related hypoventilation in the presence or absence of obstructive sleep apnea. • Obesity-hypoventilation syndrome is important to recognize because it is associated with right-sided heart failure and increased mortality compared with obstructive sleep apnea alone.[68]

Continued

Table 108-1 Key Considerations in Specific Clinical Conditions and Settings—cont'd

Clinical Condition	Key Considerations
Stroke	• The prevalence of sleep breathing disorders is approximately 50% to 70% in the setting of acute ischemic stroke, with representation by both obstructive and central sleep apnea.[69,70] • Sleep breathing disorders tend to improve over time after stroke; however, half of these individuals continue to have sleep breathing disorders 3 months after stroke.[69,70] • Moderate to severe sleep breathing disorders in stroke that remain untreated may hinder the poststroke recovery process and also impair cognitive function and have an association with depression, with apparent improvement with continuous positive airway pressure use.[76]
Renal insufficiency	• Sleep breathing disorder–related intermittent hypoxia contributes to activation of the systemic and renal-specific renin-angiotensin-aldosterone system. • Recognition of sleep breathing disorders in renal insufficiency remains a diagnostic challenge because of the often atypical presentation of symptoms. • The prevalence of sleep breathing disorders is approximately 50%, with obstructive and central disordered breathing contributions.[78] • In end-stage renal disease, there is evidence that nocturnal hemodialysis is more effective than traditional hemodialysis in the improvement of sleep breathing disorder parameters.[80]
Perioperative anesthesia	• Anesthetics may result in a particular disadvantage to a vulnerable upper airway in patients with obstructive sleep apnea owing to effects of anesthesia in terms of blunting the hypoxic and hypercapnic ventilatory drives and also reducing upper airway muscle tone.[81] • Data are accruing that suggest the importance of identifying sleep breathing disorders preoperatively in an effort to improve outcomes, including perioperative and postoperative morbidity and mortality, and of enhancing anesthesiology preparedness for handling a compromised upper airway and blunted ventilatory drive during intubation and anesthesia administration.

failure is underscored by a large study noting a 50% to 60% increase in postdischarge mortality in those with an SBD compared with those without.[58] Although treatment of SBDs in patients with heart failure has not unequivocally demonstrated mortality benefit,[59] existing data are limited given suboptimal reversibility of Cheyne-Stokes breathing pathophysiology in response to CPAP. On the other hand, improvement in secondary outcomes, including left ventricular function and exertional capacity, has been observed with CPAP, with potential advantages in terms of mitigating heart failure progression. Larger ongoing clinical trials using more sophisticated treatment modalities such as adaptive servoventilation (using variable pressure support delivery to stabilize oscillatory breathing patterns) in patients with CSA and Cheyne-Stokes breathing are underway to address this question more definitively.

Meanwhile, in the electrophysiology laboratory, an electrophysiologist prepares to perform pulmonary vein isolation ablation in a patient with atrial fibrillation. In addressing the important issue of sustainability of normal sinus rhythm after such ablation in a patient with a predisposed substrate, the possibility of an SBD should be strongly considered given numerous data that have accrued supporting reduced recurrence of atrial fibrillation with SBD treatment.[60,61] Albeit, to date, these data are based on retrospectively designed studies, the consistency and reproducibility of these findings are sufficiently compelling to lead the clinician here to assess and investigate the possibility of an SBD. Epidemiologic data indicate a stronger association of CSA compared with OSA as the SBD related to atrial fibrillation, independent of underlying self-reported heart failure,[62] with an overall prevalence of about 50%.[63] Autonomic nervous system fluctuations have a strong biologic basis in terms of arrhythmogenic propensity[64] and have been implicated in experimental models of obstructive apneas and also appear to exert effects in central apnea physiology given the accompanying sympathetic excitation observed.

Pulmonary Disease

A pulmonologist treating a patient with chronic obstructive pulmonary disease (COPD) also needs to be cognizant of the potential for a concomitant SBD, that is, the "overlap syndrome" (see Chapter 119), so coined given the prevalent merging of these two common pulmonary disorders and the distinctness of the pathophysiology occurring as a result of this overlap above and beyond each disorder alone.[65] Patients with overlap syndrome have more profound hypoxia and resultant nocturnal oxygen desaturation than those with either disorder alone and thereby theoretically have higher risk for pulmonary vasoconstriction and pulmonary hypertension. A modest degree of pulmonary hypertension in general has been observed in OSA, with immediate temporal influences of apneic events resulting in acute rises in pulmonary artery pressure from initiation to termination of the apneic event.[66] Rapid eye movement (REM) sleep, characterized by predominance during the latter part of the sleep cycle, in particular represents a state of vulnerability given the physiology of REM-related reduction in hypoxic and hypercapnic ventilatory drive, reduced muscle tone, and enhanced likelihood for bronchoconstriction due to REM-related parasympathetic tone. Consideration of nocturnal bronchodilators for the COPD and CPAP therapy for the OSA is warranted in light of data demonstrating that these modalities when used together confer a survival advantage compared with supplemental oxygen (long-term oxygen therapy) alone.[67]

Obesity-hypoventilation syndrome (OHS; see Chapter 120) is characterized by morbid obesity and awake resting hypercapnia accompanied by sleep-related hypoventilation with or without the presence of OSA. Importance of recognition of an SBD in OHS lies in its association with right-sided

heart failure and, if left untreated, increased mortality risk.[68] The degree of hypoxia is more profound and extensive in OHS with OSA than in OSA alone.

Other Comorbid States

A neurologist evaluates a middle-aged woman in the office subsequent to the patient suffering right-sided lacunar stroke with failure of improvement in functional outcome and depressed mood. An SBD (primarily OSA and/or CSA) in this setting is not only an independent predictor of ischemic stroke but also highly prevalent after stroke and accompanied by worse post-stroke outcomes. An SBD is common after ischemic stroke, affecting 50% to 70% of patients, and although over time the degree of SBD may improve, at least half of patients will have an SBD 3 months after the stroke.[69,70] Although both OSA and CSA may occur after a stroke, improvement over time is more pronounced for CSA than for OSA.[71] CSA represents a negative prognostic indicator after stroke[71] and is related to stroke severity as well as topography.[72,73] Although some epidemiologic data support a stronger association of SBDs and incident ischemic stroke in men,[74] other data in an Asian population suggest that younger women may be at higher stroke risk.[75] When of moderate or severe degree, SBDs can hinder the recovery and rehabilitation process. Identification and treatment of an SBD are important given associations of SBDs with depressed mood, cognitive dysfunction, and impairment of the ability to perform activities of daily living. Data suggest that treatment of SBDs may thwart compromise of neurologic and cognitive function in the stable poststroke phase.[76] Interestingly, stroke etiology subtypes, including cardioembolic, large artery, and small artery strokes, are not variable in terms of association with SBDs or their severity.[77]

Similarly, an elderly man with daytime sleepiness and snoring in the setting of chronic renal insufficiency is being seen by a nephrologist in clinic. Although sleep symptoms are common among patients with chronic renal insufficiency, there are challenges in terms of specificity of these symptoms in the detection of SBDs in this setting. It also appears that there are a fair number of patients with an SBD and chronic renal insufficiency who do not have standard symptoms but do have an SBD. Insults to the kidney occur because of SBD-related intermittent hypoxia, which activates not only the systemic but also the renal-specific renin-angiotensin system, resulting in progression of renal disease and thereby underscoring the importance of the detection and treatment of SBDs. In end-stage renal disease, an estimated 50% of patients are afflicted with an SBD.[78] In end-stage renal disease, both OSA and CSA are observed, with a prevalence apparently concordant to that observed in the setting of heart failure[79] and with a potential overlap in pathophysiology in terms of contribution of rostral neck edema. Nocturnal versus conventional hemodialysis has been demonstrated to improve both OSA and CSA physiology.[80]

A middle-aged obese man with hypertension and a neck circumference of 18 inches is being prepared to undergo an appendectomy under general anesthesia. He undergoes standard preoperative anesthesia testing. Per the STOP-BANG screening questionnaire, he meets high pretest probability for OSA (more than three items positive). Patients with OSA may exhibit anesthesia medication–related reduction in hypoxic and hypercapnic ventilatory, with increased sensitivity to sedatives and opioids. A detailed airway assessment should be performed in these cases, with preparedness tactics in place, including airway adjuncts such as oral, nasal, and laryngeal mask airways as well as use of the ramped-up position, which enhances visualization of the glottic aperture during intubation. Attention to postoperative management is also of critical importance; descriptive data indicate elevated residual neuromuscular blockade effect, postoperative hypoxia, and increased length of stay in patients at high risk for OSA compared with those without high pretest probability for OSA.[81]

ECOLOGIC FEATURES

Holistically speaking, SBDs will have a bearing on one's health status and life course. In relation to the project of understanding patients' personal narratives and reactions concerning SBDs and their treatment, it is evident that generalizing about the life course narratives of persons with SBDs is epidemiologically not feasible. Treating SBDs will always remain a practical art.

Patients individually fit suspected or diagnosed SBDs into personal narratives. On the population-level of understanding the effects of SBDs, one may certainly appreciate how compliance with treatment may be partially determined by social or cultural contexts, to yield different rates for morbidities and mortality across groups.

In considering the course of OSA and its treatment, for example, it is essential to appreciate that although the cardiovascular and cerebrovascular outcomes are typically "hard" (i.e., not subject to reappraisal),[82] the decision to accept an identity (here the issue is having an SBD) or a treatment (such as CPAP) is "soft" (i.e., continually subject to reevaluation).

There are many patients who have sleep studies, and who even go on to use CPAP (or other non-PAP therapies) for a period of time, but then stop using treatment, only to come back to consider treatment years later. The implication is that the goal for the clinician need not exclusively be immediate treatment compliance (although this is certainly a main issue for insurers and a worthy goal for a practitioner), but instead to place some credence on identifying with the patient's personal narrative as the patient confronts the diagnosis of an SBD. For many days' clinical work, doing the medical art well is measure enough, irrespective of the patient's decision or indecision. This is the approach taken with motivational interviewing,[83] which, curiously, turns out to be highly effective in interacting with patients across a broad range of clinical presentations, severity, and response to treatment suggestions.

It should also be appreciated that the chronic sleep deprivation and neurocognitive sequelae associated with chronic OSA have an effect on patients' comprehension and treatment compliance. The frontal lobes are the most affected by sleep deprivation. In clinical practice, lowering of frontal functioning from chronic sleep deprivation has been observed to degrade a patient's capacity for maintaining personal morale, regulating affect in response to daily events, thinking more reflectively, remembering well, and relating to others. Anecdotally, some SBD patients can thus present as having psychiatric or personality disorders in clinical encounters, leading to a delayed or missed diagnosis of an SBD. Many presenting patients have impaired comprehension of health care information and physician recommendations. Additionally, it has been long known that for some patients, sleep deprivation may increase anxiety. This anxiety might be generalized or conditioned to nighttime and sleep stimuli; for example, for OSA

patients, each breath could be a choking experience. The sleepy patient may not have sufficient judgment and free will when making health care decisions about OSA treatments. Yet, for some who are diagnosed and treated, the mental health symptoms will be substantially improved and occasionally cured, owing to the benefits of getting uninterrupted wholesome sleep and reduced oxidative stress.

The clinician should also consider that comorbid disorders may disturb sleep to such an extent that compliance with treatment for an SBD becomes difficult if not impossible. This may arise, for example, if the patient has comorbid insomnia. However, other comorbidities can likewise provide the context for impairing good sleep. These include psychiatric conditions, cardiorespiratory conditions, pain, aging, and the use of particular medications such as alerting agents (e.g., stimulants, with some such as pseudoephedrine even available over the counter), anticholinergics, corticosteroids, and adrenergic inhalers.

The aging process presents interesting clinical issues regarding the SBDs. As middle age advances into older age, populations become increasingly biologically and clinically diverse throughout this span of time in relation to the burden of comorbidities, including those that an SBD might affect. Additionally, there is the consideration that the major measure of the sleep apneas, the AHI, may have normative ranges that increase with aging in adults. Reynolds and colleagues noted that in a cohort of patients advancing though the age span from 50 to 80 years, the AHI according to 1980s scoring rules increased per decade to a plateau at about 70 years of age.[84] These studies of recruited samples of healthy aging would give the impression that aging populations have AHIs that are biologically normal for their age although elevated for younger age groups. Indeed, there is debate about the optimal approach to treatment of SBDs, particularly OSA, in aging populations, including whether there might be an acceptable AHI level, perhaps 15 events per hour, which could be medically insignificant in older patients. However, although in practice it seems likely that some increase in AHI might be overall benign in some patients; aging patients should be considered to require the same attention to clinical symptoms and the comorbid milieu as younger populations in considering the diagnosis and treatment of the various SBDs. The recent literature on reduced sleep, SBDs, and hypoxemia playing an initiating or amplifying role in the development of Alzheimer disease and other dementias (including ischemia-related stroke)[85] reinforces such a clinical strategy.

CLINICAL PEARLS

- SBDs are a modifiable risk to be considered in numerous cardiologic settings, including treatment of resistant hypertension, minimization of risk for restenosis after percutaneous coronary intervention, mitigation of progression of heart failure and its adverse outcomes, and reduction in atrial fibrillation recurrence after intervention.
- SBDs also contribute to morbidity accompanying pulmonary disorders, including COPD, pulmonary hypertension, and OHS.
- SBDs are increasingly recognized as not only an independent predictor of acute stroke but also an obstacle in the poststroke recovery and rehabilitation process.
- Assessment of an SBD in the preoperative anesthesia setting is imperative to ensure implementation of optimal preparedness strategies for the perioperative and postoperative settings, including providing availability of adjunct airways and postoperative respiratory monitoring.
- Upper airway and central physiologic characteristics that predispose to the propagation of apneas and hypopneas in OSA include the critical closing pressure of the upper airway, upper airway muscle recruitment, arousal thresholds, and loop gain.

SUMMARY

In the past 20 years, treating the SBDs has developed from a medical curiosity to an epidemiologic mandate. Such a mandate for sleep medicine encompasses not only OSA but also other numerous forms of sleep-related respiratory disturbances (see Chapter 107). Despite the evidence base, however, patients and many medical professionals still have a dim appreciation of current knowledge about SBDs, their recognition, and their response to therapy. Even as physiologic knowledge of SBDs advances, there remains a continuing need for the expansion of public and medical education about SBDs and a continuing need for developing decision supports about SBDs for nonsleep clinicians as well as sleep specialists. That said, there is every reason to be confident that the physiologies of SBDs will continue to be explored in ever-increasing depth and sophistication. The four-factor functional model for OSA as discussed earlier will motivate attempts to apply a causal, systematic analysis of SBDs, based on measurable physiologic variables. Biochemical, genetic, and epigenetic variables related to the SBDs will also need to be explored more broadly, particularly those that have tie-ins to other highly prevalent cardiovascular, respiratory, and metabolic health conditions.

Selected Readings

Dempsey JA, et al. Pathophysiology of sleep apnea. *Physiol Rev* 2010; **90**(1):47–112.

Dewan NA, Nieto FJ, Somers VK. Intermittent hypoxemia and OSA: implications for comorbidities. *Chest* 2015;**147**(1):266–74.

Institute of Medicine Committee on Sleep Medicine and Research, Colten HR, Altevolgt BM, editors. *Sleep disorders and sleep deprivation: an unmet public health problem*. Washington DC: National Academies Press; 2006. p. 3, Extent and Health Consequences of Chronic Sleep Loss and Sleep Disorders. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK19961/>.

Kapur V, et al. Underdiagnosis of sleep apnea syndrome in U.S. communities. *Sleep Breath* 2002;**6**(2):49–54.

Malhotra A, Orr JE, Owens RL. On the cutting edge of obstructive sleep apnoea: where next? *Lancet Respir Med* 2015;**3**(5):397–403.

Mansukhani MP, Wang S, Somers VK. Chemoreflex physiology and implications for sleep apnoea: insights from studies in humans. *Exp Physiol* 2015;**100**(2):130–5.

Somers VK, et al. Sympathetic neural mechanisms in obstructive sleep apnea. *J Clin Invest* 1995;**96**(4):1897–904.

Strohl KP. Sleep medicine training across the spectrum. *Chest* 2011;**139**(5):1221–31.

Strohl KP, Cherniack NS, Gothe B. Physiologic basis of therapy for sleep apnea. *Am Rev Respir Dis* 1986;**134**(4):791–802.

Wolf J, Drozdowski J, Czechowicz K, et al. Effect of beta-blocker therapy on heart rate response in patients with hypertension and newly diagnosed untreated obstructive sleep apnea syndrome. *Int J Cardiol* 2016; **202**:67–72.

A complete reference list can be found online at ExpertConsult.com.

Chapter 109

Central Sleep Apnea: Definitions, Pathophysiology, Genetics, and Epidemiology

Madalina Macrea; Eliot S. Katz; Atul Malhotra

Chapter Highlights

- The various clinical entities comprising sleep breathing disorders are the result of pathophysiologic mechanisms that frequently overlap. Defining the types of central sleep apnea (CSA) within the sleep breathing disorder spectrum is essential for a common language among clinicians, educators, and researchers.
- CSA includes several heterogeneous syndromes, many heavily represented in day-to-day medical practice. Recent scientific evidence allows a more comprehensive understanding of CSA epidemiology, genetics, pathophysiology, and associated morbidity and mortality.
- Several mechanisms participate in the control of breathing. Chemical, mechanical, and neural pathophysiology involved in CSA, and their clinical implications, are detailed.
- Considerable progress has been made in our understanding of control of breathing related to CSA, with major implications of these new findings for patient care. Only by further mechanistic research are new therapeutic strategies likely to emerge.

Sleep breathing disorders (SBDs) is characterized by repetitive periods of cessation in breathing (i.e., apneas) or reductions in breathing (i.e., hypopneas) that occur during sleep. The various clinical entities belonging to SBDs are the result of pathophysiologic mechanisms that frequently overlap; centrally driven events are primarily due to a temporary loss of output from the pontomedullary pacemaker that generates breathing rhythm, resulting in loss of the respiratory pump muscles (diaphragm, thorax, abdomen). Alternatively, obstructive events are primarily due to inward collapse of the oropharynx when the pharyngeal dilator muscles are relaxed, resulting in loss of airflow because of upper airway narrowing.[1] Both obstructive and central respiratory events converge in their symptoms of frequent nocturnal awakenings and excessive daytime sleepiness.

Defining the end points of SBDs polysomnographically (i.e., central sleep apnea [CSA] and obstructive sleep apnea [OSA]), is often a straightforward process. In CSA, both oronasal flow and thoracoabdominal excursions are absent; that is, there is an absence of respiratory effort during the cessation of airflow, whereas in OSA, there are ongoing respiratory efforts during the absence of oronasal flow. In contrast, differentiating rigorously between events within the SBD spectrum (i.e., "central" and "obstructive" hypopnea) is difficult without quantification of respiratory effort as recorded by esophageal pressure monitoring. Because esophageal manometry is mildly invasive and rarely employed clinically, thoracic and abdominal excursions assessed by respiratory inductance plethysmography are widely used to detect asynchrony of these excursions during hypopnea (consistent with obstruction) or in-phase breathing (consistent with decreased central drive). Thus, central hypopnea is characterized by a proportional and synchronous decrease in thoracic and abdominal excursions, whereas obstructive hypopnea is characterized by paradoxical inward rib cage movement or asynchronous decrease in the thoracic and abdominal excursions (Figure 109-1). Nasal pressure recordings are sometimes used as a surrogate for upper airway narrowing because inspiratory flattening has been shown to correspond with inspiratory flow limitation. Additionally, obstructive and central apneas may overlap within the same event: such "mixed" apneas have features of both conditions, when an apnea begins with loss of central drive to breathe ("central" apnea) but then proceeds with increasing effort against an occluded upper airway ("obstructive" apnea).

DEFINITIONS

As defined by the *International Classification of Sleep Disorders*, third edition (ICSD3), CSA includes six heterogeneous adult syndromes (Box 109-1).[2,3] Several of these have in common a waxing and waning ventilatory pattern.

1. *Cheyne-Stokes breathing* (CSB) is an abnormal pattern of breathing characterized by oscillations of tidal volume between apnea or hypopnea at the nadir of ventilation and hyperpnea at the height of ventilation, with a spindle-like crescendo-decrescendo pattern in the depth of breathing.[4] According to the American Academy of Sleep Medicine

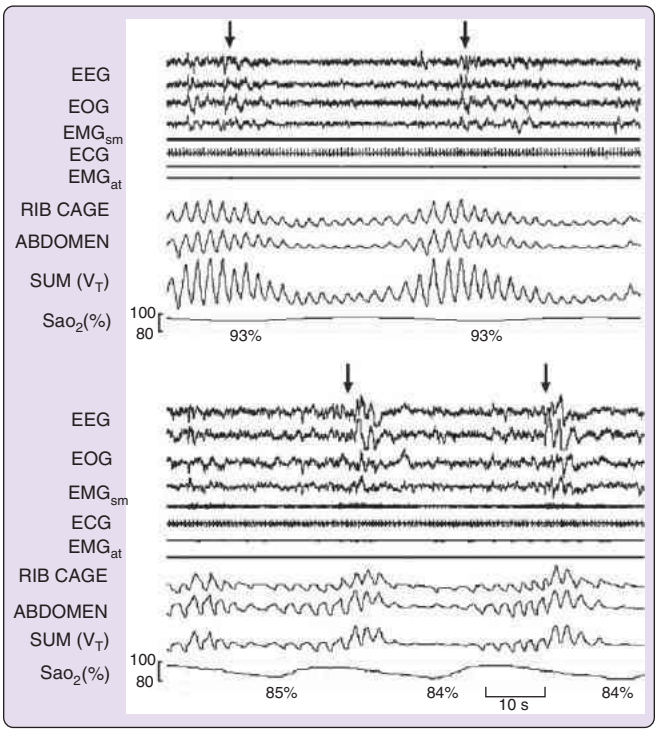


Figure 109-1 Polysomnographic recordings of central and obstructive hypopneas from patients with heart failure with use of respiratory inductance plethysmography. The *upper panel* shows a central hypopnea during stage 2 NREM sleep in a patient who has central sleep apnea with Cheyne-Stokes breathing. Note in-phase gradual waxing and waning of tidal volume during hyperpnea and only minimal O_2 desaturation during hypopnea. Arousal occurs several breaths after termination of the hypopnea. The *lower panel* shows an obstructive hypopnea in a patient with obstructive sleep apnea. Note that in contrast to central hypopnea, rib cage and abdominal motion are out-of-phase and O_2 desaturation is greater during hypopnea, and the rise in ventilation following its termination is more abrupt and hyperpneas are shorter. In addition, arousals occur earlier at hypopnea termination. ECG, Electrocardiogram; EEG, electroencephalogram; EMG_{sm}, submental electromyogram; EMG_{at}, anterior tibial EMG; EOG, electrooculogram. Arrows (↓) indicate arousals. (From Central Sleep Apnea and Cheyne-Stokes Respiration, Volume 5, Issue 2, The Proceedings of the American Thoracic Society.)

Box 109-1 HETEROGENEOUS ADULT SYNDROMES OF CENTRAL SLEEP APNEA

Central sleep apnea with Cheyne-Stokes breathing
Central sleep apnea due to a medical disorder without Cheyne-Stokes breathing
Central sleep apnea due to high-altitude periodic breathing
Central sleep apnea due to a medication or substance
Primary central sleep apnea
Treatment emergent central sleep apnea

(AASM),[5] CSB in adults is scored when both of the following are met: (1) there are episodes of three or more consecutive central apneas or central hypopneas, or both, separated by a crescendo and decrescendo change in breathing amplitude with a cycle length of at least 40 seconds (typically 45 to 90 seconds), and (2) there are five or more central apneas or central hypopneas, or both, per hour associated with the crescendo and decrescendo breathing pattern recorded over a minimum of 2 hours of monitoring. In terms of nocturnal oxygen desaturation, there is generally less desaturation during central apnea and hypopnea than during obstructive events in patients with heart failure.[6]

2. *Primary CSA* resembles CSA-CSB except that the cycle duration is shorter, arousals occur earlier (at the termination of apnea versus during or near the peak ventilatory effort), and resumption of breathing is more abrupt and not crescendo, typically with a large-volume breath. The patient must not be hypercapnic while awake ($Paco_2$ greater than 45 mm Hg). The diagnostic polysomnography (PSG) shows five or more apneic episodes per hour of sleep, the number of central apneas or central hypopneas more than 50% of the total number of apneas and hypopneas, and absence of CSB.
3. *High-altitude periodic breathing* is seen in normal persons at elevations greater than 7600 meters and in some at lower altitudes (see Chapter 122). This ventilatory pattern is characterized by periods of alternating hyperpnea and apnea,[7] the cycle length typically being between 12 and 34 seconds. PSG, if performed, demonstrates recurrent central apneas or hypopneas primarily during NREM sleep at a frequency of five or more per hour.
4. *CSA due to a medical condition, without CSB* is encountered in individuals with cardiac, renal, and neuromuscular disease who have CSA without the CSB.
5. *Central sleep apnea due to a medication or substance* is commonly seen in patients with long-term opioid use that causes respiratory depression by acting on the μ receptors of the ventral medulla. PSG demonstrates lack of CSB and five or more central apneas or central hypopneas, or both,[1] per hour of sleep, with the number of central apneas or central hypopneas, or both, greater than the total number of apneas and hypopneas.
6. *Treatment emergent central apnea* (or "complex" sleep apnea) is included in the ICSD3 and refers to CSA not explained by another CSA disorder (e.g., CSA with CSB or CSA due to a medication or substance). The diagnostic PSG shows five or more predominantly obstructive respiratory events per hour of sleep. The titration PSG without a backup rate shows resolution of obstructive events and emergence or persistence of central apnea or central hypopnea with both a central apnea-central hypopnea index (CAHI) of five or more per hour and number of central apneas and central hypopneas 50% or greater than total number of apneas and hypopneas.

PATHOPHYSIOLOGY

As Cherniack[8] noted in the early 1980s, breathing in an awake healthy person involves a smooth and regularly recurring sequence of inspiration and expiration without pauses. The rate and depth of breathing are regulated by a negative-feedback control system aimed at maintaining arterial partial pressures of carbon dioxide ($Paco_2$) and oxygen (Pao_2) at relatively constant levels. When diseases of the lung or chest wall produce hypoxemia and hypocapnia or hypercapnia, they usually do so without affecting the regularity of breathing. Several mechanisms and their corresponding controls influence the rhythmicity of breathing. A synopsis of the following

roadmap we used in the discussion of CSA pathophysiology and its clinical translation is detailed in Tables 109-1 and 109-2.

Mechanisms

Several types of receptors and their associated afferent and efferent pathways are involved in maintaining the regular normal breathing.

Chemical Aspects of Ventilation

Ventilatory responses vary widely between the awake and asleep state, as well as between rapid eye movement (REM) and non–rapid eye movement (NREM) sleep. Ventilation during sleep is largely regulated by the same mechanisms that drive breathing while awake,[9] except that behavioral influences[10] become suppressed in transition to, and during, sleep. Therefore central apneic events are rarely present during the awake state[11,12] or REM sleep.[13] During NREM sleep, however, changes in the respiratory pattern are primarily controlled chemically, being the result of a fine balance among a critical $Paco_2$ level, below which there is a central cessation of breathing (i.e., apneic threshold); its triggering factors (mainly hypocapnia); and respondent receptors (i.e., central and peripheral chemoreceptors). Additionally, the level of ventilation in respiratory dysrhythmias is augmented by arousals from sleep, resulting in transient hyperventilation with hypocapnia below the apneic threshold[14] and therefore initiation of central apneic events, primarily during NREM sleep.

Table 109-1 Common Non-cardiac Conditions Associated with Central Sleep Apnea Events

Medical Condition	Prevalence (%)	Authors
Multiple sclerosis	18	Braley et al.[147]
Central nervous system tumor survivors	12.9	Mandrell et al.[148]
Cerebrovascular accident	7	Johnson et al.[149]
Congenital muscular dystrophies	55	Pinard et al.[150]
End-stage renal disease on hemodialysis	17	Tada et al.[151]
Diabetes	3.8	Resnick et al.[152]

Table 109-2 Roadmap Used in Discussion of the CSA Pathophysiology

Mechanism	Control	Clinicopathophysiologic Translation
Chemical	Metabolic	Cheyne-Stokes breathing OHS Sleep transition Apnea CCHS
Mechanical	Metabolic Neural	Muscular degenerative Postarousal/postsigh central apnea
Neural	Neural	Stroke CCHS

Hypoxic Stimulus and Peripheral and Central Chemoreceptors. The chemoreceptors involved in the ventilatory response are both peripheral and central, each of them responding to changes in arterial Po_2 *and* Pco_2 in a complex, interactive manner. In mammals, the peripheral chemoreceptors are represented by the aortic and carotid bodies. The carotid bodies represent the main drive of the ventilatory stimulation due to acute, chronic,[15,16] and intermittent[17] hypoxia and contain the glomus cells that respond to the changes in the arterial blood oxygen concentration through several neurotransmitters, such as acetylcholine, substance P, and adenosine triphosphate.[18] The aortic bodies, on the other hand, likely become upregulated only if the carotid bodies are chronically absent, and then respond to changes in the arterial Po_2[19] through mechanisms that are less known. Notably, studies in patients with long-standing OSA demonstrated the possibility of the carotid bodies becoming desensitized with extended exposure to intermittent hypoxia.[20-22] In comparison with the peripheral chemoreceptors, the central chemoreceptors have a wide anatomic distribution in the brainstem (especially in nucleus tractus solitarius [NTS], locus coeruleus, raphe nuclei, and the retrotrapezoid nucleus [RTN]) and respond to central nervous system–specific hypoxia by augmentation of alveolar ventilation during both wakefulness[23] and sleep.[24]

Hypercapnic Stimulus and Peripheral and Central Chemoreceptors. In addition to responding to hypoxia, the carotid bodies also act as a sensitive detector of the adequacy of alveolar ventilation,[25] as seen in the prompt ventilatory response to small increases in arterial Pco_2 and insensitive feedback to decreasing arterial Po_2 until it reaches the critical value of 50 to 60 mm Hg.[26] Quantitatively, assuming a purely additive model, Forster estimated that 40% of the steady-state ventilatory CO_2 response belongs to the carotid body and 60% to the central chemoreceptors, the carotid body providing its response prompter.[27,28] In the absence of an exact biologic definition for the central chemoreceptors, most consideration is given to the possibility that these cells are glial or vascular cells that regulate the activity of surrounding neurons through paracrine mechanisms and respond promptly to the changes of the local neuronal pH.[29,30] Anatomically, the RTN is considered to be the predominant location of integration of the central chemoreceptor drive.[31]

Apneic Threshold and Implications for Central Sleep Apnea. During NREM sleep, motor output to respiratory muscles is dramatically reduced compared with wakefulness, causing mild to moderate sustained hypoventilation in all healthy subjects (+2 to +8 mm Hg $Paco_2$). If relative hyperpnea occurs and $Paco_2$ falls below a characteristic value for each individual (the apneic threshold), a central apnea occurs (see Fig. 109-2).[32] The hypocapnia-induced apneic threshold is not a constant value but usually occurs at a level very close to the eupneic $Paco_2$ present during wakefulness following a very small reduction in $Paco_2$, from 2 to 5 mm.

Mechanistically, to reach the apneic threshold, transient ventilatory overshoots are necessary, commonly provided by transient arousals with consequent brief hyperpnea with hypocapnia. Alternatively, to overcome the apneic threshold

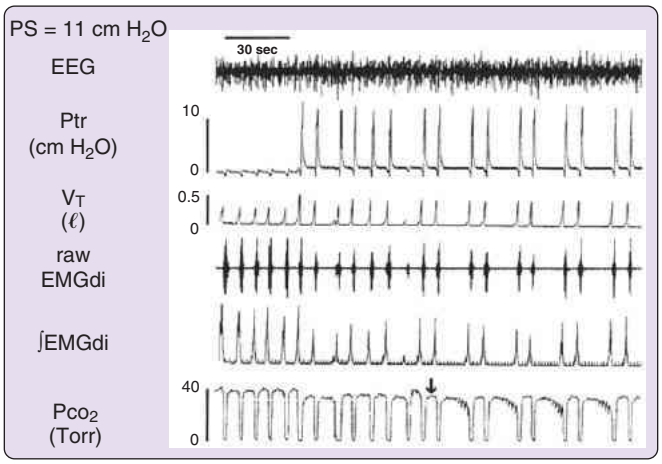


Figure 109-2 Polygraph record of one pressure support (PS) trial (11 cm H_2O) in which ventilatory instability was achieved in dogs. A reduced diaphragmatic EMG (EMGdi) and inspiratory effort on the seventh ventilator cycle was insufficient to trigger a ventilator breath. Clear periodicity developed after the ninth ventilator cycle. The *arrow* marks the petCO2 considered to be the apneic threshold. Ptr, Tracheal pressure. (Modified from Nakayama H, Smith CA, Rodman JR, et al. Effect of ventilatory drive on carbon dioxide sensitivity below eupnea during sleep. *Am J Respir Crit Care Med* 2002;165:1251–61.)

and therefore reinitiate the breathing rhythm, a $Paco_2$ higher by 1 to 4 mm Hg than the apneic threshold is needed; this difference reflects a postapneic control system termed *inertia*, aimed at enhancing the chemoreceptor stimulus after the ventilatory overshoot.

Interactions between the Central and Peripheral Chemoreceptors. Several anatomic and functional connections (e.g., the RTN receives direct input from the NTS and, thus, the carotid body)[33] serve a dual chemotactic role (peripheral and central), raising the question of interdependence between the two types of chemoreceptors. Failing to demonstrate unequivocally the existence of only one model because of variations in the experimental protocol, the literature describes three possible interactions: additive (the two responses simply sum), hyperadditive (the two responses multiply), or hypoadditive (the sum of each response is less than their mathematical sum). Regardless of the specifics of the final augmentative response, it is postulated that carotid chemoreceptors act as the immediate hypocapnic sensors,[34] given the lack of short-term response of the central chemoreceptors to systemic hypocapnia when normocapnia and normoxia are maintained at the level of the carotid bodies.[35] However, peripheral receptors do not primarily induce hypocapnic apnea by themselves, as demonstrated by experimental models involving isolated carotid body hypocapnia that fails to result in apnea.[36] Therefore it appears that both central and peripheral chemoreceptors must interact and respond to hypocapnia for the ventilatory overshoot to induce central apnea during sleep.

Mechanical Aspects of Ventilation

Dysfunction of upper airway mechanics represents the basis of OSA pathophysiology. Such dysfunction, however, is also observed in CSA because of upper airway collapsibility resulting in ventilatory instability.

The two primary collapsing forces of the upper airway are intraluminal negative pressure (generated by the diaphragm during inspiration) and extraluminal soft tissues (e.g., generated by fat deposition within bony structures surrounding the airway). These forces are opposed primarily by the pharyngeal dilator muscles, whose activity either varies from breath to breath (phasic respiratory muscles such as genioglossus) or stays similar throughout the respiratory cycle (tonic muscle, such as the tensor palatini). Additionally, activity of these muscles is dependent on mechanoreceptor and chemoreceptive influences. Studies in animals have shown that chemoreceptor activation resulted in an augmented depolarization of the inspiratory and expiratory hypoglossal motoneurons, thus providing evidence for the arterial chemoreceptors' contribution to maintaining upper airway patency throughout the respiratory cycle.[37] Additionally, the activity of the most important pharyngeal dilator muscle, the genioglossus, is accentuated by hypoxia and abolished by hyperoxia.[38] Alternatively, hypercapnia at the level of peripheral and central chemoreceptors leads to increased afferents to hypoglossal motoneurons and decreased threshold of the genioglossus activation. Comparing the quantitative participation of mechanoreceptors with chemoreceptors as modulators of upper airway activation, it has been suggested that chemoreceptors are stronger, although the two stimuli in combination may interactively augment upper airway muscle activity more than either stimulus alone.[39] Besides the upper airway muscles, the fluctuations in chemical stimuli also affect the diaphragm. Animal studies demonstrate a linear chemoreceptor-driven recruitment of the diaphragm electromyogram.[40] Endoscopy performed during both induced and naturally occurring central apnea demonstrated that upper airway obstruction occurs without evidence of an inspiratory effort in the first 10 seconds of a CSA episode.[41] Consequently, neuromuscular respiratory pathology overlaps in these different types of SBDs, making a clear adjudication between obstructive and central events difficult in many cases.

Neural Aspects of Ventilation

The respiratory neurons are divided into two groups, inspiratory and expiratory. The former belong to the dorsal respiratory group localized in the area of the NTS; the latter belong to the ventral respiratory group localized adjacent to the nucleus ambiguous.[42] Although the hypoxic and hypercapnic afferent responses of the peripheral and central chemoreceptors activate certain populations of respiratory neurons, the details of such intricate processes are still missing; likewise, the relative contribution of each neuronal population to central apnea pathogenesis is also unknown.[43] Studies of several congenital disorders have provided information on the central apnea neuronal ventilatory impairment and helped in understanding better the sudden infant death syndrome. Such rare congenital diseases include Leigh syndrome, a mitochondrial encephalopathy whose manifestations include frequent post-sigh apneic episodes due to lung stretch receptors ending their vagal afferents into abnormal NTS[44]; and Fukuyama-type congenital muscular dystrophy, in which sudden death is commonly encountered as a result of migration defects of the brainstem structures involving pathology of the arcuate nucleus that acts as a central chemoreceptor sensitive to hypercapnia.[45]

As Harper and colleagues[46] reviewed recently, however, CSB with or without CSA affects the brain structure and function beyond the rhythmicity of breathing and includes

hormonal, autonomic, and behavioral (affect, memory, and cognition) functions. Neural injuries of the ventrolateral and dorsal medullary areas are common in heart failure patients who demonstrate CSB patterns with or without CSA, affecting the final pathway of sympathetic outflow and sympathetic tone regulation.[47] Congenital central alveolar hypoventilation syndrome (CCAHS) neuropathology also involves the ventrolateral medulla, with subsequent dysfunction of the respiratory phase switch.[48] Neurotransmitter system injuries have been also recognized in CCAHS, in the raphe system, locus coeruleus (noradrenergic neurons), ventral midbrain, hypothalamus, and basal ganglia (dopaminergic fibers).[49] The cerebral cortex is not spared in either of these conditions; ischemic damage to the right insula is significantly accentuated in heart failure patients who have a high prevalence of OSA and CSB.[50] Essentially, because of hippocampal injury, short-term memory and cognitive impairments are also common in CCAHS.

Ventilatory Control in Central Sleep Apnea

Metabolic Control of Ventilation

Normal respiratory rhythmicity during sleep is maintained by a complex feedback mechanism best described by the concept of "loop gain," with CO_2 responsiveness between the eupnea and apneic threshold contributing.[51] *Loop gain* is an engineering term that describes the dynamic feedback of several stabilizing ventilatory mechanisms composed of three elements: (1) the controller gain (chemoresponsiveness, including ventilatory response to Pao_2 and $Paco_2$ above and below eupnea); (2) the plant gain (effectiveness of CO_2 excretion from such ventilatory response); and (3) mixing gain (e.g., from circulation delay between the lungs and the peripheral and central chemoreceptors). Simplistically, the loop gain could be defined as the ratio of the amplitude of the ventilatory response to a ventilatory disturbance. A loop gain of less than 1 accompanies a stable ventilatory system with low respiratory variability because disturbances lead to smaller responses, assuring a rapid return to a stable pattern. On the contrary, a loop gain of greater than 1 accompanies an unstable ventilatory system with high respiratory variability, in which disturbances lead to disproportionately large responses, resulting in a perpetual waxing and waning pattern. The ventilatory control system is dynamic, with both chemical and nonchemical inputs contributing to the breath-by-breath variability, as detailed by Khoo[52] (Figure 109-3).

Ventilatory variability is due to the feedback instability of the hypoxic and hypercapnic chemosensitivities. For example, a transient episode of hyperpnea leads to an eventual decrease in ventilation, but given the lag in the chemical response due to the circulation time, the initial correcting ventilatory response occurs well into the hyperpneic episode. This brief episode of hypopnea or apnea elicits a similarly delayed response, resulting in an ongoing oscillatory pattern of hyperpnea-hypopnea, whose magnitude and duration depends heavily on the *net effects* of the ventilatory control system. Several factors influence each of the gains: (1) the controller gain is affected by the sensitivity of the peripheral chemoreceptors to changes in gas partial pressure, sensitivity of central centers to peripheral chemoreceptors, and the excitability and integrity of the lower motor neurons supplying the respiratory muscle; (2) the plant gain is determined by the respiratory cycle frequency, $Paco_2$, Pao_2, ventilation-perfusion matching,

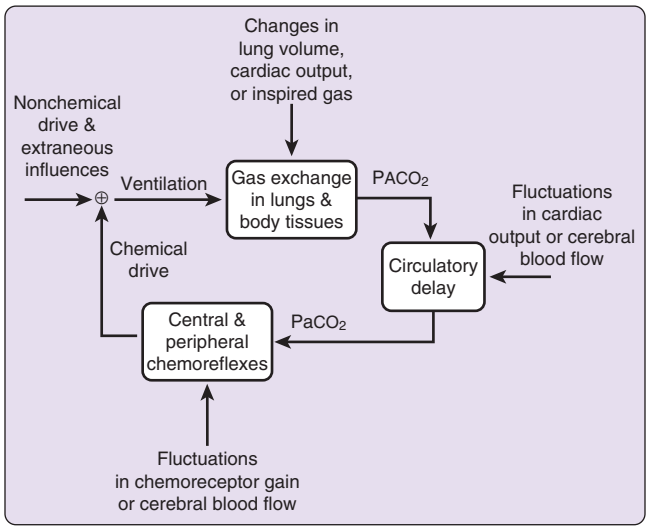


Figure 109-3 Chemical and nonchemical inputs contributing to the breath-by-breath variability.

and dead-space ventilation[53]; (3) the mixing gain is dependent on the circulatory delay time, thoracic blood volume, and brain extracellular fluid volume. In addition to the loop gain, another factor affecting ventilatory stability is the $Paco_2$ reserve, that is, the difference between the eupneic $Paco_2$ and the $Paco_2$ at the apneic threshold. The lower the $Paco_2$ reserve, the smaller the increase in the ventilation required for reaching the apneic threshold and developing central apnea. Conversely, when the apneic threshold for $Paco_2$ is far away from the eucapnic $Paco_2$, large ventilatory changes are necessary to lower the $Paco_2$ below the apneic threshold, decreasing the likelihood of developing central apnea.

The theoretical applications of the concepts of ventilatory instability and loop gain have been translated clinically by several clinical experiments. Common to all of them is the pathophysiologic relationship between alveolar ventilation (V_A) and alveolar Pco_2 described best by several authors[54-56]; a compiled explanation is given by Dempsey and colleagues[51] and depicted in Figure 109-4.

In hypoxic and normoxic acetazolamide-induced metabolic acidosis, the accompanying hyperventilation results in an increase in the V_A required to reduce the $Paco_2$ to the apneic threshold, protecting against apnea and respiratory instability. Despite a similar conceptual pathway of *reduced plant gain* for both hypoxic and normoxic hyperventilation, in hypoxia the slope of the ventilatory response increases and therefore the CO_2 reserve below eupnea is decreased, predisposing to ventilatory instability compared with nonhypoxic hyperventilation (Figure 109-5).[54]

Ventilatory instability is also promoted by *increasing plant gain,* such as is seen with $NaHCO_3$-induced metabolic alkalosis that narrows the CO_2 reserve without a change in the slope of the CO_2 response below eupnea.[58] For the controller gain, both increased and decreased controller gain have been demonstrated using pharmacologic manipulations of the peripheral chemoreceptor sensitivity; intravenous administration of dopamine resulted in a fall in ventilation and reduced O_2 sensitivity[59] (i.e., *reduced controller gain*), whereas administration of the dopamine D_2-receptor antagonist domperidone

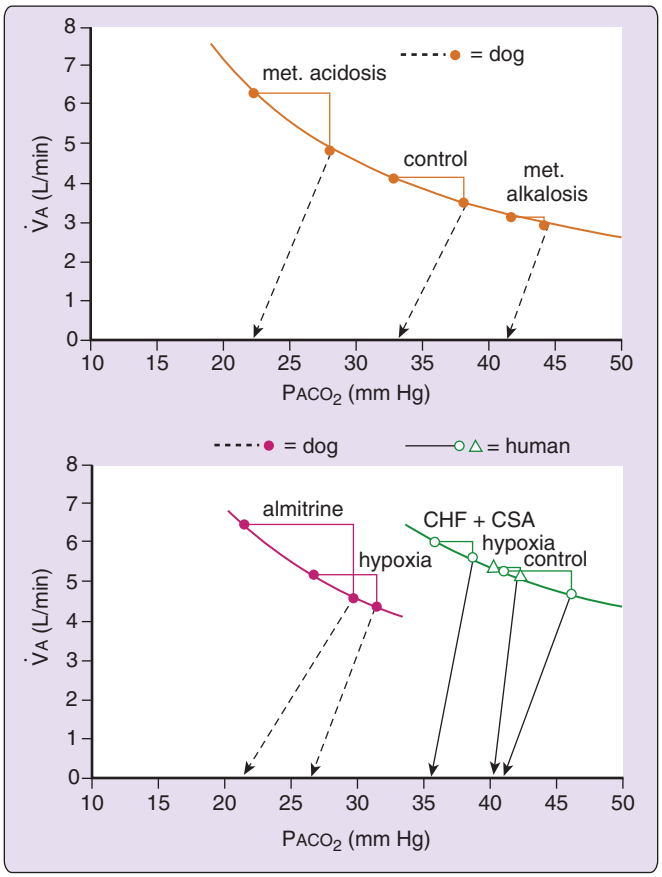


Figure 109-4 The effects of changing background ventilatory drive on the gain of the ventilatory responsiveness to CO_2 below eupnoea, on "plant gain," and on the CO_2 reserve (ΔP_{ETCO_2} eupnoea – apnea) in sleeping dogs and humans. Data are plotted on separate isometabolic lines for dogs [CO_2 flow ($\dot{V}_{CO_2}$) = 150 mL/min^{-1}] and for humans ($\dot{V}_{CO_2}$ = 250 mL/min^{-1}). The *diagonal dashed* or *continuous lines* join eupneic and apneic points, and their slopes indicate the gain below eupnoea of the ventilatory response to hypocapnia in each condition. The height of the *vertical bar* above the isometabolic line indicates the increase in $\dot{V}_A$ required to reduce the P_{aCO_2} to the apneic threshold (i.e., the inverse of plant gain). The CO_2 reserve is the difference in P_{ACO_2} between eupnoea and the apneic threshold. CHF, Congestive heart failure; CSA, central sleep apnea.

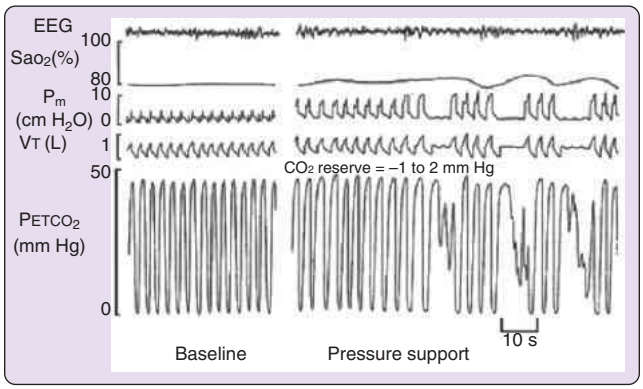


Figure 109-5 Hypoxia reduces the CO_2 reserve. A healthy human is exposed to moderate hypoxia (Pa_{O_2} 80%) for 15 to 20 minutes during NREM sleep, causing mild hyperventilation. When pressure support ventilation is subsequently applied, note that a transient reduction of only 1 or 2 mm Hg in P_{ACO_2} is required to cause apnea and periodic breathing. This effect contrasts with the 3- to 5-mm Hg ΔP_{ACO_2} required in the normoxic control condition. The CO_2 reserve is markedly reduced in hypoxia despite the reduced P_{ACO_2} and plant gain because the slope of the $\Delta\dot{V}_A$–ΔP_{ACO_2} relationship below eupnoea is significantly increased. EEG, Electroencephalogram; P_m, Mean airway pressure. (From Braley TJ, Segal BM, Chervin RD. Sleep-disordered breathing in multiple sclerosis. *Neurology* 2012;28:929–36.)

resulted in increased carotid body sensitivity to O_2 (i.e., *increased controller gain*).[60] As detailed in Figure 109-5, the steeper the slope of the V_{ACO_2}–P_{ACO_2} relationship, the smaller the change in P_{aCO_2} required to reach the apneic threshold, and therefore the higher the controller gain. Numerous clinical entities are associated with one or more abnormalities of loop gain, as summarized in Table 109-3.

Neural Control of Ventilation

Neural mechanisms of ventilatory control in central apnea are likely as important as the metabolic pathways, dictating behavior during wakefulness, affecting airway patency, and controlling respiratory plasticity. The wakefulness stimuli to breathe include tonic excitatory inputs from the so-called reticular formation, brainstem aminergic systems, and hypothalamic orexin neurons.[61] Younes,[62] using correlational analysis, demonstrated that an effective neural control of the upper airway and chest wall respiratory muscles was more important than the inherent passive collapsibility of the airway.

Respiratory plasticity translates into *long-term facilitation* (LTF), a term used to describe the increase in the respiratory activity that persists after the conclusion of an acute episode of intermittent hypoxia.[63] Mediated through several receptors, including serotonergic and *N*-methyl-D-aspartate, this response in sleeping humans is aimed at stabilizing ventilation through increased minute ventilation (i.e., ventilatory LTF),[64] decreased inspiratory upper airway resistance,[65] and increased genioglossus electromyographic activity (i.e., upper airway LTF).[66] However, concentrating on central apnea pathophysiology, Chowdhuri and colleagues[67] demonstrated that, in healthy participants undergoing nasal noninvasive ventilation for promoting hypocapnic central apnea, the increase in the hypocapnic ventilatory response resulted in a significant decrease in the CO_2 reserve, thus offsetting the protective effect of LTF.

Examples of Pathophysiologic to Clinical Applications

Cheyne-Stokes Breathing with or without Central Sleep Apnea

CSB is characterized by a destabilizing interplay of several gains: specifically, controller and plant, with CSA presenting when this destabilization is most accentuated. A *high controller gain* is due to a hypersensitive ventilatory chemoreflex response to CO_2. Although the exact mechanism for this increase in chemosensitivity is not yet known, both congestive (i.e., pulmonary edema[68] and left atrial stretch[69]) and noncongestive factors (reduced carotid arterial blood flow[70]) result in vagal afferents that stimulate central respiratory control centers and fail to allow the P_{aCO_2} to increase at sleep onset.[71] Ventilatory control is also affected in congestive heart failure (CHF) patients because of an attenuated cerebrovascular reactivity to the changes in P_{aCO_2} levels.[72] *The mixing gain* is a concept sometimes used to define how a delay in the circulation can be destabilizing for ventilation. The fact that the chemoreceptors are in the carotid bodies and brainstem rather than in the lung is one factor that can

Table 109-3 Loop Gain Abnormalities in Clinical Disorders

Increased Plant Gain	Decreased Plant Gain	Increased Controller Gain	Decreased Controller Gain	Increased Mixing Gain
Obesity-hypoventilation syndrome (OHS)	Congenital central hypoventilation syndrome	Cheyne-Stokes breathing (CSB)	OHS	CSB
Neuromuscular weakness	Hypercapnic chronic obstructive pulmonary disease	High-altitude periodic breathing Treatment emergent central apnea		Idiopathic pulmonary hypertension

contribute to instability because periodic breathing would be unlikely if chemoreceptors were in the lung. Circulatory delay was induced in classic experiments by Guyton and colleagues,[73] who showed that delays of several minutes (beyond what could occur clinically) were sometimes required to induce periodic breathing in animals. Subsequent studies suggested that circulatory delay was similar in patients with CSB with CHF compared with patients without CSB matched for the severity of heart failure. However, some studies have shown that improvements in circulatory delay are associated with improvement in loop gain (and hence improved AHI). Thus, in aggregate, the data suggest that circulatory delay is necessary but not sufficient to destabilize ventilation in most cases. Therefore the overall response in CSB is that of an increased loop gain manifesting as increased ventilatory instability.

CSB with CSA had been also noted in individuals with cerebrovascular accident (CVA) and chronic renal failure. Among patients with CVA and CSA, the presence of CSB with long hyperpnea and cycle durations, and a gradual rise to peak tidal volume during hyperpnea, was associated with left ventricular systolic dysfunction, but was not related to the location or type of stroke. The authors indicated that the presence of CSA with CSB was more closely associated with left ventricular systolic dysfunction than it was with the stroke itself.[74] Similarly, Yamamoto and Mohri,[75] studying the influence of chronic renal insufficiency on SBDs in patients with symptomatic chronic heart failure, found that most of these patients had unspecified central events, with estimated glomerular filtration rate comparable between non-SBD and SBD groups. The authors suggested that renal dysfunction played a relatively minor role in determining breathing abnormalities in chronic heart failure.[75]

High-Altitude Periodic Breathing

High-altitude periodic breathing is another example of ventilatory instability that occurs during sleep in individuals during ascent to moderate and high altitude (see Chapter 122). Individual susceptibility to high-altitude periodic breathing is driven by multiple genetic factors; polymorphisms in numerous genes, including the hypoxia-responsive transcription factor subunit EPAS1/HIF2α and additional genes in the HIF pathway linked to hemoglobin level, have been associated with differences in susceptibility to or severity of acute mountain sickness associated conditions.[76,77] In this case, high-altitude periodic breathing is the result of ambient hypoxia inducing hyperventilation (*increased controller gain*), which further leads to hypocapnia and consequently to *decreased plant gain*. Overall, however, the controller gain dominates the decreased plant gain, resulting in periodic breathing.[78]

Treatment Emergent Central Sleep Apnea

Treatment emergent central sleep apnea develops in some patients both during titration and after continuous positive airway pressure (CPAP) therapy initiation. This phenomenon is associated with *increased controller gain* due to lowering upper airway resistance with perhaps a contribution of the air leak washing out the anatomic dead space.[79] It usually resolves spontaneously with ongoing CPAP therapy. As in a pilot study, loop gain was higher in patients with treatment emergent sleep apnea in whom central apneas persisted after 1 month of CPAP therapy; loop gain measurement in these patients may enable an a priori determination of those who need alternative modes of PAP.[79a]

Obesity-Hypoventilation Syndrome

Obesity-hypoventilation syndrome (OHS) is characterized by a combination of obesity (body mass index >30 kg/m^2) and arterial hypercapnia during wakefulness ($Paco_2$ >45 mm Hg) (see Chapter 120). OHS is the result of an interplay between respiratory mechanics and ventilatory drive, with leptin, a circulating protein produced mainly by adipose tissue, playing a role. A deficiency of this adipokine, as seen in the leptin-deficient *ob/ob* mouse mode, results in impaired respiratory mechanics, depressed ventilatory responsiveness, and awake hypercapnia.[80] Because leptin replacement in these mice reverses their OHS, recent work has focused on the potential role of leptin in individuals with OHS. It is presumed that the development of central leptin resistance or relative leptin deficiency in OHS could contribute to the development of awake hypoventilation by altering respiratory drive output as well, affecting the mechanical properties of the lungs and chest wall and attenuating the normal compensatory mechanisms used by individuals to cope with obesity-related respiratory loads.[81] These patients have decreased ventilatory responsiveness to hypoxia and hypercapnia compared with similarly obese non-OHS patients and also respond with large increases in $Paco_2$ to small decreases in ventilation (*increased plant gain*), increasing overall the probability of developing central apneic events.[82]

Congenital Central Alveolar Hypoventilation Syndrome

CCAHS is a rare congenital disease caused by mutation in *PHOX2B* gene leading to lack of central drive and decreased

ventilatory response to $Paco_2$ (decreased controller gain) despite normal lungs and respiratory muscle function.

Hypercapnic Chronic Obstructive Pulmonary Disease

Although not characterized by frank central sleep apneas or as a CSA disorder, advanced chronic obstructive pulmonary disease (COPD) is associated with progressive hypercapnia due to impaired lung mechanics, with renal compensation toward a physiologic pH (by increasing serum bicarbonate). Prognosis in individuals with advanced COPD has been reported to be negatively affected by hypercapnia,[83,84] and degree of hypercapnia is not correlated with survival after hypercapnia has developed.[85] The long-term optimal management of the hypercapnia in these patients remains unclear. Recent data by Köhnlein and colleagues demonstrated that the addition of long-term noninvasive positive pressure ventilation to standard treatment improved 1 year survival of patients with hypercapnic, stable COPD when noninvasive positive pressure ventilation was targeted to reduce hypercapnia.[86] Patients with COPD and concomitant OSA may be referred to as having "overlap syndrome" (see Chapter 119), and large series have shown that patients with overlap syndrome who did not receive treatment with nocturnal CPAP had a lower survival rate than patients who suffered from either COPD[87] or OSA[88] alone. However, data regarding optimal management of overlap syndrome are lacking.

Opioid-Induced Central Sleep Apnea

The exact pathophysiologic mechanism of opioid-induced apnea remains poorly understood but is likely related to opioid-induced suppression of inspiration generated by the pre-Bötzinger complex in the brainstem.[89] Both a periodic, non-waxing waning breathing pattern, and a cluster-type breathing pattern, each with central apneas, have been reported during NREM sleep in individuals receiving chronic opiates.[90] Chronic opioid use is a risk factor for the development of central sleep apnea and ataxic breathing,[91] but it is only rarely associated with daytime hypercapnia.[92]

GENETICS

Congenital Central Alveolar Hypoventilation Syndrome

Most CSA disorders in adults have not been linked to specific genotypes. The one clear exception is CCAHS, which is a monogenetic disorder of central respiratory control associated with diffuse autonomic dysregulation[93] and, at times, Hirschsprung disease and tumors of neural crest origin.[94] CCAHS is characterized by a specific facial phenotype, such as boxy facies and an inferior inflection of the lateral segment of vermillion border on the upper lip.[95] CCAHS has a familial presentation, and the *PHOX2B* mutation located on chromosome 4p12 has been identified and confirmed as the disease-defining gene.[96-99] CCAHS, a lifelong disease, is diagnosed in the absence of other systemic pathology and a positive *PHOX2B* screening test or whole-gene *PHOX2B* sequencing test.[100] Clinically, CCAHS is defined by an inability to adapt appropriately to needed ventilatory changes; these patients have altered or absent perception of shortness of breath when awake and profound and life-threatening hypoventilation during sleep.[101] Patients with CCAHS develop apnea or severe bradypnea during NREM sleep. However, expression of the disease is highly variable, with some patients presenting as neonates and others presenting in adulthood, largely depending on the genotype. Approximately 90% of mutations involve excessive polyalanine repeats of the *PHOX2B* gene beyond the normal 20/20 pattern observed in the normal population. Polyalanine repeat patterns of 20/25 to 20/33 typically present at birth with hypoventilation. By contrast, people with a 20/24 pattern may present after the neonatal period, including as adults. Approximately 10% of CCHAS patients have nonpolyalanine repeat mutations (frameshift, missense, or nonsense), and they are typically affected at birth with hypoventilation during wakefulness and sleep. Therapeutically, CCAHS patients require intratracheal or noninvasive positive pressure ventilation during sleep, and about one third also require additional ventilator support during wakefulness, including positive pressure ventilation or diaphragmatic pacing.[102] Generally, adults present with the 20/24 CCAHS genotype, which typically involves only mild hypoventilation that can be managed with noninvasive ventilation during sleep only.

EPIDEMIOLOGY

Risk Factors

Several independent risk factors have been established for CSA-CSB. In patients with CHF and reduced left ventricular ejection fraction, risk factors for CSA-CSB include age older than 60 years, male gender, presence of atrial fibrillation, and hypocapnia.[103-105] For patients with treatment emergent CSA, a high baseline AHI or arousal index, hypertension, opioid use, coronary artery disease, stroke, and CHF all appear to be risk factors.[106]

Prevalence

CSA is estimated to account for 5% to 10% of patients with SBDs that, according to ICSD3, includes OSA, sleep-related hypoventilation disorders, and sleep-related hypoxemia disorder.[107] Additionally, variations in the hemodynamic profile of CHF patients predispose them to alterations day to day and sometimes within the same night of the predominant type of apnea—from OSA to CSA, and vice versa.[108,109]

Cheyne-Stokes Breathing

CSA-CSB is highly prevalent in patients with left ventricular dysfunction regardless of the etiology (ischemic vs. idiopathic), type (preserved or low ejection fraction), New York Heart Association (NYHA) class, and acuity of event (acute or chronic heart failure).[110] CSA-CSB can present during both sleep and wakefulness. Nocturnal CSA-CSB has been studied mostly in stable compensated heart failure and is present in up to 44% of patients who have heart failure with a reduced ejection fraction (HFREF)[111,112] and in up to 27% of patients who have heart failure with a preserved ejection fraction (HFPEF).[113] CSA-CSB during wakefulness is less common, occurring in 16%[114] of patients with HFREF NYHA class II or III; however, because it emerges in the early afternoon and evening and evidence of nocturnal CSA-CSB correlates only weakly with its presence, its reported prevalence could be underestimated. CSA-CSB has been demonstrated after myocardial infarction and unstable angina; in both situations, it is a common occurrence, being present in more than 60%[115] of these patients.

Primary Central Sleep Apnea

Primary CSA, formerly categorized as "idiopathic" CSA, is uncommon. The general population prevalence of primary CSA is not known. However, within the sleep center population, the prevalence has been reported to be 4% to 7%. A higher prevalence of idiopathic CSA has been reported in older patient populations.[116] These individuals usually complain of excessive daytime sleepiness, insomnia, or difficulty breathing during sleep.[117]

High-Altitude Periodic Breathing

Despite considerable heterogeneity in the susceptibility to altitude illness, periodic breathing in the form of cyclic central apneas and hypopneas occurs in almost all individuals at a sufficiently high altitude.[118]

Treatment Emergent Central Sleep Apnea

When treatment emergent CSA (or "complex" CSA) is simply defined as the emergence of central apneas and hypopnea both during and after the application of PAP therapy in patients with OSA, its estimated prevalence in the general sleep center patient population is between 10% and 15%.[119] Treatment emergent CSA is a dynamic process, and its prevalence decreases with ongoing PAP therapy within 2 to 3 months in most patients.[120]

Central Sleep Apnea Due to a Medical Disorder

A synopsis of several common noncardiac medical conditions associated with CSA events is described in Table 109-1. In cardiac conditions not related to left heart disease, CSA events were identified by PSG in 10.6%[121] of a cohort of such patients who developed NYHA class II or III disease due to variety of conditions such as idiopathic pulmonary hypertension, chronic thromboembolic disease with pulmonary hypertension, COPD, and interstitial lung disease. PSG and ambulatory cardiorespiratory sleep studies documented concomitant CSA in up to 39% of patients with idiopathic pulmonary hypertension and NYHA class II to IV chronic thromboembolic disease with pulmonary hypertension[122] and in up to 20% of patients with hypertrophic cardiomyopathy.[123]

Central Sleep Apnea Due to a Medication or Substance

Opioid-induced CSA has only been recognized since about 2000,[124] with a reported prevalence, for example, of 30% of patients in a methadone pain program.[125] Given the progressive increase in opioid use for symptom management in both neoplastic and chronic diseases, it is expected that such CSA will be increasingly identified in clinical sleep practice.

Age

In both general and heart failure populations, CSA-CSB seems to be more commonly encountered in patients of advanced age. In children with CHF, CSA-CSB is quite rare,[126] whereas in a random sample of men aged 20 to 100 years, using sleep laboratory evaluation subsequent to a telephonic survey, Bixler and colleagues[127] noted CSA in 0.4% of those 45 to 64 years old and in 1.1% for those 65 to 100 years old. Others[128] have reported an even higher occurrence of 17% in a population aged 71 years and older.

Gender

Among healthy middle-aged adults, CSA syndromes are overall much more common in men (7.8%) than in women (0.3%).[129] For example, a study including a large proportion of women with stable HF reported unspecified CSA in only 0.05% of those with HF and in none of those with preserved ejection fraction heart failure.[130] Although OSA is increasingly recognized in postmenopausal women, similar consistent data for CSA are lacking.

Race

No data are available on the racial distribution of CSA syndromes to our knowledge.

Morbidity

Central Sleep Apnea and Cardiac Hemodynamics

In CSA-CSB, intermittent surges in blood pressure and heart rate occur in association with oscillations in ventilation. Such surges can be precipitated by cyclic increases in sympathetic nervous system activity targeting the heart and peripheral vasculature.[131,132] Studies concentrating on these hemodynamic responses have confirmed that the frequency and peaks of heart rate and blood pressure oscillations are dependent primarily on periodic oscillations in ventilation.[133] The clinical significance of this finding is not certain, but surges in blood pressure during hyperpnea may be one factor related to the poorer prognosis in patients with heart failure with CSB compared with those without it.[134] More recently, Yumino and colleagues[135] assessed the beat-to-beat stroke volume from before until the end of central respiratory events during sleep in patients with HFREF and demonstrated an increase in stroke volume by a mean of 2.6% (P <.001 for the difference).

Central Sleep Apnea and Cardiac-Related Hospital Readmission

The only study to date[136] that prospectively evaluated cardiac readmission associated with SBD in a cohort of hospitalized patients with acutely decompensated HFREF demonstrated CSA-CSB to be a predictor of both 1-month and 6-month readmission (univariable rate ratios of 1.5 and 1.63, respectively). Ongoing studies are evaluating whether treating CSA-CSB prevents such readmission.

Central Sleep Apnea and Cerebrovascular Accident

Using near-infrared spectroscopy in individuals with acute and subacute CVA, Pizza and colleagues documented asymmetrical patterns of cerebral hypoxia during unspecified CSA events, with significantly larger changes on the unaffected compared with the affected hemisphere.[137]

Mortality

As the oxyhemoglobin desaturations, arousals, increased sympathetic output, and negative intrathoracic pressure (during hyperpnea that follows central apnea in CSA-CSB) contribute to myocardial ischemia, CSA-CSB could contribute to excess mortality in patients with heart failure. Relatively large studies looking specifically at the mortality associated with unspecified CSA and CSA-CSB have provided divergent results, some likely deriving from the lack of a strict definition for CSA or for CSA-CSB. Javaheri and associates[138] evaluated

survival in HFREF (ejection fraction <45%) over a period of 51 months and demonstrated that patients with unspecified CSA had half the survival time of those without such CSA, 45 and 90 months, respectively (P = .01), independent of systolic function, NYHA functional class, heart rate, serum digoxin and sodium concentrations, hemoglobin, and age. In contrast, Andreas and colleagues[139] noted that, in patients with HFREF, nocturnal CSA-CSB had no prognostic impact. Both Andreas and colleagues[139] and Lange and Hecht[140] reported that awake CSA-CSB was associated with a high likelihood of death within 1 to 24 months. Roebuck and associates[141] noted that systolic heart failure patients with unspecified CSA had decreased survival at 500 days but similar long-term survival compared with those without such CSA. Luo and colleagues[142] demonstrated that unspecified CSA had no effect on the prognosis of middle-aged patients with CHF, whereas Bakker and associates[143] provided contrasting evidence by demonstrating a significantly lower survival rate in patients with heart failure and unspecified CSA compared with both heart failure and OSA (mean survival time difference, 3.8 years; P = .005) and those with heart failure only (mean survival time difference, 4 years; P = .01). Additionally, within the group of patients with HFREF and unspecified CSA, mortality is reported to be significantly higher in the "severe" (AHI >22.5/hour) unspecified CSA group compared with the "mild" unspecified CSA group (AHI <22.5/hour; 38% vs. 16%; unadjusted P = .002 and adjusted for the confounders age and NYHA class P = .035).[144] Notably, all of these studies have focused on understanding mortality in untreated patients with heart failure and CSA with or without CSB, and definitive outcome data on the impact of PAP therapy on the natural progression of unspecified CSA and CSA-CSB in patients with heart failure are still lacking. To date, the largest randomized controlled multicenter trial, the Canadian Continuous Positive Airway Pressure for Patients with Central Sleep Apnea and Heart Failure Trial (CANPAP)[145] and its post hoc analysis,[146] showed that death and heart transplantation events did not differ between the control group and the CPAP group. The post hoc analysis showed improved survival without heart transplantation in the CANPAP group when CPAP therapy was associated with a reduced frequency of CSA-CSB events to fewer than 15 events per hour.

CLINICAL PEARLS

- Centrally driven respiratory events are primarily due to a temporary loss of output from the pontomedullary pacemaker that generates breathing rhythm, resulting in loss of diaphragmatic activity.
- CSA with CSB is a form of periodic breathing, commonly observed in patients with heart failure, in which central apneas alternate with hyperpneas that have a waxing-waning pattern of tidal volume.
- Nocturnal CSB has been studied mostly in stable compensated heart failure and is present in up to 44% of patients with low ejection fraction and in up to 27% of patients with preserved ejection fraction. However, variations in the hemodynamic profile of heart failure patients predispose them to day-to-day and sometimes within-night alterations of the predominant type of apnea—OSA to CSA, and vice versa.
- Ventilatory control in CSA is largely chemically driven, especially during NREM, and is the result of a fine balance between a critical $PaCO_2$ level, below which there is a central cessation of breathing (i.e., apneic threshold); its ventilatory triggering factors (mainly hypocapnia); and respondent receptors (i.e., central and peripheral chemoreceptors). This complex feedback mechanism is best described by the concept of loop gain.
- CSA-CSB is likely an independent risk for increased mortality or cardiac transplantation in patients with heart failure.
- Definitive outcome data on the effect of PAP therapy on the natural progression of CSA and CSA-CSB in patients with heart failure are still lacking.

SUMMARY

In CSA, both oronasal flow and thoracoabdominal excursions are absent; that is, there is an absence of respiratory effort during the cessation of airflow. CSB and treatment emergent (or "complex") CSA are the most common clinical CSA patterns. CSA with CSB is characterized by oscillations of ventilation between apnea and tachypnea, with a waxing and waning crescendo-decrescendo pattern in the depth of respirations, and is highly prevalent in patients with heart failure. Treatment emergent CSA most commonly refers to the development of CSA with the application of CPAP in patients with OSA; in most cases, this breathing pattern resolves spontaneously with ongoing therapy. CSA-CSB is likely associated with increased mortality in treated patients with systolic heart failure. It remains unclear whether improving the frequency of CSA-CSB in sleep improves clinical outcomes in this setting or, conversely, resolution of the CSA-CSB is simply a marker of a good prognosis.

Selected Readings

Costanzo MR, Khayat R, Ponikowski P, et al. Mechanisms and clinical consequences of untreated central sleep apnea in heart failure. *J Am Coll Cardiol* 2015;**65**(1):72–84.

Dempsey JA. Crossing the apneic threshold: causes and consequences. *Exp Physiol* 2004;**90**:13–24.

Eckert DJ, Jordan AS, Merchia P, et al. Central sleep apnea: pathophysiology and treatment. *Chest* 2007;**131**:595–607.

Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;**3**:141–63.

Jordan AS, McSharry D, Malhotra A. Adult obstructive sleep apnea. *Lancet* 2014;**383**:736–47.

Nishino T, Lahiri S. Effects of dopamine on chemoreflexes in breathing. *J Appl Physiol* 1981;**50**:892–7.

A complete reference list can be found online at ExpertConsult.com.

Central Sleep Apnea: Diagnosis and Management

Andrey V. Zinchuk; Robert Joseph Thomas

Chapter 110

Chapter Highlights

- Pathologically enhanced respiratory chemoreflexes result in a spectrum of polysomnographic breathing patterns and disorders, including central sleep apnea (CSA), periodic breathing/Cheyne-Stokes breathing, high-altitude sleep apnea, and treatment emergent CSA.
- A pathologically decreased chemoreflex can result in hypercapnic CSA.
- Opiate use causes a disintegrative CSA disorder with relatively unique polysomnographic features.
- Conventional polysomnography has limitations in accurate phenotyping the sleep apneas; thus, the true prevalence of various forms of central apnea syndromes is uncertain.
- Evidence suggests that about one third of those with obstructive sleep apnea have respiratory control as a key mediator of disease. Predominance of events in non–rapid eye movement sleep and persistent or enhanced respiratory instability during treatment with continuous positive airway pressure are key features that are readily recognized in clinical settings.
- Adaptive ventilation is a major advance in noninvasive ventilatory therapy of CSA syndromes. However, benefits in outcomes in heart failure patients are not proven, and there is potential for harm. Several other off-label approaches can be considered as primary or adjunctive therapy.
- Volume-assured positive pressure ventilation can improve oxygenation and ventilation in hypercapnic CSA and hypoventilation syndromes but may cause sleep fragmentation if pressure fluctuations are excessive.
- Residual disease during treatment is common in CSA syndromes. The long-term persistence of the residual central apnea depends on etiology and associated disorders.

The term central sleep apnea (CSA) describes both the pattern of an individual respiratory event and the clinical syndrome characterized by repeated episodes of apneas during sleep caused by an impaired respiratory drive system.[1,2] This is in contrast to obstructive apneas, in which respiratory drive remains active during the apnea, associated with upper airway occlusion (Figure 110-1).

Although central apneas are less frequent than obstructive, they present in ways and settings that are diverse. CSA breathing patterns can vary from the rhythmic sequences of apnea and recovery breaths in congestive heart failure (CHF) to the ataxic respiratory patterns in patients with opioid use.[3,4] Most humans exhibit central apneas during transitions into sleep, and they appear in travelers to high altitude. CSA patterns are associated with a range of medical conditions, from end-stage renal disease to multiple system atrophy and from opioid dependence to the central congenital hypoventilation syndrome.[5-9] CSA is important to recognize because of complications ranging from frequent nighttime awakenings and excessive sleepiness to adverse cardiovascular outcomes and mortality.[10-12]

Central respiratory events in sleep rarely occur in isolation, and many patients with sleep apnea appear to live on a phenotypic spectrum between the obstructive and central apneas. In heart failure and opioid-induced sleep apnea, central and obstructive apneas often coexist.[8,13] In treatment emergent CSA, alleviation of obstructive apneas with positive airway pressure (PAP) amplifies or unmasks a sensitive chemoreflex with resultant centrally mediated apneas and periodic breathing.[14] Identifying where the patient is on the obstructive-to-central spectrum and focusing on the factors responsible for this physiology are critical for accurate diagnosis of CSA and its management.

DEFINITIONS

Current definitions for central apnea and hypopnea are based on polygraphic data and are reviewed in Chapter 109 as well as the American Academy of Sleep Medicine (AASM) scoring manual.[15] A CSA syndrome is defined when five or more central apneas *and/or* central hypopneas are present per hour; that is, a central apnea-hypopnea index (CAHI) of greater than 5, with CAHI comprising more than 50% of all respiratory events.[16] For the various CSA syndromes, additional criteria related to signs and symptoms and specific etiologic entity are required.[16]

These diagnostic criteria can pose a challenge to investigators and clinicians alike because reliably differentiating hypopneas as central versus obstructive is difficult. Evidence of upper airway obstruction on polysomnography, including flow-limitation, does not rule out central apneas/hypopneas,[16a,16b] and esophageal manometry is rarely used in practice. Unclassified hypopneas are

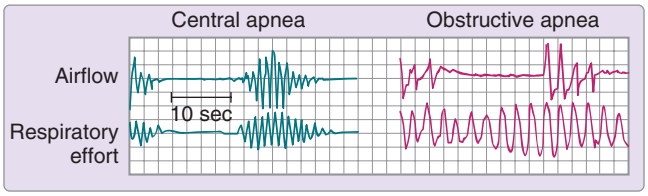


Figure 110-1 Central and obstructive sleep apnea. The relationship between airflow and respiratory effort in central and obstructive apnea. During central apnea, cessation of airflow occurs without associated ventilatory effort. Respiratory effort is present during an obstructive apnea. (From Wellman A, White DP. Central sleep apnea and periodic breathing. In: Kryger M, Dement W, editors. *Principles and Practice of Sleep Medicine*, fifth edition. Saunders: Philadelphia; 2011. p. 1140-1152.)

thus summed into the overall apnea hypopnea index (AHI), not the specific CAHI, biasing towards obstructive SDB.[16c,16d,16e] Accurately classifying SDB disorder as predominantly central or obstructive has implications for treatment.

Integrated analysis of polysomnography (PSG) features can improve identification of central hypopneas.[17] Predominance during non–rapid eye movement (NREM) rather than rapid eye movement (REM) sleep, lack of inspiratory airflow curve flattening or thoracoabdominal paradoxical breathing (chest wall moving inward with inspiration) during hypopnea, and arousal after[18] and gradual flow restoration pattern at hypopnea termination can help classify hypopneas as central.[17] Automation of hypopnea phenotyping (obstructive vs. central)[19] is possible, but accuracy in comparison to electromyography is limited (69%).

CLASSIFICATION OF CENTRAL SLEEP APNEA SYNDROMES

The *International Classification of Sleep Disorders*, third edition (ICSD3) group provides one framework for classifying CSA "syndromes"[16] (see Chapters 107 and 109). The approach in this chapter (Table 110-1) aims to demarcate physiologic and pathologic states in which CSA occurs and link mechanisms of CSA with therapies for each group. For example, it incorporates respiratory chemoreflex phenotyping based on PSG morphology or computational signal analysis. Thus "high chemosensitivity" suggests itself as a clinical category, the patterns of which could include central apneas, periodic breathing, or PAP-induced respiratory instability, signifying the need to consider chemoreflex stabilization in treatment. Our approach thus differs in some cases from the ICSD3 classification and should be considered complementary. For instance, sleep-related hypoventilation syndromes (e.g., obesity-hypoventilation syndrome [OHS]), a separate category in ICSD3, are included under the umbrella of CSA in this chapter under the hypercapnic category. Our aim is to emphasize the role impaired respiratory drive plays in the pathophysiology of these conditions and how treatment can be selected to address this impairment (e.g., bilevel ventilation with backup rate).

PATHOPHYSIOLOGY OF CENTRAL SLEEP APNEA SYNDROMES THAT AFFECT DIAGNOSIS AND TREATMENT

Pathophysiology of CSA is discussed in detail in Chapter 109, including the chemical, mechanical, and neural aspects of respiratory control; the feedback loop between the sensors and the respiratory center; the loop gain (measure of respiratory system stability) and its components (controller, plant, mixing); and other important features. Figure 110-2 summarizes the interplay between ventilatory drive (controller) and the lung's ability to excrete CO_2 (plant) in relation to normal (eupnea) and cessation (apnea) of breathing. In this section we highlight the physiologic concepts important for our approach to diagnosis and treatment and discuss them in the context of specific CSA syndromes.

Table 110-1 Pathophysiologic Classification of Central Sleep Apneas

Physiologic	Pathologic

• Sleep transition
• Phasic REM
	Nonhypercapnic
	• Medical condition related
	  • Congestive heart failure
	  • Poststroke
	  • ESRD
	  • PAH
	  • Atrial fibrillation
	• High altitude
	• Idiopathic
	Hypercapnic
	• Congenital central hypoventilation syndrome
	• Primary chronic alveolar hypoventilation syndromes
	• Other CNS disorders associated with CSA
	  • Encephalitis, tumors, strokes
	  • Anatomic abnormalities
	  • Neurodegenerative disorders
	• Muscular and PNS disorders associated with CSA (selected examples)
	  • Muscular dystrophies
	  • Acid maltase deficiency
	  • Charcot-Marie-Tooth disease and other neuropathies
	  • Postpolio syndrome
	  • Myasthenia gravis
	Disintegrative (e.g., brainstem injury, opioid induced)
	CSA with OSA or upper airway disorders (including treatment emergent CSA)

CNS, Central nervous system; CSA, central sleep apnea; ESRD, end-stage renal disease; OSA, obstructive sleep apnea; PAH, pulmonary arterial hypertension, PNS, peripheral nervous system; REM, rapid eye movements.

Pathophysiologic Changes that Lead to Central Sleep Apneas and Periodic Breathing

Interactions of three factors predispose an individual to ventilatory instability and central apneas-hypopneas during sleep[2,20]: low CO_2 reserve (CO_2 reserve = $Paco_2$ eupneic – $Paco_2$ apneic), abnormally high *or* low loop gain (a product of controller, plant, and mixing gains), and sleep state and stage instability.

CO_2 reserve is affected by changes in plant and controller gains. For example, CO_2 reserve is decreased in metabolic alkalosis (increased plant gain; see Figure 110-2), promoting the risk for central apneas and ventilatory instability, whereas it is increased with metabolic acidosis (decreased plant gain),[21]

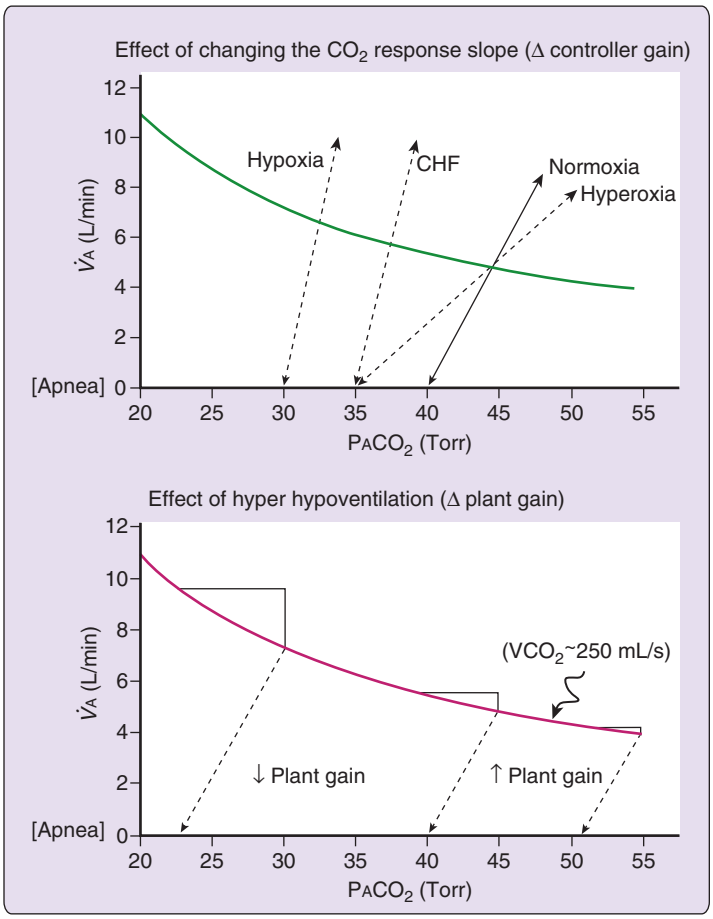


Figure 110-2 Changing plant gain (*bottom*) and controller gain (*top*) influences on CO_2 reserve. Diagrammatic representation of the steady-state relationship between alveolar ventilation and alveolar P_ACO_2 ($PaCO_2$) at a fixed resting CO_2 production (of 250 mL/min). The schematic figure shows how changing plant gain or controller gain will influence the "CO_2 reserve" or Δ $PaCO_2$ between eupnea and apnea. *Top,* Changing the background drive to breathe without changing the slope of the $\Delta\dot{V}_A$ versus Δ P_ACO_2 relationship (controller gain) above or below eupnea. For example, background hyperventilation (via metabolic acidosis or specific carotid body stimulation with almitrine) raises $\dot{V}_A$ and lowers P_ACO_2 along the isometabolic hyperbola (decreased plant gain). This means that a greater transient increase in $\dot{V}_A$ and reduction in P_ACO_2 is required to reach the apneic threshold than it would be under control, normocapnic conditions. The reverse is true for conditions that reduce the background drive to breathe and cause hypoventilation (e.g., metabolic alkalosis). *Bottom,* At any given level of background P_ACO_2, changing the slope (or responsiveness) of the relationship below eupnea would alter the CO_2 reserve or the amount of reduction in P_ACO_2 required to cause apnea. Changing the slope of the ventilatory response to CO_2 above eupnea would alter the susceptibility for transient ventilatory overshoots. Often both plant and controller gains may change together; note the reduced plant gains and increased controller gain, with hypoxia or with congestive heart failure patients. The increased controller gain dominates, and the net effect is a decreased CO_2 reserve and instability. (Modified with permission from Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;3[1]:141–63.)

which is protective against central apneas (see Figure 110-2). Administration of oxygen (which reduces hypoxic stimulus to breathing) has been shown to decrease ventilation and responsiveness to $Paco_2$ during sleep.[22] This stabilizes breathing through reduction in controller gain and increase in $Paco_2$ reserve, whereas hypoxia leads to opposite effects (see Figure 110-2). In addition, inherent delays in the negative feedback loop controlling ventilation (mixing gain) increase loop gain. This delayed recognition of blood gases by the controller (as in those with systolic CHF) predisposes to unstable and periodic breathing.

Transitions into and out of sleep, and between sleep stages, are inherently unstable in terms of respiratory control.[5] Brief central apneas and hypopneas occur during this time in normal individuals because of "unmasking" of the sleep apneic threshold that is very close to the wake eupneic threshold. In addition, upper airway and diaphragmatic muscle tone is reduced, with the associated increase in upper airway resistance and decreased inspiratory force, resulting in hypoventilation.[23,24] Then, as $Paco_2$ rises above the apneic threshold, rhythmic breathing is maintained.

During sleep, brief and abrupt transitions to wakefulness, termed *arousals*, can result in ventilatory instability, with level of ventilatory response and arousal threshold playing important roles. With a sudden arousal, the sleep eupneic $Paco_2$ (normally about 5 mm Hg higher than awake $Paco_2$) is detected as hypercapnic by the aroused respiratory control center. This signal to increase ventilatory drive, combined with the removal of the upper airway resistance induced by sleep, results in increased ventilatory response and reduction in $Paco_2$.[25,26] When sleep resumes, the current $Paco_2$ is considered to be hypocapnic for the sleeping brain, that is, below the apneic threshold, resulting in central apnea. Thus any process that leads to frequent sleep-wake transitions, such as sleep maintenance insomnia, sleep apnea, maladaptation to continuous positive airway pressure (CPAP), or periodic limb movement disorder, can increase the propensity to ventilatory overshoots, periodic breathing, and CSAs, especially in a setting of high chemosensitivity.[25,27-29]

Pathophysiologic Changes in Specific Central Sleep Apnea Syndromes

Nonhypercapnic Central Sleep Apnea

Disorders manifesting as nonhypercapnic (eupneic or hypocapnic) CSA have two physiologic phenomena in common: (1) normal or slightly low awake steady-state $Paco_2$, and (2) increased ventilatory responsiveness to $Paco_2$ or hypoxemia (increased loop gain). In the setting of arousals, CSAs are perpetuated owing to the so-called "inertial" effect (see Chapter 109).

Central Sleep Apnea and Periodic Breathing of Heart Failure. Heart failure is associated with Cheyne-Stokes breathing (CSB), which is seen at times during wakefulness and frequently during sleep. CSB is characterized by a crescendo-decrescendo pattern of tidal volumes with central apnea or hypopnea occurring at the nadir of the cycle (typically 60 to 90 seconds; Figure 110-3). It is in part a consequence of increased loop gain due to a heightened chemoreflex (sensitive controller)[30,31] and lack of the "normal" increase in $Paco_2$ with sleep onset (decreased CO_2 reserve).[32] These are superimposed on prolonged circulation time (increased mixing gain) resulting in cyclical ventilatory instability.[33]

Periodic breathing of shorter cycles occurs in other settings and conditions with "hyperactive" chemoreflex (as described later). Although ICSD3 defines CSB with specific cycle lengths and intervening apneas, we feel CSB represents one end of the chemoreflex activation severity spectrum, whereas at the other end of the clinical spectrum, although poorly recognized in practice, is nonapneic short cycle (≤30 seconds) periodic breathing.

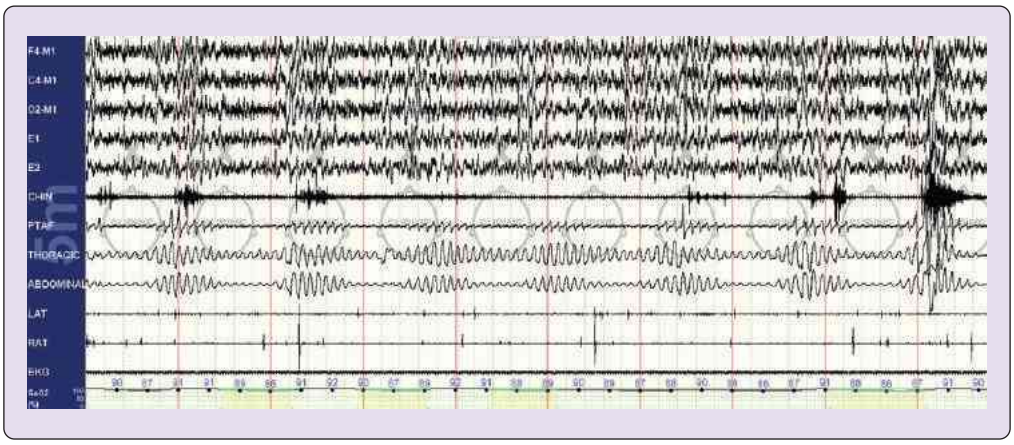


Figure 110-3 **Relatively long cycle periodic breathing and Cheyne-Stokes respiration.** Ten-minute screen compression, each vertical line is 30 seconds. A patient with congestive heart failure in NREM sleep. Note the symmetrical, concordant, waxing and waning flow and effort. Cycle lengths are about 45 to 50 seconds.

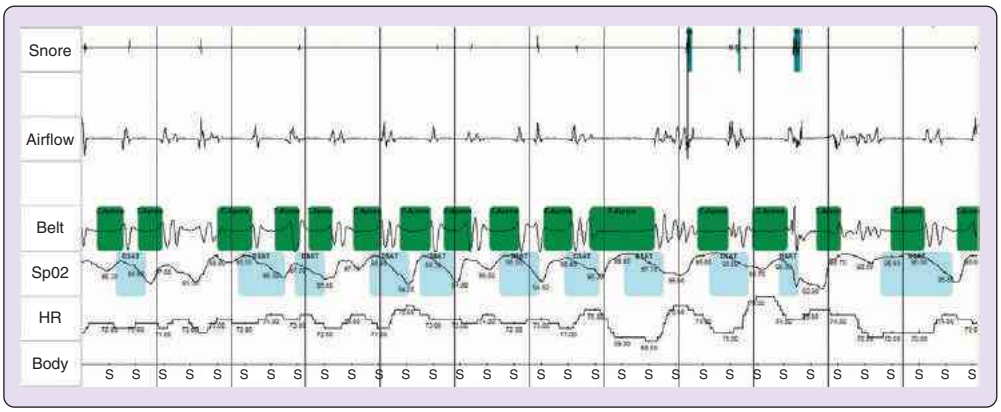


Figure 110-4 **Idiopathic sleep apnea.** A home sleep study on a medication-free 27-year-old nonobese man presenting with mild daytime sleepiness (Epworth Sleepiness Scale score of 9/24), nocturnal awakenings, and unrefreshing sleep. Note the short cycles (about 20 seconds) of pure central respiratory events.

Central Sleep Apnea Due to High-Altitude Periodic Breathing. In contrast to CSB in heart failure, the cycle time of periodic breathing at high altitude is short (probably owing to elimination of the mixing gain defect). The mechanism involves exposure to hypoxemia with resultant chemoreceptor-mediated hyperventilation during NREM and REM sleep. After approximately 10 minutes of hypoxia in a sleeping human, tidal volumes oscillate in a waxing and waning pattern. The oscillations increase in magnitude as hypoxia is maintained and $Paco_2$ falls further to the level of apneic threshold.[34] When that is reached, overt periodic breathing occurs with cycle times of 15 to 25 seconds (two to five large tidal volume breath clusters followed by apneas of 5 to 15 seconds).[2,35-37] There is wide variation in O_2 during this time. The predominant mechanism is increased loop gain manifested as a reduced $Paco_2$ reserve (1 to 2 mmHg) and increased chemosensitivity (see Figure 110-2).[21,38] The marked increase in arousals and decrease in slow wave sleep potentiate respiratory instability. Attesting to the key role for hypocapnia in this disorder of central apnea and periodic breathing, the breathing disorder can be improved with administration of small amounts of CO_2 (increase in $Paco_2$ reserve),[39] increasing dead space (increase in $Paco_2$ reserve),[40] and acetazolamide (reduction in plant gain and increase in $Paco_2$ reserve).[21] Patients with obstructive sleep apnea (OSA) and high loop gain have PSG features mimicking high-altitude periodic breathing, including NREM dominance and short cycle periodic breathing.

Primary Central Sleep Apnea. Primary (idiopathic) CSA is a rare disorder characterized by repetitive episodes of central apneas in NREM sleep, which are short and irregular (rather than periodic) and terminate with an abrupt, large breath (Figure 110-4), in contrast to what is characteristically seen in CHF. The most clearly demonstrated pathophysiology is an increased hypercapnic ventilatory response during wakefulness.[29,41] In addition, impairment of switching between expiration and inspiration has also been found in these patients.[42] It has been speculated that the long expiratory pause that typically occurs with these CSA events may be attributable to this impairment.

Other Nonhypercapnic Central Sleep Apnea Syndromes and Their Associated Medical Conditions. Common chronic medical conditions such as end-stage renal disease (ESRD), cerebrovascular accident (CVA), and pulmonary hypertension have been associated with nonhypercapnic CSA. In patients with ESRD, the central apnea index (i.e., frequency of central

apneas) inversely correlates with $Paco_2$ and cardiac silhouette enlargement,[6] and ultrafiltration increases $Paco_2$ with associated decline in the CAHI by 55%.[43] These findings suggest that a similar link may be present between volume overload and the CSA-CSB of CHF.

CSB occurs in a minority of patients (~7%) after CVA.[44] Although in many cases post CVA the CSB is associated with left ventricular dysfunction and hypocapnia,[45] some authors note an increased prevalence of CSB in patients with lacunar strokes (~20%) and without left ventricular dysfunction.[46] CSB has been reported in patients with idiopathic pulmonary arterial hypertension (PAH). The postulated mechanisms include decreased stroke volume and increased mixing gain,[2] although, as in case of CVA, no PAH-specific studies have been done.

Hypercapnic Central Sleep Apnea

Hypoventilation due to a failed or failing automatic control (and effector) system is the pathophysiologic hallmark of disorders that manifest with hypercapnic CSA. They can be broadly approached as disorders of impaired central drive ("won't breathe") or impaired respiratory muscle control ("can't breathe"). In general, the former category is due to processes involving the brainstem respiratory centers (e.g., congenital central alveolar hypoventilation syndrome), whereas the latter is due to neuromuscular weakness disorders (e.g., amyotrophic lateral sclerosis). Most of the above disorders are associated with pathologically *low* loop gain (either because of controller or plant components) and worsening of hypoventilation and apneas during REM sleep (in contrast to hypocapnic CSA, which is NREM dominant). The latter occurs primarily because of intercostal muscle atonia during REM. Most of these conditions are classified under the sleep-related hypoventilation disorders in the ICSD3 (see Chapter 107).

Congenital Central Alveolar Hypoventilation Syndrome and Idiopathic Central Alveolar Hypoventilation. Congenital alveolar central hypoventilation syndrome (CCHS, or "Ondine's curse") is a rare disorder of respiratory control and autonomic systems, first reported by Mellins in the 1970s.[47] Small tidal volumes and monotonous respiratory rates result in hypoventilation while the wakefulness and behavioral stimuli supply the respiratory drive. With sleep onset, worsened hypoventilation, hypercapnia, and hypoxemia ensue due to the impaired automatic control system. In many cases, if not identified early, this leads to asphyxia and death.[48] Mutations of the *PHOX2B* gene are disease defining.[49] This gene encodes a transcription factor responsible for the fate of early autonomic nervous system cells, including those in the respiratory control centers.[48]

Obesity-Hypoventilation Syndrome ("Pickwickian Syndrome"). "Joe was a wonderfully fat boy, standing upright with his eyes closed." This is the first depiction of OHS, found in Charles Dickens' book, *The Posthumous Papers of the Pickwick Club*. Today's medical literature requires that obesity and awake hypoventilation ($Paco_2$ >45 mm Hg) be present (without other causes for the latter) for the diagnosis of OHS.[50] Abnormalities include progressive hypoventilation and hypoxemia during NREM sleep with further impairment in REM sleep,[51] and OSA (nearly universal in OHS[52]). Pathogenetic mechanisms are complex and insufficiently investigated, even in face of the obesity epidemic. They include ventilatory abnormalities of OSA, increased work of breathing,[53] and blunted chemosensitivity (decreased loop gain).[54-56] There is a suggestion that resistance to leptin, a hormone produced by adipocytes that normally augments ventilatory response to $Paco_2$, is a possible mechanism for reduced controller gain in OHS.[57-59]

Other Central Nervous System–Related Disorders. Central neurologic processes that cause impairment of the brainstem respiratory centers, such as compression, edema, ischemia, infarct, tumor, encephalitis, and Arnold-Chiari malformations, have been associated with breathing dysrhythmias and CSA.[60-69] The specific manifestations depend on the location and the type of the insult. For instance, automatic failure of breathing control ensues following cervical cordotomy.[70] Damage to areas other than the brainstem (thalamus, basal ganglia, centrum semiovale) can lead to CSA, suggesting the importance of the descending signals for generation of the automatic breathing stimulus.[46]

Peripheral Nerve and Muscle Disorders. Neuromuscular diseases, such as muscular dystrophy, myasthenia gravis, Guillain-Barré syndrome, amyotrophic lateral sclerosis, postpolio syndrome, and Charcot-Marie-Tooth disease, can lead to awake alveolar hypoventilation, with worsening hypoventilation during sleep. This is occasionally associated with central apneas, although sleep-related hypoventilation without outright central apnea is the more prominent feature. Ventilation during sleep in patients with respiratory muscle disease often deteriorates well before awake ventilation is affected.

Disintegrative Central Sleep Apnea and Hypoventilation Associated with Opiates. Although the respiratory depressant effects of opioids are well known,[71,72] the effects of opiates on sleep are widespread and complex and affect many patients.[73] With chronic use, hypoventilation and obstructive and central apneas can occur in a single patient, fulfilling diagnostic criteria for several disorders under the ICSD3 classification. In patients with daytime hypercapnia, nocturnal hypoventilation can be profound. Two unique patterns of breathing tend to occur in patients taking chronic opioids: (1) cluster breathing characterized by cycles of deep breaths with relatively stable tidal volumes with interspersed central apneas of variable duration and (2) Biot breathing (ataxic breathing) with variable tidal volumes and rates.[2] In addition, patients taking chronic opioids with nearly pure OSA on initial PSG evaluation can develop treatment emergent central apnea.[74] In chronic opioid use, there is also a high prevalence of obstructive events.

Of the various opioid receptors, stimulation of μ and κ receptors tends to drive respiratory depression[8] primarily in the pre-Bötzinger complex.[75,76] At low doses, tidal volumes decline,[77] whereas at higher doses, respiratory rate and rhythm generation are suppressed.[8] Morphine given to normal human subjects acutely decreases hypercapnic and hypoxic controller gain[78]; however, chronic administration results in decreased hypercapnic but increased hypoxemic chemosensitivity.[79] Mechanisms of ataxic breathing (Figure 110-5), common among those taking chronic opioids (nearly 70%) and those taking higher doses (>200 mg of morphine,[4]), have not been elucidated. Similar features could occur with injury to the

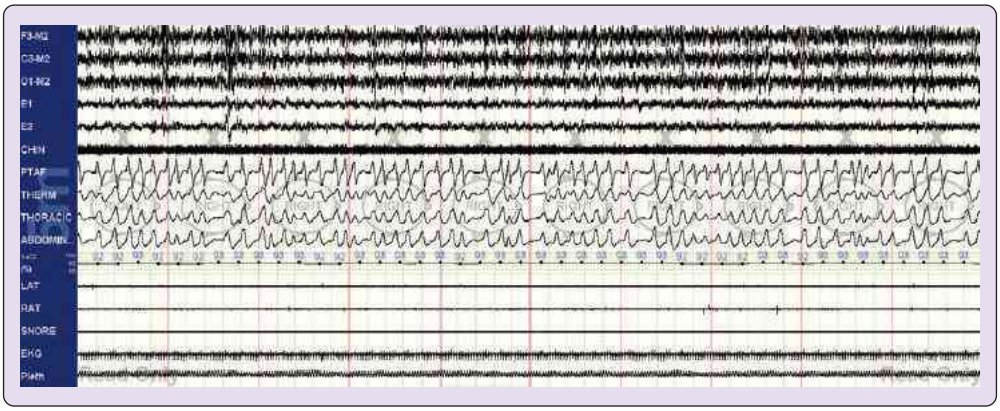


Figure 110-5 **Opiate-induced central sleep apnea and ataxic breathing.** Ten-minute screen compression; each vertical line is 30 seconds. In this methadone-treated 56-year-old woman, the most characteristic feature of opiate-related disease is the variability in expiratory duration, although tidal volumes also vary. These polysomnographic features are readily recognizable and occur in NREM sleep.

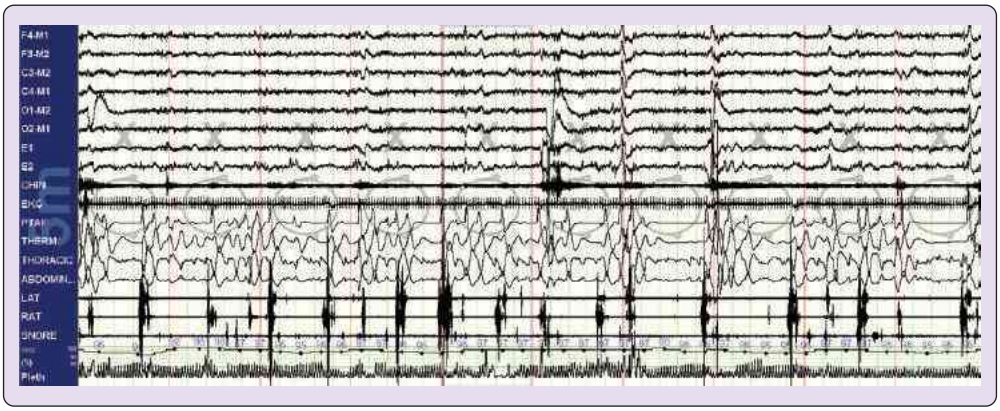


Figure 110-6 **Central sleep apnea associated with head and neck chemoradiation.** Ten-minute screen compression; each vertical line is 30 seconds. A 71-year-old man treated with radiation and platinum-based chemotherapy for laryngeal cancer. He presented with severe insomnia, multiple nocturnal arousals, and daytime fatigue. Note the variable-duration central apneas, mixed features, and sleep fragmentation.

carotid bodies, such as following head and neck chemoradiation (Figure 110-6).

Disordered Interplay Among Upper Airway Obstruction, Breathing Control, and Sleep-Arousal Propensity: Treatment Emergent Central Sleep Apnea ("Complex Sleep Apnea"). The $Paco_2$ reserve is labile during NREM sleep,[80] and arousals due to maladaptation to PAP can occur and drive instability.[14] Upper airway collapsibility as measured by Pcrit (pressure at which passive critical closing of the upper airway occurs) shows overlap between patients with OSA and controls.[81] Variations in Pcrit alone account for only a portion of variations in the apnea-hypopnea index (AHI)[82] or differences between those with pure OSA (100% of apneas obstructive) and predominant OSA (coexisting with CSA and mixed apneas).[83]

In some patients with OSA, central apneas and periodic breathing "emerge" during initiation of CPAP. This phenomenon is termed *treatment emergent CSA* in ICSD3[15] and is defined when there are five or more central apneas or hypopneas per hour of sleep, making up greater than 50% of all respiratory events during titration of CPAP in those fulfilling OSA criteria during diagnostic PSG. The existence of treatment emergent CSA (previously known as *complex sleep apnea*) as a unique entity has been controversial in the sleep world.[84] The original description noted a set of relatively unique features on *diagnostic* PSG and an incomplete treatment response that included induction of central apneas or persistent periodic breathing when CPAP was applied. The key feature is NREM-dominant central hypopneas or periodic breathing with obstruction (Figure 110-7), resolving spontaneously during REM sleep. That is, induction of central apneas was not required. However, the NREM dominance may be readily seen during positive pressure titration (Figure 110-8). A subsequent publication "defined" complex apnea, proposing the current ICSD criteria,[84] which also became the criteria used by medical insurance to qualify patients for the more expensive adaptive ventilators. The ICSD3 allows for the coexistence of periodic breathing and OSA (a "splitting" approach). The term *complex apnea* (a "lumping" approach) may have clinical utility in that a single term captures a number of ICSD categories with a common pathogenesis and identical responses to therapy. Central hypopneas are rarely scored in clinical practice, and thus reports of complex apnea or treatment-emergent sleep apnea, especially if using the ICSD3 treatment-emergent category and criteria, likely include only patients far on the spectrum of chemoreflex-driven instability than the middle ground, where a substantial minority of

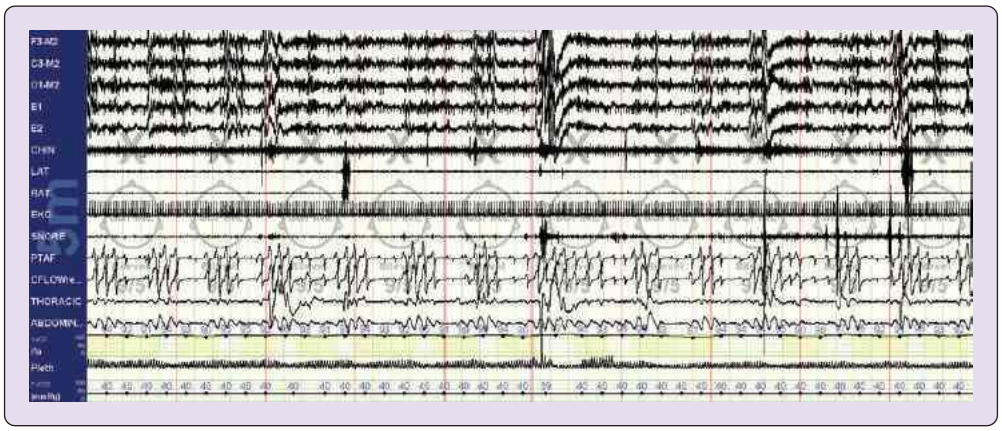


Figure 110-7 Key feature of NREM-dominant apnea. Ten-minute screen compression; each vertical line is 30 seconds. Periodic breathing with short cycles (30 seconds or less) and variable degrees of obstruction. Conventional scoring typically identifies these events as obstructive. Flow limitation is often seen, but the waxing-waning characteristic is usually evident.

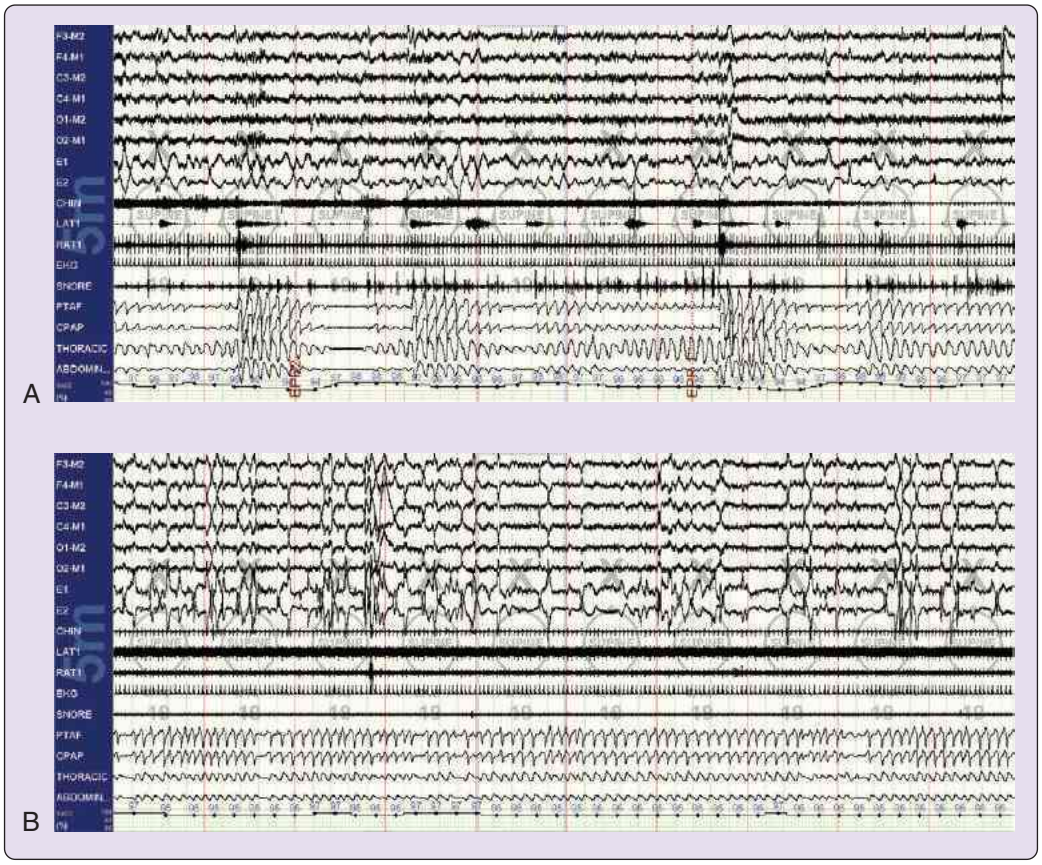


Figure 110-8 NREM-dominant sleep apnea. **A,** NREM-dominant sleep apnea, with continuous positive airway pressure (CPAP) during NREM sleep. Ten-minute screen compression; each vertical line is 30 seconds. Unresolved respiratory events occur across a range of CPAP pressures (5 to 19 cm) with long cycle events, some periodic breathing features, and clear obstructive features. **B,** NREM-dominant sleep apnea during REM sleep. Ten-minute screen compression; each vertical line is 30 seconds. The same subject as in **A** with spontaneous transition to REM sleep showing resolution of all abnormality. The CPAP pressures were progressively reduced to 10 cm with continued maintenance of stable breathing in REM sleep.

patients could fall. In these patients, short cycle (≤30 seconds) periodic breathing, with features of admixed obstruction, is highly reminiscent of high-altitude periodic breathing and was part of the original description.[80] It is possible that long cycles (≥60 seconds) may be caused by subclinical (or even subechocardiographic) cardiac diastolic dysfunction, but data are lacking on cardiac function differences between patients demonstrating purely long- versus short-cycle events.

A consistent feature of patients with treatment emergent or complex sleep apnea, documented in most publications

describing this phenomenon, is sleep fragmentation, which often persists despite reasonable respiration-targeted therapy. Because arousals amplify hypocapnic instability, inadequate cohesion of the NREM sleep–related network activity seems to be core pathology in some of these patients. This phenomenon is reminiscent of reports of CHF patients, in whom sleep fragmentation persists beyond that attributable to respiratory events.[85]

A pertinent question is whether the findings of the treatment emergent sleep apnea phenotype persist with continuous use of CPAP. Because central hypopneas or periodic breathing were not quantified in most studies, underestimation is probable and of uncertain degree. Lack of persistence may imply that it is simply a marker of the severity of OSA and dynamics of its improvement, or it may reflect an artifact of scoring approaches that ignore or misidentify central hypopneas. Some studies report resolution in 78% to 86% of patients[86,87] with 2 to 12 months of CPAP treatment, whereas others note treatment success rates of about 50%.[88] A prospective study by Cassel and colleagues,[89] who followed 675 patients with OSA with PSG at 0 and 3 months. At time zero, 12% had treatment emergent CSA, which resolved in 74% of the cases by 3 months; however, 7% of the original cohort without treatment emergent CSA were noted to have it on the 3-month study. One approach to quantifying the persistence of treatment emergent CSA is to measure residual respiratory events after several months of CPAP, using the flow data available in current generation devices. In a study (unpublished) of 217 patients after more than 6 months of therapy, the manually scored AHI_{FLOW} of 10/hour or greater was seen in 23%, and the central apnea index at the baseline sleep study was the only predictor of residual disease.

The predominant role of the CO_2 control instability in pathogenesis of treatment emergent or complex sleep apnea is supported by resolution with small increases of inhaled CO_2.[90,91] The mechanisms for improvement in chemoreflex events after prolonged use of CPAP include reduction in controller gain and increase in $Paco_2$ reserve.[92] Stabilizing central respiratory motor output through prevention of transient hypocapnia prevents most OSA in selected patients with a high chemosensitivity and a collapsible upper airway, whereas increasing respiratory motor output through moderate hypercapnia eliminates "obstructive" apnea in most patients with a wider range of chemosensitivity and CO_2 reserve.[93] Reducing chemosensitivity through hyperoxia has a limited and unpredictable effect on OSA.[93]

EPIDEMIOLOGY OF CENTRAL SLEEP APNEA AND ITS SUBTYPES

This topic is discussed in detail in Chapter 109. The epidemiology data for CSA are largely based on standard AASM definitions of respiratory events in sleep[94] and may underestimate the prevalence of CSA. This is largely due to the inability to effectively distinguish central from obstructive hypopneas without esophageal manometry, leading to classification bias toward obstructive sleep-disordered breathing (SDB; see Definitions section).

As alternate measures for detecting centrally mediated SDB are developed and automated, CSA prevalence may rise. For example, central apneas, periodic breathing, and CSB are patterns that suggest chemoreflex-mediated respiratory control dysfunction.[95] A biomarker of heightened chemoreflex activity (narrow-band elevated low-frequency coupling [$e\text{-}LFC_{NB}$]) has been described using an electrocardiogram-based analysis of heart rate variability and heart rate–respiratory coupling. This metric quantifies the metronomic self-similar oscillations that characterize nonhypercapnic CSA and periodic breathing.[96] One third of a large, community-based patient cohort with SDB (the Sleep Heart Health Study) exhibited the $e\text{-}LFC_{NB}$, which is associated with CSA and periodic breathing.[97] This proportion is roughly in keeping with detailed phenotyping experiments performed over the years.[81,98]

CLINICAL FEATURES AND DIAGNOSIS OF CENTRAL SLEEP APNEA AND ITS SUBTYPES

Clinical Presentation

The clinical presentation of patients with CSA varies by the etiology and subtype (Table 110-2). Symptoms and signs are not specific to CSA and often overlap with those of OSA as well as the underlying conditions leading to CSA (e.g., dyspnea on awakening in heart failure patients[99]). The following sections describe the PSG characteristics important in CSA, including the differences between nonhypercapnic and hypercapnic subtypes and finally the unique features of CSA associated with specific disorders.

Polysomnographic Features Important in Central Sleep Apnea

Diagnosis of CSA syndromes generally requires a full-night recording of standard PSG with special attention to inspiratory effort, to differentiate central (no inspiratory effort throughout event) versus obstructive apneas. Although this differentiation is simple for apneas, for hypopneas it requires esophageal manometry.[100,101] Respiratory inductance plethysmography excursions are present in both central and obstructive hypopneas, and determining whether decreases in effort and flow are proportionate can be arbitrary and difficult to operationalize.[102] Alternative strategies using PSG-based algorithms have been developed but show marginal accuracy compared with esophageal manometry (68% in one study).[17]

Table 110-2 Clinical Characteristics of Patients with Sleep Apnea

Central		Obstructive
Nonhypercapnic	Hypercapnic	
Insomnia	Daytime sleepiness Morning headache	Daytime sleepiness
Mild intermittent snoring	Snoring	Prominent snoring
Awakenings (choking, dyspnea)	Respiratory failure	Witnessed apneas, gasping
Normal body habitus	Normal or obese Polycythemia Cor pulmonale	Commonly obese Upper airway narrowing

Finally, cardiopulmonary coupling signal analysis as described previously may also be used to differentiate central predominant (chemoreflex-driven) versus obstructive SDB phenotypes[95] but has not been validated using esophageal manometry.

A consistent feature of nonhypercapnic, heightened chemoreflex-mediated central apnea is predominance of events during NREM sleep, especially during non–slow wave stages.[22,103-108] A metronomic self-similar appearance is typical, in contrast to opiate-induced CSA, in which variability of expiratory phase is characteristic. Additional features of chemoreflex modulation during sleep are noted in Table 110-3. In contrast, in many cases of hypercapnic CSA (and in OSA), the severity of SDB worsens markedly during REM sleep, especially if the motor neurons of the diaphragm are involved.[2,104] Notable exceptions are CCHS and opioid-induced CSA, in which SDB worsens in NREM sleep. Finally, objective measures of hypoventilation are needed, such as arterial, transcutaneous, or end-tidal P_{CO_2} (P_{ETCO_2}) to confirm hypercapnia in sleep.[15]

Propensity for sleep fragmentation and upper airway collapsibility can both worsen CSA and have specific treatment implications (see Treatment of Central Sleep Apnea). A sleep fragmentation phenotype on PSG can be suggested by prolonged sleep-wake transitional instability (>10 minutes), low sleep efficiency (<70%), persistently high N1 stage during PAP titration (>15%), and poor evolution of slow wave sleep (<1 Hz).[103] Upper airway collapsibility can be measured through the Pcrit, derived from relationships between maximal inspiratory airflow and nasal or mask pressure in OSA patients.[81,109]

Because of the challenges in diagnosing CSA without esophageal manometry, we recommend taking into account an array of the PSG features to distinguish between central and obstructive phenotypes as described previously (and in Table 110-3). Such an approach, combined with complementary measures of chemoreflex hyperactivity (e.g., cardiopulmonary coupling analysis) and identification of propensity for sleep fragmentation and airway collapsibility, can augment recognition of and treatment of CSA.

Unique Features of the Disorders

Nonhypercapnic Central Sleep Apnea

Central Sleep Apnea at High Altitude. Travelers to high altitudes often experience restlessness, frequent brief arousals, and unrefreshing sleep,[35] at least in part due to periodic breathing and CSA (see Chapter 122). Men are twice as likely to be affected as women, and at altitudes above 5000 feet it is nearly universal. PSG features are discussed in the section on Pathophysiology of Central Sleep Apnea Syndromes that Affect Diagnosis and Treatment.

Primary Central Sleep Apnea. Patients with primary CSA often present with insomnia or frequent awakenings during the night, rather than daytime sleepiness as seen in OSA.[110,111] Cycles of central apneas in idiopathic CSA are shorter (20 to 40 seconds) and not as gradual as in central sleep apnea with Cheyne-Stokes breathing (CSA-CSB) (see Pathophysiology of Central Sleep Apnea Syndromes that Affect Diagnosis and Treatment). Underlying medical conditions must be ruled out before diagnosis.

Heart Failure and Central Sleep Apnea. Clinical features, implications, and treatment of central apneas in CHF are discussed in detail in Chapters 109 and 129. In brief, age older than 60 years, male sex, and atrial fibrillation appear to be risk factors for CSA among those with CHF.[112] Patients generally present with CHF symptoms, fatigue, and weakness rather than sleepiness.[99,113-115] Symptoms and apneas improve with position changes, including lateral positioning, and are independent of the postural effects on the upper airway[116] suggesting that J-receptor activation or oxygen stores play a role in the pathogenesis of CSA-CSB. Arousals occur midcycle at the peak of the recovery.[117]

Other Medical Conditions Associated with Central Sleep Apnea. CSA is also found in patients with ESRD, CVA, PAH, and atrial fibrillation. There are no particular clinical features with high predictive value for CSA among these

Table 110-3 Recognition of Strong Chemoreflex Modulation of Sleep Breathing

Polysomnographic Feature	Relatively Pure Obstructive Sleep Apnea	Chemoreflex-Modulated Sleep Apnea
Periodic breathing, Cheyne-Stokes breathing	Rare	Typical (often short cycle, <30 sec in absence of CHF)
Respiratory event timing	Variable (each event tends to have different durations)	Self-similar, metronomic
Severity during sleep state	Greater severity in REM	Minimal severity in REM
Effort signal morphology	Well maintained during obstructed breath	Complete or partial loss between recovery breaths
Flow-effort relationship	Discordant: flow is reduced disproportionately to reduction in effort	Concordant: flow and effort follow each other in amplitude
Arousal timing	Early part of event termination	Crests event, often in the center of the sequence of recovery breaths
Oxygen desaturation	Irregular, progressive drops, V-shaped contour	Smooth, symmetrical, progressive drops rare

CHF, Congestive heart failure; REM, rapid eye movements.
From Thomas RJ. Alternative approaches to treatment of Central Sleep Apnea. *Sleep Med Clin* 2014;9(1):87–104.

patients (hypocapnia may be a clue), thus diagnosis requires heightened clinical suspicion and PSG.

Patients with ESRD and CSA (concomitant obstructive and mixed apneas are common) are generally male, older, and more frequently volume overloaded.[118,119] Ultrafiltration improves $Paco_2$ and CAHI,[43] suggesting initial avenue for management of CSA in ESRD. Higher suspicion for CSA is also warranted in those with larger territory and more severe CVAs.[46,120] Patterns of CSA include CSB with long cycle times in CVA patients with left ventricular dysfunction and periodic breathing with shorter cycle lengths in those without left ventricular dysfunction.[121-123] Central apneas improve with oxygen[121] and tend to resolve as patients recover from their stroke.[124,125] In patients with PAH, older age and sleepiness (Epworth Sleepiness Scale score of >10) are predictive of SDB,[126] and CSB is the predominant CSA pattern with cycle of about 45 seconds. Presence of atrial fibrillation should raise suspicion for CSA, and vice versa.[127-130] Among community-dwelling men, increasing central apnea index correlates with increasing prevalence of atrial fibrillation, and presence of CSA-CSB is associated with odds ratio of 4.5 for atrial fibrillation, even when controlled for cardiac comorbidities, including CHF.[128]

Hypercapnic Central Sleep Apnea

Hypercapnic CSA and hypoventilation should be considered if a condition associated with them (see Table 110-1) or certain clinical features (see Table 110-3) are present. A PSG and an assessment of sleep hypoventilation through nocturnal $Paco_2$ (or surrogate) are recommended. In cases in which hypercapnic CSA and hypoventilation are discovered during a PSG obtained for another indication (e.g., OSA), a careful clinical approach to identify the underlying cause is warranted. Our initial assessment is based on locating the lesion along an anatomic pathway that could result in hypoventilation: corticobulbar tracts, brainstem, bulbospinal tracts to cervical spinal cord, anterior horn cells, lower motor neurons, neuromuscular junction, and intercostal and diaphragmatic muscles. Lung and chest wall abnormalities are generally apparent on examination and basic diagnostic studies and can help identify the underlying medical disorder (e.g., chronic obstructive pulmonary disease) in those with hypercapnic CSA and hypoventilation. Selected conditions are discussed next.

Central Congenital Alveolar Hypoventilation Syndrome. The classic feature of presentation for CCHS is mild awake and marked sleep-related alveolar hypoventilation with hypercapnia and hypoxemia (Figure 110-9). Although classically diagnosed at birth, owing to variable penetrance of the *PHOX2B* mutations (polyalanine repeat mutations [PARMs]), some patients can present as late onset in childhood or even in adulthood.[131-135] In these individuals, alveolar hypoventilation can be unmasked by administration of CNS depressants and anesthetics, recent severe pulmonary infections, or in the setting of treatment for OSA, and CCHS should be considered in those without another explanation for hypoventilation. Clinical associations that should raise suspicion for CCHS include Hirschsprung disease, tumors of neural crest origin, autonomic dysfunction, facial dysmorphology, and dermatographism.[49,136,137] The ventilatory response and sensation of dyspnea are greatly diminished or absent in children with CCHS.[138] The respiratory pattern during sleep is characterized by markedly diminished tidal volumes and inappropriately constant respiratory rate in the face of hypercapnia and hypoxemia.[9,139] Ventilation is more stable during REM versus NREM sleep. There are variations in clinical phenotype with PARM genotype. For example, individuals with 20/25 (normal alanine repeat genotype being designated as 20/20) rarely require 24-hour ventilatory support, whereas for genotypes 20/27 to 20/33, continuous ventilatory support is needed.[140-142] Care for patients with CCHS should be provided through centers with extensive expertise in the condition.[12]

Obesity-Hypoventilation Syndrome. OHS should be suspected in an obese individual with daytime sleepiness.[143] Relative to the average OSA patient, those with OHS have a higher central fat distribution, are more likely to complain of dyspnea, and present with signs of cor pulmonale.[144-146] The diagnosis hinges on increased awake $Paco_2$ (>45 mm Hg) in the absence of other known causes of hypoventilation (e.g., chronic obstructive pulmonary disease, restrictive lung disease) and obesity (body mass index >30 kg/m^2). Serum bicarbonate and resting pulse oximetry can be used as simple screening tools for further testing. A bicarbonate level of 27 mEq/L or greater has 92% sensitivity for $Paco_2$ greater than 45 mm Hg (specificity of 50%[147]) among those with OSA. Rest O_2 saturation of less than 94% while breathing ambient air also suggests a need for blood gas measurement.[50,148] Up to 40% of patients with OHS continue to have persistent sleep-related O_2 desaturation despite elimination of upper airway obstruction, and hypoventilation can persist on autoadjusting PAP.[148] Hypercapnia may be more important than hypoxemia in mediating cognitive impairment individuals.[149,150] A recent study showed that although both CPAP and noninvasive ventilation improved $Paco_2$, sleep architecture, and SDB severity compared with lifestyle modification, with noninvasive ventilation, the decrease in $Paco_2$ was greater and was associated with improved 6-minute walk distance, spirometry, and some quality of life measures.[151]

Neurodegenerative Disorders. CSA and hypoventilation should be suspected in patients with neurodegenerative disorders. They are most common in multiple sclerosis (MS) and multiple system atrophy (MSA).[7] MS patients with brainstem involvement manifest with central apneas in contrast to those with nonbrainstem lesions and controls.[152] In MSA, central apneas, CSB, and apneustic breathing have all been reported.[153-155] CSA is uncommon in Alzheimer disease and Parkinson disease.[7,156-159] In amyotrophic lateral sclerosis, hypoventilation is the most common presenting feature of SDB, with nocturnal symptoms preceding daytime ventilatory failure.

Muscular and Peripheral Nervous System Disorders Associated with Central Sleep Apnea. In patients with muscular disorders, in addition to degeneration of the myocytes, impaired respiratory drive can contribute to hypoventilation. This abnormality was found in 20% of a myotonic dystrophy cohort.[160] In a recent study of 85 patients with myotonic dystrophy, 11% and 15% were found to have CSA and mixed sleep apneas, respectively, with 39% having OSA.[161] These patients were not sleepy but noted poor sleep quality as the most common symptom. The CAHI in this group correlated with slow oral swallowing time.[161,162]

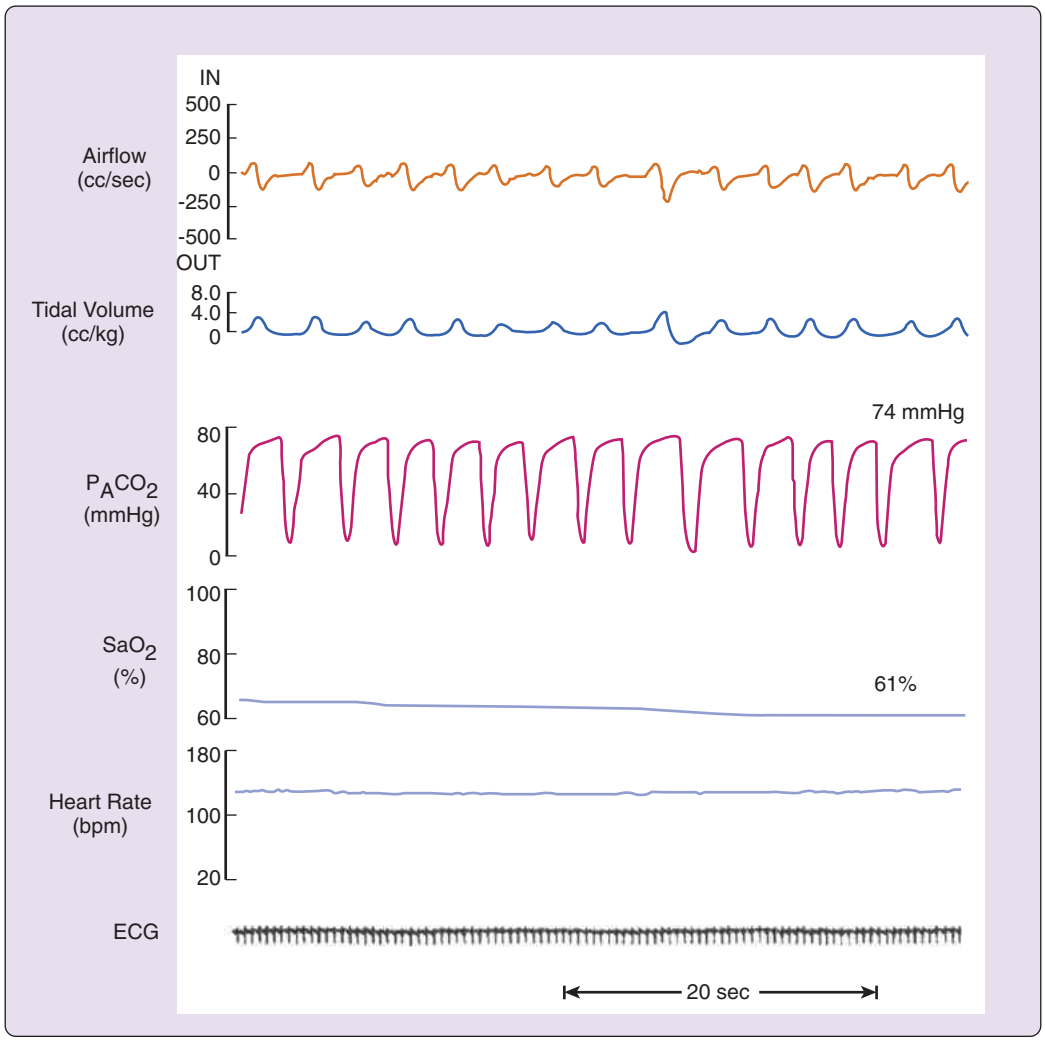


Figure 110-9 Central congenital hypoventilation. Polygraph tracing from 28-month-old girl, demonstrating typical breathing pattern during NREM sleep in congenital alveolar central hypoventilation syndrome. Note inappropriately regular (20 breaths/min), shallow breathing (tidal volumes averaging 3.5 mL/kg). Progressive hypercapnia and hypoxemia did not stimulate ventilation, arousal, or beat-to-beat heart rate variability. (Adapted with permission from Weese-Mayer et al.[9])

Disorders affecting the diaphragm or its nerve supply (Charcot-Marie-Tooth disease and other neuropathies, myasthenia gravis, and other neuromuscular junction disorders) present predominantly with sleep-related alveolar hypoventilation.

Opiate-Induced Central Apneas and Hypoventilation (Disintegrative Central Sleep Apnea). CSA should be considered in most patients on chronic opioid therapy (COT) with symptoms of disturbed sleep. Although most extensively studied among patients taking pure opioid receptor agonists (e.g., methadone, oxycodone), recent data reveal that combinations of partial agonists and antagonists (buprenorphine and naloxone)[163] also result in significant SDB, both central and obstructive. In those with CSA, case series suggest that there is increased sleep fragmentation, increased stage 2 sleep, and decreased REM and slow wave sleep,[164] consistent with NREM predominance of central events. On PSG, patients on COT can show predominantly OSA, CSA, or mixed phenotypes. CSA in COT can present during the initial polysomnogram or emerge after treatment of predominantly obstructive disease (treatment emergent CSA).[74] There is some consistency in published reports that decreased tolerance, efficacy (high residual sleep apnea), and compliance with CPAP are present in this setting of CSA (see Treatment of Central Sleep Apnea section later). Disintegrative CSA patterns may also be seen in patients with brainstem injury such as with stroke and MS. Finally, hypoventilation is not unique to use of opioids and can be seen with anesthetics, sedatives, and muscle relaxants.

Treatment-Emergent Central Sleep Apnea. Clinical features and diagnosis of treatment emergent CSA are discussed in detail under Pathophysiology of Central Sleep Apnea Syndromes that Affect Diagnosis and Treatment above. In brief, this entity is typically recognized by the appearance of central apneas, hypopneas, and periodic breathing when continuous or nonadaptive bilevel PAP is increased to control airway obstruction in patients diagnosed with OSA.[165] The most characteristic feature of such chemoreflex-driven respiration

is not the morphology of individual events but rather is the NREM sleep dominance and the timing and morphology of the sequential events (nearly identical) in a consecutive series of events.[14,95,103]

Various techniques may be used to identify patients with heightened chemoreflex and reduced CO_2 reserve. Times series analysis of electrocardiogram, described in earlier sections, can provide a map of state sleep oscillations with e-LFC_{NB} as a marker of central apneas and periodic breathing in those with treatment emergent CSA.[95] Loop gain, along with other phenotypic traits in OSA (genioglossus muscle responsiveness, arousal threshold, and Pcrit) can be assessed using dynamic flow and pressure responses to positive pressure dial-down.[81,98] In one study[81] of OSA patients, 19% had a relatively noncollapsible upper airway similar to controls, and in these patients, loop gain was almost twice as high as in patients with a collapsible airway, despite comparable AHIs, suggesting that treatment approaches other than upper airway stabilization maybe useful in these patients (see Treatment of Central Sleep Apnea). Other methods used to quantify loop gain and predict CPAP responsiveness among those with CSA-CSB are referenced for the interested reader.[30,166-168] If confirmed, respiratory chemoreflex phenotyping may become a common clinical reality.

TREATMENT OF CENTRAL SLEEP APNEA

Positive pressure, including "enhanced" positive pressure (see later) and non–positive pressure approaches, is available for the treatment of both hypocapnic and hypercapnic CSA syndromes, including idiopathic, treatment emergent or complex, periodic breathing, hypercapnic of various etiologies, and opiate-induced CSA. All these phenotypes, which may coexist and exhibit within-night and night-to-night dynamism, require an exact application of a multimode core therapeutic approach, including upper airway support, respiratory rhythm and drive modulation, and enhanced sleep consolidation as core approaches.[103]

Positive Pressure–Based Therapy

CPAP is a recommended initial option for CSA, based on the premise that upper airway obstruction is relevant for hypercapnic and nonhypercapnic types of CSA, a position endorsed by AASM guidelines.[169] However, there are now enough data to demonstrate that CPAP alone is poorly effective and tolerated in nonhypercapnic CSA syndromes, whereas adaptive servo ventilation (ASV) and enhanced CPAP (used with respiratory stabilization approaches that include hypocapnia minimization, sedatives, carbonic anhydrase inhibition, and oxygen) are superior treatment approaches for efficacious suppression of central apneas and periodic breathing patterns on the polysomnogram. Nonadaptive (fixed pressure) bilevel positive pressure ventilation alone is also suboptimal: this tends to exaggerate CSA and periodic breathing. Although using a backup rate with fixed bilevel positive pressure ventilation can reduce central apneas as the machine-delivered mandatory breaths substitute for lack of patient-derived respiratory effort, comparative studies with ASV show the latter to achieve superior elimination of central apneas. Nonrandomized evaluations show these devices to be about as effective as each other.[170] Because individual adaptive ventilator algorithms are substantially different, specific patient subsets may have differential responses (e.g., short vs. long cycle periodic breathing). Such individual differences in responses are currently not predictable through PSG features.

ASV devices provide expiratory support, inspiratory pressure support, and backup supportive responses guided by measures of ventilation or flow averaged over several minutes. These devices are primarily designed for patients with elevated loop gain and thus nonhypercapnic CSA but can be beneficial when hypoventilation is not the primary and sole abnormality, such as opiate-induced CSA. When used for CSA in patients with treatment emergent CSA and heart failure, central apneas are decreased in frequency, and numerous neurohumoral and cardiac function parameters are improved in heart failure patients.[171,172] Muscle sympathetic activity is reduced by adaptive ventilation but not CPAP in patients with CHF and CSA-CSB.[173] Treatment with ASV is better tolerated than CPAP and is effective in suppressing central apneas and improving oxygenation regardless of the presence of heart failure. Positive effects on sleep architecture are less impressive. The criteria for success and the respiratory event scoring criteria (often 4% desaturation association for hypopneas) can overestimate effectiveness. A randomized prospective trial of CPAP versus ASV used a success threshold of suppressing central AHI (essentially central apneas) below 10/hour of sleep as a criterion for success.[174] ASV was superior to CPAP in suppressing respiratory events, but sleep quality, sleepiness, and quality of life were not different between groups, raising the question of the best approach to quantify effectiveness beyond merely apnea suppression. Alternative indexes such as time in periodic breathing or stable breathing may provide useful efficacy information. However, in the intention-to-treat analysis, success (AHI <10/hour) at 90 days of therapy was achieved in 89.7% versus 64.5% of participants treated with ASV and CPAP.[174] The results also show how difficult it is to bring sleep and breathing back close to normal in these patients, even with optimal treatment of the CSA. Residual sleepiness in patients with CSA can be improved over auto-CPAP.[175] Similarly, opiate-induced CSA is difficult to treat with PAP, but ASV is superior to CPAP.[176] These patients do not have classic periodic breathing or metronomic central apneas and are characterized by ataxic breathing and central apneas of variable lengths. ASV devices can impose a rhythm in these conditions but can also induce patient-ventilator asynchrony and therefore need to be carefully titrated.

General Principles Regarding the Use of Adaptive Servo Ventilation Devices

All ASVs provide fixed or automatic expiratory airway support, adjustable minimal and maximal pressure support, and different options for backup rates (user specified, device algorithm estimated, or none). Volume and flow targets are used, and the sampling-averaging window extends over 3 to 4 minutes. Thus ASV devices track long-range respiratory patterns and make adjustments in the parameters to maintain the target within a prespecified range. Detection of apneic obstructive events results in expiratory pressure increases, whereas the difference between minimal and maximal allowable pressure support is the "adaptive space." The devices may be used with tight or relatively wide open user constraints depending on the preference of the physician and pressure tolerance of the patient. Cycle lengths likely influence

outcomes, and it is our observation that ASVs are less effective in patients with short cycle (≤30 seconds) periodic breathing. A subset of patients demonstrate immediate ASV intolerance and desynchrony, and this effect does not resolve with added time. Patients with prolonged sleep-wake transitional instability can have the pathology markedly amplified by an ASV. The normal fluctuations of respiration during REM sleep can inappropriately trigger the adaptive algorithms of ASVs and cause arousals, but this seems rare in clinical practice.

Recognition of Efficacy and Scoring of Respiratory Events During Adaptive Servo Ventilation Use

Scoring respiratory events during adaptive ventilation should use the pressure output signal from the ventilator. This is roughly equal and opposite to the patient's respiratory output. The flow and effort signals combine patient and ventilator contributions and give a false sense of success. See Figure 110-10 for excessive "pressure cycling," which is a response of an adaptive ventilator to ongoing periodic breathing. When pressure cycling persists, sleep fragmentation can be severe even if respiration is "improved." This pattern means than periodic breathing pathology is ongoing, necessitating the continued pressure response. When the ventilator enables stable respiration, cycling between the minimal and maximal pressure support zones is minimal. Further details on algorithms and titration strategies may be obtained from recent comprehensive reviews.[177,178] Bench testing of ASV algorithms show device-specific response characteristics, but stable breathing does not readily occur across a range of simulated central apnea patterns.[179]

ASVs are powerful devices, and if there is patient-ventilatory asynchrony, they can induce hypocapnia, excessive cycling of pressures, arousals, distorted flow patterns, and physical discomfort. Increased mortality through sudden cardiac death, and no benefits including quality of life, were reported in the Treatment of Predominant Central Sleep Apnoea by Adaptive Servo Ventilation in Patients with Heart Failure (SERVE-HF) study,[179a] which showed increased mortality with an ASV in patients with systolic heart failure (ejection fraction ≤45%) and AHI greater than 15/hour with 50% or more central events and a central AHI of 10/hour or greater. Hypocapnia, metabolic alkalosis, hemodynamic perturbations, and excessive sympathetic driving associated with excessive pressure cycling are speculative mechanisms of adverse outcomes.[180] The general consensus is that ASVs should be avoided in heart failure with reduced ejection fraction and CSA. Caution is recommended in the use of these devices in other vulnerable populations, such as patients with stroke or heart failure without reduced ejection fraction. Tracking of device data such as tidal volume stability, degree of pressure cycling, and surrogate signs of patient-ventilator asynchrony (e.g., wide variations in expiratory durations), regardless of underlying disease state, and not relying on manufacturer derived AHIs alone are prudent.

Considerations for the Treatment of Hypercapnic Central Sleep Apnea

Management of the sleep related breathing disorders of the hypoventilation syndromes is a complex process; a few key points are noted here.[181] Respiratory support may be provided by bilevel ventilation with a backup rate, volume target pressure-support ventilation, or invasive volume ventilation through a tracheostomy. These modes are also readily available on several home ventilators.

Volume-assured pressure support (VAPS) is an advance in management of hypoventilation syndromes and hypercapnic CSA. Besides expiratory, minimal, and maximal inspiratory support, backup rates, and various breath modulations such as inspiratory time and trigger and cycle sensitivity, a tidal volume target may be set. VAPS is most effective if there is hypoventilation without CSA, but it can provide benefits if used cautiously in hypercapnic CSA.

Sufficient expiratory pressure support to prevent major obstructive events is critical. REM sleep can demonstrate greater severity than NREM sleep and typically requires greater ventilation than NREM sleep. However, NREM dominance may also be seen, as in opiate-induced CSA. An autoexpiratory pressure function can aid in managing patients with markedly higher REM sleep settings. The backup rate is usually set slightly below the patient's native rate. However, if

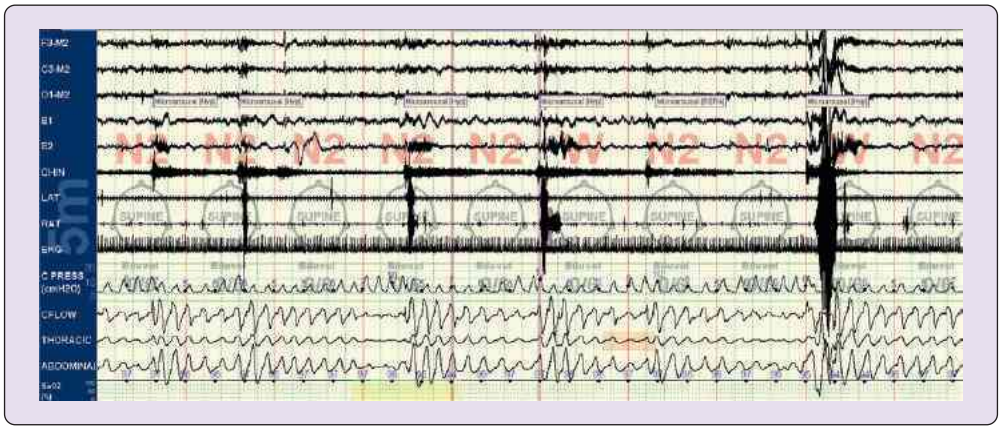


Figure 110-10 Pressure cycling during adaptive ventilation treatment. Five-minute compression snapshot; each vertical line is 30 seconds. The C-PRESS channel is the pressure output from the adaptive ventilator (Adapt SV). This 56-year-old man had predominantly central apneas, which were eliminated. However, respiratory instability, repetitive arousals, and pressure cycling continued without resolution, despite adjustments of pressure support. Persistent pressure cycling is readily recognized during home use by generating expanded night data using the device software, with attention to tidal volume and pressure traces. The device may not automatically detect respiratory events during such periods.

there is bradypnea (e.g., respiratory rate below 6 breaths/minute) or tachypnea (e.g., rate above 20 breaths/minute), entraining the patient to a different rate may be difficult and result in patient-ventilator desynchrony. Substantial inspiratory support (e.g., 25 cm H_2O) may be required to enable optimal ventilation.

Positive pressure ventilation therapy for hypercapnic CSA poses the specific challenge of inducing relative hypocapnia and respiratory instability and associated sleep fragmentation by overly aggressive ventilation.[182] There is a tradeoff between improving ventilation and oxygenation versus sleep quality because excessive volume targets and the associated pressure rises can induce sleep fragmentation. The rate of change in pressure support can be prescribed in these devices, providing one form of "brake" to prevent pressure-related sleep fragmentation. In the absence of PSG titration, when sleep versus wake and NREM versus REM sleep treatment requirements are not estimated, treatment may be less precise than possible. An iterative approach of two or three PSG titrations separated by a few months can result in greater precision of therapy and sleep quality, allowing resetting of the respiratory controller. Although empirical home ventilation can be obtained under specific diagnostic conditions (such as amyotrophic lateral sclerosis) and severities based on pulmonary function tests, PSG titration with transcutaneous CO_2 monitoring can improve precision of care.

Alternative Approaches to Positive Pressure Therapy

Minimization of Hypocapnia

That CO_2 can stabilize respiration has been known for decades, and prevention of hypocapnia is a critical stabilizing factor in sleep respiratory control. However, high concentrations of CO_2 fragment sleep by inducing arousals secondary to respiratory stimulation and sympathoexcitation.[183,184] The challenge has been delivery of CO_2 in a clinically adequate, tolerated, and precise manner. Holding the CO_2 steady and just above the NREM sleep CO_2 threshold can protect the CO_2 reserve while maintaining sleep consolidation.

In a study manipulating inhaled CO_2 in OSA patients but directly relevant to CSA treatment,[93] 26 patients with OSA (AHI 42 ± 5 events/hour with 92% of apneas obstructive) were treated with O_2 supplementation, an isocapnic rebreathing system in which CO_2 was added only during hyperpnea to prevent transient hypocapnia, and a continuous rebreathing system. With isocapnic rebreathing, 14 of 26 reduced their AHI to 31% ± 6% of control ($P < .01$) (responders); 12 of 26 did not show significant change (nonresponders). The responders versus nonresponders had a greater controller gain, a smaller CO_2 reserve, but no differences in Pcrit. Hypercapnic rebreathing (+4.2 ± 1 mm Hg P_{ETCO_2}) reduced AHI to 15 ± 4 of control ($P < .001$) in 17 of 21 subjects with a wide range of CO_2 reserve. Hyperoxia (SaO_2 ~95% to 98%) reduced AHI to 36% ± 11% of control in 7 of 19 OSA patients tested. Addition of a closed volume (dead space) to exhale increases rebreathing of exhaled air and results in a rapid increase in CO_2 levels and an increased tidal volume and respiratory rate. The concept has been used in mechanical ventilation to reduce hypocapnia for several years and more recently has been successfully used to treat CSA-CSB in heart failure.[184] Combining hypocapnia minimization with positive pressure is logical, and we have shown that keeping CO_2 above the apnea threshold with the use of enhanced expiratory rebreathing space (EERS) is an effective adjunct to PAP therapy[90]; EERS is the dead space concept applied to pressure ventilation and may be used with continuous or adaptive pressure support (Figure 110-11). There is no or a minimal increase in inspiratory CO_2 because of the positive pressure–induced washout. The physiologic target for titrations with EERS is to maintain end-tidal CO_2 ($ETCO_2$) at the low-normal range for sleep, that is, an increase of 2 to 8 mm Hg in $ETCO_2$. CO_2 manipulation can also be done by bleeding CO_2 into the circuit by a more precisely controlled flow-independent method. Successful treatment of mixed OSA and CSA using a proprietary device, the positive airway pressure gas modulator (PAPGAM), has been reported. This device delivers precisely controlled concentrations of CO_2. In a small case series, 6 patients with an average mixed apnea AHI of 43/hour on CPAP improved, with reduction of AHI to 4.5/hour and addition of 0.5% to 1% using PAPGAM.[91] Dynamic CO_2 manipulation (delivery restricted to a specific phase of the respiratory cycle) may, in future studies, improve the stabilizing effects of CO_2.[185]

Oxygen

Nasal O_2 has a long history of use to treat CSA and periodic breathing without CSA.[186,187] Effectiveness is typically partial, and residual sleep apnea and sleep fragmentation are common. High loop gain that is not primarily driven by hypoxia may not respond to O_2 at clinically safe doses, an escape from suppression of carotid body firing. Adding oxygen to CPAP may benefit CSA and treatment emergent CSA with a reduction in responsiveness of peripheral chemoreceptors and loop gain.[188,189] A study in a U.S. veterans population showed benefit in a predominantly CSA population, but the PSG changes were delayed by as much as an hour or more.[190] Respiratory event cycles can lengthen with the use of O_2. Such a change may "reduce" the respiratory event index but not imply a true stabilization of respiration. Use of O_2 also negates use of a desaturation parameter to score hypopneas. Beneficial effects of oxygen for CSA are not limited to those with oxygen saturations below thresholds used for long-term nasal oxygen therapy (e.g., ≤88%). Use of O_2 is off-label for CSA syndromes. The limitations include the long-term cost and difficulty with reimbursement in nonhypoxic patients.

Enhancing Sleep Consolidation

Arousals from sleep have a role in sleep apnea pathophysiology.[28] Sedatives can probably be used safely in minimally nonhypercapnic hypoxic CSA and NREM-dominant apnea in general because arousals further destabilize sleep and worsen sleep apnea severity For example, eszopiclone has been shown to reduce the AHI in patients with obstructive apnea with a low arousal threshold.[191] Likely mechanisms of benefit of sedatives in nonhypercapnic CSA include reduction of arousal-induced hypocapnia and increasing the proportion of NREM sleep spent in stable breathing. Triazolam, temazepam, zolpidem, and clonazepam have all been shown to reduce periodic breathing and CSA.[192-195]

Opioid-Induced Central Sleep Apnea

CSA is commonly associated with chronic opiate use and is dose related with substantial individual differences. PSG features of opiate effects may be more common than clinical symptoms, and the impact, on health, of CSA exclusively due to opiate use remains to be defined. Decreasing the dose of

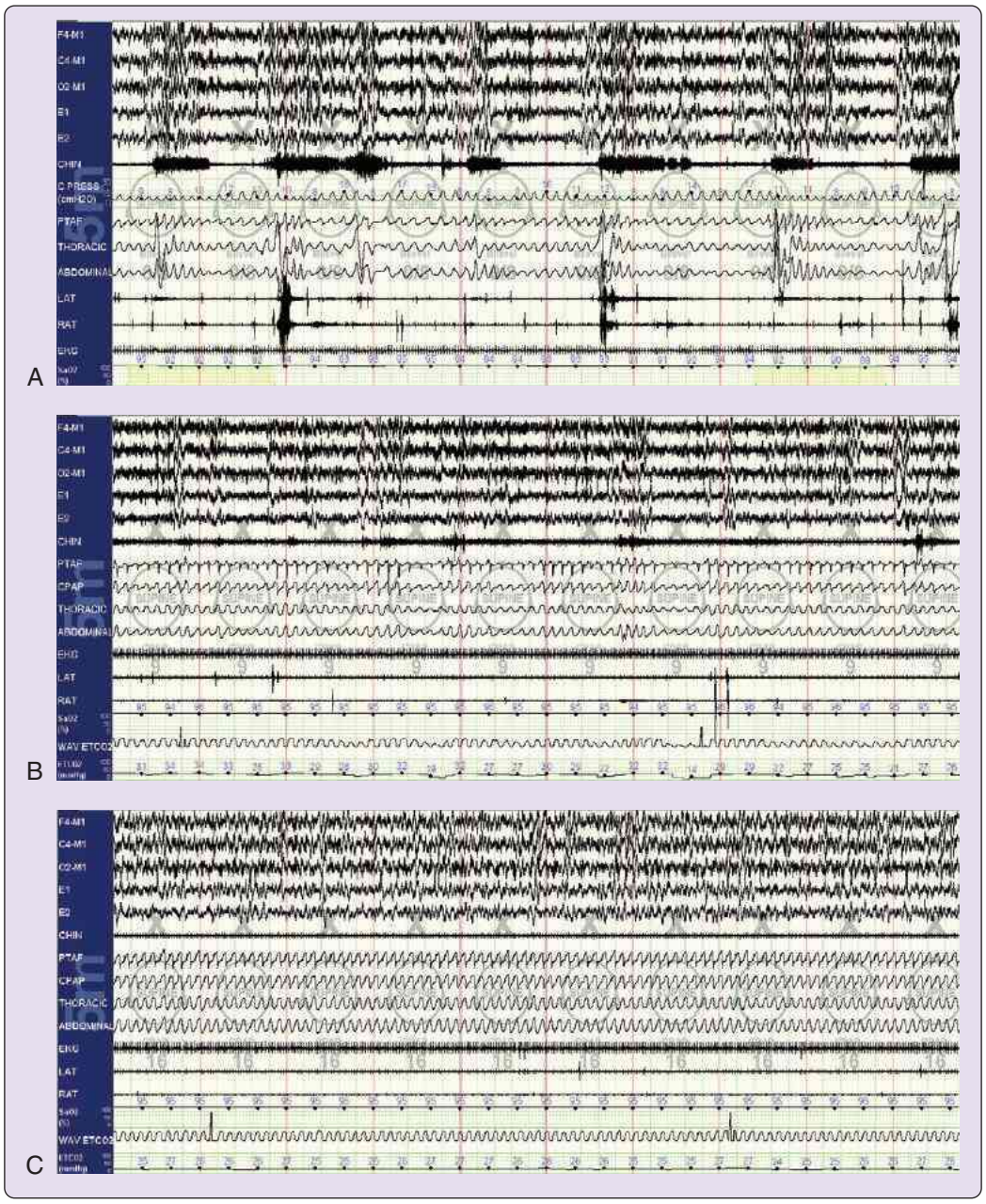


Figure 110-11 **Efficacy of CO_2 manipulation for stabilizing respiration.** A 72-year-old man with congestive heart failure (same patient as in Figure 110-3). *Top,* Adaptive servoventilation failure; note excessive cycling of pressure (the CPRESS channel) and the associated arousals. *Middle,* After addition of 50 mL enhanced expiratory rebreathing space (EERS), note stabilization of respiratory rhythm but residual flow limitation and mild residual periodic breathing. *Bottom,* With 100 mL EERS; note normalization of sleep and respiration. The end-tidal CO_2 ($ETCO_2$) signal plateau is slightly blunted and thus the CO_2 measured is falsely low. However, the resting wake $ETCO_2$ of this patient was 30 mm Hg, a level that can be readily seen in patients with congestive heart failure.

opiates may help reduce the frequency of central apneas[4] and should be routinely considered within the constraints of the disorder for which it was prescribed. In many patients, however, stopping opiates entirely may not be possible. CPAP alone will rarely be effective therapy. Opiate-induced ataxic breathing is quite sensitive to CO_2 levels—with ready induction of central apnea and worsening of dysrhythmic breathing on continuous or nonadaptive bilevel positive pressure ventilation. Although these patients tend to show mild hypercapnia, with $ETCO_2$ in the high 40- to low 50-mm Hg range, using a nonvented mask and EERS as needed to hold CO_2 in the mid 40-mm Hg range (thus preventing destabilizing degrees of hypocapnia despite supporting the upper airway) can be helpful regardless of the positive pressure mode used. We have found the use of acetazolamide to be of consistent benefit. Adaptive ventilation is a double-edged sword in these patients, being able to both enable stable breathing and markedly destabilize breathing.[176,196,197]

Carbonic Anhydrase Inhibition

Acetazolamide, a diuretic and carbonic anhydrase inhibitor, diminishes the ventilatory response of the peripheral chemoreceptors to hypoxia, decreases loop gain, and reduces the ventilatory response to arousals.[198-201] In animal models, it has been shown to lower the P_{ETCO_2} apnea threshold and widen the difference between the eupneic and P_{ETCO_2} thresholds.[21] Acetazolamide has been used in treating nonhypercapnic CSA or CSB, in patients with and without heart failure.[202] Although such results may be statistically significant, the degree of residual sleep apnea is unacceptable as sole long-term therapy. The drug may convert those with mixed OSA and CSA to mostly obstructive (the reverse of CPAP-induced CSA). Those with short cycle (≤30 seconds) periodic breathing not responding to EERS or adaptive ventilation are the best candidates. Acetazolamide has been successfully used as CPAP adjuncts at high altitude.[203,204] Zonisamide[205] and topiramate[206] have carbonic anhydrase inhibitory effects and could in theory be used in the place of acetazolamide. Acetazolamide can aid the treatment of hypercapnic CSA by reducing the propensity for worsening of respiratory instability, which pressure support ventilation can induce. The drug may have a special role in those with coexisting CSA in NREM sleep (requiring low levels of PAP) and OSA in REM sleep (requiring higher pressures that could exacerbate CSA in NREM sleep).

Other Drugs with Possible Benefits in Nonhypercapnic Central Sleep Apnea

Clonidine[207] and the 5-α reductase inhibitor finasteride[208] have been shown to improve breathing stability. Inhibition of H_2S (gaseous signal transmitted in carotid body) in a rodent model of CHF nearly normalized chemosensitivity and breathing instability, and may serve as a new therapeutic target.[209] A case report of unilateral carotid body denervation in a man with systolic heart failure and moderate CSA showed that chemosensitivity and sleep apnea severity were reduced and shifted to an obstructive phenotype 2 months after treatment, accompanied by an improvement in quality of life.[210] A recently completed trial of carotid body denervation in systolic heart failure patients may provide key risk-benefit information (ClinicalTrials.gov Identifier: NCT01653821) on pharmacologic targeting of carotid body function.

Other Treatment Options

Phrenic nerve stimulation[211,212] is an investigational approach that can improve the CAHI in heart failure patients with CSA-CSB. Sleep quality may not improve proportionately, and there is a concern that suppression of central apneas without enabling stable breathing or targeting the core pathophysiology of high loop gain may not provide long-term benefits, similar to the SERVE-HF study. A subset of CSA and treatment emergent CSA patients appear very supine position dependent, and avoidance of the supine position can markedly improve treatment efficacy.[116,213] An additional effect of body position, this time from vertical to horizontal, is on fluid redistribution from the caudal to cranial parts of the body.[214-218] The effect is rapid and is associated with increased neck circumference and hypocapnia from increased lung water in patients with central apnea. Therapeutic manipulation could include careful diuresis, a wedge pillow, or sleeping in a recliner.

CLINICAL PEARLS

- Sleep apnea caused by a pathologically activated respiratory chemoreflex results in a wide PSG spectrum of disease, with variable features of upper airway obstruction mixed with more traditionally accepted central patterns.
- Pathophysiology guides treatment. Treatment end points should aim to normalize sleep and sleep-breathing biology, not merely suppression of scored events to an arbitrary threshold.
- It is useful to consider CSA as taking hypercapnic and nonhypercapnic forms. Accurate phenotyping of sleep apnea is increasingly important because several on-label and off-label therapies, singly or in combination, have improved therapeutic options for CSA syndromes.
- Recognizing NREM sleep dominance of disease in nonhypercapnic CSA and minimizing the importance of concomitant upper airway flow limitation when identifying otherwise typical periodic breathing are applicable in clinical and research assessments. Hypercapnic CSA may be REM dominant, with the exception of opiate-induced CSA, and require ventilatory support for management. Unlike OSA, CSA is relatively difficult to treat and increases the risk for poor compliance, residual symptoms, ongoing sleep fragmentation, and high residual respiratory events despite therapy.

SUMMARY

CSA caused by pathologic activation of the respiratory chemoreflex includes idiopathic CSA, periodic breathing, high-altitude sleep apnea, and treatment emergent or complex sleep apnea. A narrow NREM sleep CO_2 reserve and propensity for arousal and sleep fragmentation are key pathophysiologic drivers. A unifying theme for nonhypercapnic CSA is predominance in NREM sleep and a metronomic appearance; cycle times, the duration of the respiratory event from peak to peak or trough to trough, can be short (≤30 seconds). Sleep fragmentation is often severe, but hypoxia is relatively moderate in nonhypercapnic CSA. The prevalence and evolution of residual CSA phenotypes with treatment remains to be accurately estimated owing to limitations of current approaches to PSG scoring; new methods need to be validated to outcomes. Non-invasive adaptive ventilation provides pressure support approximately equal and opposite to patient-generated ventilation, with substantial differences in individual device algorithms. Off-label approaches may be considered as adjuncts to improve therapeutic efficacy, including minimization of hypocapnia with dead space adapted to PAP, acetazolamide, and sedatives to reduce the arousal threshold.

Hypercapnic CSA is associated with pathologically reduced activation of the respiratory chemoreflex and may be seen in association with hypoventilation syndromes and neurologic disorders. Disease severity can be maximal in REM sleep, although successful ventilation (and reduction in hypercapnia) may result in respiratory instability in NREM sleep owing to relative hypocapnia. Non-invasive bilevel ventilation with a backup rate and volume-assured ventilation are primary therapeutic options; acetazolamide may improve respiratory drive and reduce NREM-related instability.

Selected Readings

Abraham WT, Jagielski D, Oldenburg O, et al. Phrenic nerve stimulation for the treatment of central sleep apnea. *JACC Heart Fail* 2015;**3**(5):360–9.

Correa D, Farney RJ, Chung F, et al. Chronic opioid use and central sleep apnea: a review of the prevalence, mechanisms, and perioperative considerations. *Anesth Analg* 2015;**120**(6):1273–85.

Cowie MR, Woehrle H, Wegscheider K, et al. Adaptive servo-ventilation for central sleep apnea in systolic heart failure. *N Engl J Med* 2015;**373**(12):1095–105.

Dempsey JA, Smith CA, Przybylowski T, et al. The ventilatory responsiveness to CO(2) below eupnoea as a determinant of ventilatory stability in sleep. *J Physiol* 2004;**560**:1–11.

Dempsey JA, Xie A, Patz DS, Wang D. Physiology in medicine: obstructive sleep apnea pathogenesis and treatment—considerations beyond airway anatomy. *J Appl Physiol* 2014;**116**:3–12.

Eckert DJ, White DP, Jordan AS, et al. Defining phenotypic causes of obstructive sleep apnea: identification of novel therapeutic targets. *Am J Respir Crit Care Med* 2013;**188**:996–1004.

Eckert DJ, Younes MK. Arousal from sleep: implications for obstructive sleep apnea pathogenesis and treatment. *J Appl Physiol* 2014;**116**:302–13.

Gaig C, Iranzo A. Sleep-disordered breathing in neurodegenerative diseases. *Curr Neurol Neurosci Rep* 2012;**12**:205–17.

Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;**3**:141–63.

Rao H, Thomas RJ. Complex sleep apnea. *Curr Treat Options Neurol* 2013;**15**:677–91.

Thomas RJ. Alternative approaches to treatment of central sleep apnea. *Sleep Med Clin* 2014;**9**:87–104.

Wolfe LF, Patwari PP, Mutlu GM. Sleep hypoventilation in neuromuscular and chest wall disorders. *Sleep Med Clin* 2014;**9**:409–23.

Xie A, Teodorescu M, Pegelow DF, et al. Effects of stabilizing or increasing respirator motor outputs on obstructive sleep apnea. *J Appl Physiol* 2013;**115**:22–33.

A complete reference list can be found online at ExpertConsult.com.

Chapter
111

Anatomy and Physiology of Upper Airway Obstruction

James A. Rowley; M. Safwan Badr

Chapter Highlights

- Upper airway patency is determined by craniofacial structure, surrounding tissues, intrinsic properties of the upper airway, and the neuromuscular function of the upper airway.
- With sleep, the loss of the wakefulness drive to breathe is associated with decreased neuromuscular activity of the upper airway, and in particular decreased afferent reflexes, and is further modified by such factors as lung volume, hypercapnia, and age. Upper airway caliber decreases, with a resultant increase in upper airway resistance and compliance and upper airway collapsibility.
- In patients with obstructive sleep apnea (OSA), there is evidence of histologic changes to upper airway tissues, which may explain abnormalities in sensorimotor and reflex activity in the upper airway in patients with OSA.
- This chapter describes and discusses the factors that influence the determinants of upper airway obstruction in humans. However, further research is needed to understand how these determinants interact to maintain upper airway patency.

Although investigations into the pathogenesis of obstructive sleep apnea (OSA) have been underway since the disorder was first described, the mechanisms underlying an increased propensity to sleep-related upper airway obstruction in some individuals are not well understood. This chapter reviews the anatomy and physiology of the upper airway as they relate to upper airway patency and propensity to obstruct during sleep. The occurrence of collapse during sleep and not wakefulness implicates the removal of the wakefulness drive to breathe as a key factor underlying sleep-related upper airway obstruction. The determinants of upper airway patency—upper airway neuromuscular activity and nonneuromuscular factors including craniofacial structure, surrounding tissues, and intrinsic properties of the upper airway itself—are discussed.

This chapter additionally reviews the effect of sleep on upper airway properties such as patency, resistance, compliance, and collapsibility. Within each area, the factors that influence these properties, including host factors (e.g., gender and body mass index [BMI]), disease (e.g., tonsillar hypertrophy, fluid overload), and OSA, are examined.

BASELINE DETERMINANTS OF UPPER AIRWAY PATENCY

Upper Airway Function and Structure

The human upper airway is unique in that it serves as a multipurpose passage. It both transmits air to the lungs (through the nose and mouth) and liquids and solids to the esophagus (through the mouth). The upper airway, particularly the nose, also serves as a heat exchanger. In humans, the upper airway, particularly the larynx and lips, is important for vocalization. However, because it serves multiple purposes, portions of the upper airway lack rigid support, and hence are prone to collapse.

The upper airway is classically divided into five regions based on anatomic structures (see Figure 111-1, *A*). Each of these regions is either rigid and resistant to collapse or semirigid and susceptible to collapse. Whether rigid or semirigid, however, each region can become occluded because of other anatomic variants or abnormalities. Although the nose is a rigid section of the upper airway because of its bony components, it can become obstructed owing to nasal congestion and polyps, altering upper airway mechanics. The nasopharynx is defined as the area from the posterior aspect of the nasal turbinates to the horizontal plane of the soft palate. Thus the proximal portion of the nasopharynx tends to be rigid, but the distal region is semirigid. Nasopharyngeal patency can be compromised by local mass lesions and palatal and uvular hypertrophy or edema. The oropharynx is defined as the area from the soft palate to the base of the tongue and is semirigid. It can be further divided into an area posterior to the soft palate (retropalatal) and tongue (retroglossal). Oropharyngeal patency is generally compromised from tonsil hypertrophy, palatal or uvular enlargement, or macroglossia. Because the nasopharynx and oropharynx are semirigid, these two areas are the site of collapse in most patients with OSA. The hypopharynx extends from the base of the tongue to the larynx and is relatively rigid and resistant to collapse. Finally, the larynx, the most distal portion of the upper airway, is rigid, composed of both cartilage and muscle.

Upper airway caliber can be measured by a variety of methods, including computed tomography (CT) scanning, magnetic resonance imaging (MRI), nasopharyngoscopy, and acoustic imaging. These methods allow researchers to measure

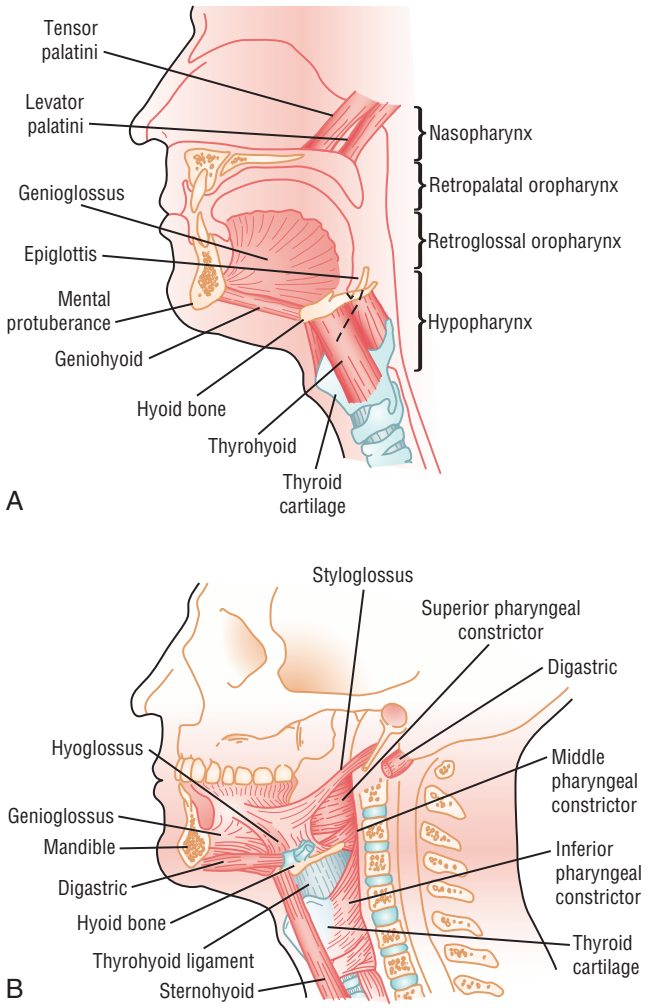


Figure 111-1 A, Schematic diagram of upper airway anatomy showing the classic divisions of the pharynx and key upper airway muscles. **B,** Schematic diagram of upper airway muscles and other key landmarks such as the hyoid.

upper airway caliber either at one point in the respiratory cycle (a static measurement) or dynamically throughout the respiratory cycle.

Neuromuscular Function of the Upper Airway

The upper airway musculature consists of 24 pairs of striated skeletal muscles extending from the nares to the larynx (see Figure 111-1, *B* for major anatomic landmarks and muscles).[1,2] These pharyngeal muscles have complex anatomic relationships but can generally be classified into groups that regulate the position of the soft palate, tongue, hyoid bone, and pharyngeal walls. The muscles are generally activated in groups to control the major functions of the upper airway such as phonation and swallowing.

There are two general patterns of electrical discharge from upper airway muscles when these are studied with multiunit electromyograms (EMGs): tonic (constant) activity, independent of phase of respiration; and phasic activity, occurring during one part of the respiratory cycle. There are at least 10 upper airway muscles that may be classified as pharyngeal "dilators," innervated by multiple cranial nerves. Some, such as the genioglossus, are classified as dilators by virtue of their phasic inspiratory activity. Others, such as the tensor palatini, do not clearly have a dilating effect but demonstrate activity throughout the respiratory cycle (tonic activity) and are presumed to "stiffen" the upper airway wall and decrease pharyngeal collapsibility. It is widely accepted that upper airway dilators play a critical role in preserving pharyngeal patency.[3] There is evidence from EMG studies that activity of upper airway dilators begins about 200 milliseconds before onset of thoracic pump activity in normal subjects.[4,5]

Upper airway narrowing or obstruction during sleep is associated with a sleep-related decrease in upper airway muscle activity. The effect of non–rapid eye movement (NREM) sleep on upper airway muscle function is complex and difficult to study because of the challenges in isolating the many other influences on upper airway muscle activity, such as changes in air flow, magnitude of negative pressure in the pharyngeal airway, and lung volume. Available evidence indicates that NREM sleep is associated with a reduction in tonic or phasic EMG activity in numerous upper airway muscles,[2] including the levator palatini,[6] tensor palatini,[7] palatoglossus,[6] and geniohyoid.[8] Studies measuring single motor unit activity of the genioglossus muscle also noted decreased activity of the phasic inspiratory motor units at sleep onset[9] and NREM stage 2 sleep with increased discharge frequencies and duration in NREM stage 3 sleep.[10] The EMG changes are accompanied by upper airway narrowing and increased upper airway resistance.

The effect of rapid eye movement (REM) sleep on upper airway muscle activity is more clearly documented. Activity of antigravity muscles is reduced during REM sleep, and there is strong evidence that activity of phasic upper airway dilating muscles, such as the genioglossus, is greatly attenuated during REM sleep,[11,12] particularly during periods of phasic rapid eye movements.[13,14] Reduced activity has also been shown for the alae nasi[13] and geniohyoid muscles.[8] Similar findings have been reported for single motor unit activity of the genioglossus.[15] In summary, the sleep state is associated with decreased upper airway muscle activity.

The response of upper airway muscles to chemical and mechanical perturbations during sleep may be relevant physiologically and clinically to such reduced activity. Negative pressure applied to the upper airway results in a brisk reflex response in upper airway muscle activity. This reflex is attenuated with application of topical lidocaine, indicating mediation through local mechanoreceptors.[16] Studies of reflex activity in the genioglossus, palatoglossus, and tensor palatini muscles show that this negative pressure reflex response is attenuated during NREM[17-19] and REM sleep[20] compared with wakefulness. Similarly, responsiveness of the genioglossus muscle to hypercapnia is attenuated during sleep.[21] This reflex response has also been shown to be attenuated with aging during both wakefulness and sleep.[22] These data suggest that upper airway dilator muscles are less able to maintain upper airway patency in the face of chemical or mechanical perturbations. Furthermore, there is evidence that lung volumes alter genioglossus muscle activity during NREM sleep, with decreases in end-expiratory lung volume being associated with increased genioglossus activity above baseline.[23]

The large number of upper airway muscles and their complex interactions mandate caution in extrapolating findings from studies focusing on the genioglossus or hypoglossal nerve activity alone, particularly because measurement of electrical activity of the muscle is not necessarily an appropriate

surrogate for muscle fiber shortening or indeed for upper airway dilation. In fact, there is evidence that upper airway muscle activation is not necessarily sufficient to dilate the upper airway under either physiologic or loading conditions, including resistive loading due to airway resistance[24] or elastic loading due to either increased soft or fat tissue[25-27] or small mandibular enclosure.[28] Nor may such activation be necessary; for example, in sleeping humans, increased end-expiratory lung volume has been found to result in decreased upper airway resistance and increased retropalatal cross-sectional area in association with *reduced* EMG activity of the genioglossus.[29] In patients with OSA, increased lung volume causes a substantial decrease in sleep-disordered breathing during NREM sleep,[30] and an inverse correlation between continuous positive airway pressure (CPAP) requirements and lung volume in patients with OSA has been found.[31]

It is also unclear whether complete atonia of the pharyngeal muscles increases upper airway collapsibility. For example, the pharyngeal airway becomes more collapsible in dead infants,[32,33] but not in paralyzed animal preparations.[34,35] In addition, REM sleep, associated with decreased neuromuscular stimulation, is not associated with changes in upper airway compliance or collapsibility in human studies of upper airway physiology.[36,37] Likewise, increased pharyngeal compliance in patients with sleep apnea cannot be attributed to decreased upper airway dilating muscle activity per se because patients with OSA show increased activity of the genioglossus muscle during wakefulness[38] and sleep,[39] perhaps as a compensation for anatomically reduced caliber. Similarly, when the pharyngeal airway is narrowed during hypocapnic central apnea,[40] more pronounced narrowing (or even closure) occurs in patients with OSA relative to normal control subjects despite complete inhibition of upper airway dilating muscle activity in both groups.

A related question is the role of upper airway dilating muscles in stabilizing the upper airway and therefore preventing upper airway obstruction, apart from activity that is directly associated with increased or decreased upper airway patency. There is evidence, for example, that stimulation of the hypoglossal nerve results in decreased collapsibility (i.e., a stiffer airway) and decrease surrounding pressure in animal models.[41,42] Existing studies indicate that supraphysiologic electrical stimulation of the hypoglossal nerve using a surgically implanted upper airway stimulation device in patients with OSA lead to significant improvements in the severity of the sleep-disordered breathing in a select group of patients.[43,44]

Nonneuromuscular Factors Contributing to Upper Airway Patency and Obstruction

Upper Airway Muscle Histology

There is a large body of research examining upper airway muscle histology in patients with OSA, based on the hypothesis that pathologic changes in upper airway muscle histology may promote upper airway obstruction by increasing propensity to upper airway muscle fatigue or delay reopening through impairment of sensorimotor function. Studies have shown a variety of histologic changes, including edema and mucosal gland hypertrophy,[45] neurogenic injury,[46] changes in muscle enzyme activity,[47,48] and leukocytic inflammation.[49,50] A consistent finding across studies is an increase in type 2 fast-twitch fibers in the genioglossus muscle of patients with OSA.[46,51-53] Because type 2 fibers are more likely to fatigue than type 1 fibers, these studies suggest that upper airway muscles in OSA patients are more susceptible to fatigue than in normal subjects. In contrast, there is no consistent finding for differences in tongue protrusive force[54,55] in patients with OSA compared with normal subjects.

Changes in sensorimotor activity of the upper airway in OSA patients have been studied based on evidence that upper airway sensory receptors contribute to apnea-terminal arousal and that topical anesthesia impairs these responses. In patients with OSA, changes in two-point discrimination, vibratory sensation, upper airway muscle reflex response to short air pulses, and sensory perception to varying air flow rates have been observed.[56-58] These changes may explain the observation that inspiratory load sensation is decreased in patients with OSA.[59,60] Such impairment of upper airway sensorimotor function may contribute to decreased upper airway muscle activity in response to upper airway obstruction, loading, or collapse.

Although the previous studies indicate changes in upper airway neuromuscular histology and sensorimotor function in patients with OSA, there is only minimal evidence linking specific histologic changes or sensorimotor changes to changes in upper airway mechanics or propensity of the upper airway to collapse. In fact, the noted histologic changes, if secondary to recurrent airway collapse, may not be contributing to an increased propensity to collapse. For instance, treatment with nasal CPAP has been associated with improvement in genioglossus muscle force production[52] and vibratory thresholds.[56] Thus further investigation is necessary to determine to what extent changes in upper airway histology and function seen in patients with OSA are primary or secondary and whether and how such changes, primary or secondary, contribute to the noted increased propensity to collapse in sleep in patients with OSA.

Craniofacial Structure

Craniofacial structure is an important determinant of upper airway patency. This is most evident in children with craniofacial abnormalities such as Pierre Robin sequence and Treacher Collins syndrome, each of which are associated with an increased prevalence of OSA.[61] In adults, several anatomic abnormalities have been associated with OSA, including retrognathia, micrognathia, overjet, and a high arched palate.[62-64]

Several investigations have used lateral cephalometry to analyze the contribution of craniofacial structure to the development of OSA (Figure 111-2).[65-73] These studies vary widely in methodology, sample size, gender ratios, and the presence and degree of obesity. In these studies, common craniofacial abnormalities that have been associated with increased severity of sleep apnea include (1) smaller airway dimensions, particularly those involving the maxilla and mandible; (2) mandibular retrognathia; (3) decreased posterior airspace; (4) an inferiorly placed hyoid bone; and (5) increased soft palate dimensions and length. These abnormalities decrease the dimensions of the nasopharynx and oropharynx, likely increasing the risk for upper airway obstruction. In one study in 57 male patients with OSA, airway collapsibility[74] was correlated with soft palate length, hyoid bone distance, and an inferiorly placed hyoid bone.

MRI has also been used to compare upper airway craniofacial structure and soft tissue between subjects with and

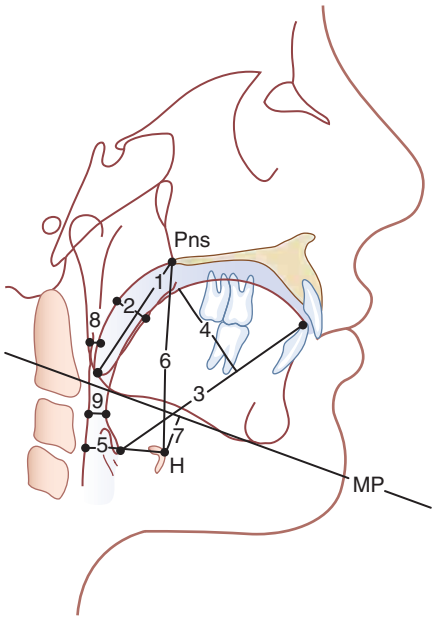


Figure 111-2 Cephalometric landmarks and soft tissue, hyoid position, and airway size variables frequently used in cephalometric studies. Landmarks: Pns, posterior nasal spine; H, hyoid bone; MP, mandibular plane (tangent line from the symphysis to the inferior border of the mandibular angle). Variables: (1) SPl, the length of the soft palate; (2) Spw, the width of the soft palate; (3) Tl, the length of the base of the tongue; (4) Tw, the width of the tongue; (5) H-Ph, the distance from the hyoid bone to the posterior wall of the pharynx; (6) H-Pns, the distance from the hyoid bone to Pns; (7) H-MP, the distance from the hyoid bone to the mandibular plane; (8) SPAS, the upper posterior pharyngeal space; (9) IPAS, the lower posterior pharyngeal space. In one study, critical closing pressure was predicted by the length of the soft palate (1), distance from hyoid bone to posterior wall of the pharynx (5), and the distance from the hyoid bone for the posterior nasal spine (6). (From Sforza E, Bacon W, Weiss T, et al. Upper airway collapsibility and cephalometric variables in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 2000;161[2 Pt. 1]:347–352.)

without OSA (Figure 111-3). Important differences have been found in men and include a wider mandibular divergence, smaller mandibular length, and smaller area at the mandibular plane being associated with OSA.[75,76] Although the hyoid bone was inferiorly placed in these studies in subjects with OSA, it was not a primary determinant of upper airway obstruction.

Craniofacial indexes derived from both cephalometric and MRI data have been used to compare OSA susceptibility between races. Redline and colleagues found that bony and soft tissue factors and brachycephaly were associated with OSA in whites, whereas only soft tissue factors were similarly associated in African Americans.[66,77] In contrast, Polynesian men with OSA were shown to have more mandibular retrognathia and larger nasal aperture width than white men, whereas neck circumference, tongue, and soft palate dimensions were associated with the respiratory disturbance index in the white subjects.[78] Finally, Japanese subjects with OSA have upper airway bony dimensions, whereas obesity and upper airway soft tissue and volume are less important.[76,78a] Taken together, these data indicate that race-specific craniofacial and neck structural factors contribute to the likelihood of having OSA.

Only one study has specifically compared gender differences in craniofacial and structural factors that could contribute to a propensity for OSA. Using MRI, Malhotra and colleagues[79] found that men had increases in airway length, soft palate cross-sectional area, and airway volume, presumably contributing to a diathesis for upper airway obstruction.

It should be noted that in many of these craniofacial studies, obesity remained the predominant etiologic factor for OSA, with abnormal craniofacial structure most important in nonobese patients with OSA. However, in a recent study in a large sample of men, obesity alone explained only 26% of the variance in the apnea-hypopnea index (AHI), and obese patients with unfavorable airway dimensions were susceptible to larger increases in OSA severity.[67]

Surrounding Tissues and Pressures

A collapsing upper airway transmural pressure can be generated either by a negative intraluminal pressure or a collapsing surrounding pressure. The role of negative intraluminal pressure in the pathogenesis of upper airway obstruction is widely hypothesized,[3] whereby a subatmospheric intraluminal pressure generated by the thoracic pump muscles causes upper airway collapse by "sucking" the hypotonic upper airway. However, there are no data showing that such subatmospheric intraluminal pressure causes upper airway obstruction in sleeping humans. In addition, upper airway narrowing and obstruction do not appear to require negative pressure. For example, studies using fiberoptic nasopharyngoscopy have shown that the upper airway narrows during hypocapnia mediated central inhibition.[40,80] Isono and colleagues[81] compared the mechanics of the pharynx in anesthetized and paralyzed normal subjects and in patients with OSA. The pharynx was patent at atmospheric intraluminal pressure in normal subjects and required negative intraluminal pressure for closure. In contrast, patients with OSA had a positive closing pressure; that is, the pharynx was occluded at atmospheric intraluminal pressure. Similarly, the critical closing pressure in patients with OSA has been generally found to be positive, as opposed to the negative critical closing pressure in normal subjects.[82, 83]

The disconnect between the occurrence of upper airway obstruction and of negative intraluminal pressure supports the possibility that upper airway patency is, in part, determined by the extrinsic or surrounding pressure contributed to by properties of soft tissue structures of the upper airway. Using MRI technology, three factors have been found to be most significantly associated with an increased risk for OSA: increased tongue size, increased size of lateral pharyngeal walls, and increased total soft tissue volume (Figure 111-4).[84] The association of increased tongue and lateral pharyngeal wall size with OSA has also been noted in CT and cephalometric studies of the upper airway[65,85] as well as in clinical studies.[62] Subsequent work has shown that these same factors show familial aggregation, even after correction for confounding factors such as gender and age.[86] Thus the known familial predisposition to OSA[87] may be in part explained by heritable soft tissue factors.

Enlarged tonsils have also been shown to be associated with an increased risk of OSA even after correction for BMI and neck circumference.[62] Enlarged tonsils are particularly noted as a causative factor in children and thin adults, who may have resolution of OSA after tonsillectomy.[88]

CT and MRI of the upper airway have also demonstrated evidence of increased soft tissue volume and pharyngeal fat at

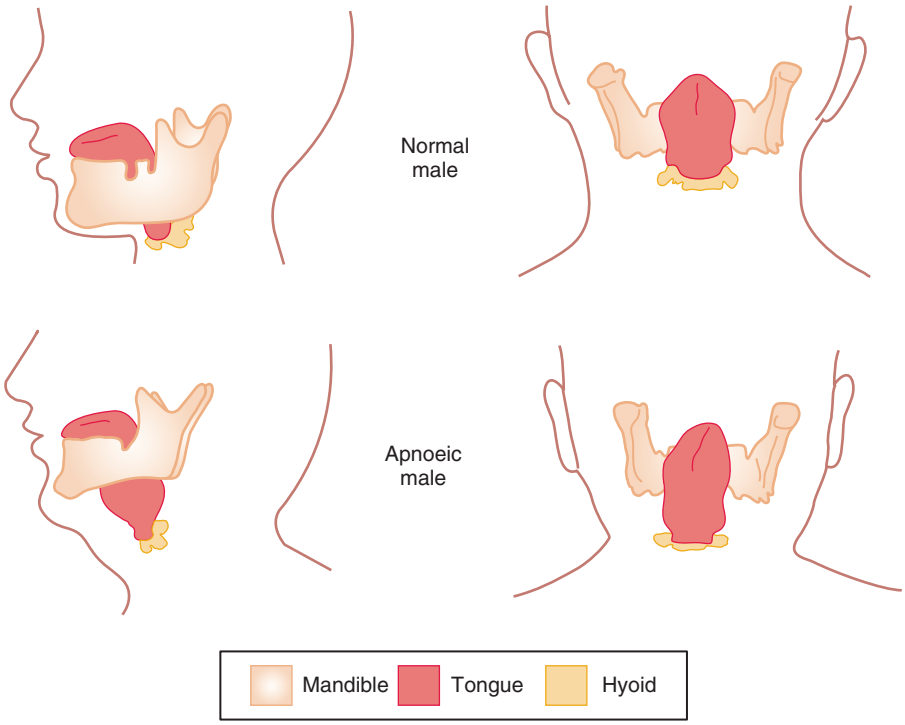


Figure 111-3 Three-dimensional reconstruction of hyoid, tongue, and mandible in a patient with obstructive sleep apnea (*bottom:* male, apnea-hypopnea index [AHI] 86 events/hr; body mass index [BMI] 31 kg/m^2; 49 years of age) and a normal subject (*top:* male, AHI 5 events/hr; BMI 25 kg/m^2; 44 years of age) illustrating the inferior-posterior positioning of hyoid and enlarged tongue volume. Note that the hyoid is more inferior-posteriorly positioned in the apneic subject than in the normal subject; tongue volume is greater in the apneic subject than in the normal subject. (From Chi L, Comyn FL, Mitra N, et al. Identification of craniofacial risk factors for obstructive sleep apnoea using three-dimensional MRI. *Eur Respir J* 2011;38[2]:348–58.)

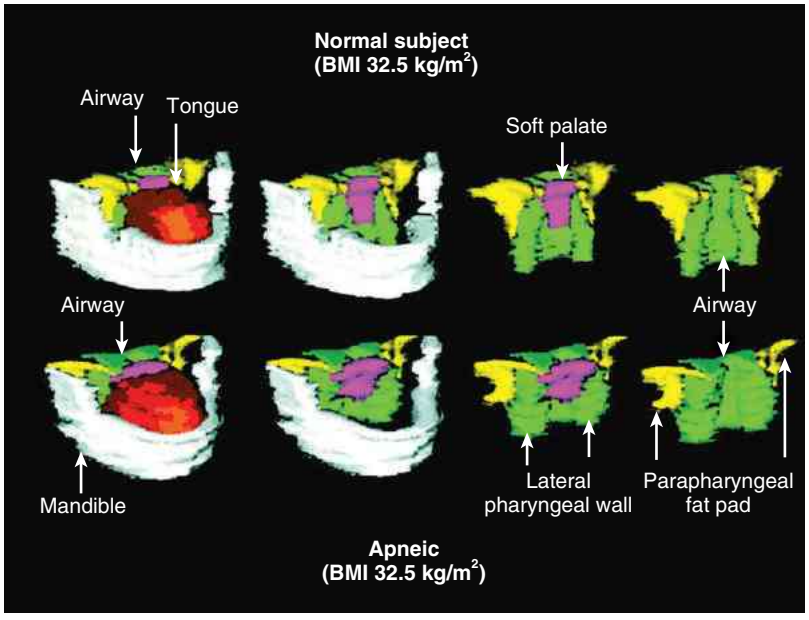


Figure 111-4 Volumetric reconstruction of axial magnetic resonance images in a normal subject and a patient with sleep apnea, both with an elevated body mass index of 32.5 kg/m^2. The mandible is depicted as *gray*, tongue as *orange/rust*, soft palate as *purple*, lateral parapharyngeal fat pads as *yellow*, and lateral/posterior pharyngeal walls as *green*. Note that the airway is larger in the normal subject than in the apneic subject. The tongue, soft palate, and lateral pharyngeal walls are larger in the patient with sleep apnea. (From Schwab RJ, Pasirstein M, Pierson R, et al. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 2003;168[5]:522–30.)

the level of the nasopharynx in males,[89] which could explain, in part, their higher prevalence of OSA. Pharyngeal fat volume was found to correlate with the AHI in one study,[27] but not in other studies.[84,85] Further investigations are needed to better determine the role of pharyngeal fat volume in particular, and extrinsic tissue volume and pressures overall, in the generation of a collapsing transmural pressure and the pathogenesis of upper airway obstruction in sleeping humans.

Intrinsic Properties of the Upper Airway

The collapsing effect of transmural pressure on upper airway patency is subject to modification by the intrinsic compliance of the pharyngeal wall. In addition to the Isono and colleagues[81] data noted previously, it has been shown that, in the isolated upper airway model of collapsibility, critical closing pressure is negative during complete paralysis, indicating that at normal atmospheric pressure, the normal upper airway remains open.[90,91] These studies suggest that the pharyngeal wall has an intrinsic "stiffness" or resistance to collapse. The determinants of such intrinsic stiffness have not been fully elucidated but likely involve complex interactions among many of the pharyngeal components already discussed, including muscles (which may have different properties in a passive state compared with a stimulated state), bony structures (particularly in the nasopharynx), soft tissues, and vascular properties, with such interactions affecting pressure and cross-sectional area relationships.

In the passive or paralyzed nasopharynx, the relationship between pharyngeal transmural pressure and cross-sectional area is curvilinear, implying that the airway becomes more compliant as the cross-sectional area decreases, that is, the "tube law" (Figure 111-5). Therefore baseline decreased airway cross-sectional area is likely a determinant of diathesis for upper airway obstruction, supported by evidence that the pharyngeal airway is smaller during wakefulness in patients with OSA relative to that of normal subjects.[92-94] However, it has been shown that under the dynamic conditions of inspiratory flow limitation, airway compliance decreases at more negative driving pressures (associated with a smaller cross-sectional area), indicating that static airway properties may not be as physiologically or clinically relevant during dynamic conditions.[94a]

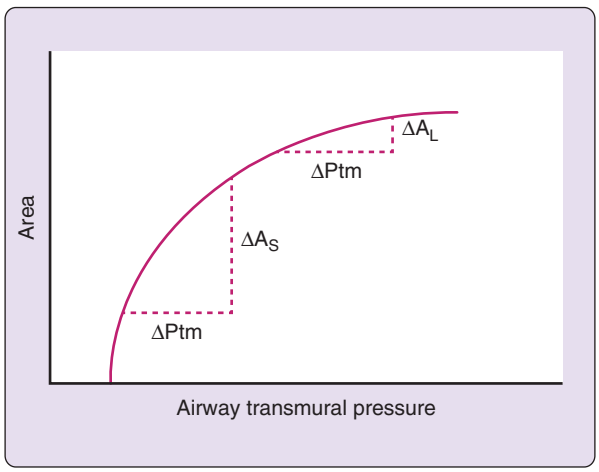


Figure 111-5 Illustration of the tube law of the pharynx. As transmural pressure (P_{tm}) increases, so does cross-sectional area (A). The slope of the tube law represents compliance of the pharynx. Note that compliance decreases as the area of the pharynx increases.

In association with such intrinsic properties of the upper airway, increased inspiratory lung volume is associated with increased upper airway caliber and decreased collapsibility. Potential mechanisms through which the increased caudal traction works include increased longitudinal tension, increased subatmospheric pressure through the trachea, and decreased transmural pressure.[42,95-97] Therefore caudal traction appears to influence upper airway collapsibility by both dilating and stiffening the pharyngeal airway as well as decreasing extramural tissue pressure. It is likely that patients with OSA are more dependent on such increased lung volume–associated dilatation or stiffening because of their relatively compliant upper airway.[81,98,99]

Vascular perfusion of the upper airway is also a potential determinant of intrinsic pharyngeal wall stiffness. Vasoconstriction and vasodilation have been shown to cause a decrease and increase in upper airway resistance, respectively.[100-102] More recently, a series of experiments have investigated the relationship between rostral fluid shifts and upper airway properties, demonstrating an association between reduction in leg fluid volume and increased neck circumference. In awake subjects, such rostral body fluid shift is associated with increased pharyngeal resistance,[103] decreased upper airway cross-sectional area,[104] and increased upper airway collapsibility.[105,106] Similar results have been found in patients with drug-resistant hypertension[107] and end-stage renal dialysis,[108,109] two groups in which there is evidence of an increased prevalence of OSA.[110] Interestingly, despite similar changes in leg total fluid volume and neck circumference with lower body positive pressure, men have a larger increase in collapsibility than women,[106] suggesting that a differential response to fluid shifts between men and women could contribute to the difference in gender prevalence in OSA.

When upper airway closure occurs, surface mucosal forces may impede subsequent upper airway opening.[35] In awake humans, surfactant and other topical lubricants have been shown to decrease the opening and closing pressures of the upper airway and to decrease upper airway resistance in sleeping normal subjects.[111,112] Mucosal lining forces may be particularly important in patients with OSA with mucosal inflammation from repeated trauma,[111] in whom the AHI in sleep decreases with the use of such soft tissue lubrication.[112,113]

In summary, upper airway patency is determined by multiple factors that are present during wakefulness and sleep, all typically further compromised during sleep compared with awake, leading to changes in upper airway function that contribute to sleep-related upper airway obstruction. The four primary factors include neuromuscular activity of the upper airway, craniofacial structure, tissues surrounding the upper airway, and the intrinsic properties of the airway (Figure 111-6). The following section reviews the sleep-specific effects on these factors and associated upper airway resistance and patency.

SLEEP EFFECTS ON UPPER AIRWAY PATENCY AND COLLAPSIBLITY

The sleep state is a challenge, rather than a period of rest, for the ventilatory system. In addition to the reduced activity

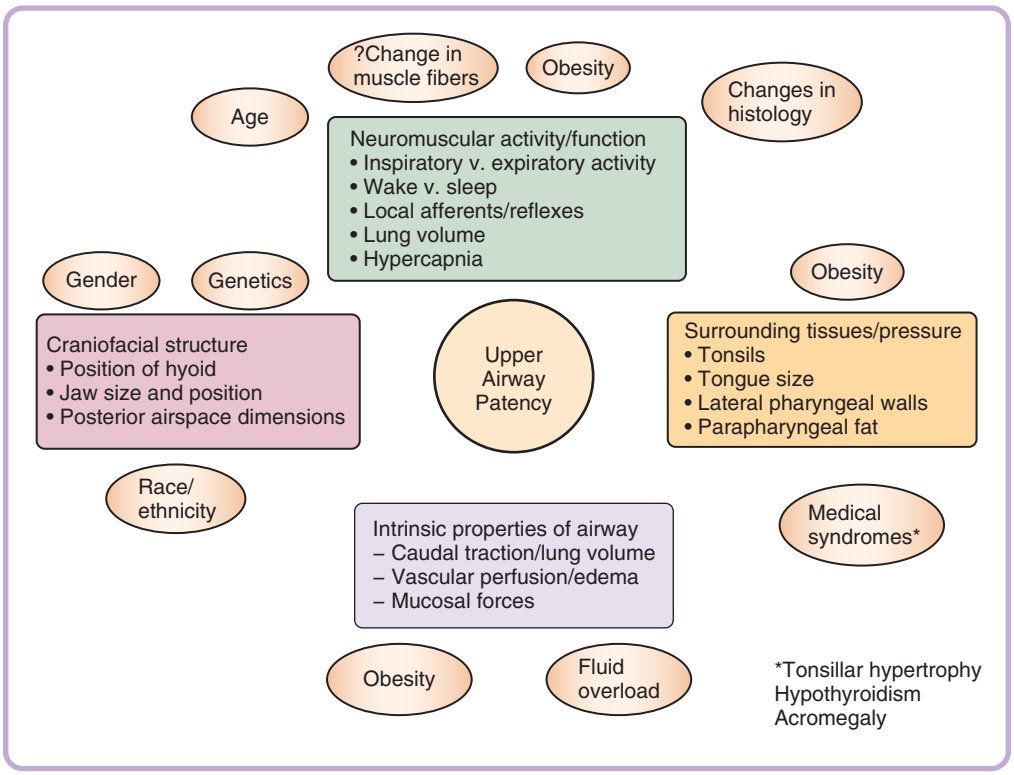


Figure 111-6 Summary showing the major determinants of upper airway patency and other factors that modify these main determinants.

of upper airway dilators discussed previously, consequences of the loss of wakefulness on the upper airway include reduced upper airway caliber, increased upper airway resistance, increased pharyngeal compliance, and collapsibility. Ultimately, these changes lead to reduced tidal volume and hypoventilation.

Upper Airway Caliber and Resistance

The sleep state is associated with upper airway narrowing and a corresponding increase in upper airway resistance. Using nasopharyngoscopy during sleep in normal subjects, Rowley and colleagues[46,126] have shown that, during NREM sleep, both retropalatal and retroglossal cross-sectional area decreases to approximately 70% of awake cross-sectional area, with further narrowing of the retroglossal airway during REM sleep. Decreased cross-sectional area corresponds with the pattern of decreased upper airway dilator muscle activity during NREM sleep and further reduction of the genioglossus during REM sleep. In REM sleep, retroglossal but not retropalatal cross-sectional area decreases further compared with NREM sleep.[36,114]

The evidence for increased upper airway resistance during sleep is compelling, even in normal subjects.[36,114-116] In fact, increased upper airway resistance occurs as early as sleep onset and continues to increase, reaching highest values in slow wave sleep. Most evidence indicates that there are no further increases in upper airway resistance during REM sleep compared with NREM sleep in normal humans.[36,114,117] In summary, the sleep state is associated with upper airway narrowing, which manifests as increased upper airway resistance and decreased pharyngeal caliber.

It is important to note that upper airway resistance is not an independent measure of the dynamic behavior of the pharyngeal airway during sleep. Many subjects exhibit inspiratory flow limitation, in which the pressure-flow graph demonstrates a changing relationship between driving pressure and inspiratory flow, culminating in complete dissociation between pressure and flow; that is, pressure continues to decrease with no further increase in flow (Figure 111-7, *A*). Thus the optimal physiologically meaningful measurement of upper airway resistance is the slope of the linear portion of the pressure-flow loop, which likely reflects upper airway caliber at the narrowest point in the upper airway at the beginning of inspiration (Figure 111-7, *B*).

Measurements and Meanings of Compliance and Collapsibility

The walls of the pharyngeal airway consist of compliant soft tissue structures, amenable to changes in pressure during the respiratory cycle. During wakefulness, upper airway caliber is constant during inspiration, with a decreased caliber during expiration, returning to inspiratory values at end-expiration. This finding has been observed in both normal subjects[92,118] and in patients with OSA[118] using either CT scanning or nasopharyngoscopy. Using nasopharyngoscopy, NREM sleep has been observed to be associated with significant dynamic within-breath changes in cross-sectional area, reaching a nadir at mid-inspiration,[118] with a rapid increase in cross-sectional area during expiration (Figure 111-8).[40] In addition, in subjects with OSA, there is progressive upper airway narrowing before the onset of apnea, which is primarily seen during expiration.[119] BMI appears to be a determinant of the degree

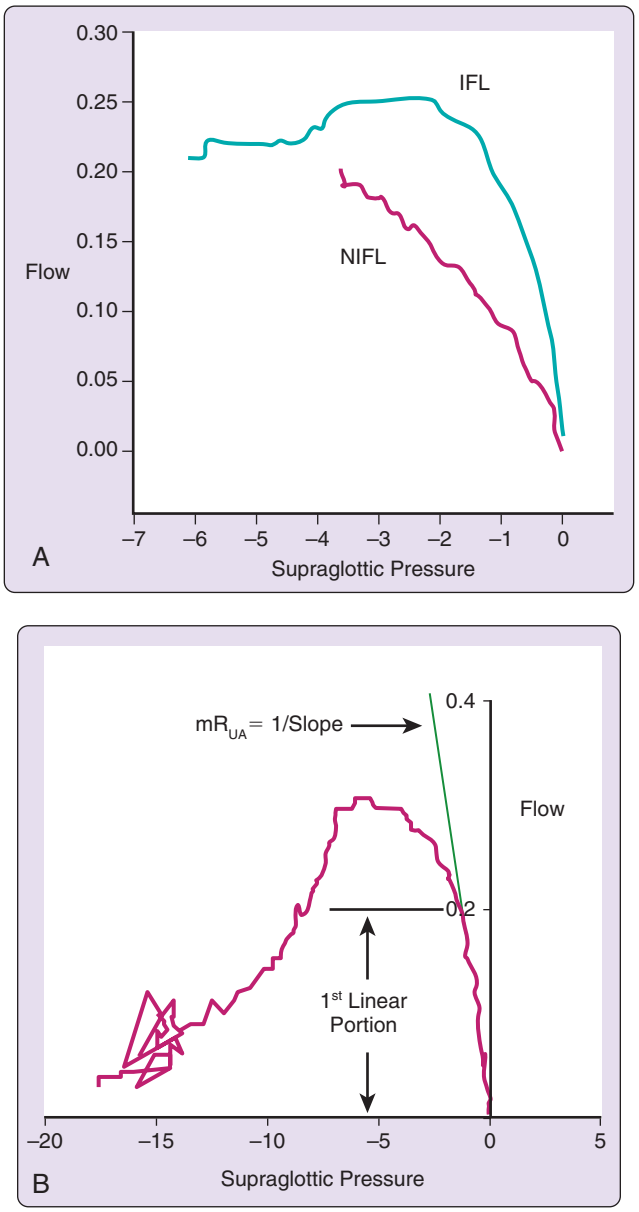


Figure 111-7 A, Pressure-flow loops illustrating a non–flow-limited (NIFL) and a flow-limited (IFL) breath. **B,** Illustration of a flow-limited breath and measurement of resistance along the first linear portion of the pressure-flow loop.

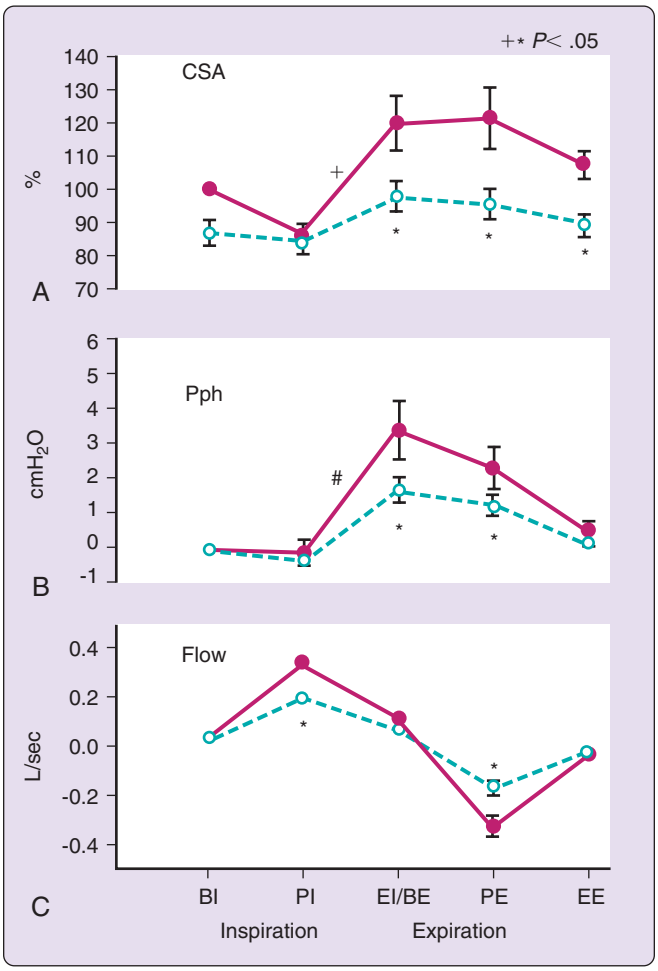


Figure 111-8 Retropalatal cross-sectional area (CSA) **(A)**, pharyngeal pressure (Pph) **(B)**, and flow **(C)** during control (*closed circles*) and hypocapnic hypopnea (*open circles*). Note the significant CSA change throughout the respiratory cycle within the control breaths compared with the hypopnea breaths. BE, Beginning expiration; BI, beginning inspiration; EE, end expiration; EI, end inspiration; PE, peak expiration; PI, peak inspiration. (Modified from Sankri-Tarbichi AG, Rowley JA, Badr MS. Expiratory pharyngeal narrowing during central hypocapnic hypopnea. *Am J Respir Crit Care Med* 2009;179:313–9.)

of airway narrowing seen in these studies.[40,118] The sleep reversal in the pattern of change in upper airway cross-sectional area is thought to be due to sleep-related increase in upper airway compliance, a decrease in pharyngeal caliber, and subsequently, decreased (more negative) inspiratory intraluminal pressure.

The changes in upper airway patency during sleep can be investigated using compliance as a measurement. Compliance of the pharyngeal wall is an important modulator of the effect of pressure changes on upper airway patency. The occurrence of pharyngeal narrowing and flow limitation suggests, although does not prove, increased pharyngeal compliance during sleep. Using a methodology that measures changes in cross-sectional area at different levels of applied pressure, it has been demonstrated that compliance is increased as the pharyngeal caliber decreases[81,120,121] and that the upper airway of patients with OSA is more compliant than that of normal subjects,[81,98,121,122] consistent with an increased propensity to collapse. Using a methodology that combines measurement of cross-sectional area using fiberoptic nasopharyngoscopy and measurement of intraluminal pressure at the same level in normal subjects has confirmed that retropalatal compliance is increased during NREM sleep compared with wakefulness, with no difference between REM sleep and wakefulness.[36] At the retroglossal level, however, compliance is not increased during either NREM or REM sleep compared with wakefulness.[114] The dissociation between compliance and reported muscle activity in these studies is consistent with studies in patients with OSA demonstrating that increased pharyngeal compliance occurs despite an increased activity of the genioglossus muscle during wakefulness[38] and sleep,[39] perhaps as a compensation for anatomically reduced caliber. This finding again suggests a major role for nonneuromuscular factors, as referred to previously, as determinants of pharyngeal compliance.

Collapsibility, which is the propensity of the upper airway to collapse or obstruct under certain conditions, increases during sleep compared with awake, likely owing to many of the factors discussed previously. Upper airway collapsibility has been primarily measured using the critical closing pressure, Pcrit, which is based on the concept of the Starling resistor,[123] whereby maximal flow through the resistor is dependent on the resistance of the upstream segment and the pressure surrounding the collapsible segment (Figure 111-9). In humans, the critical closing pressure can be partitioned between its passive mechanical properties (passive Pcrit) and active dynamic responses (active Pcrit).[124,125] Applying this model to humans, it has been shown that across the spectrum of obstructive sleep-disordered breathing, active Pcrit correlates with propensity for airway collapse.[82,83,126] For instance, Pcrit in normal subjects is generally less than 10 cm H_2O, whereas in patients with predominant hypopneas, it is between zero and −5 cm H_2O, and in patients with predominant apneas, it is more than zero cm H_2O. Although both active and passive Pcrit are increased in patients with OSA compared with control subjects, the difference between active and passive Pcrit is greater in non-OSA controls compared with those with OSA, likely associated with the greater ability of normal subjects to maintain airway patency.

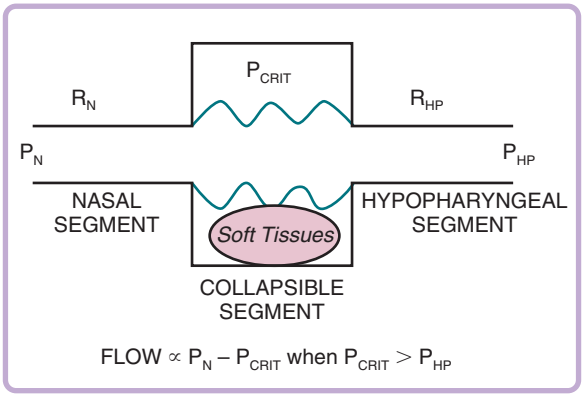


Figure 111-9 Starling resistor model of the upper airway. In this model, flow is proportion to the difference between P_N and Pcrit, with Pcrit greater than P_{HP}. P_N, Nasal (upstream) pressure; P_{HP}, hypopharyngeal (downstream) pressure; R_N, resistance in the nasal segment; R_{HP}, resistance in the hypopharyngeal segment.

The roles of mechanical loads and compensatory responses related to Pcrit are summarized in Figure 111-10. As shown in the left-hand bar with graded shading, approximate levels of Pcrit measurements define a continuum of upper airway collapsibility from health to disease. A Pcrit of approximately −5 cm H_2O represents the level above which obstructive hypopneas and apneas will occur. Structural characteristics of the upper airway impose mechanical loads and increase Pcrit, predisposing the upper airway toward collapse. Intact dynamic neuromuscular responses decrease Pcrit and maintain upper airway patency. In contrast, blunted neuromuscular responses increase Pcrit and predispose the upper airway toward obstruction.

Phase of the Respiratory Cycle: Inspiratory Versus Expiratory Narrowing

Upper airway obstruction during sleep is characteristically attributed to inspiratory narrowing owing to a collapsing subatmospheric pressure against a hypotonic pharyngeal airway. However, several lines of evidence implicate expiratory narrowing as a possible mechanism of the initial narrowing. First, ventilatory motor output is an important determinant of upper airway patency. Oscillation of ventilatory motor output, during the characteristic periodic breathing of OSA, is associated with pharyngeal narrowing or obstruction at the nadir of the motor output, especially in individuals with a highly

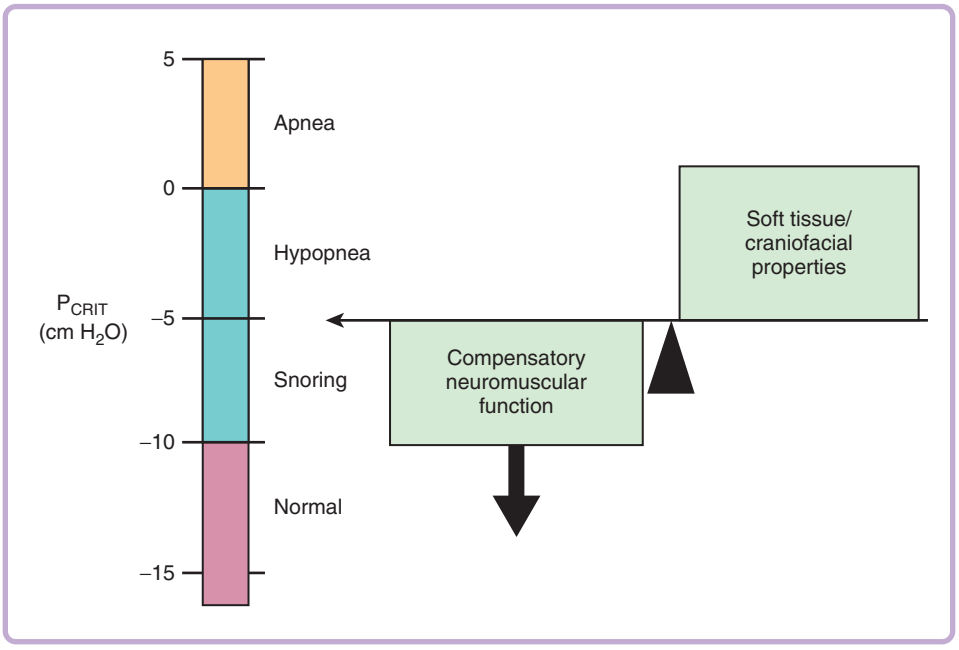


Figure 111-10 Role of mechanical loads and compensatory neuromuscular responses in the context of critical pressure measurements. See text for explanation. (From Patil SP, Schneider H, Marx JJ, et al. Neuromechanical control of upper airway patency during sleep. *J Appl Physiol* 2006;102:547–56, with permission.)

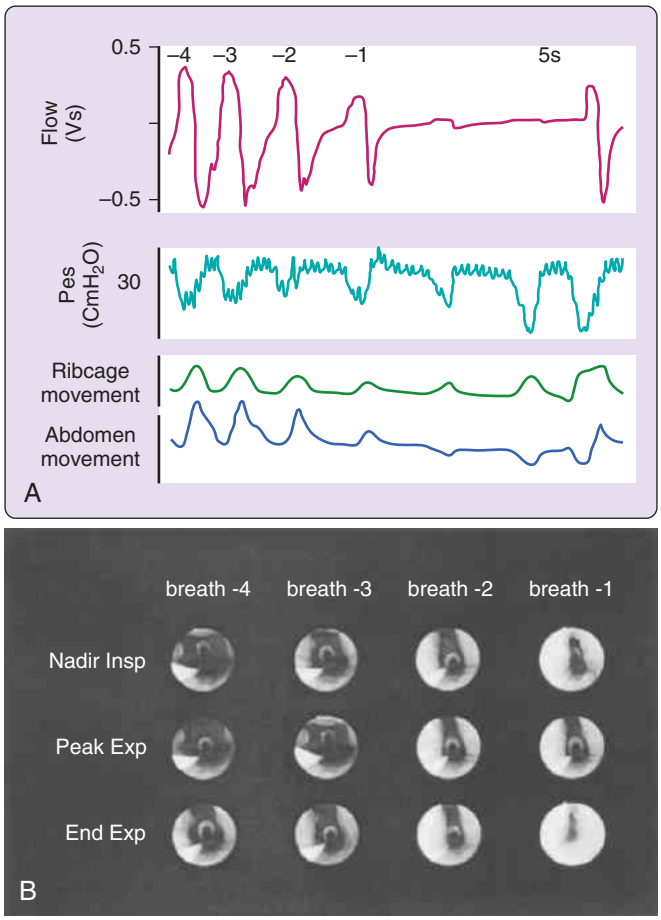


Figure 111-11 **A,** A recording of air flow (flow; inspiration positive), esophageal pressure (Pes), and rib cage and abdominal movements. Tracings show four breaths leading to an obstructive apnea (breaths -4, -3, -2, and -1). During the apnea, respiratory effort is indicated by the negative swings in the esophageal pressure and paradoxical rib cage and abdominal movements. **B,** Fiberoptic images of the retropalatal airway during the four breaths shown in **A** where breath-1 is the breath immediately preceding the apnea and breath-4 is the breath farthest away from the apnea. Within each breath the images selected correspond to the smallest cross-sectional area (CSA) that occurred during inspiration (Nadir Insp), the largest CSA during expiration (Peak Exp), and the CSA at end-expiration (End Exp). Note that progressive narrowing is occurring in both inspiration and expiration. Within each image the *dark area* is the airway lumen, the *lighter horseshoe shape* is the epiglottis, and the *white triangular shape* in the *bottom left corner* is the esophageal pressure catheter. (From Morrell MJ, Arabi Y, Zahn B, Badr MS. Progressive retropalatal narrowing preceding obstructive apnea. *Am J Respir Crit Care Med* 1998;158[6]:1974–81.)

collapsible airway.[127] Second, an obstructive apnea is often preceded by expiratory narrowing of the upper airway as evidenced by increased expiratory resistance[128] or progressive expiratory narrowing, detected by fiberoptic imaging (Figure 111-11).[119] Finally, although upper airway narrowing or occlusion occurs during a spontaneous or induced hypocapnic central apnea[80] or induced hypocapnic hypopnea,[40] pharyngeal narrowing during central hypopnea occurs during the expiratory phase only and is associated with increased expiratory upper airway compliance. Therefore upper airway obstruction may occur in either inspiration or expiration (Figure 111-12). Individuals with a high surrounding tissue pressure may be particularly susceptible to expiratory pharyngeal narrowing during such low ventilatory motor output and driving pressure.

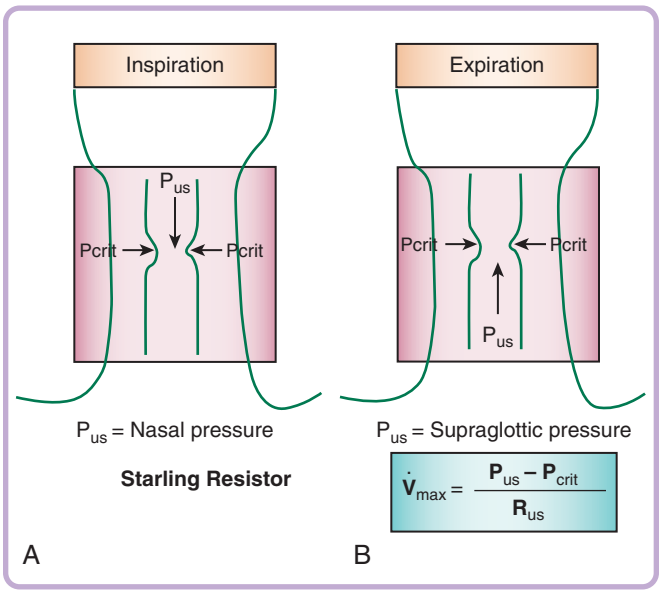


Figure 111-12 Schematic illustration for the collapsible segment of upper airway during hypocapnic hypopnea as a Starling resistor. In this model, flow is determined by the gradient between the upstream segment and critical closing pressure (Pcrit). During inspiration **(A)**, when upstream pressure (P_{us}) (i.e., nasal pressure) is below the Pcrit, the collapsible segment is closed, and no flow occurs. During expiration **(B)**, when P_{us} in the supraglottic area is below the Pcrit, the collapsible segment is closed, and no flow occurs. During hypocapnic hypopnea, expiratory flow is limited, correlating with the gradient between the supraglottic pressure and Pcrit. Hence this pressure gradient is an important determinant of pharyngeal narrowing. R_{us}, Upstream resistance; $\dot{V}_{max}$, maximal flow. (From Sankri-Tarbichi AG, Rowley JA, Badr MS. Expiratory pharyngeal narrowing during central hypocapnic hypopnea. *Am J Respir Crit Care Med* 2009;179[4]:313–9.)

Gender, Body Mass Index, and Weight Effects on Upper Airway Structure and Function

Potential determinants of upper airway mechanics during sleep include many variables known to be associated with an increased prevalence of OSA, such as gender, BMI, and age.

Most studies indicate no consistent difference in upper airway caliber, or compliance, between men and women without OSA. Upper airway resistance during NREM sleep is also similar in both genders,[117,129] although one study[130] demonstrated higher upper airway resistance in men during slow wave NREM sleep. REM sleep has not been similarly studied. Likewise, studies during wakefulness demonstrate no significant difference in upper airway cross-sectional area or smaller airway in women.[92,131-133] In addition, sleep-related narrowing is similar in men and in women (approximately a 40% decrease in cross-sectional area for both genders) from wakefulness to NREM sleep.[134] However, it appears that men have increased retropalatal compliance compared with women[134] because of gender difference in neck circumference, again indicating that factors other than gender are important in explaining such gender differences in upper airway function. Finally, there is no demonstrated gender difference in Pcrit under active conditions.[129] Thus the available studies taken together do not suggest a gender difference alone in upper airway mechanics during wakefulness or sleep, except for higher retropalatal compliance in men compared with women, in subjects without OSA. However, gender has been found to modulate the effect of BMI on Pcrit, and thus upper airway collapsibility, with men, with or without OSA,

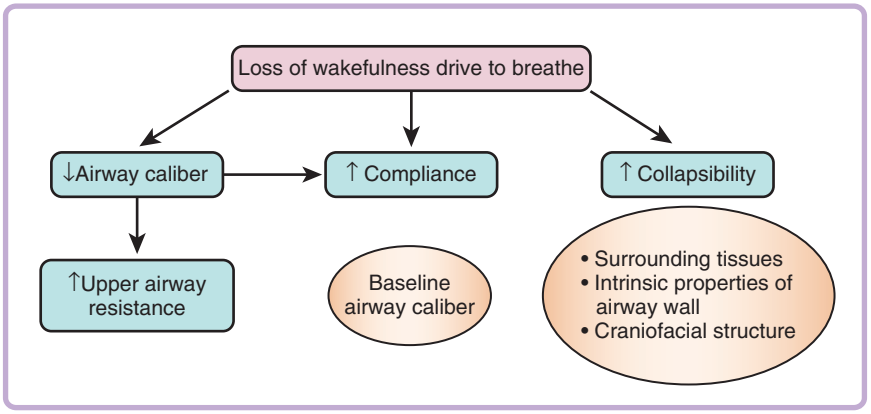


Figure 111-13 Effect of the loss of wakefulness drive to breathe on the upper airway.

increasing passive Pcrit in relation to increasing BMI more than women.[124]

The effect of age on upper airway resistance during sleep is variable across different studies. Browne and colleagues[135] and Thurnheer and associates[117] found no difference in upper airway resistance between young (<40 years) and older (>40 years) subjects. In contrast, in a group of 60 subjects without OSA, Rowley and colleagues found age to be the only independent predictor of upper airway resistance, with increased age associated with increased upper airway resistance. BMI was not a predictor of resistance.[129] More recently, however, increasing age was associated with increased upper airway resistance during sleep in a linear fashion.[136] Overall, it appears that aging is associated with increased upper airway resistance, and thus possibly increased diathesis for pharyngeal narrowing, during sleep. However, because age does not appear to be a predictor of Pcrit,[124] the significance of age-related increased upper airway resistance is unclear.

Hormonal Activity and Upper Airway Activity

There is also evidence that upper airway collapsibility can be influenced by hormonal activity, particularly leptin activity, in humans. For example, Shapiro and colleagues found in obese subjects that, although leptin levels were not associated with passive Pcrit or the severity of sleep apnea (presumably OSA), increased leptin levels were associated with an increased difference between active and passive Pcrit, independent of BMI and neck circumference.[137] Thus leptin appears to be associated with a decreased propensity to upper airway collapse.

In summary, upper airway patency during sleep is compromised by the loss of wakefulness drive of breathing (Figure 111-13). The loss of the wakefulness drive to breathe results in decreased upper airway neuromuscular activity and reflex activity, leading to decreased upper airway caliber and increased in upper airway resistance. The loss of wakefulness drive to breathe also results in increased airway compliance and collapsibility.

CLINICAL PEARL

The upper airway may be compromised and an individual put at an increased risk for OSA because of enlargement of soft tissue structures such as the tonsils, tongue, and lateral pharyngeal walls as a result of disease or obesity. Craniofacial structure, which is determined by genetics, race, and ethnicity, is an important determinant of upper airway patency; clinically significant changes can include micrognathia and retrognathia, overjet, and a high arched palate. Obesity, through a decrease in lung volume (particularly functional residual capacity), and fluid overload can each lead to changes in the intrinsic properties of the upper airway, increasing propensity to collapse. These factors, in association with sleep-related alterations in both neural control of the upper airway and central neural control of breathing, increase the propensity to upper airway obstruction or collapse in sleep in some individuals, leading to the clinical disorder of OSA.

SUMMARY

Upper airway patency is determined by multiple factors that are present during wakefulness and sleep, with sleep generally associated with compromise of these factors, leading to changes in upper airway function that contribute to upper airway obstruction. The four major determinants of upper airway patency are neuromuscular activity, craniofacial structure, tissues surrounding the upper airway, and the intrinsic properties of the airway. These determinants are modified by other factors, including age, obesity, gender, ethnicity, fluid overload, and other medical disorders, such as tonsillar hypertrophy (see Figure 111-6). During sleep, the loss of the wakefulness drive to breathing results in decreased upper airway neuromuscular activity and reflex activity, leading to decreased upper airway caliber and increased upper airway resistance (Figure 111-13). In addition, there is associated increased airway compliance and increased collapsibility, both of which are influenced by factors that determine baseline upper airway caliber, such as surrounding tissues, craniofacial structure, and the intrinsic properties of the upper airway.

This chapter has discussed the important structural aspects of upper airway structure, function, and patency. However, it is important to note that, although the upper airway narrows in humans during sleep, whether any given individual develops sufficient obstruction to develop the clinical disorder of OSA is likely an interplay between both upper airway structure and function and neural control of breathing during sleep.

Selected Readings

Genta PR, Edwards BA, Sands SA, et al. Tube law of the pharyngeal airway in sleeping patients with obstructive sleep apnea. *Sleep* 2015 Oct 5. pii: sp-00004-15. [Epub ahead of print].

Kirkness JP, Schwartz AR, Schneider H, et al. Contribution of male sex, age, and obesity to mechanical instability of the upper airway during sleep. *J Appl Physiol* 2008;**104**(6):1618–24.

Malhotra A, Pillar G, Fogel RB, et al. Genioglossal but not palatal muscle activity relates closely to pharyngeal pressure. *Am J Respir Crit Care Med* 2000;**162**(3 Pt 1):1058–62.

Morrell MJ, Arabi Y, Zahn B, Badr MS. Progressive retropalatal narrowing preceding obstructive apnea. *Am J Respir Crit Care Med* 1998;**158**(6): 1974–81.

Patil SP, Schneider H, Marx JJ, et al. Neuromechanical control of upper airway patency during sleep. *J Appl Physiol* 2007;**102**(2):547–56.

Rowley JA, Sanders CS, Zahn BR, Badr MS. Gender differences in upper airway compliance during NREM sleep: role of neck circumference. *J Appl Physiol* 2002;**92**(6):2535–41.

Saboisky JP, Stashuk DW, Hamilton-Wright A, et al. Neurogenic changes in the upper airway of patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 2012;**185**(3):322–9.

Sankri-Tarbichi AG, Rowley JA, Badr MS. Expiratory pharyngeal narrowing during central hypocapnic hypopnea. *Am J Respir Crit Care Med* 2009; **179**(4):313–19.

Schorr F, Kayamori F, Hirata RP, et al. Different craniofacial characteristics predict upper airway collapsibility in Japanese-Brazilian and caucasian males. *Chest* 2015 Aug 20. doi: 10.1378/chest.15-0638. [Epub ahead of print].

Schwab RJ, Pasirstein M, Kaplan L, et al. Family aggregation of upper airway soft tissue structures in normal subjects and patients with sleep apnea. *Am J Respir Crit Care Med* 2006;**173**(4):453–63.

Schwab RJ, Pasirstein M, Pierson R, et al. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 2003;**168**(5):522–30.

Stanchina ML, Malhotra A, Fogel RB, et al. The influence of lung volume on pharyngeal mechanics, collapsibility, and genioglossus muscle activation during sleep. *Sleep* 2003;**26**(7):851–6.

van Lunteren E, Strohl KP. The muscles of the upper airways. *Clin Chest Med* 1986;**7**(2):171–88.

A complete reference list can be found online at ExpertConsult.com.

Chapter
112

Snoring and Pathologic Upper Airway Resistance Syndromes

Riccardo Stoohs; Avram R. Gold

Chapter Highlights

- Over the past two decades, knowledge of pathologic pharyngeal collapse during sleep has expanded from apnea and hypopnea to include even the mildest, silent inspiratory airflow limitation (IFL) during sleep. Both the clinical researcher and the sleep medicine practitioner of today must be able to recognize the mildest IFL on a polysomnogram and its clinical implications. This chapter discusses the physiologic and clinical features of IFL during sleep.
- Although habitual snoring is very common, the prevalence of isolated snoring (snoring in the absence of apnea and hypopnea, oxygen desaturations, arousals from sleep, and symptoms of obstructive sleep apnea) is unknown. Recent clinical investigation has led to uncertainty about whether such snoring can be considered benign. This chapter presents the issues involved in the diagnosis and management of isolated snoring.
- The paradigm of IFL during sleep leading to recurrent respiratory effort–related arousals is inadequate to explain the varied signs and symptoms of upper airway resistance syndrome (UARS) that have been recognized during the past decade or to distinguish between UARS patients and asymptomatic, healthy individuals whose polysomnograms are remarkably similar. This chapter discusses the evolving paradigm of UARS.

Snoring and upper airway resistance syndrome (UARS) represent obstructed breathing during sleep too mild to cause more than slight sleep fragmentation but with potential pathologic significance. In the past, the chief question to be answered was whether snoring, in the absence of obstructive sleep apnea (OSA), causes hypersomnolence, metabolic disorders, or cardiovascular disease. Currently there is growing evidence that mild inspiratory airflow limitation (IFL) during sleep, even in the absence of audible snoring or increased sleep fragmentation, may have a causative role in a variety of disabling somatic and affective disorders whose available treatment options are of limited benefit. After providing some background on the terms used in this chapter, we will first cover the physiology, recognition, and definition of IFL during sleep, the phenomenon known as increased *upper airway resistance*. We will next present the clinical manifestations of increased upper airway resistance, snoring, and UARS, attempting to organize the growing body of knowledge into a plausible pathophysiology and clinical paradigm of increased upper airway resistance in sleep.

BACKGROUND

Glossary of Terms Central to This Chapter

The terms defined below are described in detail and in context in this chapter:

Inspiratory airflow limitation. IFL describes a state of the upper airway (the pharynx) during sleep in which inspiratory airflow plateaus at a maximal level despite a continued increase in the pressure gradient between the nostrils and the hypopharynx. The failure of inspiratory airflow to increase despite the continued increase in the pressure gradient across the upper airway is caused by fluttering of the upper airway that prevents further increase in airflow. IFL can be divided into two subgroups based on whether or not it is *audible:* (1) *snoring* and (2) *silent IFL.* In this chapter, *snoring* is further divided into two subgroups: (a) *habitual snoring* and (b) *isolated snoring*

Snoring or inspiratory snoring. Audible inspiratory fluttering of the upper airway. It can occur during obstructive hypopnea when hypopnea is associated with a decrease in inspiratory airflow by 30% lasting at least 10 seconds accompanied by either an arousal from sleep or a 3% decrease in oxygen saturation. Alternatively it can occur in the absence of the above criteria for hypopnea, with higher levels of airflow, or with shorter duration or absence of arousals or oxygen desaturation. The presence of inspiratory snoring *always indicates the presence of IFL.* Although expiratory snoring exists (and will be discussed later in this chapter), the term *snoring* used without a modifier in this chapter refers to inspiratory snoring.

Habitual snoring: This term describes an observation (often a complaint) by a bed partner or roommate that a person consistently snores when asleep.

Isolated snoring: Following polysomnography, if an otherwise healthy, asymptomatic habitual snorer does not meet the current third *International Classification of Sleep Disorders*, third edition[1] (ICSD3) criteria for OSA, the patient is described as having *isolated snoring*.

Specific criteria for being "otherwise healthy" in the context of being a habitual snorer are described later in this chapter.

Silent inspiratory airflow limitation. Silent IFL is defined, and characterized by, the same *fluttering* of the upper airway that characterizes snoring; however, the frequency of the fluttering during silent IFL is, by definition, inaudible by humans.

Respiratory effort–related arousal (RERA): RERAs are transient arousals from sleep that follow a period of nonhypopneic IFL (either snoring or silent IFL) and are presumed to be caused by the inspiratory effort required to move air across a fluttering airway. As the name suggests, RERAs are *not* respiratory events per se, as are apneas and hypopneas. Whether an arousal following a period of IFL is in fact caused by the IFL cannot be definitively ascertained during clinical polysomnography; it is a presumption. To label an arousal a RERA, the authors of ICSD3 require 10 seconds of recognizable IFL preceding the arousal. A standard time requirement for IFL preceding an RERA, however, is not a feature of RERAs in the medical literature; a single flow-limited inspiration before arousal is the definition in some research.

Respiratory disturbance index (RDI): ICSD3 has replaced the apnea-hypopnea index (AHI) as a measure of the severity of OSA with the frequency of apneas, hypopneas, and RERAs. In this chapter, this new measure of the severity of OSA will be termed the RDI.

Upper airway resistance syndrome: The UARS does not exist in the ICSD3. It should be thought of as a syndrome coined by Dr. Christian Guilleminault and used by researchers who have broken away from the paradigm that hypersomnolence in patients with sleep-disordered breathing requires the presence of sleep fragmentation by apneas and hypopneas. In this chapter, UARS is defined as the symptom of either hypersomnolence or fatigue together with the presence of IFL during sleep adjudicated by polysomnography (as per the definition of IFL given previously) and an AHI of less than 5 per hour; the latter is the threshold of an OSA diagnosis in the ICSD2. In this chapter, symptoms and signs of UARS will be introduced to differentiate between UARS, OSA, and isolated snoring.

Upper Airway Resistance Versus Pharyngeal Collapse

Two terms used to describe the behavior of the upper airway (or pharynx) during sleep among snorers and patients with UARS are increased upper airway *resistance* and upper airway *collapse*. Many sleep researchers consider IFL during sleep to result from narrowing of the pharyngeal airway and increased resistance caused by the relaxation of pharyngeal dilator muscles, together with subatmospheric upper airway pressures during inspiration. As they measure increasingly negative esophageal or supraglottic pressures during inspiratory snoring, they think of upper airway *resistance* increasing. From this reasoning the clinical term *upper airway resistance syndrome* (UARS) was derived (as discussed later).

In contrast to this intuitive model of increasing upper airway resistance during sleep is the experimentally validated Starling resistor model of IFL[2] (see Chapter 17). The Starling resistor model postulates that the pharyngeal airway during sleep is a collapsible tube that will in fact collapse whenever the pressure within falls below a critical level, the pharyngeal "critical pressure" (Pcrit). It has been shown experimentally that as the severity of sleep-disordered breathing increases from isolated snoring to severe OSA, the pharyngeal Pcrit progressively increases from negative (subatmospheric) levels to positive levels.[3,4] Collapse of the pharynx, however, is not synonymous with apnea. When the pharynx collapses during sleep, one might experience either persistent apnea (no inspiratory airflow) or IFL (inspiratory airflow that has reached its maximum). When the pressure at the *upstream* end of the pharynx (the nares during inspiration) falls below Pcrit, the pharynx collapses, with resulting persistent apnea. When the pressure at the nares is above Pcrit, but the pressure at the downstream end of the pharynx (supraglottic pressure during inspiration) falls below Pcrit, as in a snorer, the pharynx also collapses. Because pharyngeal collapse leads to cessation of inspiratory airflow, pharyngeal pressure immediately equilibrates with nasal pressure opening the airway, with resumption of inspiratory airflow. The result is cyclical collapse and opening (fluttering) of the pharyngeal airway *limiting* inspiratory airflow to a fixed, maximal level (with the driving pressure fixed at nasal pressure minus Pcrit, no matter how low supraglottic pressure descends). Therefore, according to the Starling resistor model, the upper airway does not experience increased *resistance* during sleep, but a *fixed driving pressure* that limits airflow to a maximal level.

The language subsuming upper airway resistance and upper airway collapse therefore is derived from two different models of IFL. In this chapter, we will allude to *pharyngeal collapse* in the section that follows describing the polysomnographic appearance of IFL but use the term *upper airway resistance* for the remainder of this chapter, which does not require modeling of IFL.

Upper Airway Resistance Syndrome

As introduced in the glossary, we use the term UARS in this chapter, and it is found as a diagnosis in the current medical literature. However, the ICSD3[4] does not include UARS in its classification of sleep related breathing disorders but rather incorporates the polysomnographic manifestations of UARS into OSA. A brief discussion of the history of UARS will help the reader understand this dichotomy.

UARS came to public attention following the publication of a case series in 1993 by Dr. Christian Guilleminault and associates.[5] From among 48 patients with a diagnosis of idiopathic hypersomnolence, they selected 15 with the following characteristics:

- Intermittent or continuous snoring during sleep at home
- An AHI below the threshold for OSA by in-laboratory polysomnography (5/hour at Stanford University)
- More than 10 arousals per hour of sleep (a threshold they chose, attempting to limit their selected patients to those with an *increased* frequency of arousals)
- The presence of *upper airway resistive events* associated with arousals. These events were associated with reductions in airflow below the threshold to qualify as hypopnea and were identified using a pneumotachograph measurement of airflow and esophageal manometry to quantify effort.

Treatment of these 15 patients with nasal continuous positive airway pressure (CPAP) eliminated their resistive events and their associated arousals and relieved the patients' hypersomnolence (measured objectively by multiple sleep

latency testing). Because the patients did not meet diagnostic criteria for OSA (their resistive events were not hypopneas), the investigators designated a new syndrome: UARS. They hypothesized that UARS is a disorder of hypersomnolence related to sleep fragmentation by upper airway resistive events too mild to meet the diagnostic criteria of hypopnea. They further hypothesized that UARS patients have increased sensitivity to the respiratory effort related to these resistive events, giving rise to repetitive arousals (compared with OSA patients who typically arouse in response to higher degrees of obstruction: i.e., apneas and hypopneas). The hypothesis that UARS patients exhibit increased sensitivity to respiratory effort during sleep led to their arousals being termed RERAs.

Almost from the start, the establishment of a new syndrome of sleep-disordered breathing based on sleep fragmentation by RERAs created controversy.[6,7] Many believed that RERAs and hypopneas were essentially the same phenomenon and that both OSA and UARS described sleep fragmentation caused by upper airway resistive events giving rise to hypersomnolence. To eliminate the need for an additional syndrome of sleep-disordered breathing based on sleep fragmentation by RERAs, the authors of ICSD3 incorporated RERAs into the diagnostic criteria for OSA, creating diagnostic thresholds for OSA based on the combined frequency of obstructive events: apneas, hypopneas, and RERAs—the RDI. Therefore, by the clinical criteria of ICSD3, UARS has been "absorbed" into OSA.

The significance of UARS, however, does not end with its *absorption* into OSA. Since first being described by Guilleminault and associates as a syndrome of sleep fragmentation leading to hypersomnolence,[5] the paradigm of UARS has evolved. Investigators have observed that the sleep of UARS patients is characterized not only by the presence of RERAs but also by electroencephalographic differences and differences in sleep architecture that distinguish it from the sleep of healthy individuals. These *qualitative* differences in the sleep of UARS patients resolve with treatments that eliminate IFL during sleep and the hypersomnolence of UARS patients. Therefore, although the ICSD3 clinical criteria for OSA will result in many patients being treated for OSA who previously were diagnosed with UARS, it is not established that these former UARS patients are hypersomnolent because of sleep fragmentation by RERAs. Consequently, UARS continues on as a syndrome being studied by researchers examining alternatives to the OSA pathophysiologic paradigm of sleep fragmentation by apneas, hypopneas, and RERAs.

In our discussion of UARS later in this chapter, we will describe more fully the clinical presentation of UARS and the findings of investigators that have led them away from a sleep fragmentation paradigm of this disorder. We will also consider the alternative pathophysiologic paradigms of UARS that continue to evolve.

Inspiratory Airflow Limitation

Classically, the term *snoring*, the audible fluttering of the pharynx during inspiration, has been used to describe IFL during sleep. The term *snoring*, however, implies that pharyngeal fluttering is present only when it can be heard by a listener (the word *snore* itself resembles the sound of snoring). Hearing, however, is an insensitive means of detecting inspiratory fluttering of the pharyngeal airway during sleep. Because using the term *snoring* may lead one to believe that IFL is only present when audible, which is not the case, we have chosen to describe the characteristic inspiratory airflow through a fluttering upper airway during sleep as a state of *inspiratory airflow limitation* (as defined in the glossary above). The term IFL was first used by Schwartz and associates in their study of pharyngeal collapsibility during sleep in healthy humans[8] and was derived from the parallel term *expiratory airflow limitation*, used to describe expiratory airflow from the lungs of patients with asthma, chronic obstructive bronchitis, and emphysema whose bronchi flutter on expiration, limiting airflow.[9]

Recognizing Inspiratory Airflow Limitation with Physiologic Testing

With the incorporation of RERAs into the diagnostic criteria for OSA, recognizing the presence of IFL preceding an arousal and differentiating its appearance from that of non-flow-limited breathing during polysomnography is an important skill for practitioners of sleep medicine to develop. Similarly, for clinicians and polysomnographic technologists involved in titrating nasal CPAP to treat OSA, recognizing IFL during sleep is an important aspect of polysomnography to understand. In this section we begin by illustrating the appearance of IFL using airflow and supraglottic pressure tracings. We then demonstrate how IFL is recognized using the airflow and effort tracings available during clinical polysomnography.

Figure 112-1 illustrates five breaths during continuous non–rapid eye movement stage (NREM) 2 (N2) sleep at atmospheric pressure, all characterized by IFL. The individual being monitored is a 24 year-old woman with a body mass index of 19.9 kg/m^2 who does not snore audibly and has an AHI of 0.3/hour. Because Figure 112-1 has both an airflow tracing and a supraglottic pressure tracing, it precisely demonstrates the presence of IFL. Specifically, it shows that inspiratory airflow is limited to a maximal level (intersected by the vertical lines) despite the observation that the pressure gradient across the pharyngeal airway (atmospheric pressure minus supraglottic pressure) continues to increase (atmospheric pressure remains the same while supraglottic pressure continues to decrease beyond the vertical line). This defines IFL.

Figure 112-2 demonstrates both IFL and non-flow-limited inspiration in the same individual diagnosed with UARS during nasal CPAP titration. Although the left panel of the figure clearly demonstrates IFL at atmospheric pressure similar to that observed in Figure 112-1, the right panel, recorded at the therapeutic nasal CPAP level of 4 cm H_2O, presents airflow and supraglottic pressure tracings that parallel each other through four inspiratory cycles. The parallel tracings demonstrate that inspiratory airflow is continuously proportional to the driving pressure, 4 cm H_2O minus supraglottic pressure, and thus, according to the above definition of IFL, is not flow limited.

Figures 112-1 and 112-2 illustrate that, when one is provided with both an airflow signal and a supraglottic pressure signal, recognizing IFL is not difficult. It is emphasized that IFL is not defined by any specific decrease in inspiratory airflow (e.g., a 30% or 50% decrease in airflow) relative to *non*-flow-limited inspiration. Rather, IFL is defined by a specific relationship of airflow to driving pressure (nasal pressure minus supraglottic pressure). IFL can be more difficult to recognize in the absence of a supraglottic pressure signal

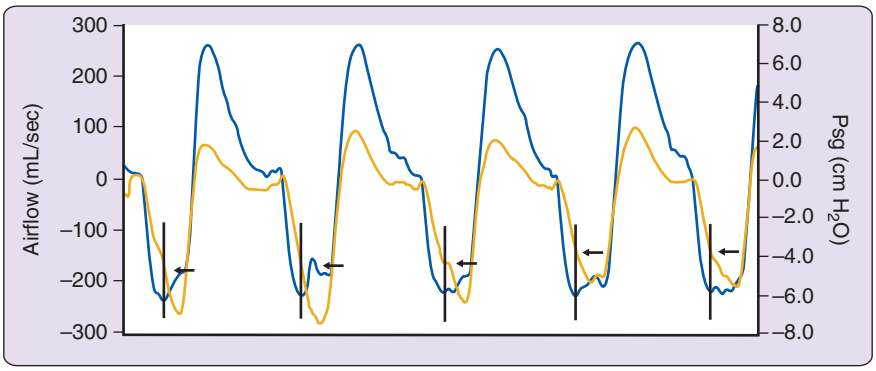


Figure 112-1 This figure illustrates inspiratory airflow limitation (IFL) in a sleeping research participant wearing a nasal mask attached to a pneumotachograph measuring airflow, with a pressure catheter placed through her nose to just above her vocal cords measuring supraglottic pressure (Psg). Airflow is the *blue tracing* with the units indicated on the left axis (inspiration is downgoing). Effort, represented by Psg, is the *yellow tracing* with the units indicated on the right axis. For each inspiration, a plateau in airflow during early inspiration is intersected by a *vertical line*. Beyond the line, there is no further increase in inspiratory airflow despite the continued decrease in Psg and a continued increase in the inspiratory pressure gradient Patm—Psg. Indeed, not only does the inspiratory airflow not increase, but also in the first four breaths it appears to decrease, a phenomenon known as *negative effort dependence* of airflow. IFL occurs when Psg decreases below this participant's pharyngeal Pcrit. The *horizontal arrows* mark the Psg at the onset of maximal flow (intersected by the *vertical line*) and suggest that this participant's pharyngeal Pcrit is approximately −4 cm H_2O, a common value for primary snorers or individuals who have upper airway resistance syndrome.[3]

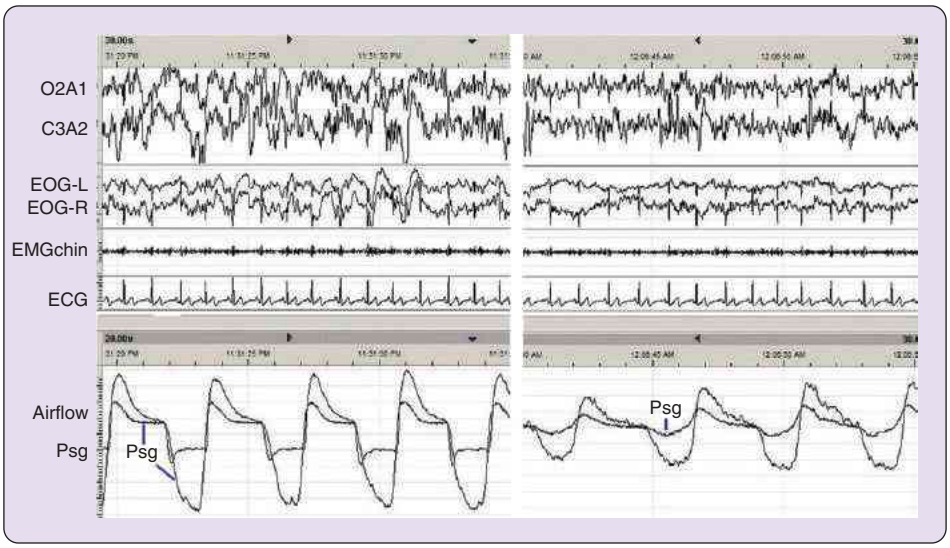


Figure 112-2 This figure demonstrates both inspiratory airflow limitation (IFL) and *non*-flow-limited inspiration in the same individual during nasal continuous positive airway pressure (CPAP) titration. The two polysomnographic tracings are obtained in stage N2 sleep, ½ hour apart. Below the sleep monitoring channels recording electroencephalograms (O2A1, C3A2), electrooculograms (EOG-L & R), superficial electromyograms of the chin (EMGchin), and electrocardiogram (ECG) are recordings of airflow (a pneumotachograph tracing) and supraglottic pressure (Psg). The *left panel* demonstrates four breaths at atmospheric pressure, whereas the *right panel* demonstrates four breaths with nasal CPAP at 4 cm H_2O. In each panel, airflow (*black tracing*) and Psg (*blue tracing*) are superimposed. The *left panel* demonstrates the plateau of inspiratory airflow (downgoing) at a maximal level occurring as Psg continues to decrease, which defines IFL. In the *right panel*, because pharyngeal pressure and Psg do not fall much below 4 cm H_2O (the CPAP applied to the nasal mask), Psg always remains above pharyngeal Pcrit, and the airflow and pressure tracings parallel each other (airflow is always determined by the pressure gradient: 4 minus Psg).

because one is then missing driving pressure; indeed, the presence of IFL can only be *assumed* in the absence of a supraglottic pressure tracing). To enable clinicians to recognize IFL during clinical polysomnography without the recording of supraglottic pressure, researchers have investigated the possibility of identifying IFL from the airflow signal alone.

In 1998, two studies evaluated the utility of a plateau of inspiratory airflow measured as a nasal pressure signal (pressure transducer airflow [PTAF]) to identify IFL.[10,11] Hosselet and associates[10] studied more than 47,000 breaths during polysomnography in 10 symptomatic OSA patients and 4 asymptomatic individuals without OSA, classifying the shape

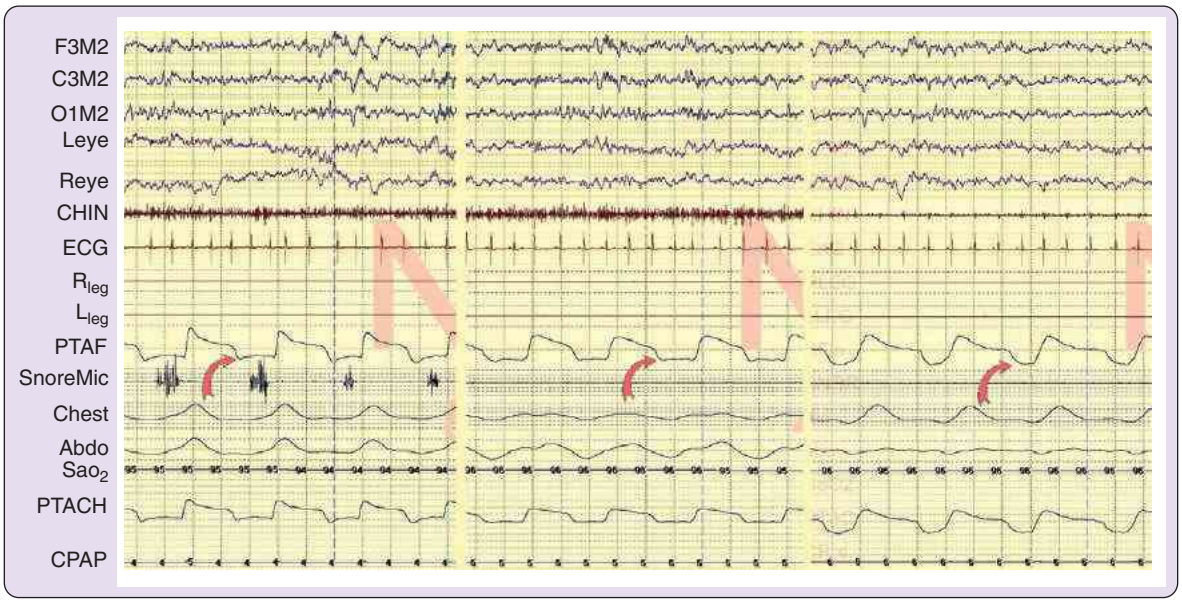


Figure 112-3 This figure's three panels, from left to right, represent three 12-second intervals at nasal continuous positive airway pressure (CPAP) levels of 4 cm H_2O, 5 cm H_2O, and 6 cm H_2O. Below the sleep monitoring channels recording electroencephalograms (F3M2, C3M2, O1M2), electrooculograms (L & Reye), superficial electromyograms of the chin (CHIN), and right and left tibialis anterior (R & L_{leg}) and ECG are several channels recording respiratory parameters. The respiratory channels include a pressure transducer-generated airflow signal (PTAF), a microphone placed on the neck to record snoring (SnoreMic), impedance plethysmography of the chest and abdomen (Chest and Abdo; movement), oxyhemoglobin saturation (SaO_2), a pneumotachograph airflow signal (PTACH), and a pressure transducer-recorded CPAP level (CPAP). At 4 cm H_2O, the *left panel*, the patient's inspirations all demonstrate IFL with audible snoring. Inspiratory airflow (downgoing) is seen to increase rapidly and then to plateau with a prolonged time spent at maximal inspiratory airflow (highlighted by the *arrow*). At 5 cm H_2O, IFL persists without audible snoring. Inspiratory airflow, again, increases rapidly and then demonstrates a prolonged plateau at maximal flow (highlighted by the *arrow*). At 6 cm H_2O, the airflow tracing no longer demonstrates inspiratory airflow limitation. Inspiratory airflow increases to its maximum more gradually and then immediately decreases, spending only a short time at maximal airflow. Expiratory time is prolonged relative to flow-limited conditions (the two *left panels*), and inspiration is a smaller percentage of the respiratory cycle.

of each breath's PTAF signal while also recording a pneumotachograph and supraglottic pressure signal (a gold standard assessment of airflow vs. driving pressure). They used a computer algorithm to classify each PTAF inspiration as non-flow-limited (sinusoidal in shape and resembling the airflow signal in the right panel of Figure 112-2), flow-limited (having a clear plateau and resembling the airflow signal in Figure 112-1 and the left panel of Figure 112-2), or intermediate (not sinusoidal but not fulfilling their program's criteria for an inspiratory plateau). The PTAF signal clearly separated their asymptomatic controls without OSA from their OSA patients, with the former having fewer flow-limited events. In a similar study of seven habitual snorers, Clark and associates[11] found that an inspiratory airflow plateau determined by PTAF identified flow-limited inspirations with a sensitivity and specificity of approximately 80%. Thus PTAF evidence of a clear inspiratory airflow plateau is a reasonably reliable method for identifying IFL during clinical polysomnography.

The ability to recognize IFL during diagnostic polysomnography, whether in-laboratory or during out-of-center sleep testing (OCST), can also be aided by the findings of Schneider and associates: During IFL, the ratio of the inspiratory time to the time of the entire respiratory cycle (i.e., the "duty cycle") is prolonged.[12] This distinction is illustrated in Figure 112-2, where four flow-limited inspirations (left panel) take up a larger portion of the respiratory cycle time than the non-flow-limited inspirations (right panel), where expiratory time is more prolonged. During IFL, the inspiratory airflow increases rapidly and remains near maximum throughout most of inspiration (left panel), maximizing the tidal volume under the flow-limited conditions. During the *non*-flow-limited breaths (right panel), the increase in inspiratory airflow is more gradual, and airflow remains near maximal for a shorter portion of inspiration.

Figure 112-3 represents both IFL and *non*-flow-limited breathing in a single patient with UARS undergoing nasal CPAP titration. In the absence of a supraglottic pressure signal, IFL can be recognized by the change in the airflow tracing between the two left panels recorded at CPAP levels of 4 and 5 cm H_2O demonstrating IFL, and the right panel recorded at a CPAP of 6 cm H_2O illustrating non-flow-limited airflow at therapeutic CPAP. At 4 and 5 cm H_2O, the flow-limited inspiratory airflow tracing is characterized by a rapid increase in airflow to a maximum followed by a prolonged plateau at maximal flow. At 6 cm H_2O, the non-flow-limited inspiratory airflow increases more gradually without a subsequent plateau, but a rapid decrease of airflow and a shorter ratio of inspiratory time to respiratory cycle time (exhalation is prolonged relative to flow-limited conditions). Thus, in the absence of a supraglottic pressure signal, both the shape and the relative duration of the inspiratory airflow tracing provide evidence for the presence of IFL.

Figure 112-4, a 30-second epoch of sleep from a patient with UARS, provides examples of overt IFL, more subtle IFL,

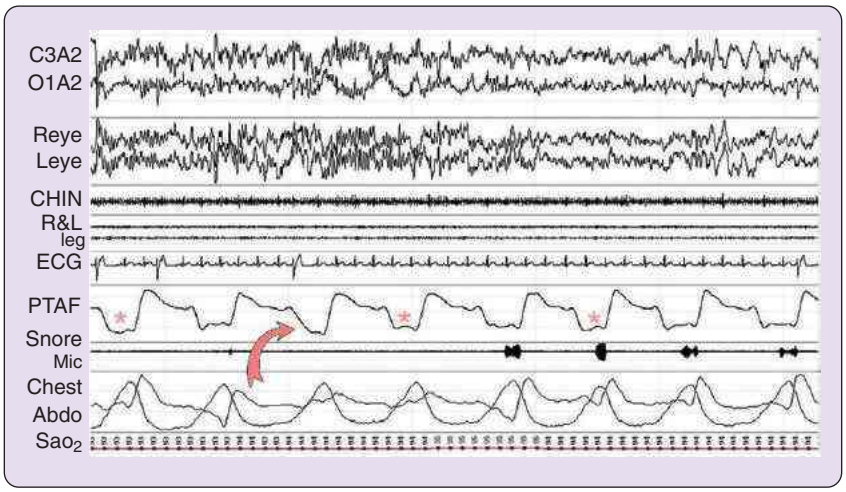


Figure 112-4 This figure demonstrates a 30-second epoch of NREM stage 2 (N2) sleep containing eight consecutive breaths representing both overt and subtle (*asterisks*) inspiratory airflow limitation and a *non*-flow-limited breath (*arrow*). The sleep parameters recorded include electroencephalograms (EEGs; C3A2, O1A2), electrooculograms (EOGs; Reye & Leye), superficial electromyogram of the chin (CHIN), superficial electromyograms of the right and left tibialis anterior (R & L_{leg}), and ECG. The respiratory parameters recorded are labeled similarly to those in Figure 112-3. Refer to the text for a complete characterization of the breathing.

and *non*-flow-limited breathing. Although several breaths in Figure 112-5 demonstrate the rapid increase in inspiratory airflow and long inspiratory airflow plateau of IFL, several (marked by an asterisk) demonstrate a less prolonged plateau of inspiratory airflow. The presence in these breaths, however, of a rapid increase in inspiratory airflow followed by a short plateau, as well as the accompanying snoring (recorded by microphone) for one of the breaths, can be used to identify all of these as examples of subtle IFL compared with the one *non*-flow-limited inspiration (marked by an arrow). Viewed from the perspective of respiratory cycle time, one can also appreciate that the *non*-flow-limited breath is preceded by the longest expiration and the ratio of inspiratory time to respiratory cycle time for the breath is lower than for the flow-limited breaths in the figure.

To summarize, in the absence of a supraglottic pressure catheter, IFL can be recognized, using a nasal pressure signal, as a plateau in inspiratory airflow and a prolongation of inspiratory time relative to total respiratory cycle time. Inspiratory airflow can be observed to rise rapidly to a maximum and to remain there for most of inspiration. From the figures provided, it can also be inferred that snoring, the audible manifestation of the fluttering pharyngeal airway characterizing IFL, is not as sensitive an indicator of IFL as the combined airflow and driving pressure criteria (Figures 112-1 and 112-2) or the airflow tracing alone. Figure 112-1, which in fact is a tracing of a lean female with no history of snoring, demonstrates definitive evidence of IFL determined by her airflow and supraglottic pressure recordings. In Figure 112-3, nasal CPAP of 5 cm H_2O resolves the patient's snoring before the airflow tracing demonstrates resolution of IFL at 6 cm H_2O. Figure 112-4 demonstrates five breaths clearly characterized by the airflow plateau associated with IFL, only four of which demonstrate snoring. For this reason, the presence of snoring should not be relied on to determine whether a patient with sleepiness or fatigue has sleep-disordered breathing. Even in the absence of apneas and hypopneas, silent IFL (defined in the glossary) associated with arousals (RERAs) may be prevalent enough to establish a diagnosis of OSA when using ICSD3 criteria, or UARS, using the criteria of sleepiness or fatigue in the presence of IFL as presented in the glossary. Similarly, the absence of snoring should not be relied on to determine whether a nasal CPAP level is therapeutic (i.e., has eliminated IFL). Rather, the polysomnographer, sleep medicine physician, and polysomnographic technologist should differentiate IFL during sleep from non-flow-limited inspiration using an airflow signal generated by either a pneumotachograph or a PTAF signal and determine therapeutic nasal CPAP as the pressure that eliminates IFL (as demonstrated in Figures 112-2 and 112-3). A CPAP level that eliminates IFL will, of necessity, eliminate all apneas, hypopneas, and RERAs.

PATHOLOGIC UPPER AIRWAY RESISTANCE SYNDROMES: CLINICAL ASPECTS

Snoring

Habitual snoring, as described early in the glossary, can be observed in patients with OSA complaining of daytime sleepiness, fatigue, and insomnia. Habitual snoring may also occur in the absence of symptoms and signs of OSA and without an RDI (a frequency of apneas, hypopneas, and RERAs) adequate to establish a diagnosis of OSA in the absence of symptoms—that is, an RDI of 15 per hour. In the latter instance, according to the ICSD3, it is regarded as "isolated snoring," listed in the category of sleep-related breathing disorders. This isolated snoring was previously referred to as habitual, simple, or primary snoring in the category of other parasomnias in the ICSD2.

As already noted, snoring is a sleep-related sound caused by vibration of soft tissue in the upper airway under conditions of IFL. In most individuals with isolated snoring, the snoring is limited to inspiration, although early expiratory snoring or snoring throughout expiration can occur.[13] Whether occurring during inspiration or expiration, snoring is generated by high-frequency opening and closing (fluttering) of

upper airway structures, including the tongue base and soft palate, aided by the adhesive properties of mucosal secretions. Acoustic studies have shown that the major frequency content of snoring is below 2000 Hz, with peak power below 500 Hz[14] (thus in the frequency range able to be heard by humans). Snorers experience increased total pulmonary resistance during sleep related to reduced upper airway muscle tone causing IFL and leading to increased inspiratory effort.[15]

There is, understandably, considerable variation in the prevalence figures reported for habitual snoring. Studies differ in how the study population is selected, with some determining a prevalence of isolated snoring and others including habitual snorers with OSA. Studies also differ in how snoring is assessed. S*ubjective* assessment has been conducted using the report of bed partners, but clinical experience indicates that a few snorers hear their own snoring, especially during sleep-wake transitions. *Objective* measurement relies on the use of calibrated monitoring devices either during polysomnography or in a setting outside a sleep disorders center. Patients also document their own snoring using recording applications for smart devices such as cellular phones. In addition to differences in methodology regarding diagnoses and snoring assessment, differences in gender and obesity distribution between studies may affect snoring prevalence significantly. Both gender and obesity can affect upper airway resistance (alternatively, collapsibility, assessed as the pharyngeal Pcrit) either by structural changes or neuromuscular mechanisms. Therefore the variability of study design is one reason for the varied prevalence figures for habitual snoring seen in the literature.

Further complicating a determination of the prevalence of snoring is the observation that, within individual patients, the severity of subjective snoring reported by the bed partner does not correspond with either objectively assessed snoring or the subjective assessment of the sleep technician monitoring the patient.[16] Data from Somnolab Sleep Disorders Center in Dortmund, Germany show that 28% of individuals reporting habitual snoring at home fail to present significant snoring during in-laboratory polysomnographic recordings (unpublished data). This may, in part, be due to considerable night-to-night variability of snoring intensity. Interventional studies have demonstrated that time spent snoring and snoring volume within one individual can vary from night to night depending on factors such as sleeping position, medications, alcohol intake, and cumulative or acute sleep debt. Alternatively, the discrepancy observed between a bed partner's report of snoring severity and that observed during in-laboratory polysomnography could be due to allergens in the home environment altering upper airway pressure-flow relationships. Also, discrepancies between snoring reported by a bed partner and that measured during polysomnography can be observed in patients who have had a recent change in bed partner, suggesting either differences among bed partners in sensitivity to the noise or in willingness to complain.[17]

As discussed earlier, prevalence data for snoring often come from study samples containing both individuals with isolated snoring and those with OSA. The 2011 Centers for Disease Control and Prevention report on unhealthy sleep behaviors, one such study, reports a snoring prevalence of 48% based on a telephone survey.[18] The report does not indicate how many of the snorers in this survey complained of hypersomnolence or other symptoms of OSA. Furthermore, the report does not specify the snoring severity of those individuals labeled as snorers: intermittent versus habitual. Based on data from the Sleep Heart Health Study, a sample of 5615 community-dwelling adults between the ages of 40 and 98 years, 13% of the participants had an AHI of less than 5/hour and reported *habitual* snoring (3 to 7 nights/week).[19] These data estimate a prevalence of habitual snoring without OSA of less than 15% in a community sample. Symptom data, however, are not provided and so one cannot determine a prevalence of isolated snoring. Of note, 29% of the 5615 men and women did not know whether they snored (perhaps because of absence of bed partner). Although the previous two examples illustrate the difficulty investigators have in determining the prevalence of habitual and isolated snoring, it remains clear that snoring is a common phenomenon that frequently prompts a referral for a sleep evaluation to establish a diagnosis of OSA.

When a habitual snorer presents for a sleep evaluation, polysomnography is warranted if witnessed apnea, hypersomnolence, fatigue, insomnia, somatic syndromes typically described among UARS patients (discussed in the next section), or comorbidities such as metabolic syndrome, cardiac dysrhythmia, or atrial fibrillation are present. In this case, polysomnography may lead to treatment of OSA when the ICSD3 diagnostic criteria for OSA are met. Habitual snoring in the absence of witnessed apnea, symptoms or syndromes, or comorbidities (after appropriate screening for comorbidities) does not automatically warrant a polysomnogram. Habitual snoring is a common occurrence among middle-aged, overweight men, and polysomnographic evaluations of *all* habitual snorers carries a very high cost-to-benefit ratio. A more practical approach would be to monitor asymptomatic, healthy habitual snorers over time for the development of signs and symptoms that would support obtaining polysomnography. Alternatively, OCST can be used to rule out moderate to severe OSA in asymptomatic, healthy habitual snorers in need of reassurance.

The rationale for not obtaining polysomnography in asymptomatic, healthy habitual snorers extends beyond the issue of costs, to a consideration of benefit. Specifically, even if such an individual fulfills ICSD3 criteria for OSA, the question is whether such an asymptomatic, healthy individual is in fact in need of treatment. To the contrary, cross-sectional polysomnographic data from an investigation by Pavlova and associates[20] of 163 asymptomatic, nonobese individuals screened for the absence of metabolic syndrome and cardiovascular disease (25% reporting "some" snoring) demonstrate that many such individuals have RDIs above 15/hour, fulfilling ICSD3 criteria for OSA. Indeed, the mean RDI for individuals older than 65 years was 22 per hour in Pavlova's study.[20] Similar data exist in three studies comparing inspiratory airflow dynamics during sleep between patients with somatic syndromes[21,22] and UARS[23] with those of rigorously screened healthy controls. For the three studies, 4 (11%) of 35 healthy controls (14 men and 21 women) met the ICSD3 threshold for OSA that would justify their treatment without symptoms or comorbidities (RDI ≥15/hour). Another 4 of the healthy controls had values of RDI between 10/hour and 15/hour, approximating the threshold for treatment. Thus, in the absence of data demonstrating a health risk from habitual snoring alone, a prudent approach to polysomnographic evaluation can be justified.

Before leaving the subject of whether to treat asymptomatic, habitual snorers without comorbidities, a word of caution is appropriate. In 110 overweight volunteers with snoring and mild OSA (27% smokers, 27% with hypertension, and 69% with hyperlipidemia), Lee and associates[24] demonstrated that the amount of time spent snoring was correlated with the extent of asymptomatic carotid artery stenosis independent of AHI and histories of other comorbidities. These findings suggest that in individuals predisposed to atherosclerosis (by smoking, hypertension, or hyperlipidemia), habitual snoring may be an *additional* risk factor for developing carotid artery atherosclerosis. On the other hand, data from a recently published study with a 17-year follow-up of 380 community-dwelling adults failed to document a significant relationship between objectively measured nocturnal time spent snoring and all-cause mortality from cardiovascular disease.[25] In the absence of certainty about the effects of habitual snoring, one should evaluate (beginning with noninvasive methods) an asymptomatic, habitual snorer without metabolic syndrome or atrial fibrillation for evidence of atherosclerosis before deciding that the patient is not in need of treatment and can be followed over time.

Asymptomatic, healthy individuals seeking treatment for habitual snoring or isolated snoring (following polysomnography because of reports of witnessed, or patient-perceived, apnea) will usually do so because they are concerned about the disruption of their bed partner's sleep. Any treatment that will lower the pharyngeal Pcrit, reducing the occurrence of IFL, will also have a beneficial effect on audible snoring. A wide variety of over-the-counter remedies are available for snoring, but they are of limited efficacy. A report of the American Academy of Sleep Medicine Clinical Practice Review Committee published in 2003 summarizes the absent or limited benefits of products such as nasal dilators, lubricants, oral dietary supplements, and magnetic pillows and mattresses.[26] In contrast to these ineffective treatments, any effective treatment used for OSA will be effective for asymptomatic snoring. Among these treatments, few isolated snorers choose nasal CPAP, considering it a burden to use and to maintain.

Successful or partially successful treatment of isolated snoring has been reported using lifestyle modifications. A lifestyle modification like weight reduction (by diet or bariatric surgery) can be an effective treatment for snoring because it can substantially lower pharyngeal Pcrit.[27] Because there are no "dose-response" data concerning the effect of weight loss on snoring intensity (loudness), the weight loss target should be based on factors like the pretreatment body mass index and the weight loss needed to obtain other expected health benefits. Although weight loss can effectively decrease the intensity of snoring, long-term maintenance of reduced weight is often unsatisfactory. For this reason, weight loss in combination with increased physical activity may be a more desirable approach. There is an independent, beneficial effect of physical activity on self-reported snoring in obese women.[28] Another lifestyle alteration that can decrease the intensity of snoring is avoiding alcohol consumption before going to bed. In a small group of otherwise asymptomatic snorers, Riemann and colleagues demonstrated that presleep alcohol ingestion increased the objectively measured incidence and loudness of snoring in a dose-dependent manner.[29] Other lifestyle modifications that can reduce snoring include avoiding sleep deprivation and the use of sedative-hypnotic medications.

Oral mandibular advancement appliances have been used successfully for the treatment of mild to moderate OSA and asymptomatic snoring in patients with a healthy dentition. Good results can be achieved with 50% to 75% of maximal voluntary protrusion. For patients with an insufficient number of healthy teeth, a tongue-retaining device may be a good alternative. Patients should be advised that snoring may not be completely abolished, but significant reductions in the time spent snoring and snoring intensity can be obtained.

Surgery can be performed for isolated snoring to decrease its occurrence and intensity. Surgical targets include the nasal turbinates and septum, the nasopharynx, oropharynx, tongue base, and hypopharynx. Sleep nasendoscopy with a flexible endoscope is increasingly used to perform a preoperative assessment of possible surgical targets. For this procedure, anesthesia is used to simulate sleep. At this time the data regarding the value of nasendoscopy before surgical treatment of snoring are indeterminate. The surgical method depends on the surgeon's preference and the availability of equipment, but procedures are performed using a scalpel, radiofrequency ablation, and YAG laser. Studies assessing the efficacy of these procedures have typically shown good immediate and short-term results. However, many of these studies have relied only on subjective assessments of snoring. A study on the subjective versus the objective improvement of snoring following palatal surgery published in 1994 did not find any objective improvement in snoring despite a subjective improvement in more than 75% of the participants.[30] In a more recent study, palatal surgery for isolated snoring improved subjective (questionnaire) and objective (sound analysis of 100 supine snorers before and after surgery) evaluations of snoring. However, the objective improvement was short-lived and correlated poorly with the subjective improvement on an individual basis.[31] A recent, long-term study evaluating patients treated with palatal surgery between 1985 and 1991 found a substantial rebound of snoring even in the absence of weight gain. In addition, 38% of the patients continued to experience surgical side effects (swallowing dysfunction, altered voice, and pain) that left them dissatisfied with the decision to have palatal surgery.[32]

Surgery intended to relieve nasal obstruction alone does not produce a significant improvement of objectively assessed snoring intensity and snoring time, nor does it decrease the AHI, despite improvement in nasal resistance.[33]

The consequences of leaving isolated snoring untreated relate specifically to the concern: will untreated isolated snoring progress to OSA over time? According to a study that followed individuals with isolated snoring over 5 years with polysomnography, isolated snoring does not progress to OSA over 5 years in the absence of a significant change in body weight.[34] Thus, to date, there is no evidence that isolated snoring progresses to OSA in the intermediate term.

In summary, isolated snoring is a diagnosis of exclusion reserved for habitual snorers who are otherwise asymptomatic without metabolic syndrome and cardiovascular disease and who do not meet polysomnographic or OCST criteria for OSA. The potential for adverse long-term cardiovascular outcomes in isolated snoring remains uncertain at this time. Treatment of isolated snoring is currently limited to attempting to improve the sleep quality of the bed partner. Available treatments include lifestyle modifications, oral appliances, and soft tissue surgery. Most available treatment options lead to short-term success but fail in the long-term.

Upper Airway Resistance Syndrome

UARS is defined as the symptom of either hypersomnolence or fatigue together with the presence of IFL during sleep by in-laboratory polysomnography and an AHI of less than 5/hour (see the glossary earlier in this chapter). As a movement away from the paradigm that hypersomnolence among patients with sleep-disordered breathing requires the presence of sleep fragmentation by apneas and hypopneas, UARS was originally accompanied by a new paradigm that sleep fragmentation by RERAs can also lead to hypersomnolence in individuals with milder resistive events[5] (discussed previously under Background). In line with this new paradigm, ICSD3 absorbs UARS into OSA by including RERAs into the severity assessment of sleep fragmentation in OSA. ICSD3 criteria for OSA now classify any patient fulfilling the previously noted UARS definition with an RDI above 5/hour as having OSA. Clearly a portion of UARS has been absorbed into OSA by the clinical criteria of ICSD3. However, there are still patients meeting the definition of UARS elaborated in the chapter with an RDI of less than 5/hour who are not included within the ICSD3 definition of OSA and are not considered, clinically, to have sleep-disordered breathing. Nevertheless, to investigators of UARS and to clinicians attempting to treat the hypersomnolence of a patient without a clear diagnosis because of too few RERAs, the recognition that sleep-disordered breathing may, in fact, exist outside the limits of ICSD3 is important and worthy of consideration. In this section, we discuss the varied clinical presentation of UARS, its polysomnographic appearance, and its evolving paradigm. To facilitate this discussion, when we refer to OSA, we will use the ICSD2 definition of OSA—an AHI of at least 5/hour–to match the definition used in the research to be presented.

Anthropometric Features and Risk Factors

Compared with patients with OSA, UARS patients are younger, leaner, and more frequently female. Published studies of UARS patients as defined by the previous criteria have established a mean age of 40 years with the average body mass index between 23 and 30 kg/m^2 (normal weight or overweight; less often, obese) and approximately 50% female.[35-37] Although craniofacial abnormalities such as a narrow, elongated face characterized by a high arched palate, reduced upper and lower intermolar distances, and a narrow anterior nasal aperture (adenoid facies) have been reported in patients with UARS, these same findings are also commonly observed in patients with OSA[38] and so cannot be considered specific for UARS. The presence of these abnormalities suggests a disturbance of facial development caused by increased nasal resistance during early childhood with mouth breathing.[39]

Signs and Symptoms

The most commonly observed polysomnographic feature of UARS patients is nonapneic, habitual snoring or silent IFL with relatively few adjudicated apneic or hypopneic events (AHI <5/hour). In clinical practice, these patients will seek medical attention for their condition because they also suffer from nonrestorative sleep, fatigue, sleepiness, or insomnia. In fact, UARS patients are more commonly referred to a sleep disorders center for their symptoms than for their snoring. Before referring these patients for cognitive behavior therapy (CBT) for insomnia, a careful sleep-related history revealing snoring without witnessed apnea will prompt polysomnographic investigation with documentation of IFL during sleep. It is emphasized that a report of witnessed apnea does not preclude a diagnosis of UARS because about one third of UARS patients are reported to have witnessed apnea but an AHI below the threshold for OSA.[40] Similarly, the absence of audible snoring does not preclude a diagnosis of UARS because inaudible IFL is observed in about 10% of patients diagnosed with UARS.[40,41] Typically, these patients have been diagnosed with insomnia and, in the absence of a sleep-related history of habitual snoring, are referred for CBT without performing polysomnography. When CBT fails to improve their condition and a polysomnogram is performed to exclude an intrinsic sleep disorder, IFL in the absence of audible snoring can be demonstrated.

The earliest reports of UARS emphasized the importance of hypersomnolence as a diagnostic criterion distinguishing it from isolated snoring.[5,13] Before those reports, polysomnographic technology used a thermistor or thermocouple to generate a qualitative airflow signal that could not be used to recognize IFL. Thus the link between hypersomnolence and IFL could not be made, and patients with UARS often received a diagnosis of idiopathic hypersomnolence. The earliest reports of UARS substituted a pneumotachograph recording of airflow for the qualitative airflow signal together with an esophageal pressure catheter measurement of inspiratory effort to establish the presence of IFL during sleep in UARS patients.[5,13] With time and the growth of clinical experience evaluating UARS patients, the diagnostic criteria for UARS have been expanded to include complaints of hypersomnolence or fatigue.[6,7,40]

Hypersomnolence and fatigue are not synonymous. Hypersomnolence indicates increased sleep pressure expressed by short sleep latency, a state that is inconsistent with a complaint of insomnia. Fatigue, on the other hand, is generally associated with longer sleep latencies, reflecting a state of hyperarousal commonly observed in patients with insomnia. Indeed, about one third of UARS patients complain of sleep-onset insomnia, and nearly two thirds report sleep maintenance insomnia.[35] Characteristically, the complaints of fatigue and insomnia among UARS patients are associated with the complaint of nonrestorative sleep.

Interestingly, UARS patients complain of more subjective sleep disturbance than OSA patients who have much more disrupted sleep.[41a] UARS patients can also experience a variety of parasomnias. Among these are sleep-related bruxism,[40] chronic sleepwalking in children[42] and catathrenia.[43]

Currently, there is not enough evidence to conclude that UARS is an independent cardiovascular risk factor. An increased prevalence of arterial hypertension among nonapneic snorers has been reported,[44] and borderline arterial hypertension has been lowered with nasal CPAP in a small series of UARS patients.[45] Hypotension and orthostatic intolerance have also been documented in about 20% of patients with UARS.[46]

Psychiatric symptoms such as depression[36,37,47,48] and anxiety[47-49] have been demonstrated among UARS patients and have responded dramatically to treatment using nasal CPAP and rapid palatal expansion in case reports.[47,48] Conversely, failure to diagnose and treat UARS is associated with a worsening of these symptoms over time.[36]

Currently there are limited data available regarding cognitive function among patients with UARS. Using a psychomotor vigilance task, Stoohs and associates[50] have reported increased reaction times among UARS patients compared with OSA patients. Research from Broderick and associates[51] suggests that, although UARS patients perceive themselves to have impaired cognitive function compared with healthy controls, objective testing fails to demonstrate such a difference.

UARS patients also commonly present with a variety of symptoms characteristic of the functional somatic syndromes.[40] In addition to insomnia, sleepiness, and fatigue, and the affective symptoms of depression and anxiety already mentioned, UARS patients may experience headaches and functional gastrointestinal symptoms and alpha-delta sleep, all common symptoms and signs of the functional somatic syndromes.[40] These functional somatic syndrome symptoms and signs (specifically, sleep-onset insomnia, headache, irritable bowel syndrome, and alpha-delta sleep) decrease in prevalence among sleep-disordered breathing patients as the AHI increases.[40] Conversely, when patients with functional somatic syndromes undergo polysomnography, IFL during sleep is commonly observed (fibromyalgia, temporomandibular joint syndrome, Gulf War illness, and irritable bowel syndrome have been studied).[21,22,52,53] In this setting, nasal CPAP has been shown to relieve the symptoms of functional somatic syndrome patients by relieving their IFL during sleep.[53,54]

Polysomnographic Findings

Polysomnographic findings among UARS patients can be subdivided into those characterizing breathing with associated arousals and those characterizing sleep architecture (electroencephalographic frequencies, sleep staging). Concerning sleep architecture, researchers have observed findings consistent with unstable, nonrestorative sleep among UARS patients.

Polysomnographic Findings Characterizing Breathing. Breathing in UARS is, by definition, characterized by an AHI below 5/hour of sleep and periods of IFL during sleep with flows greater than 50% of waking levels (exemplified in Figures 112-2 to 112-4), terminated by arousals or changes in the background electroencephalographic rhythm associated with a return of airflow to a non-flow-limited state (i.e., RERAs).[36] Oxyhemoglobin saturation generally remains above 90% throughout sleep.[35,37] In several large studies, the mean AHI for UARS patients is consistently 2/hour, and the frequency of RERAs is between 5/hour and 20/hour.[35-37] Describing the prevalence of flow-limited breaths during sleep among UARS patients has received little attention. One large study that used a snore microphone to determine the prevalence of breaths associated with audible snoring among 424 UARS patients observed a 21 ± 23% (mean ± standard deviation) prevalence of such breaths during sleep.[37] It is likely therefore that if a study were performed that included both a snore microphone and a pressure transducer for airflow measurement to identify both snoring and inaudible IFL (as explained previously), the prevalence of such breaths would be considerably higher and not a sporadic occurrence.

The preceding description of breathing during sleep in UARS does not define the syndrome based on thresholds for IFL or RERAs. Empirically, periods of IFL during sleep in UARS may last a few breaths or be continuous for many polysomnographic epochs. The presence of IFL has not been defined by a consensus frequency of *resistive* events, but it is a characteristic of breathing during sleep that can be described in a polysomnographic report based on the sleep stages in which it occurs and an impression of the prevalence of flow-limited breaths in those sleep stages (e.g., continuous, intermittent, or uncommon; Figure 112-4 is one 30-second epoch of continuous IFL in a UARS patient). Similarly, in the UARS literature, RERAs have not been defined by a consensus length of the preceding period of IFL (as has been done in ICSD3). Rather, the duration of IFL preceding a RERA has been undefined,[5] 10 seconds[36] or one flow-limited breath,[23] depending on the study. Because the diagnosis of UARS is not dependent on thresholds for resistive events or RERAs, UARS cannot be classified as mild, moderate, or severe based on these events. Indeed, there are no published data relating the severity of hypersomnolence among UARS patients to RERA frequency or prevalence of IFL.

Polysomnographic Findings Characterizing Sleep Architecture. Polysomnography of UARS patients demonstrates findings consistent with unstable, nonrestorative sleep (the altered sleep quality referred to earlier under Background). Among these findings is alpha frequency intruding into sleep, increased sleep stage shifts, and cyclic alternating pattern (CAP).

Patients with UARS experience increased alpha frequency, a frequency observed during quiet wakefulness, within their sleep electroencephalogram.[40,55] This increased alpha frequency may be seen in stage N3 sleep, where it has been termed *alpha–delta sleep*[40,56] (Figure 112-5) or in N1 and N2 sleep[48] (Figure 112-6; also observed in Figure 112-4). It is emphasized that this alpha frequency occurs during continuous sleep and is not the consequence of an electroencephalographic arousal. Among patients with fibromyalgia and chronic fatigue syndrome, the intrusion of waking alpha frequency into the sleep electroencephalogram is hypothesized to reflect a state of *aroused*, nonrestorative sleep.[57,58] Although resolution of alpha frequency intrusion in sleep has never been described among fibromyalgia or chronic fatigue syndrome patients, such resolution has been observed among adolescent UARS patients when sleep quality improves following rapid palatal expansion[48] (Figures 112-5 and 112-6).

UARS patients also demonstrate sleep stage instability. This instability may be recognized as frequent shifting from deeper to lighter sleep stages or to wakefulness, with decreasing depth of sleep designated as the stage sequence: REM, N3, N2, N1, and wake. The frequency of sleep stage shifting in UARS patients is decreased by treatment with nasal CPAP[54] (Figure 112-7) and rapid palatal expansion,[48] which overcome upper airway resistance. The mechanism by which nasal CPAP eliminates sleep stage shifts is not simply elimination of sleep fragmentation associated with RERAs. Although shifts between stages N2, N1, and wake require an intervening arousal, stage shifts between REM, N3, and N2 do not require an arousal. Indeed the N3-to-N2 sleep stage shifts in Figure 112-7 that decrease in frequency with nasal CPAP all occur during continuous sleep (the difference between N3 and N2 being determined by the prevalence of delta waves) and do not represent the elimination of RERAs by nasal CPAP. The occurrence of frequent shifts from deeper to lighter sleep is hypothesized to be an adaptive response to a danger or *stressor*, lightening the individual's sleep and allowing for a quicker response to an emergency.[59] Increased shifts from deeper to

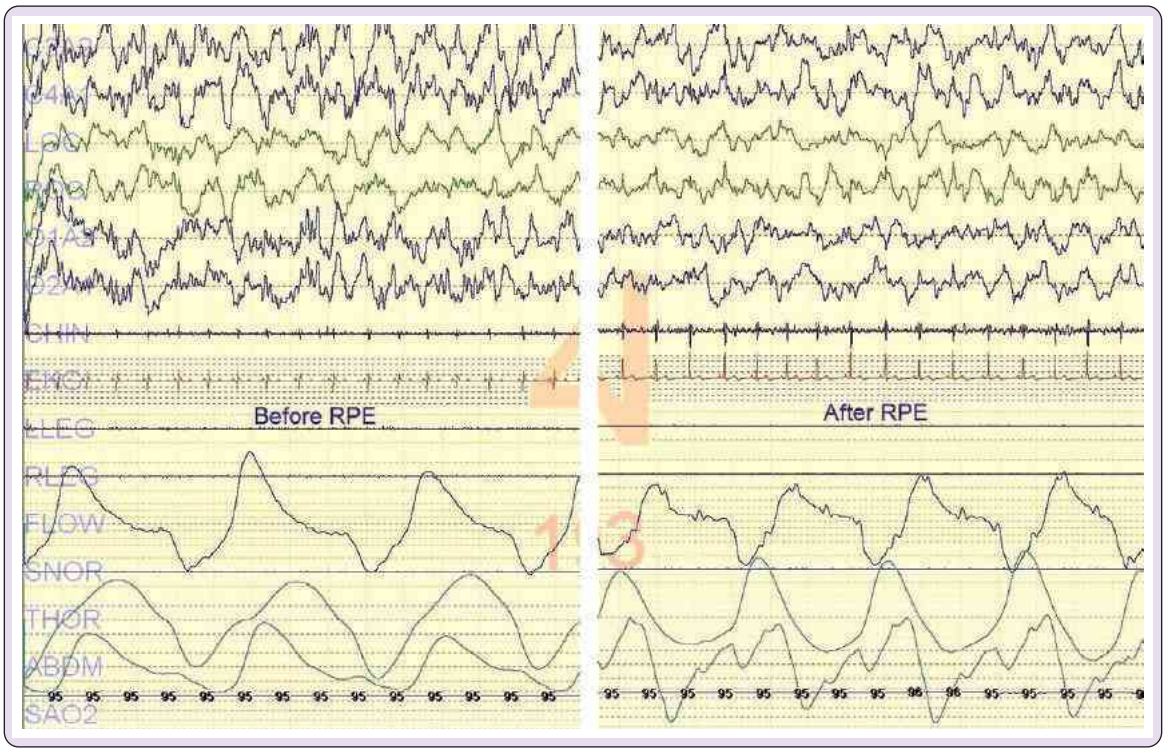


Figure 112-5 This figure demonstrates two 15-second periods of NREM stage 3 (N3) sleep recorded at the same time of night, before and after rapid palatal expansion (RPE; 13 months between studies) in a 16 year-old boy with severe chronic fatigue who was diagnosed with upper airway resistance syndrome. Recording includes four electroencephalographic channels (*purple*; C3A2, C4A1, O1A2, O2A1), left and right electrooculograms (*green*; LOC, ROC), electromyograms of the chin (CHIN) and left and right tibialis anterior muscle (LLEG, RLEG), and an electrocardiogram (EKG). Respiratory channels include pressure transducer airflow (FLOW), a snore microphone (SNOR), thoracic and abdominal wall movement (THOR, ABDO), and oxygen saturation (SaO_2). Before RPE, the patient demonstrates alpha-delta sleep characterized by low-frequency, high-amplitude delta waves with superimposed prominent 7- to 11-Hz alpha waves observed best in electroencephalographic leads C3A2 and C4A1. After RPE, the alpha frequency is greatly decreased in amplitude or gone. Associated with this change, The EKG demonstrates a decrease in heart rate between studies from 72/minute before RPE to 64/minute after RPE, suggesting decreased sympathetic nervous system tone between studies.

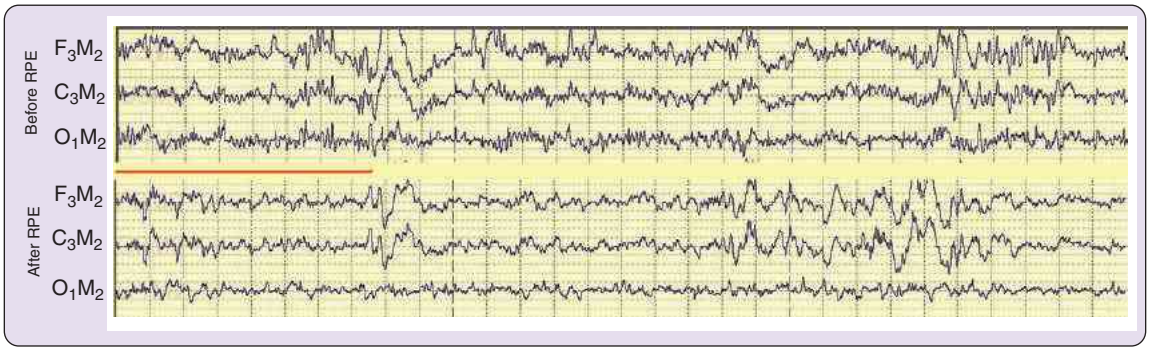


Figure 112-6 This figure demonstrates 30 seconds of stage N2 sleep recorded at the same time of night, before and after rapid palatal expansion (RPE), from an 18-year-old man with severe depression who was diagnosed with upper airway resistance syndrome. The two recordings each include three electroencephalographic channels (F_3M_2, C_3M_2, O_1M_2). As in Figure 112-5, the recording before RPE shows prominent alpha frequency (at approximately 7 Hz; seen well above the *orange line*). After RPE, the alpha frequency is greatly reduced in amplitude and the underlying theta frequency of 3 to 5 Hz is seen more clearly. (Reproduced with permission from Miller P, Iyer M, Gold AR. Treatment resistant adolescent depression with upper airway resistance syndrome treated with rapid palatal expansion: a case report. *J Med Case Rep* 2012;6[1]:415.)

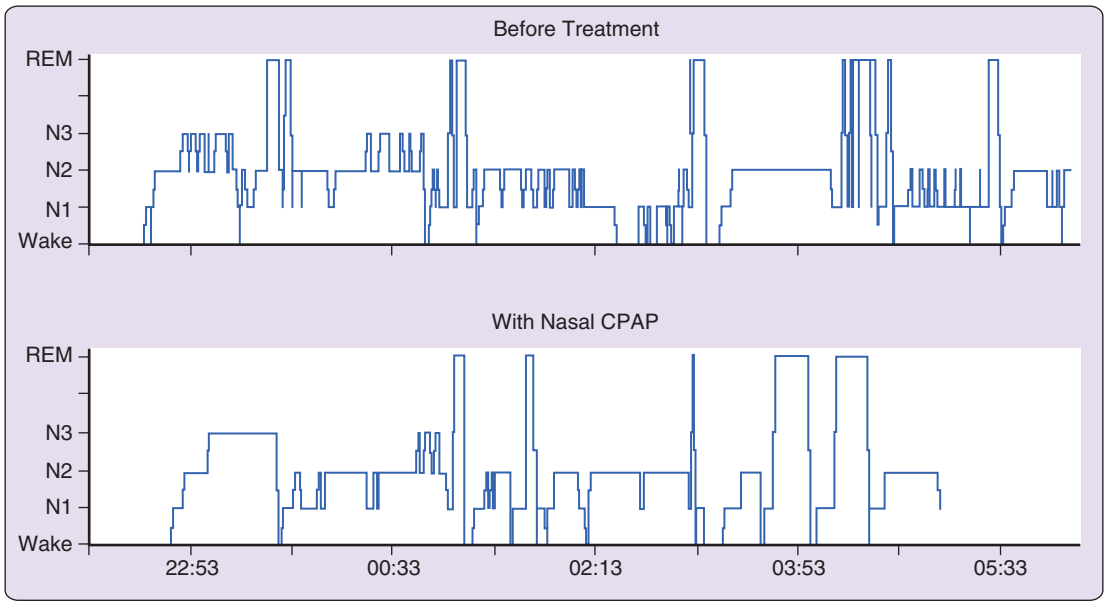


Figure 112-7 This figure demonstrates two hypnograms (plots of sleep stages against time of night with increasing depth of sleep staged as: wake, N1 [NREM stage 1], N2, N3, REM) from a 43-year-old veteran of the first Gulf War (1990–1991) who returned with complaints of moderate fatigue and severely impaired sleep quality (symptoms of Gulf War illness) and was found to have an apnea hypopnea index of 5/hour.[21] The upper hypnogram is derived from his polysomnogram before treatment, and the lower hypnogram was obtained from a polysomnogram performed (while sleeping with nasal CPAP at 9 cm H_2O) after the veteran slept with nasal CPAP nightly for 3 weeks and experienced improvement of his fatigue and sleep quality. The initial hypnogram demonstrates frequent shifts from deeper to lighter sleep stages throughout the night. The hypnogram obtained following symptomatic improvement demonstrates fewer sleep stage shifts. Frequent shifts from deeper to lighter sleep are thought to be an adaptive response to stress that enables the individual to respond more quickly to an emergency. (Reproduced with permission from Amin MM, Gold MS, Broderick JE, Gold AR. The effect of nasal continuous positive airway pressure on the symptoms of Gulf War illness. *Sleep Breath* 2011;15[3]:579–87.)

lighter sleep commonly occur in healthy people sleeping for the first night in a new location, such as a sleep laboratory.[60]

A second manifestation of sleep stage instability among UARS patients is the occurrence of CAP, which is defined by a periodic disruption of NREM sleep by electroencephalographic events that do not meet the threshold for an arousal by conventional sleep staging criteria.[61] Indeed, these electroencephalographic events constitute the changes in background rhythm associated with the return of airflow to a non-flow-limited state, referred to earlier concerning RERAs. Among UARS patients, increasing levels of these nonarousal electroencephalographic events correlate with increasing levels of sleepiness and fatigue.[61] Furthermore, the presence of CAP is a marker for increased sympathetic nervous system tone[62] commonly found under conditions of stress (see later). Therefore, CAP is one of the manifestations of unstable, nonrestorative sleep that leads to daytime sleepiness and fatigue among UARS patients.

Pathophysiology and Clinical Correlates

Hypotheses concerning the pathophysiology of UARS continue to evolve as the manifestations of the disorder and the body systems affected increase in number. The first case report of UARS patients published by Guilleminault and associates presented the disorder as one of sleep fragmentation by RERAs associated with hypersomnolence that improved with nasal CPAP treatment.[5,13] The new insight was that apneas and hypopneas associated with pronounced arousals from sleep, characteristics of OSA, were not necessary to produce hypersomnolence. Rather, the investigators hypothesized that mildly increased upper airway resistance, with its associated increase in inspiratory effort, could, episodically, produce brief periods of alpha frequency intrusion into sleep (alpha arousals), also producing hypersomnolence. Although this paradigm of UARS (henceforth termed the *RERA paradigm*) provided an explanation for the hypersomnolence associated with UARS, it did not provide an explanation for the somatic, cognitive, and affective complaints, such as insomnia, fatigue, body pain, depression, anxiety, cognitive dysfunction, and gastrointestinal dysfunction, or for the parasomnias, such as bruxism, sleepwalking, and catathrenia, that have subsequently been associated with UARS.[40,42,43,47-49,51,63,64]

As clinical experience with UARS patients increased, investigators came to postulate that the hypersomnolence of UARS patients is not simply the consequence of sleep fragmentation but also of altered sleep quality that affects its restorative properties. The alpha frequency intrusion into sleep[40,55] and the unstable sleep stages characterized by increased shifts from deeper to lighter sleep[48,54] and CAP,[61] as described earlier, were seen as alternative responses to pharyngeal collapse during sleep that maintain a more patent pharyngeal airway while maintaining sleep continuity. In contrast, OSA patients, whose only response to pharyngeal collapse during sleep is episodic arousals terminating apneas and hypopneas, experience more sleep fragmentation.[35] Bao and Guilleminault[65] have further hypothesized that UARS evolves into OSA over time because of upper airway trauma related to snoring. According to this hypothesis, because of the effect of snoring on the upper airway, UARS patients eventually lose their increased sensitivity to pharyngeal collapse and their

sleep-maintaining response. As a consequence, their sleep deepens and their mild resistive events become hypopneas and apneas terminated by arousal. This proposed *sleep quality* paradigm of UARS provides an explanation for the alpha frequency intrusion into sleep and sleep stage instability characterizing UARS patients; however, it remains focused on nonrestorative sleep leading to hypersomnolence without providing an explanation for the spectrum of somatic, cognitive, and affective disorders also associated with UARS.

A third paradigm of UARS, the *chronic stress* paradigm, builds on the *sleep quality* paradigm of UARS and provides a more complete explanation for the varied symptoms associated with the syndrome.[66] The paradigm postulates that some individuals can become sensitized to upper airway resistance as a stimulus that activates the stress response (activation of the hypothalamic-pituitary-adrenal axis and sympathetic nervous system by the brain's limbic system) as if it were an existential threat. Because upper airway resistance during sleep occurs for at least several hours daily, in these individuals, it constitutes a chronic stress with associated symptoms including sleep onset and sleep maintenance insomnia, headaches, gastrointestinal and bladder irritability, body pain, anxiety, and depression. In addition to these symptoms, so prevalent among UARS patients, chronic stress is associated with hypertension, type 2 diabetes mellitus, gonadotrophic hormone deficiency leading to sexual dysfunction (erectile dysfunction in men and polycystic ovarian syndrome in women), and growth hormone deficiency leading to diminished growth in children. These are all prominent medical conditions associated with OSA. According to such a chronic stress paradigm, the sleep fragmentation by arousals and altered sleep quality caused by alpha frequency intrusion and sleep stage instability observed among UARS patients is not a direct effect of upper airway resistance on sleep continuity, but an adaptive response of the brain to the existence of a disturbance or threat. Having sleep continuously interrupted, or having a state of vigilance maintained during sleep through alpha frequency intrusion and sleep stage instability theoretically enables the individual to respond more quickly to a danger, an apparent survival advantage.[59] This advantage, however, is accompanied by the disadvantages of parasomnias and daytime sleepiness resulting from the chronically altered sleep. The chronic stress paradigm of UARS explains not only the hypersomnolence associated with UARS but also the somatic complaints, affective disorders, cognitive dysfunction, and parasomnias observed among UARS patients. A more complete discussion of this paradigm can be found in the review by Gold.[66]

In summary, the pathophysiologic and associated clinical paradigm of UARS has evolved from sleep fragmentation by RERAs through altered sleep quality as a direct response to upper airway resistance, leading to milder resistive events than occur among OSA patients, to recent consideration of upper airway resistance, provoking chronic stress with sleep-related, somatic, cognitive, and affective consequences. The pathophysiologic paradigms of UARS will continue to evolve as new data accumulate. However, the recognition that altered sleep *quality* contributes to the hypersomnolence and fatigue of UARS patients supports the idea that UARS also exists below the RDI threshold for a diagnosis of OSA—an important possibility when one contemplates making the diagnosis of idiopathic hypersomnolence.

Treatment

The treatment of UARS uses the same treatments that have been discussed earlier for snoring. Chief among these treatments is nasal CPAP, which is highly effective and can be precisely titrated by the prescribing physician to eliminate IFL during sleep. To titrate nasal CPAP for UARS patients, one must titrate to convert IFL during sleep into non-flow-limited breathing, as illustrated in Figures 112-2 and 112-3. In a large, published clinical series of sleep-disordered breathing patients all titrated in this manner (using only nasal masks), mean therapeutic level of nasal CPAP for 22 UARS patients was found to be 7 cm H_2O with a range of 4 to 9 cm H_2O.[3] Although autotitrating positive airway pressure has not been studied specifically in UARS, the algorithms used attempt to eliminate IFL during sleep and should be acceptable for treatment of UARS as they are for OSA. For patients unable (or unwilling) to breathe with the mouth closed or to wear a nasal mask, alternative forms of treatment such as mandibular advancement appliances or tongue-retaining devices, weight loss, and surgical procedures may be considered as previously described. A growing body of literature suggests that applying positive airway pressure through an oronasal mask is not a reliable method for eliminating IFL during sleep[67,68] and anesthesia.[69] Among pediatric patients, rapid palatal expansion performed by an orthodontist has been used effectively to treat UARS (e.g., the patients in Figures 112-5 and 112-6[48]) and mild OSA.

CLINICAL PEARLS

- Silent inspiratory airflow limitation (IFL) during sleep is characterized by either an inspiratory airflow plateau or an increase in the ratio of inspiratory time to the respiratory cycle time with a prolongation of the time near maximal inspiratory airflow.
- In a patient consulting the clinician for habitual snoring that disturbs his or her bed partner, with normal alertness, no somatic or metabolic disorders, and no known cardiovascular disease, a polysomnogram will likely reveal either isolated snoring or asymptomatic OSA. Consider evaluating the results of a carotid ultrasound, looking for evidence of atherosclerosis, before foregoing specific OSA treatment in this setting.
- For patients with functional somatic syndromes complaining of insomnia, fatigue, headache, body pain, gastrointestinal or bladder irritability, anxiety and depression, with or without audible snoring, consider performing polysomnography to diagnose UARS (OSA by ICSD3 criteria) because prevention of IFL during sleep may be an effective treatment not only for fatigue and insomnia but also for somatic symptoms.

SUMMARY

Our understanding of pathologic pharyngeal collapse during sleep has progressed from recognizing obstructive apneas and hypopneas associated with arousal from sleep and oxygen desaturation (clinically, obstructive sleep apnea) to recognizing the mildest IFL without audible snoring, arousal, or oxygen desaturation. At the same time, our understanding of the consequences of pathologic pharyngeal collapse during sleep has expanded from hypersomnolence and cardiovascular

and metabolic disorders to include associations with somatic syndromes, affective disorders, and carotid artery atherosclerosis independent of metabolic syndrome. Underlying this evolution is a new paradigm of *sleep related breathing disorders* (often referred to as "sleep-disordered breathing") in which pharyngeal collapse during sleep acts not only directly, causing oxygen desaturation and arousal from sleep, but also indirectly, with even the mildest IFL during sleep serving as a chronic activator of the body's stress response. In this context, one can appreciate the evolving understanding of what constitutes clinically significant sleep related breathing disorders associated with sleep-related upper airway pathophysiology.

Selected Readings

Broderick JE, Gold MS, Amin MM, et al. The association of somatic arousal with the symptoms of upper airway resistance syndrome. *Sleep Med* 2014;**15**:436–43.

de Godoy LB, Palombini LO, Guilleminault C, et al. Treatment of upper airway resistance syndrome in adults: where do we stand? *Sleep Sci* 2015;**8**(1):42–8.

Dubrovsky B, Raphael KG, Lavigne GJ, et al. Polysomnographic investigation of sleep and respiratory parameters in women with temporomandibular pain disorders. *J Clin Sleep Med* 2014;**10**:195–201.

Gold AR, Dipalo F, Gold MS, et al. Inspiratory airflow dynamics during sleep in female fibromyalgia patients. *Sleep* 2004;**27**:459–66.

Gold AR, Schwartz AR. The pharyngeal critical pressure: the whys and hows of using nasal continuous positive airway pressure diagnostically. *Chest* 1996;**110**:1077–88.

Guilleminault C, Kirisoglu C, Poyares D, et al. Upper airway resistance syndrome: a long-term outcome study. *J Psychiatr Res* 2006;**40**:273–9.

Guilleminault C, Stoohs R, Clerk A, et al. A cause of excessive daytime sleepiness: the upper airway resistance syndrome. *Chest* 1993;**104**:781–7.

Hosselet JJ, Norman RG, Ayappa I, et al. Detection of flow limitation with a nasal cannula/pressure transducer system. *Am J Respir Crit Care Med* 1998;**157**:1461–7.

Lee SA, Amis TC, Byth K, et al. Heavy snoring as a cause of carotid artery atherosclerosis. *Sleep* 2008;**31**:1207–13.

Marshall NS, Wong KK, Cullen SR, et al. Snoring is not associated with all-cause mortality, incident cardiovascular disease, or stroke in the Busselton Health Study. *Sleep* 2012;**35**:1235–40.

Schneider H, Krishnan V, Pichard LE, et al. Inspiratory duty cycle responses to flow limitation predict nocturnal hypoventilation. *Eur Respir J* 2009;**33**:1068–76.

So SJ, Lee HJ, Kang SG, et al. A comparison of personality characteristics and psychiatric symptomatology between upper airway resistance syndrome and obstructive sleep apnea syndrome. *Psychiatry Investig* 2015;**12**(2):183–9.

Stoohs RA, Knaack L, Blum HC, et al. Differences in clinical features of upper airway resistance syndrome, primary snoring, and obstructive sleep apnea/hypopnea syndrome. *Sleep Med* 2008;**9**:121–8.

A complete reference list can be found online at ExpertConsult.com.

Chapter
113

Obstructive Sleep Apnea: Phenotypes and Genetics

Susan Redline

Chapter Highlights

- There is clear evidence that a positive family history of obstructive sleep apnea (OSA) is an important risk factor for an elevated apnea-hypopnea index (AHI) and associated symptoms such as snoring, daytime sleepiness, and apneas. Patients with OSA often have relatives—parents, siblings, and children—who have similar symptoms, a diagnosis of OSA, or both. The familial aggregation of OSA has been quantified through use of twin and cohort studies, which conservatively estimate that risk for OSA is increased by 50% in individuals with an affected first-degree relative.
- Overall heritability estimates (the proportion of the variance in a trait attributable to genetic factors) for the AHI are 0.30 to 0.40. Although there is a strong correlation between OSA and obesity, only 35% of the genetic variance in the AHI may be accounted for by genes that influence obesity (with 65% of the genetic variance likely due to genetic variants in other etiologic pathways). Other potentially inherited risk factors for OSA include craniofacial structural traits that influence upper airway patency; body fat distribution, including propensity for airway fat deposition; chemoreflex ventilatory control; and arousability to ventilator stimuli.
- Pedigree studies analyzed using linkage analysis have identified several areas where biologically plausible candidate genes are located, including candidates for ventilatory control and obesity. A number of association studies and emerging meta-analyses of candidate and genome-wide association studies also provide evidence for increased susceptibility to OSA in persons who inherit variants for genes in pathways implicated in ventilatory control, inflammation, body fat distribution, and craniofacial structure.

DEFINITION OF THE OBSTRUCTIVE SLEEP APNEA PHENOTYPE

As is the case with many other complex disorders, variable definitions have been used to characterize obstructive sleep apnea (OSA) in genetic analyses. Clinically, OSA is recognized by the occurrence of repetitive episodes of complete or partial upper airway obstruction during sleep that result in oxygen desaturation or arousal and that are usually accompanied by symptoms of loud snoring and daytime sleepiness. However, there can be substantial variability in identification and characterization of OSA because of variations in how specific respiratory events are identified and defined, the threshold level for frequency of events during sleep considered pathologic, and to what extent other clinical and polysomnographic data are thought necessary for characterizing disease status.

Most family and genetic studies of OSA have used the apnea-hypopnea index (AHI) to define phenotype. The advantages of using the AHI include its ability to be readily calculated from data obtained from overnight sleep apnea tests with moderate to high night-to-night reproducibility[1] and from widespread clinical use. In addition, it is often followed as a key outcome in OSA treatment research studies. Because the AHI has been shown to be moderately correlated with other indexes of OSA severity, such as nighttime oxygen desaturation and sleep fragmentation, it may provide information about several correlated traits that are important in disease expression. All genetic studies of OSA that use the AHI as the outcome measure have demonstrated significant familial aggregation, suggesting that this measure captures useful information for quantifying genetic associations. All candidate gene studies conducted to date also have defined "case" and "controls" on the basis of threshold levels of AHI (either >5 or >15/hour of sleep).

Additional metrics of OSA that provide more specific information on patterns of respiratory disturbances during sleep have been reported to be heritable and thus have potential for use in genetic studies. These include measures of AHI specific to rapid eye movement (REM) and non–rapid eye movement (NREM) sleep, average levels of nocturnal oxygen desaturation, and average duration of sleep-related respiratory disturbances. The latter metric, which is related to respiratory arousability, has been reported to have a heritability of almost 0.60[2] (indicating as much as 60% of the variance in this trait is explained by familial factors acting additively) and thus may be useful for identifying genes influencing ventilatory control.

A multidimensional OSA phenotype can be derived by combining polysomnographic data with information related to symptoms, signs, and outcome data. In the Cleveland Family Study, a stronger relationship between familial risk and OSA was observed when OSA was defined by an AHI greater than 15/hour plus reported daytime sleepiness than when disease was defined by AHI alone.[3] Additional power may be gained in future genetic studies of OSA that use multidimensional phenotypes. For example, alternatives to the AHI, such as indexes of flow limitation during sleep and critical airway closing pressure, may prove to be superior markers for genetic studies. However, a phenotype must be feasible for use in the large numbers of subjects who are needed for genetic epidemiologic studies of complex traits. The choice of phenotype for use in genetic and other research studies will be influenced by the cost, degree of invasiveness, and individual burden required for identifying and quantifying the phenotype and by applicability across the spectrum of age and body mass index (BMI), as well as by its accuracy and reliability.

INTERMEDIATE DISEASE PATHWAYS AND PHENOTYPES

OSA is a complex disorder that is defined using a combination of clinical and physiologic measures, such as symptoms and data from overnight sleep apnea testing. However useful such an approach may be for clinical diagnosis, such definitions may be insufficiently specific for use in genetic analyses. An alternative approach for studying the genetic basis of OSA is to study intermediate traits that confer increased risk for the disorder. Such intermediate traits may be more closely associated with specific gene products and may be less influenced by environmental modification than more complex (and downstream) phenotypes (Figure 113-1).

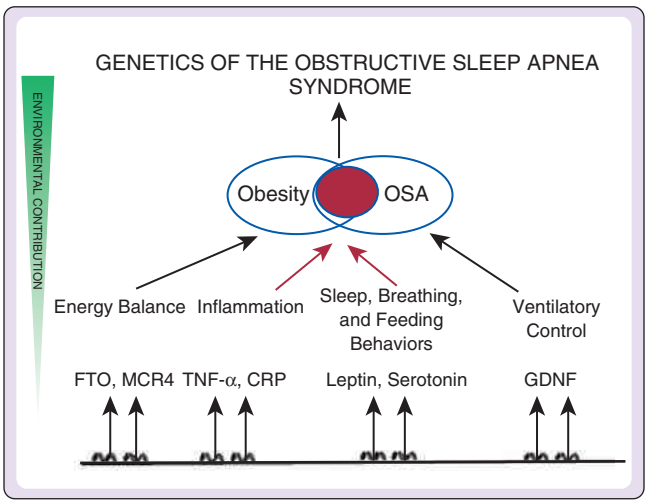


Figure 113-1 Schema showing the influence of genes on obstructive sleep apnea (OSA) from four intermediate pathways (energy balance, inflammation, sleep-feeding overlap, and ventilatory control) may be influenced by specific genes, and individually or together may influence obesity and/or OSAH. CRP, C-reactive protein gene; FTO, fat mass and obesity–associated gene; GDNF, glia-derived growth factor; Leptin: leptin or leptin receptor genes; MCR4, melanocortin-4 receptor gene; Serotonin: genes in the serotonin pathways (e.g., *HTR2A*); TNF-α, tumor necrosis factor-α gene. As one moves from the gene to the complex phenotype, the relative influence of specific genes decreases while the influence of environmental factors increases.

A number of risk factors likely interact to increase propensity for the repetitive upper airway collapse that occurs during sleep in patients with OSA. In a given person, the relevant attributes may be determined by anatomic and neuromuscular factors that influence upper airway size and function. Strong OSA risk factors are obesity and male gender. Although it has been argued that the genetics of obesity cannot be separated from the genetics of OSA, careful statistical modeling of AHI and BMI indicates that only about 35% of the genetic variance in AHI is shared with BMI, suggesting that a substantial portion of the genetic basis for OSA is in fact independent of obesity.[4] Other pathogenic pathways include those that influence upper airway size, ventilatory control mechanisms, and possibly elements of sleep and circadian rhythm control.

Thus it is useful to consider at least four primary intermediate pathogenic pathways through which genes might act to increase susceptibility to OSA: obesity and body fat distribution and related metabolic syndrome and inflammatory phenotypes, craniofacial and upper airway morphology, control of ventilation, and control of sleep and circadian rhythm[5]; these are discussed in more detail in the following sections. The limitation of this approach is that the genes so identified might not be sufficient to describe the clinically important phenotype, which might only occur in the context of other genetic and environmental factors. Specifically, susceptibility genes for intermediate traits associated with OSA might not be equivalent to the susceptibility genes for OSA.

Obesity and Body Fat Distribution

Obesity increases risk for OSA by 2- to 10-fold, with the strongest associations observed in middle age.[6] There are several pathways through which obesity predisposes to OSA. Fat deposition in the parapharyngeal fat pads may directly narrow the upper airway and predispose it to collapse when neuromuscular activation of upper airway muscles declines with sleep (see Chapters 17 and 111). Fat deposition in the thorax and abdomen (i.e., visceral fat) can increase the mechanical work of breathing, which can produce hypoventilation and reduce lung volumes, which in turn reduces parenchymal traction on the trachea, making the airway more collapsible. Reduced lung volumes also can increase propensity for oxygen desaturation to occur, increasing the likelihood that any given reduction in airflow may be classified as a "hypopnea," thus operationally increasing the severity of the adjudicated AHI, as well as physiologically increasing the severity of a hypopnea-associated disturbance. In addition, low lung volumes can reduce oxygen stores and alter "loop gain," which in turn can promote ventilatory instability. Finally, adipose tissue secretes hormones such as leptin that can influence ventilatory drive (see later).

Heritability estimates for obesity-associated phenotypes such as BMI, skinfold thickness, regional body fat distribution, fat mass, and leptin levels range between 40% and 70%, consistent with moderate to strong influences of genetic factors on these traits.[6-8] Approximately 7% of cases of early-onset obesity have been estimated to be attributable to the effects of mutations in a small number of genes involved in the leptin-melanocortin signaling pathway (i.e., melanocortin-4 receptor, leptin, leptin receptor, and pro-opiomelanocortin [POMC]), which are believed to influence weight largely through alterations in appetite regulation.[7] Mutations in melanocortin-4

receptor, the most common mutation, increase risk for severe childhood obesity by approximately 30% and also have been implicated in 0.5% to 1% of adult cases of obesity.[8]

The genetic etiology of obesity in the general population has been studied intensively in large populations that have undergone genotyping. Meta-analyses of genome-wide association studies, which examine the variation of frequency of thousands of alleles with disease status, have led to the discovery and replication of 36 genetic loci that associate with BMI.[9] However, all risk alleles together explain only 6% to 11% of the phenotypic variation.[10] The locus with the largest effect size is in the *FTO* (fat mass and obesity–associated gene),[11] which explains 0.34% of the variation in BMI among adult populations. An association between *FTO* and BMI has been replicated across populations and indicates that homozygotes for the risk allele weigh on average 3 to 4 kg more than persons without the risk allele and have an associated 1.67-fold increased risk for obesity compared with persons without the allele. Although the functioning of this gene is not well understood, *FTO* is expressed in the hypothalamus, and there is some evidence that it confers a risk for increased obesity through regulation of food intake and possibly through mechanisms that influence stress responses. Distinct genetic variants have been associated with waist-to-hip ratio,[12] which may be particularly relevant to OSA given the strong association between central obesity and OSA.

Craniofacial Morphology

Craniofacial morphology, which encompasses both bony and soft tissues, predisposes to OSA by reducing upper airway dimensions. Soft tissue structures that vary with OSA include elongation of the soft palate, macroglossia, and hypertrophy of adenoids and tonsils. Magnetic resonance imaging (MRI) has specifically shown that the lateral pharyngeal wall and tongue are larger in OSA patients compared with matched controls.[13] Cephalometry also has shown that patients with OSA compared with those without OSA have reduction of the anterior-posterior dimension of the cranial base, increased lower facial height, mandibular retrognathia or micrognathia, and inferior displacement of the hyoid[14-16] (see Chapters 111 and 143). A brachycephalic head form, measured by anthropometry, is often found in association with reduced upper airway dimensions. This head form is associated with a small but significant increased risk for OSA in those of European ancestry, and it also identifies families at risk for both OSA and sudden infant death.[17] In African Americans, this head form is uncommon and does not appear to increase risk for OSA. It is possible that a brachycephalic head form may contribute to increase risk for OSA among individuals of Asian ancestry.

In the Cleveland Family Study, both hard tissue (e.g., head form, intermaxillary length) and soft tissue (e.g., soft palate length, tongue volume) factors predicted the AHI level in European Americans. In African Americans, soft tissue factors also predicted AHI levels, but hard tissue anatomic features appeared to be only weakly associated with OSA.[18] These data support the importance of structural features in increasing susceptibility to OSA, but they also suggest that the anatomic underpinnings and the genes for upper airway anatomy might differ among ethnic groups.

Facial morphogenesis and patterning are complex processes that involve multiple signaling pathways. In humans, the genetic basis for craniofacial features is supported by both twin and family studies.[19,20] Heritability estimates (the proportion of the variance in the trait explained by additive genetic factors) for facial features such as facial height and mandibular position have been reported to be high as 0.80. A genome-wide association study has identified loci in five genes associated with facial morphology,[21] the most robust of which was in the paired box 3 gene (*PAX3*), which encodes a developmentally important transcription factor expressed in neural crest cells. There are also at least 50 syndromes in which congenital malformations of mandibular and maxillary structure occur, many of which also are associated with respiratory impairment and upper airway obstruction. These include Pierre-Robin syndrome and Treacher Collins syndrome.[22,23] Studies of various syndromes and genetic defects suggest potential roles of genes belonging to the fibroblast growth factor (e.g., *FGFR1, FGFR2, FGFR3*), transforming growth factor-β (e.g., *TGFBR1, TGFBR2*), homeobox (e.g., *MSX1, MSX2*), and sonic hedgehog (e.g., *PTCH, SHH*) pathways. Other potentially relevant candidate genes are those that have been implicated in craniofacial development, including genes on the endothelin pathway (e.g., *ECE1, EDN1, EDNRA*),[24-26] and *TCOF1*, the cause of Treacher Collins syndrome.[27] Further understanding of homeobox genes and genes controlling growth factors might contribute to our clarifying the origins of craniofacial dysmorphisms found in OSA.

Inherited abnormalities of craniofacial structure appear to explain at least some of the familial aggregation of OSA. Relatives of patients with OSA have been shown to have a more retropositioned mandible and smaller posterior-superior airway space compared with normative data.[28] Relatives of OSA probands also have been shown to have decreased total pharyngeal volume and glottic cross-sectional area, retropositioned maxilla and mandible, and a longer soft palate compared with relatives of controls.[29] In subjects both with and without OSA, acoustic reflectometry has demonstrated that more than 30% of the variance in the minimal cross-sectional area of the pharynx is heritable.[30] MRI has further demonstrated significant heritability for both the volume of soft tissue airway structures (including the tongue and lateral pharyngeal walls)[20] and hard tissue craniofacial dimensions (e.g., mandibular length and width).[31]

Although MRI precisely describes anatomic characteristics, its cost limits its utility for large-scale genetic epidemiology studies. Future research needs to assess the ability of reproducible and noninvasive techniques to identify subgroups of persons who inherit polymorphisms associated with genes relevant to craniofacial compared with other etiologic OSA pathways. There is also a need to further investigate how anatomic measurements performed awake and in the sitting position predict collapsibility during sleep and how anatomy interacts with physiology to influence susceptibility to OSA. Anatomic compromise may also be inferred from measurement of the pharyngeal critical closing pressure (Pcrit),[32] an index that can be derived during attended polysomnography with measurement of airway responses to progressive decreases in the delivery of therapeutic continuous positive airway pressure.

Ventilatory Control

Potentially inherited abnormalities of ventilatory control may predispose to OSA or central sleep apnea by affecting

ventilation, ventilatory drive, and upper airway patency. These inherited ventilatory control abnormalities may include neuromuscular responses to the influences of state (sleep-wake), chemical drive (e.g., ventilatory response to hypoxia and hypercapnia), sensitivity of ventilatory load compensation (the degree to which an individual defends the tidal volume or minute ventilation in the presence of an imposed mechanical load to breathing such as an increased resistance or elastance), and arousal threshold. They can result in different ventilatory responses to sleep-related stresses and shape both the magnitude of ventilation and ventilatory pattern and the propensity for respiratory oscillations in sleep. The relative contribution of these ventilatory control factors varies among individuals, and such variability likely contributes to genetic heterogeneity in OSA. According to a multiple-risk factor model, OSA is likely to manifest as severe anatomic compromise, regardless of nonanatomic risk factors. In the presence of a lesser degree of anatomic compromise, OSA occurs when there are coexistent abnormalities in arousal threshold, loop gain (sensitivity of the ventilatory control system to feedback loops, such as due to changes in CO_2), or muscle responses.

Experimental data indicate that there is substantial interindividual variation in the contributions of these physiologic factors to OSA.[33] Further, the magnitude of ventilatory chemoresponsiveness appears to be subject to major genetic control; for example, heritability estimates for chemoresponsiveness to oxygen saturation levels range from approximately 30% to 75%.[34] Ventilatory responses are more strongly correlated between monozygotic than dizygotic twins.[35-37,38] Population differences in ventilatory patterns and hypoxic sensitivity have been identified for populations that have adapted to living at high altitude.[39,40] Abnormalities in hypoxic or hypercapnic ventilatory responsiveness have been described in the first-degree relatives of probands with unexplained respiratory failure,[41] chronic obstructive pulmonary disease,[42,43] and asthma.[44] There is a growing understanding of the molecular bases for ventilatory responses, with identification of a number of respiratory chemoreceptors in the carotid body and lower brainstem, including the nucleus solitaries, retrotrapezoid nucleus, locus coeruleus, and raphe.[45]

Although the contributions of the previously noted population and genetic data to OSA-related phenotypes are not clear, there is some evidence that absence of the retrotrapezoid nucleus causes severe central apneas in congenital central hypoventilation syndrome.[46] Additionally, a potential role for inherited impairments of ventilatory control in influencing susceptibility to OSA has been suggested by several studies of carefully characterized families of OSA patients, which have demonstrated blunted hypoxic responses and impairment in load compensation compared with controls.[47-51]

As referred to earlier, the potential impact of deficits in ventilatory control on OSA susceptibility is likely magnified in persons with anatomically compromised upper airways. With sleep onset, the central inspiratory drive to upper airway motor neurons, a major determinant of airway patency, is reduced or fluctuates.[52,53] Any given reduction in central inspiratory drive results in greater increases in upper airway resistance in persons with anatomically compromised airways than in others.[54] Conversely, persons with greater degrees of upper airway resistance (due to craniofacial or obesity risk factors) can require a high level of compensatory drive to overcome sleep-associated airway collapse, and thus they may be especially vulnerable to the influence of genetically determined ventilatory control deficits.

These observations underscore the potential importance of considering the interaction of genetic risk factors that influence more than one etiologic pathway. Similarly, systematic characterization of ventilatory control pathophysiology and clinical phenotypes could accelerate discovery of genes that influence specific mechanistic pathways in humans, similar to mouse models which have allowed genes to be identified that determine respiratory timing, frequency, awake ventilation, chemosensitivity, and load responses. Clear strain differences have been observed for many of these phenotypes, with evidence of quantitative trait loci near plausible candidate genes. Knockout and transgenic mice also have helped identify the role of specific proteins and receptors in ventilatory chemoreception, neuromuscular transmission, and neural integration. Candidate genes identified from such studies include genes that sense O_2[55-57] and CO_2[58,59]; genes that modulate serotonin signaling[60]; genes on the endothelin pathway,[61,62] which also are important in craniofacial development; and genes that regulate neural crest migration, including *PHOX2B,* mutations of which are associated with congenital central hypoventilation.[63,64]

Control of Sleep and Circadian Rhythm

Given the effect of sleep-wake state on respiratory motor neuron activation, insights into the susceptibility of upper airway muscles to collapse during sleep may require delineation of the genetics of sleep-wake control. Orexins, neuropeptides that play a fundamental role in the regulation of appetite and sleep-wake states,[65,66] also influence arousal and muscle tone.[65] Orexin A levels are reported to be reduced in OSA.[67] Thus abnormalities in orexin genes may be relevant to OSA because of their influence on arousal, muscle tone, ventilatory control, and weight.

It may similarly be useful to consider how respiratory motor neuron control is influenced by genetic processes that determine circadian clocks, which are known to drive important metabolic and behavioral rhythms. Genes influencing circadian rhythm have been identified in animal models and in humans[68] and have been shown to influence metabolism, inflammation, and aging.[69] The relevance of these findings to OSA is unclear. However, genetic variation in circadian rhythm determination may influence apnea number and duration by affecting the distribution of REM and NREM sleep and the associated neuromuscular responses across the sleep period. Genes that influence regulation of sleep-wake rhythm may also influence the phenotypic expression of OSA (e.g., ability to compensate for sleepiness in response to recurrent apneas and sleep disruption).

FAMILIAL AGGREGATION OF OBSTRUCTIVE SLEEP APNEA

In addition to the specific heritability and genetic associations of OSA discussed to this point, significant familial aggregation of AHI or of symptoms of OSA have been observed in studies from the United States, Finland, Denmark, the United Kingdom, Israel, and Iceland.[28,29,70-72] Such studies have used a variety of designs, including cohorts, small and large pedigrees, twins, and case-control studies; they have included adults and children; and they have employed varying

Table 113-1 Familial Correlations for Apnea-Hypopnea Index

Relationship	Partially Adjusted* Familial Correlation Coefficient	BMI-Adjusted† Familial Correlation Relationship *P* Value	Coefficient	*P* Value
Parent-offspring	0.21	.002	0.17	.017
Sibling-sibling	0.21	.003	0.18	.008

*Adjusted for age, age squared (age^2), ethnic group, and gender.
†Adjusted for body mass index (BMI), age, age^2, ethnic group, and gender.
From Redline S, Tishler PV, Tosteson TD, et al. The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med* 1995;151:682–7.

approaches for assessing phenotype. Despite study design and population differences, these studies have consistently shown familial aggregation of the AHI level and symptoms of OSA in children and adults and in obese and nonobese subjects. These studies have provided clear evidence that a positive family history of OSA is an important risk factor for an elevated AHI and for associated symptoms such as snoring and daytime sleepiness, although the estimated magnitude of effects has varied greatly.

Several large twin studies have shown that concordance rates for snoring, a cardinal symptom of OSA, were significantly higher in monozygotic twins than in dizygotic twins.[71,73,74] A study of adult male twins has shown significant genetic correlations for daytime sleepiness as well as snoring, with models consistent with common genes underlying both symptoms.[71] A subsequent report from this cohort showed significant heritability for objectively measured AHI levels in this twin population.[75] A large Danish cohort study showed that the age, BMI, and comorbidity-adjusted risk for snoring were increased threefold when one first-degree relative was a snorer and were increased fourfold when both parents were snorers.[76]

The prevalence of objectively measured OSA among first-degree relatives of OSA probands has been reported to vary from 22% to 84%.[28,29,70-72] Among the studies that included controls, the odds ratio, which relates the odds of a person with OSA in a family with affected relatives to that for someone without an affected relative, has varied from 2 to 46.[3,28,29,70] Pedigree studies from the United States and Iceland have shown consistent associations; the overall risk for OSA in a family member of an affected proband compared with an individual without affected relatives is approximately 2. This is lower than that reported from case-control studies, which may be subject to biases depending on the appropriateness of the selection of cases and controls. Heritability estimates for the AHI from both pedigree[77,78] and twin studies[75] are approximately 35% to 40%. Similar parent-offspring (correlations of approximately 0.20) have been observed, and they are greater than spouse-spouse correlations.[3]

OSA has been described as occurring more commonly as a multiplex (affecting at least two members) than as a simplex (occurring in a single family member) disorder. Further evidence for a genetic basis for OSA is derived from the observation that the odds of OSA syndrome, defined as AHI greater than 15/hour and self-reported daytime sleepiness, increases with increasing numbers of affected relatives.[3] Table 113-1 show the odds for OSA syndrome given one, two, or three affected relatives with these findings, adjusted for age, gender, ethnicity, and BMI, compared with OSA patients who have no affected relatives, These results support the utility of ascertaining family history as part of the evaluation of the patient for OSA. Information on snoring, apneas, and sleepiness among first-degree relatives can be used to refine the likelihood of OSA in a given patient. Such information can also be used to help identify the need for other family members to seek sleep evaluations.

Several studies have reported a coaggregation of OSA with sudden infant death syndrome (SIDS) and acute life-threatening events.[70,79,80] Members of families with both OSA and SIDS cases have been reported to have a relatively increased prevalence of brachycephaly, an anatomic feature that is associated with upper airway narrowing, as well as reduced hypoxic ventilatory responsiveness.[80] These observations suggest that the two sleep-related breathing disorders have a shared genetic predisposition acting through ventilatory control or craniofacial structure pathways. The demonstration of widespread serotoninergic brainstem abnormalities in SIDS victims and the putative role of this pathway in respiratory drive[81] suggest a biologic basis for the potential genetic link between these disorders.

In children, both OSA and adenotonsillar hypertrophy (the chief risk factor for pediatric OSA) have been elevated in the siblings of children with OSA.[82] Pedigree studies show that the disease is transmitted across generations,[83] suggesting that common risk factors might influence OSA susceptibility in children and adults. Although hypertrophy of the tonsils is a major risk factor for childhood OSA, children of OSA probands more often have residual OSA after tonsillectomy compared with the offspring of adults without OSA,[84] suggesting the importance of underlying genetic susceptibility as a determinant of treatment response as well.

GENETIC ANALYSES

Candidate Gene Studies

The molecular genetics of OSA have been investigated using candidate gene approaches. In these association studies, the frequency of genetic variants thought to relate to disease susceptibility are compared in groups with and without OSA or are assessed in relationship to the severity of a quantitative phenotype (e.g., AHI). A number of plausible candidate genes are also found in pathways that influence the intermediate traits of obesity, craniofacial structure, and ventilatory control (Box 113-1). Candidate genes that have been examined in relationship to OSA in humans include those for apolipoprotein E (*APOE*), angiotensin-converting enzyme (*ACE*), serotoninergic pathways, leptin pathways, obesity, and inflammation. The largest single multiple-candidate gene

Box 113-1 CANDIDATE GENES* FOR INTERMEDIATE PHENOTYPES FOR OBSTRUCTIVE SLEEP APNEA

Obesity

FTO (fat mass and obesity–associated gene)
Melanocortin-4 receptor
Leptin
Pro-opiomelanocortin
Melanocyte-stimulating hormone
Neuropeptidase Y
Prohormone convertase
Neutrophic receptor TrkB
Insulin-like growth factor
Glucokinase
Adenosine deaminase
Tumor necrosis factor-α
Glucose regulatory protein
Agouti signaling protein
β-Adrenergic receptor
Carboxypeptidase E
Insulin-signaling protein
Resistin
Ghrelin
Adiponectin
Gamma-aminobutyric acid transporter
Orexin

Ventilatory Control

RET protooncogene
PHOX2B
HOX IIL2
KROX-20
Receptor tyrosine kinase
Neurotrophic growth factors
- Brain-derived neurotrophic factor
- Glia-derived neurotrophic factor
- Neurotrophic factor-4
- Platelet-derived growth factor

Neuronal synthase
Acetylcholine receptor
Dopaminergic receptor
Substance P
Glutamyl transpeptidase
Endothelin-1
Endothelin-3
Leptin
EN-1
GSH-2
Orexin

Craniofacial Structure

Class I homeobox genes
Growth hormone receptors
Growth factor receptors
Retinoic acid
Endothelin-1
Collagen types I and II
Tumor necrosis factor-α

*Includes related proteins and receptors.

study analyzed more than 1000 single nucleotide polymorphisms (SNPs) from 53 candidate genes representing key intermediate pathways in approximately 1500 individuals of European or African ancestry.[85] In European Americans, variants within the C-reactive protein (*CRP*) and glia-derived neurotrophic factor (*GDNF*) were significantly associated with OSA, with suggestive associations observed for several SNPs in the 5-hydrotrypamine receptor 2A (*5-HTR2A*) gene and endothelin-1 (*EDN1*) gene. In African Americans, a variant in the serotonin receptor 2a (*5-HTR2A*) gene was associated with approximately twofold increased odds of OSA, with suggestive associations observed for variants in the leptin receptor and hypocretin receptor 2. Genetic associations frequently reflect false-positive findings and require replication in independent samples. However, smaller candidate gene studies, summarized later, provide additional support implicating serotoninergic, leptin signaling, and inflammatory pathways.

Serotoninergic Pathways

Serotonin (5-hydroxytryptanmine [5-HT]) receptors are found in the carotid body and in the brainstem near ventilatory control centers important for chemoreception, as well as in hypoglossal neurons. Research in animals suggests that serotoninergic neurotransmission, through peripheral actions at the level of the carotid body or hypoglossal nerve, or centrally, at medullary respiratory control centers, influences a wide range of functions relevant to OSA, including upper airway reflexes, ventilation, and arousal, as well as sleep-wake cycling.[60] Although the pharmacology is complex, with at least 14 receptor subtypes, this pathway has been implicated in the pathogenesis of SIDS, which, as discussed earlier, might share common genetically determined risk factors with OSA.

Polymorphisms in three genes—*5-HTT* (5-hydroxytryptamine transporter; encoding a serotonin transporter protein that clears serotonin from the synaptic space), *HTR2A* (encoding the 5-HT_{2A} receptor), and *HTR2C* (encoding the 5-HT_{2C} receptor)—each have been studied in relationship to OSA.[86-88] Several meta-analyses have been conducted that have pooled data on variants in these genes for approximately 500 to 700 cases and controls, each from three to six studies conducted in Japan, China, Turkey, and Brazil.[89-91] These analyses indicate than an approximately twofold increased OSA risk is associated with a variant in *5-HT2A* (an allele of 5-HT2A 148G/A) and 20% and 200% increased risks are associated with variants in the *5-HTT* gene (intron-2 variable numbers of tandem repeats and *5-HTT* gene–linked polymorphic region, respectively). Although requiring further replication, these findings are noteworthy given that the HTR2A receptor appears to be the predominant excitatory receptor subtype at the hypoglossal motor neuron and thus is a strong biologic candidate for an association with OSA.

Leptin Signaling

Animal and human studies suggest that leptin, an adipose-derived circulating hormone that influences appetite regulation and energy expenditure, not only influences body weight but also has important effects on central ventilatory drive mediated by brainstem receptors in the nucleus tractus solitarius and hypoglossal motor nucleus.[92,93] Mice homozygous for a knockout mutation in leptin hypoventilate and have a blunted ventilatory response to hypercapnia. Leptin replacement improves the ventilatory responses to hypercapnia in both wakefulness and sleep in leptin-deficient mice.[94] In obese women, increases in circulating levels of leptin have been shown to correlate with the magnitude of compensatory upper

airway neuromuscular responses to experimental airway occlusion.[59] The stimulatory effects of leptin on hypercapnic ventilatory response appear to be mediated through melanocortin, which is produced from a precursor polyprotein, POMC. As described earlier, the Cleveland Family Study reported suggestive evidence for linkage to an area on chromosome 2p that houses the POMC locus,[77] an area also reported by others to be strongly linked to serum leptin levels.[95] An association of OSA with the leptin receptor LEPR also has been reported in candidate gene studies.[85,91] Thus hypothalamic and pituitary pathways involved in leptin signaling may influence OSA susceptibility.

Inflammatory Pathways

Genes in inflammatory pathways may contribute to OSA by influencing upper airway patency through effects on pharyngeal edema, tonsillar hypertrophy, and pharyngeal neuropathic changes[96] or through effects of adipokines such as leptin that influence central respiratory drive. Given the reported associations between tumor necrosis factor-α (TNF-α) levels and OSA severity and sleepiness, a functional polymorphism in the TNF-α gene has been examined in several case-control studies of OSA[97,98] (including one study that compared genetic variants in affected and unaffected sibling pairs[98]) that have reported an elevated risk for OSA in association with a variant associated with higher TNF-α levels. A more recent population-control study by some of the same investigators did not report a significant difference between controls and subjects with OSA in the frequency of that single nucleotide polymorphism of TNF-α or other alleles of TNF-α.[98a] Other studies have reported associations with variants in genes for interleukin-6 and CRP.[85,99] Variants in the nitric oxide synthase (*NOS*) and endothelin (*EDN*) pathways have been reported to be elevated in children with OSA compared with snoring controls.[100] These genes have been implicated in cardiovascular disease, which is common in OSA. These findings suggest that either there are common genetic mechanisms which predispose to both OSA and cardiovascular disease or that individuals with OSA who harbor these variants may be at increased risk for cardiovascular disease.

Apolipoprotein E. An allele of the apolipoprotein E ε4 gene (*APOE*) gene associated with increased risk for both cardiovascular disease and Alzheimer disease was reported to be associated with OSA in two cohort studies of predominantly white subjects.[101,102] Two other studies, however, did not replicate this finding.[103,104] The Cleveland Family Study reported evidence for linkage to AHI near the *APOE* locus on chromosome 19.[86] However, the *APOE* genotype did not explain the linkage findings and was not associated with OSA status. These findings suggested that the susceptibility locus for OSA is not to *APOE* but another locus close to it. A candidate gene in this area is hypoxia-inducible factor 3, which plays a role in oxygen sensing.

APOE ε4 has also been examined as a disease-modifying risk factor. Several studies have shown that individuals with moderate to severe OSA who carry one or more *APOE ε4* variants have greater cognitive impairment than individuals with OSA without such a risk allele.[87,88] It has been hypothesized that the *APOE ε4* allele increases the likelihood of brain injury to oxidative or other stresses.

Angiotensin II Converting Enzyme. Angiotensin II, an important vasoconstrictor, also appears to modulate afferent activity from the carotid body chemoreceptor and thus might influence ventilatory drive.[105] Angiotensin II levels are regulated by the actions of ACE, which is encoded by the *ACE* gene. Several studies of Chinese cohorts have reported an association between polymorphisms in the *ACE* gene and OSA, particularly in persons with hypertension.[106-110] Data from the Wisconsin Sleep Cohort and the Cleveland Family Study did not show an association between ACE genotype and OSA but did show an association between hypertension and OSA severity, which varied in strength by ACE genotype.[111,112]

Linkage Analysis

Linkage analysis quantifies the cosegregation of a disease locus and a marker locus among family members. Typically, the strength of genetic associations is expressed as an LOD score (the log-odds quantifying the probability of receiving alleles at two loci). A LOD score of 3 or more is considered strong evidence for linkage. By identifying alleles that cosegregate in related individuals, areas of the genome are identified that have an increased probability of harboring risk alleles for a given trait. Although linkage analysis has limited resolution to identify specific genetic variants, linkage signals can help prioritize areas of the genome likely to harbor risk variants, and this information can then be integrated into tests of genetic association to increase the statistical power for discovering genetic variants. Linkage analysis can also be used to identify families likely to carry risk alleles, particularly rare mutations that may have large effects. A whole-genome screen for OSA-related traits has been performed in the Cleveland Family Study.[77,78] In one set of analyses including 1275 members of 237 families, linkage analysis was used to identify genetic regions that were uniquely associated with OSA (modeling the AHI) and other regions that associated with AHI through genetic associations with BMI.[113] Several areas of significant linkage to AHI were identified that were not associated with coincident linkage for BMI. Notably, significant linkage was observed on chromosome 6 for the BMI-adjusted AHI level (LOD score of 3.5). A linkage peak on chromosome 13 near the serotonin 2a receptor was observed in African Americans for both AHI and BMI, providing supportive evidence that variants in *HTR2A* may influence both OSA and obesity.

Genome-Wide Association Analyses

Marked advances in technology permit dense mapping of genetic markers across the genome, with some assays providing coverage of more than 1 million genetic variants (SNPs), providing the opportunity to discover genetic variants for a trait without prior knowledge of candidate genes. Such whole-genome scans can be applied to family members and analyzed with linkage analysis. The first study to report a broad analysis of genetic variants for OSA used an assay ("chip") that contained 45,237 SNPs from more than 2000 genes selected to be relevant to heart, lung, blood, and sleep phenotypes.[114] In 3551 participants from three cohort studies, significant associations were identified for several novel loci with OSA. Evidence of replication in independent cohorts was found for a variant in the lysophosphatidic acid receptor (*LPAR1*), a gene expressed in the embryonic cortex with proinflammatory

effects. Craniofacial abnormalities have been observed in an *LPAR1* knockout mouse. Another replicated association was in the prostaglandin E_2 receptor (*PTGER2*), which also is expressed in neuronal tissues and previously was associated with hypertension. Preliminary findings from more recent genome-wide studies of more than 20,000 individuals studied with assays of more than 500,000 SNPs have been reported. It is expected that, as such findings are replicated, genetic loci and their corresponding pathophysiologic pathways, relevant in the pathogenesis of OSA, will be identified.

CLINICAL PEARL

A positive family history of OSA (or of related symptoms) is useful in identifying patients at increased risk for the disorder. Craniofacial abnormalities and obesity can each have a genetic basis and are risk factors for OSA. Clinicians should ask about OSA symptoms in family members, including offspring. Individuals from families with more than one affected member may harbor genetic variants for OSA and may benefit from close follow-up after interventions to ensure their OSA is adequately treated.

SUMMARY

Despite the challenges in studying an inherently complex trait, there is strong evidence from clinical and epidemiologic studies supporting the importance of familial, and specifically genetic, factors in influencing OSA susceptibility. The largest pedigree and twin studies consistently estimate heritability for the AHI to be between 35% and 40%, with recurrent risk factors of approximately 2. Although obesity is the strongest risk factor for OSA and has a clear genetic basis, causal modeling suggests that only 35% of the genetic variance in the AHI of persons with OSA is shared with pathways that determine body weight. Thus most of the genetic variance for the AHI is likely due to the influence of genes that influence other pathways, including those that influence craniofacial structure, ventilatory control, and possibly sleep-wake patterns. Molecular studies of OSA still lag behind those of other chronic diseases. However, data from candidate gene studies, linkage analyses, and emerging genome-wide association analyses implicate variants in genes in the serotonin and leptin pathways, as well as genes in novel inflammatory and development pathways, relevant to the pathogenesis and possible treatment of OSA. Further investigations of the genetic etiology of OSA should provide a means of better understanding its pathogenesis, with the goal of improving preventive strategies, diagnostic tools, and therapies.

Selected Readings

Bielicki P, MacLeod AK, Douglas NJ, Riha RL. Cytokine gene polymorphisms in obstructive sleep apnoea/hypopnoea syndrome. *Sleep Med* 2015;**16**(6):792–5.

Carmelli D, Bliwise DL, Swan GE, Reed T. Genetic factors in self-reported snoring and excessive daytime sleepiness: a twin study. *Am J Respir Crit Care Med* 2001;**164**:949–52.

Eckert DJ, White DP, Jordan AS, et al. Defining phenotypic causes of obstructive sleep apnea: identification of novel therapeutic targets. *Am J Respir Crit Care Med* 2013;**188**:996–1004.

Gislason T, Johannsson JH, Haraldsson A, et al. Familial predisposition and cosegregation analysis of adult obstructive sleep apnea and the sudden infant death syndrome. *Am J Respir Crit Care Med* 2002;**166**:833–8.

Larkin EK, Patel SR, Elston RC, et al. Using linkage analysis to identify quantitative trait loci for sleep apnea in relationship to body mass index. *Ann Hum Genet* 2008;**72**:762–73.

Larkin EK, Patel SR, Goodloe RJ, et al. A Candidate gene study of obstructive sleep apnea in European Americans and African Americans. *Am J Respir Crit Care Med* 2010;**182**:947–53.

Patel SR, Goodloe R, De G, et al. Association of genetic loci with sleep apnea in European Americans and African-Americans: the Candidate Gene Association Resource (Care). *PLoS ONE* 2012;**7**:e48836.

Patel SR, Larkin EK, Redline S. Shared genetic basis for obstructive sleep apnea and adiposity measures. *Int J Obes (Lond)* 2008;**32**(5):795–800.

Qin B, Sun Z, Liang Y, et al. The Association of 5-Ht2a, 5-Htt, and Lepr polymorphisms with obstructive sleep apnea syndrome: a systematic review and meta-analysis. *PLoS ONE* 2014;**9**:e95856.

Redline S, Tishler PV, Tosteson TD, et al. The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med* 1995;**151**:682–7.

Schwab RJ, Pasirstein M, Kaplan L, et al. Family aggregation of upper airway soft tissue structures in normal subjects and patients with sleep apnea. *Am J Respir Crit Care Med* 2006;**173**:453–63.

Yalcmkaya M, Erbek SS, Babakurban ST, et al. Lack of association of matrix metalloproteinase-9 promoter gene polymorphism in obstructive sleep apnea syndrome. *J Cranio-Maxillo-Facial Surgery* 2015;**43**:1099–103.

A complete reference list can be found online at ExpertConsult.com.

Chapter
114

Obstructive Sleep Apnea: Clinical Features, Evaluation, and Principles of Management

Harly Greenberg; Viera Lakticova; Steven M. Scharf

Chapter Highlights

- Obstructive sleep apnea (OSA) is the most common respiratory disorder of sleep, with a high prevalence that is linked to the increase in obesity.
- OSA is frequently comorbid with cardiovascular, cerebrovascular, and metabolic diseases and is commonly observed in populations with these comorbidities. The relationship of OSA with these multisystem disorders may be bidirectional.
- The pathogenesis of OSA is complex, with contributions from mechanical factors that increase collapsibility of the upper airway as well as factors that lead to instability of ventilatory control during sleep.
- Clinical assessment, in addition to various screening questionnaires, is useful to identify patients at risk for OSA. However, accurate diagnosis requires monitoring of sleep.
- In-laboratory polysomnographic testing for OSA, as well as various technologies for ambulatory or out-of-center-sleep testing, are presented. The utility and limitations of these techniques are discussed.
- Various treatment modalities for OSA are described. An individualized treatment approach that emphasizes chronic disease management that improves sleep-related health outcomes is necessary to optimize care.

DEFINITION

Obstructive sleep apnea (OSA) is a common disorder that is recognized as a major risk factor for a number of important chronic medical conditions and is responsible for poor quality of life. Ample evidence exists that patients with untreated OSA consume more health care resources than matched patients without OSA, leading to considerably increased health care use costs.[1,2] Most studies agree that treatment reduces health care use to that of matched controls. Further, estimates are that OSA remains underdiagnosed.[3]

The classic signs and symptoms of OSA include excessive daytime sleepiness (EDS), loud snoring, snorting, and gasping at night (associated with apnea termination). Physical signs commonly associated with OSA include obesity, large neck circumference, and crowding of the oropharynx (Table 114-1).

BRIEF HISTORY

Burwell and colleagues are often given credit for the first medical description of a patient with probable OSA.[4] These authors described an obese, sleepy, hypercapnic patient, demonstrating periodic breathing, who reminded them of the character Joe in the Charles Dickens 1836 novel *The Posthumous Papers of the Pickwick Club* (see also Chapter 120). However, others used this term previously to describe similar patients.[5] Although the prevailing thought was that hypoventilation, possibly owing to excess weight, contributed to somnolence, Kuhl and associates described breathing cessations at night during polysomnography (PSG) and attributed sleepiness to the resultant sleep fragmentation.[6] As a result of these findings, the world renowned neurologist Gastaut added measurements of airflow and chest wall motion to other PSG measures to document obstruction of the upper airway (UA) at night.[7,8] The Bologna group of Lugaresi and Coccagna demonstrated large swings in arterial and pulmonary pressures during apneas, thus documenting that nocturnal breathing disorders had major adverse consequences and required treatment.[9] Following the observation of Kuhl and associates, the Bologna group reported that tracheostomy improved symptoms in a group of "pickwickian" patients.[10,11] In 1981 Sullivan and colleagues published their seminal paper demonstrating that nasally applied continuous positive airway pressure (CPAP) could alleviate UA obstruction in OSA.[12] Shortly thereafter, Rapoport and associates demonstrated that the "pickwickian syndrome" could be reversed with long-term use of nocturnal CPAP.[13] Additional therapies for the disorder, now termed OSA, including UA surgeries, mandibular advancement, and even stimulation of the hypoglossal nerve, have evolved and allow the physician to offer a variety of therapies tailored to the individual patient.

PHYSIOLOGIC EFFECTS

It is now well established that OSA has a number of acute physiologic effects that are thought to contribute to adverse multisystem consequences. These include intermittent hypoxia

Table 114-1 Signs and Symptoms of Obstructive Sleep Apnea

Severe snoring, snoring, gasping, or choking in sleep
Witnessed apneas in sleep
Excessive daytime sleepiness; tendency to fall asleep in inappropriate situations (e.g., while driving, attending lectures)
Lack of energy
Morning headaches
Large neck size: 17 inches in men, 16 inches in women
Crowding of the oropharynx: Mallampati score of 3 or greater, large tonsils, large tongue, elongated uvula
Facial abnormalities: retrognathia, midface deformities
Obesity (body mass index >30)
Nocturnal gastroesophageal reflux
Impotence; erectile dysfunction

Note: Male gender and postmenopausal state in women confers increased risk.

Table 114-2 Medical and Mental Health Conditions for Which There Is Evidence of Association with Obstructive Sleep Apnea*

Hypertension[155] (25% to 50% all hypertension; as high as 83% in drug-resistant hypertension)
Myocardial infarction[156] (as high as 70%)
Stroke[157] (as high as 68%)
Depression[158]
Congestive heart failure[159] (as high as 76%)
Asthma[51]
Chronic obstructive pulmonary disease[48] (as high as 50%)
Atrial fibrillation (and other dysrhythmia)[38] (49%)
Type 2 diabetes[46] (as high as 48%)
Traffic and industrial accidents[83] (as high as fivefold increase)
Overall mortality[160] (increased risk by 46%)

*Approximate risks listed where available.

(IH) occurring as a result of apneas and hypopneas, exaggerated negative swings in intrathoracic pressure (ITP), and terminal arousals. In animal models, IH leads to increased sympathoadrenal tone, a feature well demonstrated in humans with OSA.[14,15] Further, IH leads to oxidant stress in the brain[16] and in the myocardium, a finding associated with poor left ventricular function, apoptosis of myocardial cells,[17,18] and endothelial dysfunction.[19] Further, arousals associated with termination of apneas, hypopneas, and periods of inspiratory flow limitation contribute to heightened sympathetic tone.[20] In animal models, IH has been shown to be associated with release of pro-inflammatory mediators, at least partially mediated by NF-κB–related pathways.[21,22] Exaggerated swings in ITP, primarily during inspiration against an occluded airway, lead to increased venous return and stress on the right ventricle.[23] The latter appears to be primarily responsible for pulmonary hypertension, a common finding in OSA.[24] Sleepiness is thought to be, at least in part, a result of the terminal arousals and associated sleep fragmentation.

OSA is an independent major risk factor for a number of associated medical conditions (Table 114-2). Chief among these are cardiovascular disease, including hypertension, stroke, myocardial infarction, and congestive heart failure. Heightened sympathoadrenal tone, oxidant stress, and pro-inflammatory cytokines appear to be involved in the pathogenesis of these conditions.

EPIDEMIOLOGY

There are varying estimates of the prevalence of OSA, largely owing to differences in diagnostic methods, definitions of disease, and differences in age, gender, and body mass index (BMI). As reviewed by Young and colleagues in 2002, prevalence estimates range from 2% to 26% depending on gender, definition of "disease," and population studied.[25] These prevalence rates were mostly derived from epidemiologic studies performed in the 1990s. Data from the Wisconsin Sleep Cohort were used to derive the estimate that symptomatic OSA affected 2% to 4% of middle-aged adults in the United States.[26] However, the increase in prevalence of obesity necessitated revision of these initial estimates. Using data from the National Health and Nutrition Examination Survey on BMI in U.S. populations, as well as data from the Wisconsin Sleep Cohort, Peppard and associates now estimate that among adults 30 to 70 years of age, approximately 13% of men and 6% of women have an apnea-hypopnea index (AHI) of more than 15/hour, whereas 14% of men and 5% of women have an AHI of more than 5/hour with symptoms of daytime sleepiness.[27] Both definitions meet *International Classification of Sleep Disorders*, third edition (ICSD3) criteria for OSA.[28] In the U.S. working age population, males have a twofold to threefold greater prevalence of OSA than females. A meta-analysis of community and sleep center referral–based cohorts showed that male gender is more common among patients with diagnosed OSA (odds ratio, 3.1; 95% confidence interval [CI], 2.5 to 3.8). However, among women, postmenopausal status is associated with an increased risk for OSA that equals that observed in men.[29,30] The prevalence of OSA in the United States also increases with advancing age, from about 35 to 60 years, after which there is less of an age-related increase.[31]

OSA prevalence has also been assessed in many regions of the world other than the United States and among various ethnic groups. Estimates of OSA prevalence in white European and Australian populations are similar to those observed in North America. Further, comparable OSA prevalence rates have been found in studies of Korean, Chinese, and Indian populations. Most studies have also shown similar OSA prevalence between North American white, African American, and Hispanic cohorts.[32]

PREVALENCE IN DISEASE-SPECIFIC COHORTS

Because OSA is associated with cardiac, cerebrovascular, pulmonary, metabolic, and other comorbid diseases, it is worthwhile to consider the prevalence of OSA in relevant disease specific cohorts. The pathophysiologic features of OSA, including IH, increases in sympathoadrenal tone, large swings

in ITP, and sleep fragmentation, with associated increases in oxidative stress, systemic inflammation, and endothelial dysfunction, among other factors, may contribute to many of these associated comorbid conditions. It is also worthwhile to consider that the relationship of OSA with some of these diseases may be bidirectional.

One of the most studied cardiovascular comorbidities of OSA is systemic hypertension. Most observational studies have shown that up to 50% of patients with systemic hypertension have OSA.[33] Further, OSA is a common cause of "resistant" hypertension, with a prevalence of 64% in one cohort of patients with difficult-to-control hypertension.[34] Atrial fibrillation (AF) is another cardiac disorder associated with OSA. Data from the Sleep Heart Health Study (SHHS) demonstrated a higher prevalence of AF in subjects with OSA than in those without such sleep-disordered breathing (4.8% vs. 0.9%; $P = .003$).[35] Conversely, a high prevalence of OSA (32% to 49%) has been demonstrated in various cohorts of patients with AF.[36] The association between these disorders is further documented by data that demonstrate an increasing prevalence of AF with increasing severity of OSA as assessed by the nocturnal oxygen desaturation index.[37] Other studies have shown that OSA is associated with development of AF after cardiac surgery and with an increased risk for recurrence of AF after cardioversion or ablation therapy.[38] OSA is also common in congestive cardiomyopathy and may adversely affect outcomes in this disease. Eleven percent of a cohort of patients with cardiomyopathy (defined as left ventricular ejection fraction <45%) was found to have OSA, although central sleep apnea (CSA) was more frequently observed in this group.[39] In another cohort of systolic heart failure patients, 61% were found to have a sleep related breathing disorder; approximately half had OSA, whereas CSA was present in the remainder.[40] OSA is also an independent risk factor for cerebrovascular disease with a threefold to fourfold increased odds of incident stroke in moderate to severe OSA (AHI >20/hour).[41,42] In accord with this finding, the prevalence of OSA in post–cerebrovascular accident cohorts is notably high, ranging from 38% to 74%.[43,44] The relationship of these cardiovascular and cerebrovascular disorders with OSA may be bidirectional. The adverse consequences of OSA may contribute to the development of these conditions; conversely, cardiac and cerebral dysfunction may contribute to OSA by promoting pathophysiologic factors that promote apneas and hypopneas during sleep. The association of OSA with type 2 diabetes and the metabolic syndrome is also well described. Data from the SHHS, for example, demonstrated that a remarkable 58% of subjects with type 2 diabetes had an elevated AHI.[45] A study of obese adults with type 2 diabetes showed that 87 % of that cohort had OSA, with a mean AHI in the moderate range.[46]

Although some studies have suggested that the occurrence of OSA in chronic obstructive pulmonary disease (COPD) patients is not greater than expected based on the prevalence of each disease in the population,[47] others have shown an increased prevalence of COPD among OSA patients compared with matched controls.[48] A case-control study of 1497 patients with PSG-proven OSA compared with 1489 age- and gender-matched controls showed that COPD was more prevalent in the OSA group (7.6 vs. 3.7 %; $P < .0001$).[49] OSA was also found to be highly prevalent among a cohort of patients moderate to severe COPD referred for pulmonary rehabilitation.[50] Regardless of whether the concomitant prevalence of COPD and OSA is greater than can be expected by chance occurrence, the coexistence of these disorders, termed the *overlap syndrome*, is associated with more severe nocturnal hypoxemia, hypoventilation, pulmonary hypertension, comorbid obesity, and diabetes than that which is observed in isolated OSA or COPD (see Chapter 119).

Recent studies have also suggested an association between OSA and asthma.[51] Although most data regarding a possible link between these two diseases comes from cross-sectional studies performed in asthma clinic cohorts, results consistently demonstrate that the prevalence of OSA is approximately double in these cohorts compared with the general population. Asthma severity, BMI, gastroesophageal reflux, and female gender have all been associated with increased OSA risk in asthma. An analysis of data from the Wisconsin Sleep Cohort showed that a diagnosis of asthma was associated with increased risk for incident OSA.[52] Whether these disorders are mechanistically linked and whether OSA treatment alters asthma outcomes remains to be investigated.

OSA is also frequently observed in patients with chronic kidney disease, with prevalence rates reported to be as high as 50% in patients with end-stage renal disease (ESRD).[53] An increasing prevalence of OSA has been associated with declining kidney function, ranging from 41% in patients with chronic kidney disease to 48% in those with ESRD.[54] The relationship between renal dysfunction and OSA may also be bidirectional. OSA might contribute to renal dysfunction by exacerbating hypertension, diabetes, endothelial dysfunction, sympathetic neural activity, systemic inflammation, and oxidative stress.[55,55a] Conversely, renal dysfunction may contribute to OSA, possibly owing to fluid overload that may increase UA edema as well as other factors. Support for the possibility of such a bidirectional relationship comes from data that demonstrate improvement in the AHI with intensive nocturnal hemodialysis in ESRD.[56] In addition, removal of fluid by ultrafiltration, without affecting uremia, was shown to improve the AHI in association with a decline in fluid volume of the neck.[57]

It is clear from these studies that OSA is frequently comorbid with several important and common diseases. Whether treatment of OSA alters disease-specific outcomes when coexistent with other major medical disorders remains a matter of ongoing investigation for most of these conditions.

RISK FACTORS

Anatomy

OSA is associated with anatomic risk factors that narrow the UA. As mentioned previously, the most widely recognized factor is central obesity, with a direct relationship observed between BMI and apnea severity. OSA is attributable to obesity in up to 58% of subjects.[58] Further, weight loss typically results in improvement in OSA. Linear regression modeling from the Wisconsin Sleep Cohort showed that in individuals with OSA, after adjustment for sex, age, and cigarette smoking, an approximate 1% increase or decrease in body weight was associated with a corresponding 3% increase or decrease in the AHI.[59] Obesity can contribute to airway narrowing by depositing adipose tissue around collapsible segments of the UA, increasing the size of the parapharyngeal fat pads and increasing fat content and volume of the base of

the tongue.[60,61] Obesity may also indirectly contribute to UA collapsibility by reducing lung volume. Lower lung volume is associated with reduced tracheal caudal traction on the UA, which decreases stiffness of the lateral pharyngeal walls and promotes airway collapse.[62]

In addition to deposition of adipose tissue, overall increases in UA soft tissue volume contribute to pharyngeal narrowing and predispose the UA to collapse during sleep. Numerous imaging studies have demonstrated that the UA lumen is narrower in patients with OSA compared with control subjects. Such narrowing is largely due to an increase in volume of the surrounding soft tissues. Although the cross-sectional dimension of the OSA airway is smaller at several levels, the greatest difference compared with controls is typically in the retropalatal or velopharyngeal region.[61] Magnetic resonance imaging volumetric analyses have demonstrated increased volume of the lateral pharyngeal walls, soft palate, tongue, and parapharyngeal fat pads that compromise the airway lumen. In addition, differences in the shape and length of the UA are evident. Although the UA is normally largest in its lateral dimension, its anterior-posterior dimension is greatest in OSA; this decreases efficiency of UA dilator muscles. Further, UA length is increased in OSA, which contributes to its collapsibility.[61] In addition to soft tissues, decreased size of the maxilla, with a narrowed and high palatal arch and a small retropositioned mandible can also narrow the UA.[62,63]

Other factors that contribute to a narrow airway lumen in OSA include edema and inflammation of UA soft tissue. Histologic studies of resected palatal tissue showed edema and lymphocytic infiltration of mucosal as well as muscular layers in OSA, possibly owing to vibratory trauma from snoring.[64] In addition, rostral shift of blood volume and edema fluid from the lower extremities when subjects with lower extremity edema assume a recumbent position during sleep has been associated with increased neck circumference, UA resistance, and increased AHI.[65]

Relationship of Upper Airway Anatomic Factors to Development of Inspiratory Flow Limitation and Obstruction

The UA can be modeled as a collapsible tube through which air flows (Figure 114-1). The propensity of the airway to collapse is determined by the elastic properties of the airway structure itself as well as activity of the UA dilator muscles. During inspiration, pressure in the pharyngeal lumen is negative relative to atmospheric pressure; otherwise air could not flow in the inspiratory direction. With increased inspiratory effort or narrowing of the pharyngeal lumen, pressure becomes more negative within the UA. This can lead to obstruction in collapsible segments of the UA if airway luminal pressure decreases to a value below what is termed the critical closing pressure (Pcrit) and the UA dilator muscles do not respond sufficiently. At this point, inspiratory flow ceases and an *obstructive apnea* occurs. Intraluminal pressure downstream (toward the thorax) of the closed segment is equal to intrathoracic pressure during no-flow conditions. Because the

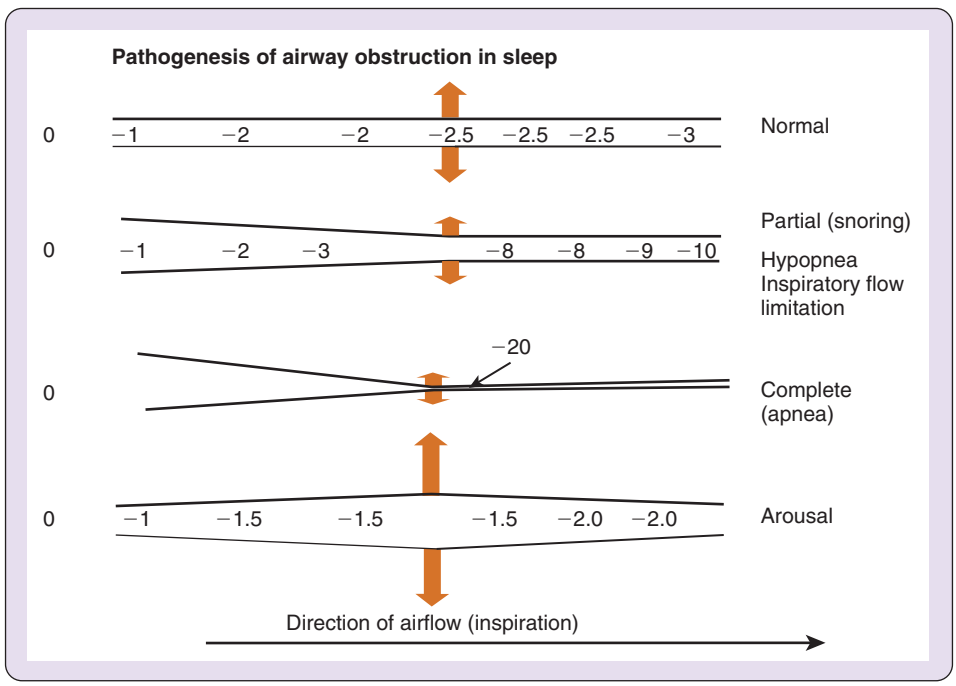


Figure 114-1 Pathogenesis of upper airway closure in OSA. The upper airway (UA) is depicted as a tube. With inspiration, there is a small gradient of pressure in the direction of flow (*upper panel*). Even though pressure is slightly negative, the airway is held open by UA dilator muscles (*orange arrow*, force represented by length of the arrow). With narrowing of the UA and some decrease in abductor force, the gradient of airway pressure is greater; flow may become limited. Vibrations in the airway produce snoring. With complete closure of the UA, no air can flow. Pressure in the airway is negative and will be equal to intrathoracic pressure (no flow condition). Closure of the UA occurs because of decreased activity of UA dilator muscles. When brainstem and other appropriate receptors sense no airflow, with hypoxia and hypercapnia, UA dilators are activated and open the airway (*bottom panel*).

patient typically generates increased inspiratory efforts against the occluded UA, exaggerated negative swings in intrathoracic pressure occur that may be quite large. If UA collapse is partial, further decreases in intraluminal pressure may not be able to overcome the increased resistance and do not result in increases in inspiratory flow, resulting in a condition of *hypopnea* or *inspiratory flow limitation* (see Chapter 112). After complete airway closure or flow limitation occurs, inspiratory flow can only be restored if UA dilator muscles respond by increasing luminal diameter, thereby decreasing resistance and restoring flow.[66]

Pcrit of the passive oropharynx varies widely among individuals. The mean value for Pcrit in a group of normal subjects has been found to be −4.35 ± 4.15 cm H_2O, whereas it was near atmospheric pressure for subjects with mild to moderate OSA (0.56 ± 1.54 cm H_2O) and above atmospheric pressure for those with severe OSA (2.23 ± 2.96 cm H_2O), although there was considerable overlap among groups.[66] For subjects with a Pcrit at or above atmospheric pressure, pharyngeal dilator muscle activity is required to maintain airway patency during wakefulness and sleep.

If anatomic and mechanical characteristics of the UA were the sole factors responsible for the occurrence of obstructive apneas and hypopneas, a strong linear relationship should exist between Pcrit and the severity of OSA as measured by AHI. However, the correlation between Pcrit and AHI is moderate at best. This indicates that factors other than mechanical or structural properties of the UA contribute to the occurrence and severity of OSA.[66,67] Primary among these nonanatomic factors is unstable or inadequate ventilatory drive during sleep to the UA dilator and ventilatory pump muscles, which can contribute to the development and perpetuation of apneas.

CLINICAL IDENTIFICATION AND ASSESSMENT

Daytime Symptoms and Functional Consequences

EDS, which is associated with persistent somnolence that may cause inappropriate or unintentional sleep episodes, is one of the most frequent symptoms of OSA, and one that adversely affects daytime function and quality of life. The presence and impact of EDS may be subtle, such as drowsiness occurring during periods of relative inactivity, or it may be more severe, with episodes of falling asleep during activities such as driving. EDS is considered to be a consequence of sleep fragmentation and has been associated with loss of vigilance. However, EDS is not universally present in all patients with OSA.[68] Further, the correlation between severity of EDS and the AHI is relatively weak.[68,69] Other factors such as nocturnal IH, autonomic dysregulation, and OSA related comorbidities such as obesity, cardiovascular disease, diabetes, and depression may also contribute to EDS.[70-73] It is worth emphasizing that the absence of a complaint of EDS does not reliably discriminate between patients with and without OSA.[74] Sleepiness may not be directly recognized by many patients who instead perceive their symptoms as fatigue. In clinical practice, EDS is often subjectively quantified using the Epworth Sleepiness Scale.[75]

EDS increases the risk for cognitive dysfunction, poor performance, injury, and motor vehicle accidents (MVAs). Despite objective evidence of EDS, some patients might not recognize impairment of performance, including driving ability, because of sleepiness.[76] It is worth noting that the term *excessive daytime sleepiness* (EDS) is somewhat of a misnomer because excessive somnolence can be very problematic in night-shift workers, with similar adverse consequences as somnolence during the day.

OSA is also associated with reduced quality of life as measured by general and disease specific QOL scales. The Functional Outcomes of Sleep Questionnaire is a QOL assessment tool sensitive to the impact of sleep disorders and excessive sleepiness and is often used to measure the impact of OSA and its treatment.[77] Objective evidence of sleepiness, including reduced sleep latency on the Maintenance of Wakefulness Test or the Multiple Sleep Latency Test, may also be demonstrated, especially in patients with moderate to severe OSA.[78]

Deficits in Cognition, Vigilance, and Executive Function

Impairment of cognitive function may occur in OSA, possibly owing to cortical arousals, sleep fragmentation, excessive somnolence, and nocturnal IH. Deficits have also been observed in sustained attention, or vigilance, which is important for prolonged complex tasks such as driving a motor vehicle.[79-81] Further, monitoring of information, reaction time, distractibility, and processing capacity are impaired in OSA.[82] These deficits may, in part, be responsible for the twofold to sevenfold increased risk for MVAs observed in OSA.[83] Crashes, or near misses, may be caused by sleepiness, inattention, fatigue, or micro-sleep episodes that result in failure to respond rapidly and appropriately.[84,85] Some studies have found a dose-response relationship between the severity of OSA and MVA risk. However, prediction of driving risk in individual patients is not precise.[85] Deficits in cognitive function, vigilance, somnolence, and other sequelae of OSA may also contribute to poor performance in the workplace (absenteeism, presenteeism) and work-related injuries.[86]

Executive function, which encompasses cognitive processes responsible for problem solving, flexibility, decision making, and initiating appropriate responses and actions, may also be impaired in OSA. In addition, deficits in memory have been associated with OSA, although not all studies have demonstrated such an effect. The underlying mechanisms responsible for impairment in executive function and memory may be related to sleepiness and attention deficits as well as to neuronal damage from oxidant stress related to IH in the prefrontal cortex and hippocampus.[87] Improvements in memory, attention, and executive function observed after CPAP therapy for OSA have been correlated with increases in gray matter volume in these regions.[87]

Mood Disorders

A higher than expected prevalence of mood disorders has been observed in OSA. A Veterans Administration study showed that 21.8% of OSA patients had major depression, whereas 16.7% had anxiety disorder.[88] Newly diagnosed OSA patients are twice as likely to develop depression within 1 year compared with controls.[89] Conversely, investigations of patients with major depression demonstrated an increased prevalence of OSA. The association between OSA and depression is strengthened by studies which showed that CPAP therapy results in sustained improvement in depression scores, particularly in patients with moderate to severe OSA.[90,91] However, cross-sectional studies have generally not

demonstrated an association of measures of apnea severity with depressive symptoms.[92]

Sleep-Related Signs and Symptoms

The classic sleep-related sign of OSA, as observed by a bed partner, is loud snoring alternating with periods of silence, associated with paradoxical movement of the chest and abdomen, terminated by a loud gasp or snort. A meta-analysis of community and sleep center cohorts with documented OSA showed that a complaint of gasping or choking during sleep was the most useful clinical predictor of OSA (likelihood ratio, 3.3; 95% CI, 2.1 to 4.6).[74] Snoring is also very common, with a prevalence of 35% in a population-based survey of persons 30 to 70 years old in Spain; importantly, an isolated complaint of snoring is not a useful predictor of OSA.[93] In contrast, the absence of snoring makes OSA less likely. Other common symptoms of OSA include awakening with a dry mouth, which may reflect mouth breathing, restless sleep, and nocturnal diaphoresis.

Nocturia may also be seen in OSA.[94] A cross-sectional analysis of the SHHS demonstrated an independent association between the AHI and prevalence of nocturia. Nocturia was also associated with disturbed sleep and with subjective complaints of daytime somnolence.[95] Increased intraabdominal pressure during obstructive apneas, confusion associated with arousals, and increased secretion of atrial natriuretic peptide are proposed mechanisms contributing to nocturia and nocturnal enuresis.[96]

Complaints of nocturnal gastroesophageal reflux (GER) are often reported in OSA. In support of an association between GER and OSA, 24-hour esophageal pH monitoring demonstrated episodes of decreased esophageal pH in 80% of OSA subjects during sleep; CPAP therapy reduced reflux events.[97] The common association of OSA with GER may be a result of large decreases in intrathoracic pressure occurring during obstructed inspiratory efforts, with increases in intraabdominal pressure, that may contribute to GER; alternatively, obesity, which increases risk for hiatal hernia, may be the primary factor leading to this association.[98]

Morning headaches are sometimes reported in OSA. The International Classification of Headache Disorders II describes OSA-related headaches as "bilateral, with a pressing quality, not accompanied by nausea, photophobia or phonophobia."[99] Headache is present on awakening and usually resolves within 30 minutes; morning headaches are eliminated with effective treatment of OSA.[99] A recent study demonstrated that 11.8% of OSA patients had morning headache. However, morning headache without OSA is also common, with a prevalence of 4.6% in this cohort.[100]

Physical Findings

As previously mentioned, obesity is the most commonly observed risk factor for OSA. Suspicion for OSA should be raised if the BMI is more than 30 kg/m^2.[101] Patients with OSA have a larger neck circumference that those without this disorder. The average neck circumference (measured at the superior border of the cricothyroid membrane in the upright position) was 43.7 ± 4.5 cm in a series of patients with OSA and 39.6 ± 4.5 cm in those without OSA (P = .0001).[102] A neck circumference at least 40 cm has a sensitivity of 61% and a specificity of 93% for OSA regardless of gender.[103] Neck circumference, as well as neck circumference corrected for height, is a correlate of increased visceral fat, which is associated with OSA.

Physical examination should assess nasal patency, oropharyngeal anatomy, and craniofacial structure. Increased nasopharyngeal resistance, due to nasal septal deviation, turbinate hypertrophy, polyps, or other obstructing lesions, is associated with OSA.[104] The soft palate, uvula, base of tongue, and tonsils should be observed with attention to their size, length, and overall volume in relation to the oropharynx. A low-lying or redundant soft palate and uvula, often with edema or erythema due to vibratory trauma and inflammation from snoring, are frequently present in OSA. Both the Mallampati classification, which assesses oropharyngeal anatomy with the tongue protruded, and the Friedman classification, which is a similar assessment but without tongue protrusion, are commonly used to stage oropharyngeal crowding.[105] These scoring systems provide a numeric scale that grades the size of, and relationship among, the soft palate and uvula, lateral tonsillar pillars, and base of tongue. In addition, tonsil size should also be assessed. Anatomic crowding of the oropharynx has been shown to be related to the presence and severity of OSA, but this finding in and of itself has limited value for predicting presence of OSA (likelihood ratio range, 1.4 to 1.6).[103]

It is also important to assess craniofacial anatomy. In particular, mandibular retrognathia, or retrusion of the mandible, narrows the posterior air space and can increase the propensity for airway collapse during sleep. A high arched and narrow hard palate may also predispose to OSA. Assessment of dentition may also be useful to identify a narrowed posterior air space. In particular, the presence of *overjet*, defined as displacement of the mandibular teeth posteriorly compared with the maxillary teeth, is indicative of a small oral cavity that may result in posterior displacement of the base of the tongue, which narrows the retroglossal airway.[106]

ASSESSMENT

In-Laboratory or Full Polysomnographic Sleep Testing

The diagnosis of sleep apnea should be confirmed objectively by sleep testing. The "gold standard" test is PSG, a multichannel assessment of physiologic variables performed in a laboratory equipped with proper sensors, trained personnel, and a standardized way of recording results, with a qualified individual to interpret the study. Recorded variables usually include electroencephalogram (EEG), electromyogram (EMG) of the submentalis muscle, electrooculogram (EOG), a measure of airflow (usually sensors by the nose and mouth), a measure of respiratory effort (chest wall and abdominal movement, EMG of parasternal muscles, or changes in esophageal pressure), a measure of oxygen saturation (pulse oximetry), pulse rate, electrocardiogram (ECG), body position, EMG of legs (anterior tibialis muscle, for leg movements), and snoring (usually by microphone). Many laboratories also record continuous video imaging of the patient both for medical-legal reasons and to observe parasomnias. Table 114-3 contains a list of some of the most common variables recorded during in-laboratory PSG testing; a typical PSG epoch with recordings of these parameters is presented in Figure 114-2. More recently, efforts have been made to simplify the use of PSG for the diagnosis of sleep related breathing disorders. These systems measure fewer variables and are

Table 114-3 Commonly Recorded Parameters During In-Laboratory Polysomnography

Parameter Recorded	Purpose	Other
EEG (several channels) EMG of submentalis (other facial as indicated) EOG	Sleep staging	Usually standard 10 to 20 system for EEG—may also be used for seizure detection (EEG) if suitable montage used
Oronasal airflow Tidal swings in CO_2 measured at the mouth Respiratory effort (rib cage and abdominal movement)	Sleep-related abnormal breathing events	Nasal cannula pressure for hypopnea and flow limitation detection Thermistor for apnea detection
Pulse oximetry	O_2 saturation, pulse	Some definitions of DBEs depend on saturation, quantifies oxygenation at night
Microphone	Snoring	Some systems use perturbation in airflow
Body position sensor	Body position	
EMG: submental and pretibial	Sleep onset, REM onset, abnormal limb movements	
ECG	Rate and rhythm abnormalities	Usually precordial only; may detect rate of rhythm disturbances or changes in ST-T segments
CO_2: end-tidal, transcutaneous	Changes in alveolar ventilation	

DBE, Disordered breathing event; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram. Other variables often recorded during in-laboratory sleep studies include audio-videography and actigraphy.

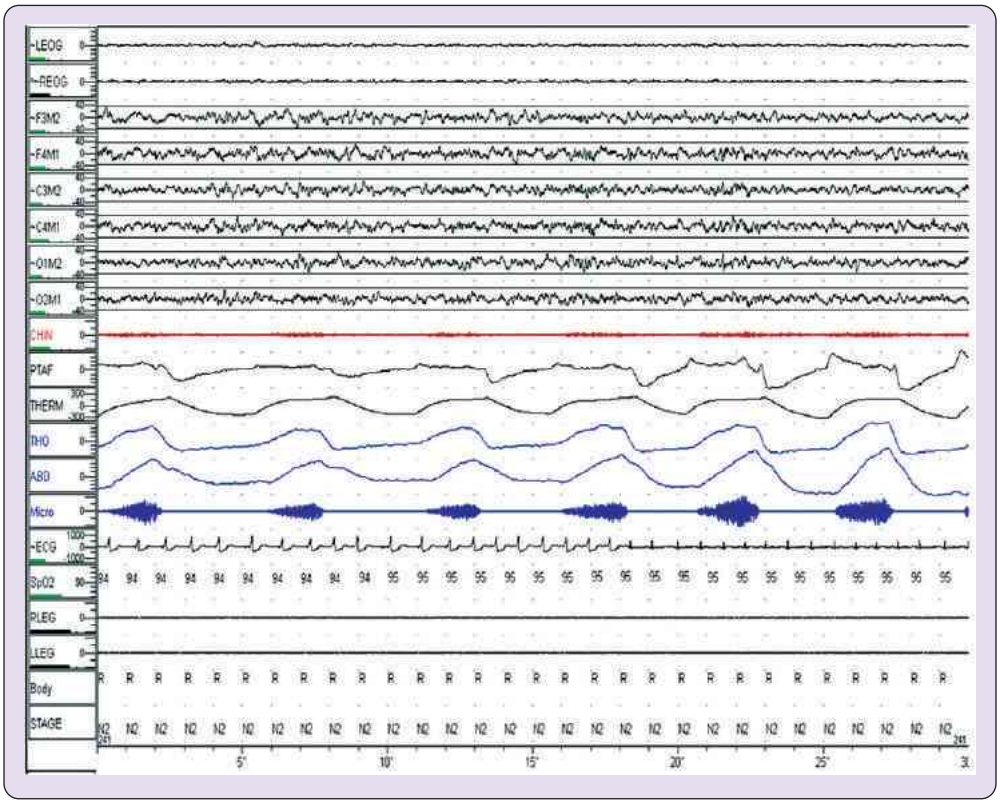


Figure 114-2 Sample montage for polysomnography. LEOG and REOG, Left and right electrooculogram; F3M2–O2M1, 6 electroencephalogram leads; Chin, electromyogram of submentalis muscle; PTAF, nasal airflow by nasal pressure; THERM, oronasal airflow estimated using a thermistor; THO and ABD, rib cage and abdominal movement by respiratory inductance plethysmograph belts; MICRO, snoring detected by a microphone taped to the neck (note the snoring); ECG, electrocardiogram; Spo_2, blood oxygen saturation of hemoglobin by pulse oximetry; RLEG and LLEG, electromyogram of the right and left legs (anterior tibialis); BODY, body position; Stage, stage of sleep (hand scored). There are timing marks at the bottom; this is a 30-second epoch.

suitable for portable or out-of-center sleep testing (OCST) under certain conditions.

Technical specifications and acceptable derivations for in-laboratory PSG testing are specified in the American Academy of Sleep Medicine (AASM) scoring manual,[107] as well as elsewhere in this volume, as are the visual rules for sleep staging, respiratory events, limb movement events, and other "events" of note. We briefly review some of the specifications for recording respiratory "events" during PSG. Airflow is usually recorded using oronasal thermistor and nasal air pressure. Nasal air pressure, measured using small cannulas in the nose, is a more accurate approximation of airflow.[108] Thermistors placed by the nose and mouth do not measure flow but indicate airflow as a change in temperature. Thus they cannot quantify flow and instead record the presence or absence of airflow. In addition, if the thermistor moves toward or away from the nose and mouth, it will record changes in the signal that do not reflect actual flow. Tidal CO_2 is also sometimes used to assess flow. It is possible to get a true quantitative measurement of airflow with a tight-fitting facemask with a pneumotachograph. However, because of discomfort, this technique is not useful on a routine basis. Further, detection of respiratory events called respiratory event–related arousals (RERAs; discussed later; see Chapter 112) depends on detection of inspiratory flow limitation. As can be seen in Figure 114-3, this is not possible with thermistors or any other sensor that simply records the presence or absence of flow (e.g., tidal CO_2 monitoring). Therefore preference should be given to using nasal air pressure recording. Alternatives to measurement of airflow include the summed chest and abdomen signal from respiratory impedance belts, use of the mask pressure signal during positive airway pressure (PAP) titrations, and various indexes of respiration derived from pulse and finger plethysmographic techniques for OCST.

The AASM scoring manual[107] recommends that the primary means for detecting apnea is absence of the oronasal thermistor signal, although alternates are suggested. The manual recommends that the primary method for identifying hypopnea should be nasal pressure, although alternates are also suggested. During PAP titrations, it is recommended to use the change in PAP mask pressure as the primary flow signal. The manual recommends the use of esophageal manometry or respiratory impedance plethysmography thoracoabdominal belts as the primary measure of respiratory effort, although belts that measure changes in resistance are also acceptable. For detection of snoring, microphones, piezoelectric sensors, or perturbations in nasal air pressure are all considered acceptable. The reader is referred to this publication for more detailed analysis.

The severity of sleep apnea is usually defined in terms of frequency of respiratory "events" (see later). This implies that counting discrete occurrences adequately characterizes the severity of disease. In general, a greater frequency of such events per hour of sleep is associated with a more severe clinical syndrome. However, this is not a tight correlation because many patients with a "severe" respiratory event index have minimal symptoms, whereas many with a "mild" index have severe sleepiness. In general, severity of OSA as estimated

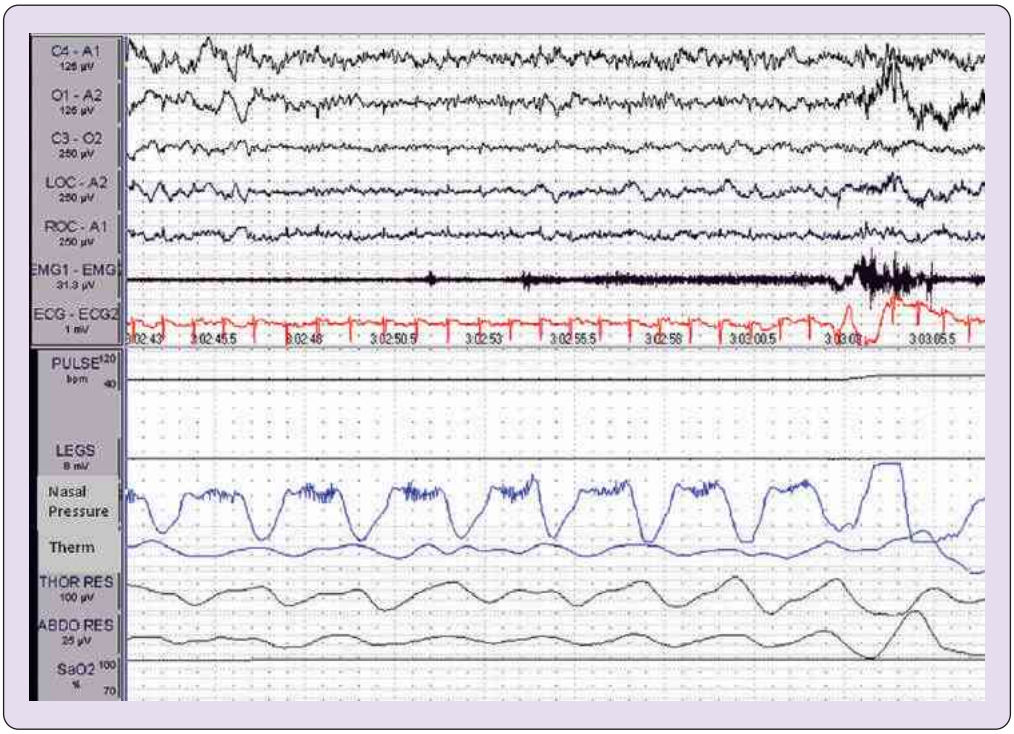


Figure 114-3 Respiratory event–related arousal (RERA). Comparison of airflow measured by oronasal thermistor (Therm) with that measured by nasal pressure (NP). For the nasal pressure signal the inspiratory direction is up, expiratory direction down. Note that with the nasal pressure transducer, inspiratory flow limitation (flattening of the signal and snoring) is readily detected. Note the arousal following this event (electroencephalogram [EEG] and electromyogram [EMG]). The first three channels are EEG, left EOG (LOC), and right EOG (ROC). EMG1-EMG; submentalis EMG; ECG, electrocardiogram; pulse is derived from the pulse oximeter; Legs, pretibial EMG.

Table 114-4 Definitions of Disordered Breathing Events in Sleep[107]

Duration	Event Type	Change in Airflow	Respiratory Effort	Associated Phenomena	Other
10-second duration measured from nadir preceding the first reduced breath to beginning of first breath approximating baseline	Obstructive apnea	Decrease ≥90%	Continues or increases throughout the entire period	NA	Oronasal thermistor (diagnostic study), PAP device (titration)
	Central apnea		Absent inspiratory effort throughout the entire period		
	Mixed apnea		Absent inspiratory effort initially with resumption in latter part of DBE		
	Hypopnea	Decrease by 30% to 90%	Continues	AASM: 3% desaturation *or* terminal arousal CMS: 4% desaturation	May score as: "Obstructive" = snoring, inspiratory flattening of flow (nasal pressure), thoracoabdominal paradox "Central" = none of above obstructive criteria
	RERA (AASM only)	Inspiratory flattening (<30%)	Increased effort	Terminal arousal	Crescendo snoring common
NA	Hypoventilation	NA	NA	NA	Increased P_{CO_2} to >55 torr for ≥10 minutes *or* ≥10 torr increased in P_{CO_2} c/w awake to >50 torr for ≥10 minutes
NA	Cheyne-Stokes breathing	NA	NA	NA	≥3 consecutive central apneas/ hypopneas separated by crescendo/ decrescendo change in airflow with cycle length >40 seconds *and* ≥5 central apneas/ hypopneas per hour sleep with crescendo/ decrescendo pattern over ≥2 hours monitoring

AASM, American Academy of Sleep Medicine; CMS, Centers for Medicare and Medicare Services; DBE, disordered breathing event; PAP, positive airway pressure; P_{CO_2} measured from arterial line, end-tidal, or transcutaneously; RERA, respiratory event–related arousal.

from accepted techniques of respiratory event indexes appears to be a reliable predictor of neurocognitive changes such as sleepiness and vigilance.

Sleep-related respiratory events have been defined in various ways. In adults, a respiratory event must last at least 10 seconds. This follows the definition of the original workers in the field who reasoned that this interval would encompass at least two breaths for the average adult.[4-11] There are other requirements for scoring specific types of sleep-related respiratory events. Table 114-4 summarizes the criteria for scoring disordered breathing events in adults as elaborated by the current version of the AASM scoring.[107] Table 114-5 lists commonly used definitions of OSA severity. The definitions are those of the Centers for Medicare and Medicaid Services.[3] Examples of PSG recordings of disordered breathing events are presented in Figure 114-3 (RERAs), Figure 114-4

Table 114-5 Severity Criteria for Obstructive Sleep Apnea*

Severity Definition	Apnea-Hypopnea Index	Coverage (CMS)
Mild	5–14	With comorbidities and symptoms†
Moderate	15–30	Yes
Severe	>30	Yes

*Definitions per Centers for Medicare and Medicaid Services (CMS).[3]
†Comorbidities that allow for coverage of "mild" disease in Medicare recipients include documentation of excessive sleepiness, impaired cognition, mood disorders, insomnia, hypertension, ischemic heart disease, history of stroke.
Note: Medicare will cover treatment for OSA when the diagnosis and severity classification were determined by home sleep testing as well as in-laboratory testing.

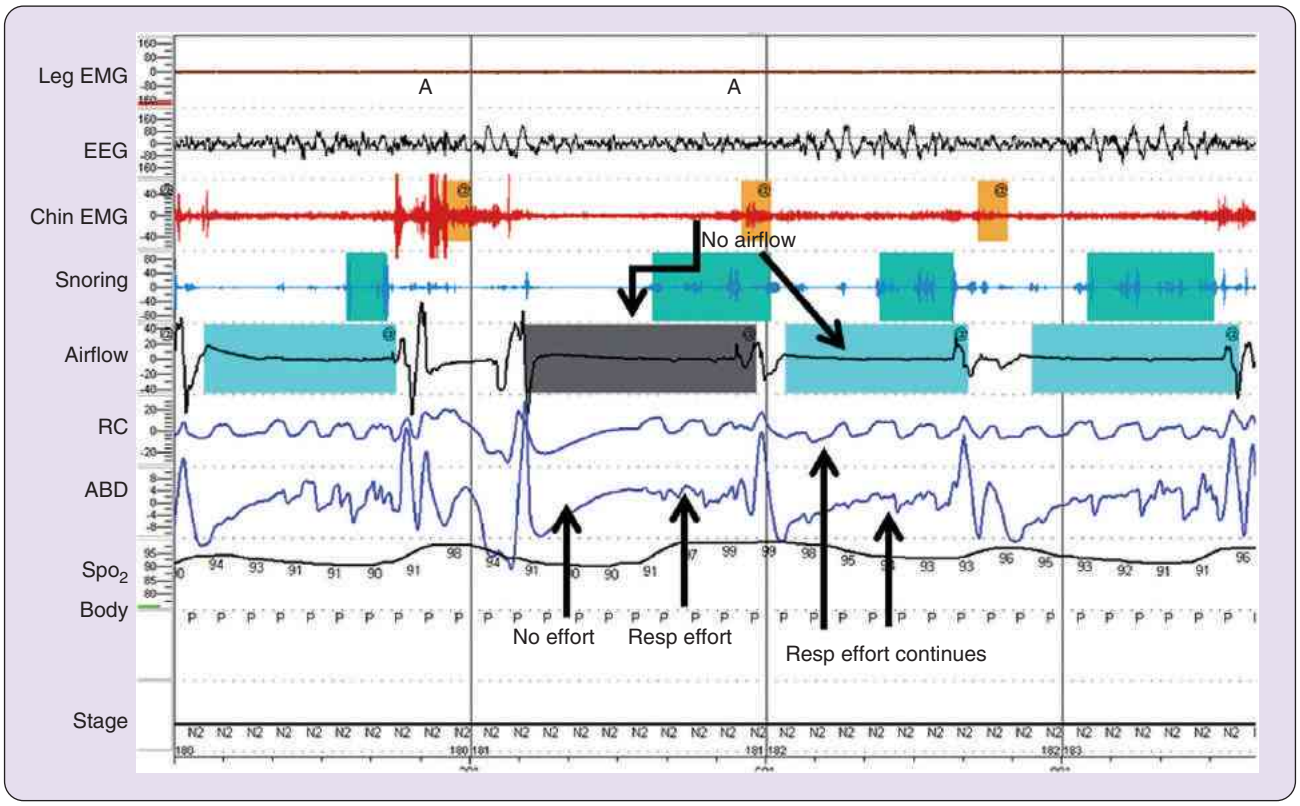


Figure 114-4 Obstructive and mixed apneas. Three obstructive apneas are outlined in blue-green, one mixed apnea in gray. Note for the obstructive apneas and the latter part of the mixed apnea, there is total cessation of airflow but chest wall movement continues. Note that the initial part of the mixed apnea shows no chest wall movement and resembles a central apnea (see Figure 114-5). Postapnea arousals are clearly seen (A). Airflow is measured by nasal air pressure. RC/ABD, Rib cage and abdominal movement by respiratory impedance pneumography; Spo_2, oxygen saturation of hemoglobin (pulse oximeter); Body, body position; Stage, stage of sleep. *Vertical lines* represent 30-second epochs.

(obstructive and mixed apneas), Figure 114-5 (central apneas), and Figure 114-6 (hypopneas).

Out-of-Center Sleep Testing

Although the previous parameters for assessing sleep-related respiratory events primarily pertain to in-laboratory, technologist attended PSG, a number of techniques have been developed to optimize the detection of sleep-related respiratory events outside of the laboratory, usually using a limited number of recording channels. Out-of-center sleep testing (OCST) offers a number of advantages compared with in-laboratory PSG. First, the initial costs are generally less than those of the state-of-the-art in-laboratory study, and primarily for this reason numerous insurance carriers require OCST for reimbursement in the initial evaluation of many patients with suspected OSA. Further, OCST offers a more rapid method of assessing the many patients with undiagnosed OSA who have limited access to, or who are reluctant to undergo, in-laboratory PSG.[109] However, Chervin and colleagues[110] performed a careful cost utility analysis, comparing in-laboratory PSG, OCST, and no testing (with treatment based on clinical characteristics). Their outcomes were based on costs per quality-adjusted life years over 5 years. These authors concluded that standard in-lab PSG provides greater quality-adjusted life years over 5 years than either OCST or no testing. Reuveni modeled costs of in-laboratory PSG versus OCST, accounting for the published technical failure rate of OCST and the published European costs for PSG.[111] They demonstrated that there was no long-term cost saving using OCST versus in-laboratory PSG.

Recent studies have directly compared health outcomes between OCST and in-laboratory PSG in selected patients. In 2010, Skomro and colleagues performed a randomized clinical trial of OCST compared with in-laboratory PSG in diagnosis and management of OSA in patients with a high clinical suspicion for the disorder.[112] All patients were evaluated and treated by physicians facile in interpretation of sleep studies as well as in clinical identification and management of sleep disorders. Exclusion criteria included concomitant cardiopulmonary morbidities and suspicion of other sleep disorders. Clinically relevant outcomes included quality of life, CPAP treatment adherence, blood pressure, and sleep quality after 4 weeks of therapy. There were no significant differences in any of these outcomes between OCST and in-laboratory PSG.

Overall, the relative advantages and disadvantages of OCST compared with in-laboratory testing in the evaluation and management of OSA require further definition. Approximately one third of OSA patients have a concomitant sleep disorder, of which approximately two thirds require treatment.[113] Thus patients need to be evaluated by a professional skilled in the evaluation and management of sleep disorders,

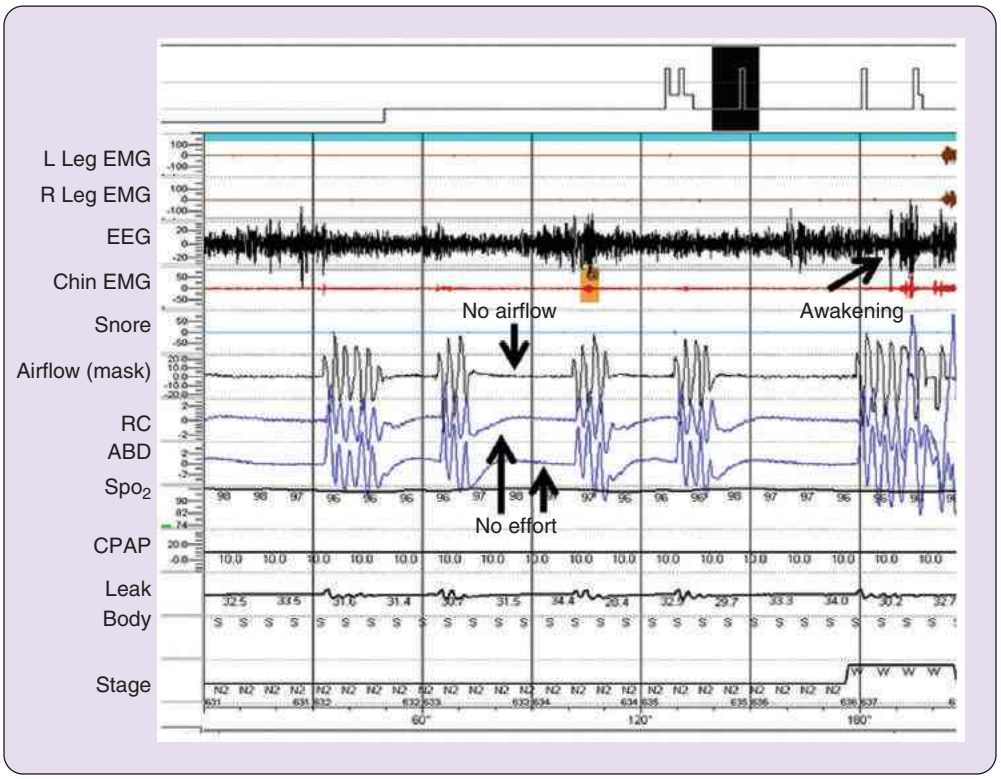


Figure 114-5 **Central apneas.** Central apneas developed during titration of continuous positive airway pressure (CPAP). Note absence of airflow (measured from the CPAP mask) and absence of respiratory effort (rib cage, abdominal movement measured by respiratory impedance pneumograph belts). RC, Rib cage signal; ABD, abdominal signal, snoring (absent here) measured from a microphone taped to the neck; Spo_2, oxygen saturation of hemoglobin measured from a pulse oximeter; Leak, estimation of leak from the CPAP mask; Body, body position; Stage, stage of sleep (hand scored). The *vertical lines* represent 30-second epochs. Note that after the series of central apneas, there is an arousal, and respiration resumes.

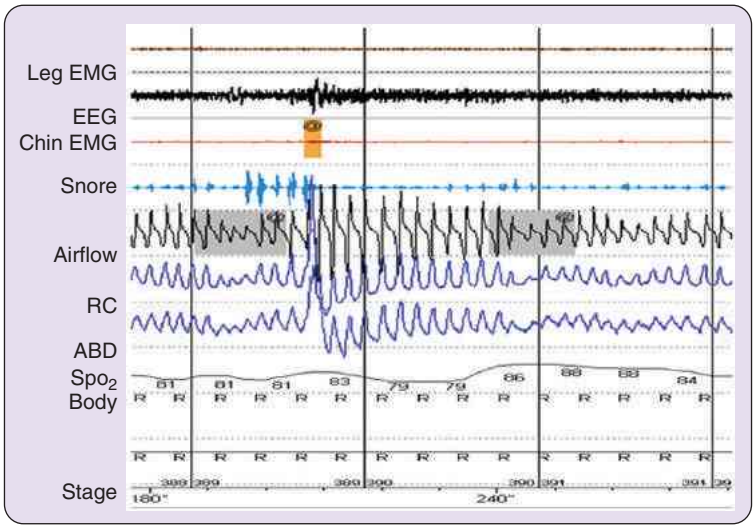


Figure 114-6 **Hypopnea.** In this diagnostic study airflow is measured using a nasal pressure cannula. The hypopneas are outlined in gray on the airflow signal. The first event shows a terminal arousal and 4% O_2 desaturation (the nadir of the event-associated desaturation is seen approximately 25 seconds after the arousal). The second event does not show an arousal, but there is an O_2 desaturation of 4%. The patient is in REM sleep. Body, Body position; RC/ABD, rib cage and abdominal movement by respiratory impedance pneumography; Spo_2, oxygen saturation of hemoglobin (pulse oximeter); Stage, stage of sleep. *Vertical lines* represent 30-second epochs.

whether OCST, in-laboratory PSG, or both are used in the diagnostic assessment. Incorporation of test results into a complete evaluation and management plan that encompasses all sleep-related complaints is essential. The clinician should also be aware that sleep testing is currently rated into levels of complexity. Type 1 is the classic in-laboratory full PSG as discussed previously, including measures of airflow, respiratory effort, oxygenation, EEG, EOG, and EMG to allow for sleep staging. Type 2 is an out-of-laboratory portable study essentially equivalent to the in-laboratory study (minimum of seven parameters). Type 3 is an unattended portable recording measuring at least four channels: heart rate, oxygen saturation, respiratory airflow, respiratory effort, but no sleep staging. Type 4 is an unattended portable study, measuring a minimum of three channels such as heart rate, oxygen saturation, and respiratory analysis. Appropriate documentation from a regional Medicare carrier should be consulted for details regarding coding, and the policies of specific insurance carriers should be consulted regarding requirements for allowing classical in-laboratory PSG versus portable OCST.

Even with the advent of reliable OCST, such testing still requires a relatively high level of instrumentation and analysis. Prescreening for OSA has the potential to improve sensitivity and specificity of both OCST and in-laboratory PSG, or even eliminate the need for such testing, taking into account pretest probability accruing from the previously noted demographic, physiologic, and clinical symptoms, signs, and conditions associated with OSA. A number of workers have therefore attempted to develop screening tools and techniques, some assessing physiologic parameters and some using self-administered questionnaires, with or without inclusion of key physical traits.

Early attempts at physiologic screening for OSA were based on simple pulse oximetry.[114] These devices generally relied on estimating the number of 3% or 4% drops in oxygen saturation. However, problems with validity and sensitivity remained. Newer techniques make use of the known coupling between cardiovascular (pulse, autonomic function, arterial tone) and respiratory phenomena. For example, Liu and colleagues used a sophisticated transform to analyze signals from a single ECG lead to detect breathing and sleep-related disordered breathing events.[115] Their analyses resulted in acceptable sensitivity and specificity (area under the receiver-operator curve of 0.79) for OSA screening using a single ECG electrode. An index called the temporal variability of the dominant frequency was well correlated to AHI and could distinguish among sleep-related breathing events of various severities.

Peripheral arterial tonometry (PAT) has also been developed as an OSA screening tool. Bar and colleagues evaluated such a device (WatchPAT, Itamar Medical, Caesarea, Israel) placed on the finger that measured pulse oximetry and volume of the finger.[116] The abrupt arousals associated with termination of the obstructive events are associated with bursts of sympathetic discharge causing vasoconstriction that decreases volume of the digit. In addition to PAT and standard oximetry, the device recorded pulse rate and movement (actigraphy). The respiratory disturbance index (RDI) measured using PAT was highly correlated with RDI measured during in-laboratory PSG (the area under the receiver-operating curve was 0.82 and 0.87 for thresholds of RDI = 10/hour and RDI = 20/hour, respectively).

Another device that relies on measures of pulse, oxygen saturation and peripheral digital volume is the photoplethysmograph (PPG; Morpheus Ox, WideMed, Herzliya, Israel). Digital volume, pulse, and oxygenation signals are recorded and imputed into proprietary algorithms that generate clinically relevant respiratory waveforms and approximation of the sleep-wake state. A recent study using the AASM 2012 apnea-hypopnea detection scoring parameters validated this device against standard in-laboratory PSG.[117] A unique feature of the study was that among the 65 subjects, 19 had significant cardiopulmonary comorbidities. There was excellent correlation between the PPG- and PSG-derived AHI. For AHI of more than 5/hour, sensitivity was 80%, specificity 86%, and positive likelihood ratio 5.9. For AHI of more than 15/hour, sensitivity was 70%, specificity 91%, and positive likelihood ratio 7.83. Further, results in patients with cardiopulmonary morbidities were not different from the rest of the subjects.

Numerous questionnaires have also been designed in an attempt to use patient symptoms to improve the pretest probability for OSA. The Harvard sleep apnea screening questionnaire was one such attempt.[118] Subjects with essential hypertension were screened. No one symptom was found to be predictive of an AHI greater than 10/hour, although loud snoring was predictive of oxygen desaturation in sleep. Numerous OSA screening questionnaires have been developed that incorporate symptoms (e.g., snoring, snorting, witnessed apneas, excessive sleepiness), demographics (e.g., gender, age), physical traits (e.g., BMI, crowding of the oropharynx, neck circumference), and important c-morbidities (e.g., hypertension). Clinical prediction formulas, some requiring sophisticated calculations and even computer assistance, have been developed that assign weights to the various factors. Questionnaires that have been validated in specific populations include the Berlin Questionnaire, designed for use in primary care settings,[119] and the STOP-BANG questionnaire, designed for preoperative screening.[120] Other such instruments include the Wisconsin questionnaire[121] and the questionnaire of Haraldsson and colleagues.[122] Overall, it appears that such questionnaires have the greatest potential to screen patients at high risk for OSA. Further, validation studies indicate that the applicability of any such questionnaire is limited to the specific populations studied.

The predictive value of the clinical examination has also been assessed in an attempt to identify features on the clinical examination with the greatest value for predicting a "positive" PSG.[73] No one single symptom or sign appears sufficiently predictive of OSA, if defined as AHI greater than or equal to 10/hour, including snoring, subjective sleepiness, and morning headache. Rather, a combination of signs and symptoms, including neck circumference, habitual snoring, systemic hypertension, and bed partner report of nocturnal gasping or choking appears to optimally identify patients most likely to have sleep-study documented OSA.[123]

PRINCIPLES OF MANAGEMENT

The adverse consequences of OSA pose an enormous health and economic burden.[124,125] Effective treatment requires a patient-centered chronic disease management approach that goes beyond initial diagnosis and therapy prescription. Monitoring and enhancing adherence to therapy, providing

alternative therapeutic modalities when needed, and managing comorbid sleep disorders are necessary to optimize long-term sleep-related health outcomes.

CPAP, which pressurizes the UA to prevent its collapse during sleep, remains the mainstay of therapy for OSA. Initiation and prescription of CPAP is usually accomplished by in-laboratory PSG CPAP titration, which determines CPAP pressure requirements during all stages of sleep in all sleep positions. Autotitrating positive airway pressure (APAP) devices assess inspiratory airflow and adjust positive airway pressure automatically to maintain normal inspiratory flow patterns. Data supporting noninferiority of APAP for initiation of CPAP therapy are limited.[126] Patients with comorbid cardiopulmonary disorders, especially those with obesity hypoventilation syndrome, central sleep apnea, Cheyne-Stokes breathing, or COPD, are not candidates for APAP therapy.

Multiple studies have evaluated the effect of CPAP on EDS. Most placebo-controlled studies demonstrated improvement in subjective measures of daytime somnolence, whereas data are mixed with regard to objective measures. A randomized placebo-controlled study with more than 1000 participants demonstrated that CPAP improved both subjective and objective measures of daytime sleepiness, especially in severe OSA (AHI >30/hour).[127] Similarly, clinical trials of the effects of CPAP on neurobehavioral and cognitive performance, as well as on overall quality of life, are mixed, with some studies demonstrating benefit.[128] A recent randomized placebo-controlled clinical trial evaluated the effects of CPAP on daytime sleepiness and quality of life in patients with mild to moderate OSA with a complaint of daytime somnolence, a patient type that represents a large portion of the OSA population. CPAP resulted in greater improvement in functional outcomes, including quality of life, subjective daytime sleepiness, and mood, compared with sham CPAP.[127] The optimal duration of nightly CPAP use necessary to achieve improvement in functional outcomes has also been a matter of investigation. In a multicenter effectiveness study that used both subjective and objective measures of daytime somnolence and quality of life as outcome measures, a greater percentage of patients achieved improvements in outcomes with longer nightly duration of CPAP use, up to 7 hours/night. Although the mean nightly duration of CPAP use in this trial was 4.7 $\pm$ 2.2 hours, a substantial minority of patients demonstrated benefit with a shorter duration of nightly CPAP use (even <2 hours/night), whereas some subjects exhibited residual sleepiness with more than 7 hours of use per night. Thus assessment of optimal CPAP use should rely not only on hours of nightly use but also on assessment of relevant treatment outcomes.[128]

The efficacy of CPAP is limited by suboptimal adherence. Many studies have evaluated measures to improve adherence to CPAP therapy. Technologic improvements in CPAP delivery, such as APAP, expiratory pressure reduction, heated tube humidification, and other modalities that adjust the contour of the pressure waveform, have contributed somewhat to improvements in comfort but have not solved the overall adherence problem with CPAP.[129] Other measures to enhance adherence include educational and supportive efforts as well as cognitive behavioral therapy, that may help to acclimatize patients to CPAP and encourage nightly use. A recent review indicated that these interventions can improve CPAP adherence.[130]

The identification of *treatment emergent central sleep apnea*, which is a form of sleep breathing disorder in which central apneas emerge during CPAP therapy disrupting sleep, may also limit efficacy and tolerability of CPAP. Although the pathophysiology of this form of central sleep apnea has not been fully elucidated, it may occur as a result of ventilatory instability during sleep. Cheyne-Stokes breathing, which is not well treated with CPAP, is common in patients with systolic heart failure and may contribute to failure of CPAP therapy in this population. However, while effective in reducing disordered breathing events, the newer modality of adaptive servo ventilation is no longer recommend in patients with heart failure and predominantly central sleep apnea due to an observed increase in mortality with this therapy.[132]

Mandibular advancement oral appliance therapy can be considered either as first-line therapy for mild to moderate OSA or as an alternative to CPAP. Improvement in somnolence, vigilance, and neurocognitive performance has been observed with this modality. Interestingly, the improvement in functional outcomes achieved with oral appliances is similar in magnitude to that seen with CPAP, despite persistence of mild degrees of OSA.[133,134] It is important to perform follow-up PSG to assess efficacy of the appliance because subjective reports of improvement may not reliably predict the AHI with therapy. In a study of patients with mild to severe OSA, only 65% of the cohort who had subjective symptomatic improvement with oral appliance therapy achieved an AHI of 10/hour or less on follow-up PSG. The degree of mandibular advancement was then increased in the incomplete responders. An additional 30% of the cohort achieved an AHI of 10/hour after this secondary adjustment, indicating the importance of the follow-up PSG.[135] Better predictors of response to oral appliances, as well as accurate determination of the optimal degree of mandibular advancement, may improve utility and efficacy of this modality. Remotely titratable appliances have been developed, in which the degree of mandibular advancement can be titrated during PSG, that may help to identify patients who respond to this modality and accurately determine the optimal degree of mandibular advancement.[136] In addition, recent data suggest that response to oral appliance therapy may be predicted by visualization of velopharyngeal widening with mandibular advancement during awake nasopharyngoscopy.[137]

Nasal expiratory positive pressure (nEPAP) is another alternative therapy for OSA that provides positive pressure during end expiration when the cross-sectional area of the UA and dilator muscle activity are at their nadir. By increasing expiratory airflow resistance using a small valve taped to the nostrils, nEPAP is created that prevents airway collapse. In addition, increases in lung volume have been observed with nEPAP, which may reduce UA collapsibility by increasing caudal traction on the UA. Significant decreases in the AHI have been observed in efficacy trials, although tolerability and compliance remain to be fully established.[138,139]

UA surgery, including nasal septoplasty, uvulopalatopharyngoplasty (UPPP), tonsillectomy, and tongue advancement procedures, as well as maxillomandibular advancement surgery, are alternative therapeutic modalities for OSA. In addition to patient preference and consideration of medical comorbidities, UA anatomy should be evaluated when assessing suitability of a surgical approach. A clinical staging system, based on observation of the oropharynx, such as that offered by

Friedman,[105] which classifies palate position, tonsil size, and BMI, can be a useful predictor of outcome of UPPP and tonsillectomy. These authors showed that the best response is achieved when the inferior border of the palate is above the base of the tongue, with enlarged tonsils and BMI of 40 kg/m^2 or lower.

Imaging with nasopharyngoscopy can more directly assess UA anatomy than clinical observation of the oropharynx. Dynamic behavior of the UA can also be observed during a Müller maneuver. Although identification of a velopharyngeal site of collapse during a Müller maneuver was originally shown to predict response to UPPP, subsequent studies demonstrated a low predictive value.[140,141] Recently, drug-induced sleep endoscopy has been introduced in which the UA is evaluated during propofol-induced sedation in an operating room setting. The relationship of drug-induced sleep endoscopy findings to UA behavior during sleep and their relevance to UA surgical operative planning and outcomes has not been established.[142,143]

Clinical experience has shown that although nasal surgical procedures including septoplasty and turbinectomy usually do not achieve resolution of OSA, they may be useful to improve tolerability of CPAP in patients with nasal obstruction. Although response to a single UA surgical procedure may be limited, a multilevel approach to the UA, including UPPP followed by genioglossal suspension or advancement with hyoid myotomy, has achieved improved success rates.[144] Modifications of current UA soft tissue surgical procedures, as well as development of new approaches, may ultimately improve outcomes. Maxillary-mandibular advancement osteotomy is also performed in the surgical management of OSA with reported high success rates. The resultant increase in volume of the UA has been correlated with reduction in the AHI. Surgical planning using three-dimensional UA imaging may also improve the success of these surgical procedures.[145,146]

Positional therapy may also be useful in selected patients in whom disordered breathing events occur predominantly during supine sleep. Avoidance of supine sleep can lead to reduction in the AHI in such cases, particularly with use of positioning devices, which are typically worn around the chest and prevent inadvertent supine sleep.[147]

Because OSA severity is sensitive to weight loss, weight reduction programs and bariatric surgery are valuable treatment modalities. Although most studies have shown that bariatric surgery is associated with resolution or improvement of OSA in most cases, it is important to recognize that OSA can persist in some cases after substantial weight loss. Because the correlation of subjective daytime sleepiness with AHI severity is relatively weak, follow-up PSG, rather than reliance on symptoms, is important to determine whether further OSA therapy is needed despite weight loss.[148,149]

Electrical stimulation of the hypoglossal nerve is a recently introduced modality that was developed in response to the observation that inadequate neural activation of the UA dilator muscles is a key factor in the pathophysiology of OSA. The stimulator can deliver phasic electrical pulses to the hypoglossal nerve at the onset of inspiration augmenting genioglossal inspiratory activity. Efficacy trials have included patients with moderate to severe OSA with BMI of less than 40 kg/m^2. Some of the studies excluded patients with concentric retropalatal collapse. Most of the trials demonstrated more than 50% improvement in the AHI, which was sustained during long-term follow-up.[150-152] Although experience remains limited, hypoglossal nerve stimulation may be a useful alternative for selected patients.

Future approaches to treatment may be developed based on identification of individual "phenotypes" of OSA. Recent studies have demonstrated interindividual differences in the pathophyphysiology of OSA. In some patients, instability of ventilatory control during sleep, measured as elevated "loop gain," is a predominant factor. Loop gain is a dimensionless value that quantifies response of the ventilatory system to decrements in ventilation such as those induced by UA obstruction. High loop gain is associated with an excessive ventilatory response to apnea (or hypopnea) that contributes to ventilatory instability and perpetuates recurrent apneas. A low arousal threshold, which destabilizes sleep, has also been shown to contribute to ventilatory instability during sleep. In other patients, inadequate activation of UA dilator muscles in response to obstruction is a predominant pathophysiologic feature. Anatomic factors, with a highly collapsible UA, indicated by a high (less negative or positive) Pcrit, is identified as the major factor contributing to OSA in other patients.[151]

Understanding OSA phenotypes may help to individualize treatment. Mechanically based therapies including CPAP, mandibular advancement oral appliances, or surgical interventions are the best therapeutic options for patients in whom anatomic factors, with a highly collapsible UA, play a major role. Patients with high loop gain, or a low arousal threshold, in whom ventilatory control instability is the predominant factor, may respond to novel therapeutic measures to stabilize sleep and ventilatory control. Approaches that increase UA dilator muscle activity during sleep, such as hypoglossal nerve stimulation, may be most useful in patients that have reduced UA dilator muscle responsiveness to UA collapse. Thus thinking about OSA as a heterogeneous disorder with multiple "phenotypes" might facilitate development of new and individualized therapies.[153,154]

SUMMARY

OSA is a highly prevalent disorder and is overrepresented in populations with cardiovascular, cerebrovascular, and metabolic disease. The pathophysiologic basis of OSA is complex, with contributions from anatomic factors that narrow the UA as well as ventilatory control instability that affects neural drive to the UA and ventilatory pump muscles during sleep. The acute, repetitive physiologic perturbations during sleep that occur as a result of obstructive apneas and hypopneas include sleep fragmentation, large swings in intrathoracic pressure, increased sympathoadrenal tone, and intermittent hypoxia and reoxygenation. OSA is associated with multiple adverse systemic consequences, including excessive sleepiness, impairment of cognitive function, mood, vigilance, and performance, including driving ability. OSA is also associated with multiple cardiovascular, cerebrovascular, and metabolic disorders as well as other medical conditions. The relationship between OSA and these comorbidities may be bidirectional. Although clinical presentation can identify patients at risk for OSA, diagnosis requires objective monitoring of sleep. In-laboratory PSG testing remains the gold standard for accurate identification of OSA; however, OCST using ambulatory technology is useful in selected patients. Skilled assessment and management of this chronic disorder are necessary to

ensure optimal long-term outcomes. Most studies have demonstrated improvement in daytime somnolence and quality of life with CPAP therapy for OSA. Other therapeutic modalities, including mandibular advancement oral appliances and surgical approaches to the UA, are useful alternative modalities in selected patients. Newer approaches such as hypoglossal nerve stimulation, nasal expiratory positive pressure devices, and other treatments require further investigation to establish their clinical utility. Investigational approaches that define an individual's predominant OSA phenotype may ultimately guide treatment decisions.

CLINICAL PEARL

OSA is highly prevalent, particularly in patients with comorbid cardiovascular, cerebrovascular, and metabolic disease. Because OSA independently contributes to morbidity and mortality, with multisystem consequences, clinical suspicion for OSA should be a priority for clinicians, especially in patients with these comorbidities. Clinical assessment, as well as sleep testing with in-laboratory PSG, or OCST in appropriately selected and managed patients, is necessary for accurate diagnosis. Effective treatment of OSA and any comorbid sleep disorders requires a focus on optimizing sleep-related health outcomes with a chronic disease management approach.

Selected Readings

Al Mawed S, Unruh M. Diabetic kidney disease and obstructive sleep apnea: a new frontier? *Curr Opin Pulm Med* 2016;**22**(1):80–8.

Berry RB, Brooks R, Gamaldo CE, et al. for the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Version 2.2. <www.aasmnet.org>, Darien, IL: American Academy of Sleep Medicine; 2015.

Dempsey JA, Smith CA, Blain GM, et al. Role of central/peripheral chemoreceptors and their interdependence in the pathophysiology of sleep apnea. *Adv Exp Med Biol* 2012;**758**:343–9.

Owens RL, Edwards BA, Eckert DJ, et al. An integrative model of physiological traits can be used to predict obstructive sleep apnea and response to non positive airway pressure therapy. *Sleep* 2015;**38**(6):961–70.

Peppard PE, Young T, Barnet JH, et al. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol* 2013;**177**(9):1006–14.

Punjabi NM, Caffo BS, Goodwin JL, et al. Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med* 2009;**6**(8):e1000132.

Rapoport DM, Sorkin B, Garay SM, et al. Reversal of the "pickwickian syndrome" by long-term use of nocturnal nasal-airway pressure. *N Engl J Med* 1982;**307**(15):931–3.

Schwab RJ, Pasirstein M, Pierson R, et al. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 2003;**168**(5):522–30.

Strohl KP, Brown DB, Collop N, et al. An official American Thoracic Society Clinical Practice Guideline: sleep apnea, sleepiness, and driving risk in noncommercial drivers. An update of a 1994 statement. *Am J Respir Crit Care Med* 2013;**187**(11):1259–66.

Sullivan CE, Issa FG, Berthon-Jones M, et al. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. *Lancet* 1981;**1**(8225):862–5.

Terrill PI, Edwards BA, Nemati S, et al. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J* 2015;**45**(2):408–18.

Weaver TE, Mancini C, Maislin G, et al. Continuous positive airway pressure treatment of sleepy patients with milder obstructive sleep apnea: results of the CPAP Apnea Trial North American Program (CATNAP) randomized clinical trial. *Am J Respir Crit Care Med* 2012;**186**(7):677–83.

Younes M. Contributions of upper airway mechanics and control mechanisms to severity of obstructive apnea. *Am J Respir Crit Care Med* 2003;**168**(6):645–58.

A complete reference list can be found online at ExpertConsult.com.

Positive Airway Pressure Treatment for Obstructive Sleep Apnea

Chapter 115

Neil Freedman

Chapter Highlights

- Continuous positive airway pressure (CPAP) therapy is indicated for patients with moderate to severe obstructive sleep apnea (OSA) with or without symptoms and for patients with mild OSA with associated symptoms or comorbid illnesses.
- CPAP consistently improves or resolves respiratory events across the spectrum of disease severity and improves symptoms of daytime sleepiness, especially for patients with moderate to severe OSA. Improvements in blood pressure are relatively small, with reductions in blood pressure tending to be greatest in patients with untreated hypertension and in those with better compliance with therapy. Improvements in other outcomes are inconsistent across the spectrum of disease severity.
- Compliance with positive airway pressure (PAP) therapy is far from perfect. Systematic education through several approaches, with or without behavioral therapy, have been the only interventions that have been associated with consistent improvements in compliance with PAP therapy. The roles of other interventions, including heated humidification, prescription hypnotics, telemedicine, and sleep specialist care, for most patients with OSA are not clear.
- Advanced PAP technologies, including bilevel PAP and expiratory pressure relief devices, have not been associated with better compliance or improvements in other important outcomes compared with standard CPAP therapy. The roles for bilevel PAP and devices with expiratory pressure relief technology in the management of most patients with OSA are not clear.
- Autotitrated positive airway pressure (APAP) used in an unattended setting, either to determine a fixed CPAP pressure or as a primary treatment, is reasonable therapy for most patients with uncomplicated moderate to severe OSA syndrome. APAP therapy has been shown to result in similar compliance and improvements in other important outcomes compared with conventionally titrated CPAP therapy. In appropriate OSA patients, an ambulatory approach using portable testing and APAP therapy should lead to reductions in the cost of management of OSA while not adversely affecting patient outcomes.

Treatment with positive airway pressure (PAP) remains the primary therapy for most patients with obstructive sleep apnea (OSA), especially those with moderate to severe OSA. This chapter reviews various forms of PAP therapy for the treatment of OSA, highlighting the indications for treatment, methods for determining an effective pressure prescription, treatment outcomes, and methods that may improve compliance with therapy. The initial part of the chapter focuses on continuous positive airway pressure (CPAP) therapy, and the later portion of the chapter emphasizes the technologic advancements in the delivery of PAP therapy, including autotitrating positive airway pressure (APAP), bilevel PAP, and expiratory pressure reduction (EPR) technologies.[1]

CONTINUOUS POSITIVE AIRWAY PRESSURE TREATMENT

CPAP therapy was initially described as a treatment for OSA by Sullivan and colleagues in 1981.[1a] Since its initial description, CPAP has become the predominant therapy for the treatment of patients with OSA because it has been demonstrated to resolve sleep-disordered breathing events and improve several clinical outcomes.[2,3] Treatment with CPAP is typically indicated for patients with moderate to severe OSA by apnea-hypopnea index (AHI) of 15 or more events/hour with or without associated symptoms or comorbid diseases and for patients with mild OSA (AHI ≥5 to ≤14 events/hour) with associated symptoms or comorbid diseases (Box 115-1).

CPAP is conventionally delivered through a nasal mask at a fixed pressure that remains constant throughout the respiratory cycle. The proposed mechanism of action of CPAP therapy is that it acts as a pneumatic splint that maintains the patency of the upper airway in a dose-dependent fashion. It does not exert its effects by increasing upper airway muscle activity[4] and acts only as a treatment, and not a cure, for the disorder. Several studies have demonstrated that withdrawing CPAP therapy in patients with OSA across the spectrum of disease severity results in the recurrence of OSA and associated daytime symptoms in most patients within 1 day to several days.[5-7]

Box 115-1 TYPICAL INDICATIONS FOR CONTINUOUS POSITIVE AIRWAY PRESSURE THERAPY FOR OBSTRUCTIVE SLEEP APNEA

- Moderate to severe obstructive sleep apnea (≥15 events per hour of sleep) with or without associated symptoms or comorbid diseases
- Mild obstructive sleep apnea (≥5 to ≤14 events per hour of sleep) with symptoms or associated comorbid diseases:
  - Symptoms: excessive daytime sleepiness, impaired cognition, mood disorders, or insomnia
  - Comorbid diseases: hypertension, ischemic heart disease, or history of stroke

Determining the Optimal Setting

The optimal CPAP settings for home use may be defined as the minimal pressure required to resolve all apneas, hypopneas, snoring, and arousals related to these events in all stages of sleep and in all sleep positions.[2,8-10] Simply, the optimal CPAP setting should resolve all sleep-disordered breathing in supine rapid eye movement (REM) sleep to account for the effects of gravity and changes in muscle tone that may occur in different sleep stages and positions.[8] The optimal pressure should also maintain oxygen saturation at or above 90% and should minimize mask leak, allowing and maintaining only mask leak that that is appropriate for the given pressure. The most current American Academy of Sleep Medicine (AASM) Practice Parameters recommend a full night of CPAP titration based on the criteria outlined previously.[9,10] A repeat CPAP titration need only be performed if symptoms of OSA reappear despite compliance with CPAP therapy, if a patient sustains a significant weight loss either through diet or bariatric surgery, or if CPAP compliance and benefits remain suboptimum by current standards.

A "split-night" sleep study, in which the initial portion of the study is used to objectively document an individual's sleep related breathing disorder followed by a CPAP titration during the second portion of the night, may be indicated in certain situations.[8-10] A split-night sleep study may be considered when the following criteria have been met: (1) an AHI of 40 or more events/hour is recorded during the initial 2 hours of the polysomnography (PSG) study and (2) at least 3 hours remain during the PSG study to conduct an adequate CPAP titration. A second full night of CPAP titration should be considered if an optimal CPAP pressure setting could not be achieved during the second portion of the split-night study. Split-night studies can also be considered for individuals who demonstrate less severe OSA, with an AHI of 20 to 40 events/hour, during the initial 2 hours of a sleep study, although data suggest that CPAP titrations in this subgroup of patients may be less accurate when performed in the split-night protocol setting. Although split-night studies potentially reduce waiting times to initiate home CPAP therapy, especially in areas with long sleep laboratory waiting times, a significant portion of patients with OSA may undergo suboptimal CPAP titrations using this format.[11]

The use of home sleep testing (HST) is currently not recommended for the titration of CPAP or other PAP therapies because there are few data on the reliability of HST for this indication. Given the absence of data regarding HST for CPAP titration, the Centers for Medicare and Medicaid Services and commercial insurance companies in the United States typically will not reimburse providers who use HST for this indication.

Table 115-1 Adequacy of Continuous Positive Airway Pressure Titration Definitions

Optimal	Reduces the RDI <5 for at least a 15-minute duration and should include supine REM sleep at the selected pressure that is not continually interrupted by spontaneous arousals or awakenings
Good	Reduces the RDI ≤10 or by 50% if the baseline RDI <15 and should include supine REM sleep that is not continually interrupted by spontaneous arousals or awakenings at the selected pressure
Adequate	Does not reduce the RDI ≤10 but reduces the RDI by 75% from baseline (especially in severe OSA patients), or in which the titration grading criteria for optimal or good are met with the exception that supine REM sleep did not occur at the selected pressure
Inadequate	Does not meet any one of the above grading criteria

RDI, respiratory disturbance index, which accounts for apneas, hypopneas, and respiratory effort–related arousals.
Adapted from Kushida CA, Chediak A, Berry RB, et al. Clinical guidelines for the manual titration of positive airway pressure in patients with obstructive sleep apnea. *J Clin Sleep Med* 2008;4:157–71.

Although there are current recommendations to help guide clinicians on how to manually titrate CPAP therapy in an attended laboratory-based setting, these recommendations largely serve as guidelines because they are principally based on the consensus of expert opinion and not on randomized trials demonstrating their superiority over other methods of manual titration.[8] The AASM guidelines classify the adequacy of a CPAP titration as delineated in Table 115-1. The AASM guidelines currently recommend considering a repeat titration study for patients who do not achieve an optimal or good PAP titration.

There are actually few data on the quality or efficacy of CPAP titrations, as defined by the AASM clinical guidelines, in clinical practice. Specifically, few data exist on how often patients actually achieve an optimal PAP titration, despite having their PAP pressures determined in an attended setting. Furthermore, there are few data examining the outcomes of patients who are initiated on CPAP therapy after undergoing PAP titrations that are less than optimum. Many of the randomized controlled trials evaluating the effect of CPAP on various outcomes have shown a mean residual AHI or respiratory disturbance index (RDI) of 5 or greater, indicating that more than 50% of these patients underwent CPAP titrations that did not achieve optimal results. Of the limited existing data from clinical settings, approximately only 50% to 60% of patients with OSA achieve an "optimal" titration and up to 30% to 40% achieve only an "adequate" or "inadequate" titration.[12,13] Thus many patients on CPAP therapy, even those who undergo attended in-laboratory titrations, may currently

be treated with suboptimal pressures settings. More data are needed to better define the optimal clinical and physiologic benchmarks for the various levels of PAP titration adequacy as well as to determine the outcomes and proper management for patients who do not achieve optimal or good PAP titrations during an in-laboratory titration study. The important point is that the clinician should not assume that a given patient is on an adequate CPAP setting simply because the patient's CPAP pressure was determined during an in-laboratory titration study.

Although current recommendations warrant that CPAP titrations occur during a full overnight in-laboratory PSG study, some data suggest that a fixed-pressure CPAP can be successfully initiated in an unattended home setting using various approaches.[14-21] Specifically, several studies confirm that CPAP therapy initiated in an unattended home setting (without HST or PSG monitoring to confirm the efficacy of treatment) can be successful in many patients with uncomplicated OSA (OSA without associated chronic obstructive pulmonary disease [COPD], congestive heart failure [CHF], or hypoventilation syndromes) when CPAP settings are determined by a clinical prediction formula,[20] by CPAP self-adjusted to resolve snoring and daytime symptoms,[17,18] or APAP therapy[14-16,21] (Table 115-2).

It is important to note that all of these methods typically offer only a starting pressure for initiating CPAP therapy. As observed in several of the study protocols, many patients may require pressure adjustments based on symptoms and problems with therapy. Because results from several studies continue to support the efficacy of PAP titrations in unattended nonlaboratory settings for patients with uncomplicated OSA, specifically with the use of APAP devices, PAP therapy may potentially be initiated in the home for many patients with uncomplicated OSA. It is possible that in the future, in-laboratory attended CPAP titrations may be reserved for patients with OSA and concomitant cardiac or respiratory disease, those with obesity hypoventilation syndrome, and those who are having difficulty with CPAP initiated in an unattended setting. If such an approach to CPAP treatment comes to fruition, patients with OSA may benefit by realizing shorter waiting times for CPAP therapy, and health care dollars should be saved by reducing the need for unnecessary PSG studies.[15,16,18,19,22]

Table 115-2 Clinical Prediction Formulas Used to Determine an Effective Continuous Positive Airway Pressure Setting

Study	Clinical Prediction Formula
Miljeteig & Hoffstein, 1993[166]	P(eff) = 0.13 (BMI) + 0.16 (NC) + 0.04 (RDI) – 5.12
Lin et al, 2003[167]	P(eff) = 0.52 + 0.174 (BMI) + 0.042 (AHI)
Stradling, 2004[168]	P(eff) = 2.1 + 0.048 (ODI) + 0.128 (NC)
Hukins, 2005[169]	BMI < 30 = 8 cm H_2O BMI 30–35 = 10 cm H_2O BMI > 35 = 12 cm H_2O
Loredo, 2007[170]	P(eff) = 30.8 + RDI (0.03) – nadir SaO_2 (0.05) – mean SaO_2 (0.2)

AHI, Apnea-hypopnea index; BMI, body mass index; NC, neck circumference; ODI, oxygen desaturation index; P(eff), effective continuous positive airway pressure; RDI, respiratory disturbance index.

Benefits of Therapy

It is the perception of many non–sleep practitioners and the lay public that CPAP treatment consistently resolves or improves several important outcomes, including sleep architecture, daytime sleepiness, neurocognitive function, mood, quality of life, and cardiovascular disease in all patients with OSA. When titrated appropriately, CPAP therapy has been demonstrated to resolve most sleep breathing disorder across the spectrum of disease severity and has been demonstrated to be superior to placebo, conservative management, and positional therapy with regard to this outcome.[9,23] Randomized controlled trials have also shown CPAP therapy to be superior to placebo at increasing the percentage of time and total time in stages N3 (non–rapid eye movement [NREM] sleep stage 3) and REM sleep. The effects of CPAP on other sleep parameters, including stages N1 and N2 sleep, total sleep time, and the arousal index, have been inconsistent across studies.[9,23]

Effect on Daytime Sleepiness

Several randomized controlled studies have shown that CPAP therapy significantly improves or resolves subjective symptoms of daytime sleepiness in OSA patients who suffer from this complaint, predominantly in those who suffer from severe OSA (AHI >30 events/hour).[18,21,24-31] The minimal and optimal amounts of nocturnal use necessary to improve symptoms of daytime sleepiness are, however, not well defined because even partial nocturnal use (as little as 2 hours per night) has been associated with significant improvements in daytime symptoms in some patients.[32,33] Although the minimal amount of time required on a nightly basis to improve symptoms of daytime sleepiness is not well established, it is clear that CPAP therapy is required for a least a portion of each night because symptoms of daytime sleepiness reappear when CPAP therapy is discontinued for as little as one to two nights.[6,34,35] Reoccurrence of daytime symptoms on CPAP withdrawal has been observed across the spectrum of OSA severity. As mentioned previously, a specific threshold for nightly use of CPAP, in terms of improvements in symptoms of daytime sleepiness, does not exist and is likely dependent on the individual.[32,33] In general, greater adherence to CPAP therapy on a nightly basis has been associated with greater improvements in symptoms of daytime sleepiness.

The data regarding the effects of CPAP on more objective measures of daytime sleepiness are more inconclusive across the spectrum of disease severity.[23,24] A large meta-analysis of randomized controlled trials comparing CPAP therapy with placebo or conservative management demonstrated only a small, although statistically significant, improvement in the mean sleep latency as measured on either the Multiple Sleep Latency Test (MSLT) or Maintenance of Wakefulness Test (MWT). Across all studies, the mean sleep latency improved by 0.93 minutes ($P = .04$). Whether this small improvement in objective sleepiness is clinically significant is unclear.

Although most patients with daytime sleepiness related to OSA will achieve significant improvements in symptoms after CPAP therapy has been instituted, this is not the case for all patients. There remains a subgroup of OSA patients who continue to suffer from symptoms of residual daytime

sleepiness despite adequate compliance with CPAP therapy,[32,33,36,37] although the actual prevalence of residual daytime sleepiness in CPAP-compliant patients remains undefined. Prospective observational data have demonstrated that as many as 20% to 30% of patients who are compliant with CPAP therapy for 7 hours or longer per night may still complain of subjective sleepiness (Epworth Sleepiness Scale score of >10) after 3 months of treatment.[32,33] In addition, many patients also may not achieve a normal level of objective alertness (as defined by the MSLT or MWT) or associated functional outcomes (as defined by the Functional Outcomes of Sleep Questionnaire [FOSQ]), despite seemingly adequate nightly use of CPAP therapy. The mechanisms responsible for this syndrome of residual daytime sleepiness also remain unclear but may in part be related to the oxidative injury effects of long-term intermittent hypoxemia on the sleep-wake cycle–promoting regions in the brain.[38]

Effect on Neurocognitive Function, Mood, and Quality of Life

Numerous studies have assessed the effects of OSA on neurocognitive functioning, mood, and quality of life.[9,30,39-50] Most randomized controlled studies demonstrate inconsistent improvements in several neurobehavioral performance parameters across the spectrum of disease severity.[23,30,31,35,39-41,51] For example, large-scale randomized controlled trials have demonstrated mild, transient improvements in several measures of executive function in patients with severe OSA, but similar improvements have not been consistently demonstrated in patients with less severe disease.[31] The data regarding the therapeutic effects of CPAP treatment on mood and quality of life are also variable and inconsistent, with many randomized trials demonstrating no clear benefits of CPAP therapy compared with placebo or conservative treatments in these parameters.[23]

One reason for the inconsistent improvements in neurocognitive function demonstrated with CPAP therapy is that the impact of OSA on neurocognitive function for most patients with OSA may be relatively small across the spectrum of disease severity. The Apnea Positive Pressure Long-Term Efficacy Study (APPLES) trial demonstrated that most patients with OSA did not have significant neurocognitive deficits and that the degree of deficit was only weakly associated with the degree of oxygen desaturation and not associated with the AHI.[52] Another possible explanation for the inconsistent effect of CPAP in improving outcomes associated with neurocognition, mood, and quality of life is the use of multiple, different measures of function to assess similar parameters. For example, there is nearly universal use of the Epworth Sleepiness Scale when assessing improvements in subjective sleepiness, yet there are multiple tests that are used across several studies to assess for improvements in mood, neurocognitive function, and quality of life. Further research is required to better define the role of CPAP therapy in alleviating these symptoms and deficits in susceptible OSA patients.

Despite the inconsistent data regarding improvements in neurocognitive function with CPAP use, several observational studies support a significant reduction in the incidence of motor vehicle accidents in patients with OSA following the initiation of CPAP therapy.[53] Although the actual time course to improved driving performance in real-life situations is not clear, driving simulator performance can improve in as little as two to seven nights of therapy. Similar to other aspects of neurobehavioral performance that may be adversely affected by OSA, many patients with OSA may continue to demonstrate impaired driving simulator performance despite several months of high adherence to CPAP therapy.[54] The explanation for this last finding is not completely clear, although it is likely that many patients may still not be adhering to PAP therapy enough on a nightly basis or achieving enough sleep on a regular basis to normalize their driving skills. Unfortunately, there is no specific threshold of CPAP use or duration of treatment that can accurately predict a given individual's fitness to safely drive a vehicle. Because the severity of OSA alone is not a reliable predictor of motor vehicle accident risk, the clinician must take into account several factors, including improvements in subjective symptoms and compliance with therapy, before determining a driver's ability to safely operate a motor vehicle.

Effect on Cardiovascular Disease

Although untreated OSA has been associated with an increased risk for hypertension and other cardiovascular diseases in certain populations, the literature and outcomes data supporting the beneficial effects of CPAP on cardiovascular outcomes have been inconsistent.[9,23,55,56] Several randomized clinical trials and meta-analyses have assessed the effects of CPAP on blood pressure.[57-59] Overall, CPAP treatment appears to attenuate the adverse effects of untreated OSA on daytime and nocturnal systolic and diastolic blood pressure and on 24-hour mean blood pressure. These data demonstrate that, compared with placebo, sham CPAP, or supportive therapy alone, CPAP treatment is associated with small (–1.8 to –3.0 mm Hg) but statistically significant improvements in diurnal mean arterial systolic and diastolic blood pressures. When considering pooled data, improvements in systolic and diastolic blood pressures have been observed both during the daytime (2.2 ± 0.7 mm Hg and 1.9 ± 0.6 mm Hg, respectively) and nighttime (3.8 ± 0.8 mm Hg and 1.8 ± 0.6 mm Hg, respectively).[57] In general, improvements in blood pressure with CPAP therapy have been associated with greater severity of baseline OSA (higher AHI), the presence of subjective daytime sleepiness, younger age, and greater adherence with CPAP use on a nightly basis.

One of the main limitations of the current studies evaluating the effect of CPAP use on blood pressure in patients with OSAS is that although these studies evaluated blood pressure as an outcome measure, several of the studies either did not include patients with hypertension or included patients with hypertension who were already adequately controlled on antihypertensive medications. More robust reductions and clinical improvements in blood pressure with CPAP therapy have been observed when evaluating data from studies that included patients with uncontrolled hypertension.[60] In patients with uncontrolled hypertension at baseline, the use of CPAP has been associated with significantly greater reductions in awake systolic and diastolic blood pressure (7.1 mm Hg and 4.3 mm Hg, respectively) compared with placebo or sham PAP therapy. These improvements have been observed even after controlling for several potential confounders, including severity of disease, daytime sleepiness, patient demographics, use of antihypertensive medications, CPAP adherence, and duration of CPAP therapy.

Few studies have compared the effects of CPAP and antihypertensive medication on blood pressure reduction in patients with OSA and hypertension. In one randomized controlled trial, medical treatment with valsartan (160 mg daily) alone without CPAP therapy reduced several parameters of blood pressure significantly more than CPAP therapy alone over an 8-week period.[61] Specifically, valsartan therapy demonstrated superior reductions in 24-hour mean arterial pressure (–2.1 ± 4.9 mm Hg with CPAP vs. 9.1 ± 7.2 mm Hg with valsartan; $P < .001$) as well as mean arterial blood pressures during the daytime and throughout the night compared with CPAP therapy. The addition of CPAP therapy to antihypertensive medication does appear to improve blood pressure control in some patients with resistant hypertension and moderate to severe OSA.[62] Based on limited data, the addition of CPAP therapy to a regimen of several antihypertensive medications improved 24-hour mean blood pressure (3.1 mm Hg; 95% confidence interval [CI], 0.6 to 5.6; $P = .02$) and 24-hour diastolic blood pressure (3.2 mm Hg; 95% CI, 1.0 to 5.4; $P = .005$), but did not result in a significant improvement in 24-hour systolic blood pressure (3.1 mm Hg; 95% CI, –0.6 to 6.7]; $P = .10$) compared with the control group. In addition, CPAP therapy resulted in a greater proportion of patients demonstrating a normal nocturnal "dip" in blood pressure compared with controls (35.9% vs. 21.6%; adjusted odds ratio [OR], 2.4; 95% CI, 1.2 to 5.1; $P = .02$). Similar to other studies, there was a significant dose-response effect, with greater nightly CPAP use being associated with greater improvements in 24-hour mean blood pressure, systolic blood pressure, and diastolic blood pressure. Although reductions in nocturnal blood have been observed more consistently in his patient population, improvements in daytime blood pressure with CPAP therapy have been less consistent.[63]

As noted previously, the presence of subjective daytime sleepiness has generally been associated with a more robust improvement in blood pressure with CPAP therapy. There is some bias in these data because most of the studies that have evaluated various outcomes have in fact predominantly assessed patients with OSA and such associated daytime sleepiness. Because more than half of all patients with OSA, including those with severe disease (AHI ≥30 events/hour) do not have associated daytime sleepiness, it would be important to determine whether treating patients with OSA who do not complain of subjective sleepiness improves blood pressure or reduces the incidence of hypertension and other cardiovascular morbidities. One large randomized controlled trial assessing CPAP therapy compared with conservative therapy in patients with moderate to severe OSA without daytime sleepiness found that CPAP therapy did not result in a statistically significant reduction in incident hypertension or cardiovascular events (nonfatal myocardial infarction or stroke, transient ischemic attack, congestive heart failure, or cardiovascular death) over a period of 4 years of follow-up.[64] When the data were stratified by CPAP adherence, however, patients using prescribed CPAP therapy for more than 4 hours per night did demonstrate a small but statistically significant ($P = 0.04$) reduction in the incidence of hypertension over the 4-year study period. Thus the benefit of treating patients with moderate to severe OSA who do not have symptoms of daytime sleepiness or cardiovascular disease remains to be better defined regarding the risk for future cardiovascular morbidity and mortality. In fact, however, the role of CPAP therapy in reducing the incidence of hypertension and other cardiovascular morbidity in OSA patients even with daytime sleepiness is also unclear because there are no large-scale long-term prospective data addressing this at this time.

The role of CPAP therapy in resolving or reducing the occurrence or reoccurrence of cardiac arrhythmias is uncertain. Several observational studies have demonstrated an association between OSA and atrial fibrillation as well as a higher risk for recurrence of atrial fibrillation after electrical cardioversion or catheter ablation therapy. These studies also have shown an association between increased adherence with CPAP therapy and a lower reoccurrence rate of atrial fibrillation after these procedures.[65-68] Because all of the current data regarding CPAP therapy and atrial fibrillation are based on observational studies, the role of CPAP as an adjunct treatment to improve atrial arrhythmia control remains uncertain. Although there may be an increased risk for ventricular arrhythmias (tachycardia and fibrillation) in some patients with untreated OSA, there are limited data on the effect of PAP therapy for reducing the incidence and prevalence of these events.[69] Thus the role of PAP therapy in reducing ventricular arrhythmias in patients with OSA is not clear.

There are several possible explanations for why CPAP therapy has not been demonstrated to result in more consistent and greater improvements in blood pressure and other cardiovascular outcomes in patients with OSA. First, most of the literature assessing the effects of CPAP on blood pressure has been based on small trials of relatively short duration (≤3 months). This duration of treatment, even in patients with underlying hypertension, may not be a long enough treatment time to improve blood pressure. Second, as mentioned previously, although several of the studies used blood pressure as an outcome measure, many of the studies enrolled patients without hypertension at baseline. Thus one would not necessarily expect to observe a change in blood pressure if hypertension was not present at the initiation of the studies. Also, most of the patients who had hypertension at enrollment were on antihypertensive medications during most of the studies, which is likely to attenuate the effect of CPAP on blood pressure decreases. Third, although improvements in blood pressure tend to be associated with better CPAP adherence, overall adherence in most studies have typically averaged between 4 and 5 hours per night. Thus inadequate nighttime CPAP use may limit the beneficial effect of therapy on blood pressure in those with and without hypertension. Fourth, even though OSA is found to be an independent risk factor for hypertension in many populations, hypertension is typically associated with several comorbid conditions that are also related to OSA. Thus treating OSA without treating the other comorbid conditions may not result in significant improvements in blood pressure or other cardiovascular outcomes. Fifth, it is possible that many patients with longstanding hypertension have fixed disease that CPAP therapy may not improve. Sixth, many studies use different definitions for hypopneas (i.e., associated with a 4 percentage-point oxygen saturation, 3 percentage-point oxygen desaturation, or 3 percentage-point oxygen desaturation with or without an associated electroencephalogram arousal), and it is difficult to compare such studies because changing the definition of the hypopneas also changes the definition of OSA as well as the severity of disease for many patients. Finally, not all untreated OSA patients are necessarily at similar risk for the development of hypertension.

Thus trials that are unable to stratify patients deemed to be at higher risk for hypertension may only demonstrate mean results across a given population. When, and if, biomarkers are discovered that may identify patients at higher risk for hypertension and other cardiovascular diseases, therapies such as CPAP may be targeted to at-risk populations that would be deemed to derive greater benefits from therapy.

Aside from the one randomized controlled trial evaluating the effects of CPAP on incident hypertension and cardiovascular disease in patients with moderate to severe OSA without daytime sleepiness, there are currently limited long-term randomized controlled data evaluating the effect of CPAP on any cardiovascular outcomes, including mortality. The most convincing long-term data regarding the potential beneficial effects of CPAP therapy on cardiovascular outcomes comes from Marin and colleagues,[70] who followed a large group of male OSA patients with a spectrum of OSA severity and associated daytime sleepiness in a prospective observational study over a period of 10 years. Their results demonstrated two important findings: (1) Compared with normal nonsnoring controls, patients with untreated severe OSA (defined as an AHI >30 events/hour) had a significantly increased incidence of both fatal and nonfatal cardiovascular events; and (2) CPAP treatment (>4 hours/night) in patients with severe OSA (AHI ≥30 events/hour) reduced the incidence of adverse cardiovascular outcomes and improved survival, demonstrating outcomes similar to normal controls. Similar improvements in outcomes with CPAP therapy were not observed in OSA patients with less severe disease because untreated mild to moderate OSA was not observed to be associated with increased risk for cardiovascular morbidity or mortality in this study. Another observational study also demonstrated improvements in cardiovascular mortality across a spectrum of OSA severity, although the data are limited by absence of a control group.[71]

Given the inconclusive nature of CPAP therapy on cardiovascular outcomes in general, the AASM Practice Parameters recommend CPAP therapy only as an adjunctive therapy to lower blood pressure in hypertensive patients with OSA.[9] Several other authorities and professional societies have recommended that further supporting data are required to better determine the role of CPAP therapy on improving cardiovascular outcomes before making recommendations for its use in various populations.[55,56]

Effect on Mild Obstructive Sleep Apnea

Most of the literature assessing the effects of CPAP on various outcomes has predominantly evaluated OSA patients with moderate to severe disease. Although approximately 28% of patients with mild disease (AHI = 5 to 14 events/hour) complain of subjective daytime sleepiness,[72] it remains unclear whether treating this group of patients with CPAP therapy improves their daytime symptoms. Results from the CPAP Apnea Trial North American Program (CATNAP) demonstrated that CPAP therapy significantly improved daytime symptoms as measured by the FOSQ compared with sham CPAP therapy in patients with mild to moderate OSA over an 8-week period of follow-up.[73] APPLES was a large multicenter randomized controlled trial comparing the neurocognitive effects of therapeutic CPAP with sham CPAP across the spectrum of OSA severity.[31] As expected, subjective daytime sleepiness and objective alertness as assessed by the MWT was improved by CPAP therapy at 6 months, but significant improvements in both of these parameters were only observed in patients with severe OSA (AHI ≥30 events/hour). In patients with moderate disease (AHI = 15 to 29 events/hour), improvements in subjective sleepiness, but not objective alertness, were observed after 6 months of therapy. In patients with mild disease, there were no significant improvements in objective alertness or subjective sleepiness after 2 and 6 months of CPAP therapy. Thus the role of CPAP therapy for this indication in patients with mild disease remains unclear based on the current data. It appears reasonable to initiate CPAP therapy in patients with daytime symptoms, but the decision to continue chronic therapy in this patients group should be based on a response to therapy. For patients with mild disease without daytime symptoms, it is not clear that treating these patients is beneficial or should be recommended based on the current data.

Effect on REM-Predominant Obstructive Sleep Apnea

The prevalence of REM sleep–related or REM-predominant OSA is unclear, in part because of the absence of a standard definition for this entity. This OSA variant tends to be more common in women, although it may affect adult patients of both genders across the age spectrum.[12,74] The association of this OSA variant with daytime or nighttime symptoms is not clear, but it appears that a subgroup of patients are affected. For patients who demonstrate this type of OSA and complain of daytime symptoms or nighttime sleep disturbance, it is unclear whether treatment with CPAP consistently improves daytime or nighttime symptoms. Limited observational data of CPAP therapy in symptomatic patients with such REM-predominant OSA have demonstrated significant improvements in daytime sleepiness, fatigue, and the FOSQ. These improvements with CPAP therapy were similar to those in patients with OSA not limited to REM sleep.[12] However, there are no randomized controlled data assessing any outcomes in this subgroup of patients, including cardiovascular disease outcomes.

Effect on Obstructive Sleep Apnea and Comorbid Diseases

CHF is a common disease with an estimated prevalence of concomitant OSA of approximately 33%. Two small randomized controlled trials demonstrated a beneficial effect of CPAP therapy on left ventricular ejection fraction (LVEF) in patients with concomitant OSA and CHF with systolic dysfunction.[75,76] Compared with optimal medical management alone, CHF patients with moderate to severe OSA showed left LVEF improvements of 5% to 9% over 1 to 3 months.[75,76] Since these earlier studies, several additional randomized controlled studies have assessed the effects of CPAP therapy on LVEF in CHF patients with and without systolic dysfunction.[77] Overall, CPAP therapy has shown statistically significant improvements in LVEF in patients with OSA and concomitant systolic dysfunction, with an average improvement in LVEF across studies of approximately 5%. In patients with diastolic CHF and concomitant OSA, CPAP therapy has not been associated with significant improvements in LVEF (1%). For patients with CHF and systolic dysfunction, one would expect this degree of improvement in LVEF to be associated with improvements in other outcomes based on trials of medical therapies for CHF. However, it is uncertain

whether the improvements in LVEF in patients with OSA and concomitant CHF translate into improvements in other important outcomes, such as reductions in hospitalizations and mortality. Most of the studies evaluating this patient population have been limited by small sample sizes and relatively short durations of follow-up (typically 12 weeks or less). Currently, two large randomized trials are evaluating the role of advanced PAP therapies in this population to determine whether PAP treatment can be used to enhance these important outcomes. Until these studies are completed, the role of CPAP therapy in patients with CHF and OSA to improve important outcomes beyond LVEF remains unclear. Patients with severe heart failure may also have concomitant central and obstructive sleep apnea. An interim analysis of this therapy suggests that advanced PAP therapies (such as adaptive servo ventilation) may have a negative effect on cardiovascular mortality in heart failure patients with predominantly Cheyne-Stokes breathing and a left ventricular ejection fraction of 45% or less (see Chapters 123 and 129).[78]

The *overlap syndrome* refers to the coexistence of OSA with COPD (see Chapter 119). The prevalence of OSA in patients with COPD appears to be similar to that of the general population. Prospective observational and retrospective studies have shown that untreated OSA in this patient group is associated with an increased risk for death and severe COPD exacerbations leading to hospitalizations compared with groups of COPD patients without concomitant OSA.[79,80] Observational data have shown that CPAP therapy in OSA patients with COPD has been associated with significant reductions in both acute exacerbations of COPD requiring hospitalizations and death, with outcomes similar to COPD patients without OSA. Increased adherence to CPAP therapy has been independently associated with reduced mortality in this patient population, whereas decreased CPAP adherence and increased age have been independently associated with increased mortality.[80] Observational data suggest that adherence to CPAP therapy for as little as 2 hours per night has been associated with a reduction in mortality in this group of patients. Given the current observational data, it is reasonable to recommend CPAP therapy in patients with the overlap syndrome, although given the absence of randomized controlled data in this patient population, the role of CPAP therapy to reduce exacerbations or improve mortality remains undefined.

The role of CPAP therapy in improving important outcomes associated with diabetes mellitus (short-term and long-term glucose control) in patients with concomitant OSA is unclear because most of the trials evaluating the use of CPAP in this patient population have yielded inconsistent results.[81,82] The role of CPAP as an adjunct therapy to improve weight loss is also uncertain, and adequate treatment of OSA has not been observed to result in enhanced weight loss in most studies.[83]

Comparison with Other Obstructive Sleep Apnea Treatments

Oral appliances (nonadjustable mandibular advancement devices and tongue-retaining devices) are typically recommended for patients with mild to moderate OSA as well as for patients with severe disease who fail or do not tolerate CPAP therapy. In general, CPAP therapy results in greater improvements in the AHI and degree of oxygen desaturation compared with oral appliance treatment. Despite these findings, improvements in daytime sleepiness tend to be similar between the two therapies. This may be related to greater overall compliance with oral appliance therapy compared with CPAP.[84,85] Comparisons of CPAP with oral appliance therapy for improvements in blood pressure are difficult. Although most of the pooled data suggest a favorable effect of oral appliance therapy on many parameters of blood pressure, most of the studies have been observational, with few head-to-head comparisons between the two treatments.[86,87] Thus, based on the current data, is it difficult to draw conclusions or make recommendations between the two therapies concerning the outcome of blood pressure control.

Oxygen Therapy

The risk for cardiovascular disease related to untreated OSA is dependent on numerous factors, including the severity of disease, as defined by the AHI, and the degree of associated oxygen desaturation. Several small studies have shown that nocturnal oxygen therapy alone can in fact improve both the AHI and degree of oxygen desaturation, although such therapy may be associated with a prolongation of apneas and hypopneas. CPAP, however, has been associated with greater improvements in the AHI compared with nocturnal oxygen therapy alone.[88] Further, a short-term (12 weeks) randomized controlled trial has shown that CPAP results in greater reductions in 24-hour mean arterial blood pressure compared with nocturnal oxygen (2 liters/minute) or supportive therapy without CPAP in patients with moderate to severe OSA and cardiovascular disease or multiple cardiovascular risk factors. Similar to other studies, decreases in blood pressure with CPAP therapy were relatively small compared with baseline or the control group (–2.4 mm Hg).[89] Oxygen therapy alone was not associated with any changes in blood pressure compared with baseline or the control group over the study period.

Outcomes Summary

CPAP consistently improves or resolves OSA events across the spectrum of OSA severity and improves symptoms of daytime sleepiness predominantly in patients with moderate to severe OSA. Improvements in other outcomes are inconsistent. Treatment with CPAP has been associated with small reductions in blood pressure, with greater reductions being observed in patients with poorly controlled or resistant hypertension. The role of CPAP therapy in reducing long-term cardiovascular risk or mortality in OSA is uncertain based on the current data. Finally, the role of CPAP in patients without daytime symptoms, cardiovascular disease, or cardiovascular risk factors across the spectrum of OSA severity is undefined.

ADHERENCE AND PROBLEMS WITH CONTINUOUS POSITIVE AIRWAY PRESSURE TREATMENT

In a perfect world, all patients with OSA would use their CPAP therapy all night, every night. Unfortunately, just like many therapies associated with other chronic diseases, adherence with CPAP therapy for OSA is far from perfect. Although there are no formal definitions of what constitutes adherence with CPAP therapy, most studies have arbitrarily defined adherence as use of CPAP greater than or equal to 4 hours

per night for 70% of the observed nights.[90] Using this definition, subjective adherence ranges between 65% and 90%, whereas objective measures of CPAP adherence have demonstrated use in the range of 40% to 83%.[91] Most studies have shown that patients usually overestimate their CPAP use by approximately 1 hour per night, a pattern that is observed in both new and long-term OSA patients.[90,92]

Short-term follow-up of OSA patients demonstrates that CPAP use patterns typically fall into two groups: (1) use of CPAP on more than 90% of the nights, with an average use time of greater than 6 hours per night, and (2) use of CPAP intermittently, with an average use of less than 3½ hours per night.[93] Early follow-up for patients newly initiated on CPAP therapy is important because these patterns of use can typically be identified within the first several days to several months of CPAP therapy.[86-89] Long-term objective follow-up has demonstrated that approximately 68% of OSA patients continue to use their CPAP therapy after 5 years.[92]

Some studies have suggested certain parameters that may predict greater short- and long-term adherence to therapy. Improved adherence has been associated with symptoms of subjective sleepiness (Epworth Sleepiness Scale score of >10), severity of OSA (AHI >30 events/hour), and average nightly adherence within the first 3 months of therapy. Reduced short- and long-term adherence has been observed in patients reporting problems during their initial night with CPAP therapy in the sleep laboratory.[92,94] Interestingly, although one might expect higher levels of CPAP pressure to predict poorer adherence, neither high nor low CPAP pressures have been shown to reliably predict CPAP use. Several studies have also associated African American race and lower socioeconomic status with poorer adherence with CPAP therapy, even in patients with standardized access to care and treatment.[95-97] The reasons for this last observation are not clear.

Role of Objective Adherence Monitoring and Limitations with Current Technology

Unfortunately, when taken together, most studies have not been able to identify factors that consistently predict short- or long-term adherence with CPAP therapy.[26,90,98-101] Because adherence with PAP therapy tends to be suboptimal, subjective adherence tends to overestimate objective PAP use, there are no consistent early predictors of PAP adherence, and PAP adherence patterns tend to be determined early in most patients, professional societies currently recommend, and many payer policies require, objective adherence data review to document adherence with therapy and identify problems that can be addressed.[3] Although most randomized controlled trials have used objective adherence data to monitor outcomes related to PAP therapy, the overall effect of assessing objective compliance data for all patients on PAP therapy is uncertain.

Most of the PAP manufacturers have developed sophisticated online software programs for monitoring several parameters of PAP therapy, including nightly adherence, efficacy of therapy (residual AHI), and problems with mask fit (primarily amount of air leak). Although there are several potential advantages to these programs, the technologies also have several potential limitations. To improve the effect of these technologies on meaningful patient outcomes, several improvements will be required, including (1) standardization of respiratory event and leak definitions among manufacturers as well as validation of the device outputs compared with PSG; (2) improved access of PAP adherence data for frontline providers, including determining ways to more easily integrate PAP adherence data into the various electronic medical record software programs; and (3) education of non–sleep specialists on interpretation of the available adherence information.

Interventions to Promote Adherence

Typical problems that may lead to reduced adherence with CPAP therapy include claustrophobia, nasal congestion, and poor mask fit, leading to leaks and skin irritation. Several interventions have been proposed and instituted in an attempt to improve adherence with CPAP therapy (Table 115-3).

The most consistent intervention that has been associated with improved CPAP adherence in most PAP-naïve patients is systematic education. Several approaches, including provider and home-based education of the patient and spouse, supportive care at therapy initiation or follow-up, phone calls, home-based videos, and daylong educational programs, have been associated with improved adherence, although no one intervention has been demonstrated to be consistently beneficial in all patient groups. In general, increased intensity of patient education or frequency of health provider contact have been associated with improved CPAP adherence.[9,23] Overall these educational interventions tend to improve CPAP adherence by approximately 35 to 50 minutes per night,[102] although the effects of these interventions on other important outcomes such as daytime sleepiness, quality of life, and cardiovascular disease and risk are unclear. Several behavioral approaches have also been associated with improved adherence. In general, these behavioral approaches, including motivational interviewing and cognitive behavioral therapy delivered in individual or group settings, have been associated with an average improvement in adherence of 1.5 hours per night. The overall effect of these behavioral approaches on CPAP adherence is not well defined because the data supporting these approaches are of lower quality than the data supporting the previously discussed educational interventions.

The data evaluating the effects of heated humidification on adherence to CPAP therapy remain controversial. Although there are some studies that demonstrate that the addition of heated humidification can improve adherence to CPAP therapy, there are several studies demonstrating no improvement in adherence with this intervention.[9,103-106] Patients who tend to benefit the most from the addition of heated humidification are those with symptoms of nasal congestion or rhinitis. Limited data evaluating the role of heated tubing have shown no improvements in adherence in patients with and without nasopharyngeal complaints.[107] The role of nasal steroids with or without heated humidification therapy, especially in unselected CPAP-naïve patients with OSA, remains unclear because many studies have demonstrated little benefit of this intervention in improving CPAP use.[106,108]

CPAP delivery interfaces, or masks, come in several shapes and sizes, including nasal masks, full-face (oronasal) masks that cover both the nose and the mouth, nasal pillows that fit into the nostrils, and oral interfaces that fit into the mouth. Some studies have observed a negative effect of oronasal masks on CPAP compliance, whereas other studies have not confirmed these findings.[109] Oronasal masks may be better for patients with chronic nasal congestion or obstruction, for those patients who are predominantly mouth breathers, and

Table 115-3 Effects of Interventions on Positive Airway Pressure Adherence

Intervention	Effect on PAP Adherence	Comments
Education and supportive care	Beneficial	Various approaches helpful, including phone calls, office and home visits, and individual and group sessions Best intervention, or combination, unclear
Behavioral therapies	Beneficial	Various therapies helpful, including motivational interviewing and CBT Most interventions studied in addition to education Best intervention, or combination, unclear
Heated humidification	Beneficial	Some data support improved adherence Most helpful for patients with nasal congestion or rhinitis Addition of nasal steroids not helpful
Advanced PAP (bilevel, EPR, and APAP)	No benefit	Not associated with improved compliance BiFlex may be the exception in CPAP noncompliant patients
Mask type	Unclear	Best mask type unclear Changing masks may alter effective PAP pressure
Hypnotics	Unclear	Eszopiclone may improve PAP titration efficacy and 6-month compliance Data do not support other hypnotics
Telemedicine	Unclear	Limited data suggest benefit, whereas other data do not support approach
Adherence monitoring	Unclear	Objective adherence monitoring recommended, but no clear data that the intervention itself improves compliance
Sleep specialist care	Unclear	Observational studies support approach RCTs show no advantage in uncomplicated OSA

APAP, Autotitrated positive airway pressure; CBT, cognitive behavioral therapy; EPR, expiratory pressure relief; OSA obstructive sleep apnea; PAP, positive airway pressure; RCT, randomized controlled trial.

for patients requiring higher CPAP pressures when mask leak is an issue. Nasal pillows have typically not been recommended for CPAP settings of more than 12 cm H_2O owing to the potential for interface leak, although more recent data show that select patients may do well with a nasal pillows interface even with higher PAP settings.[110] Overall, although proper mask fit may be crucial to the initial and ongoing acceptance of CPAP therapy, the optimal form and type of CPAP delivery interface remain unclear.[111,112] In general, the best interface for a given patient (which tends to correlate to the best adherence with therapy) is the one that the patient is most comfortable wearing.

Changing interfaces after a problem has developed has not been shown to consistently improve long-term adherence in various studies, although from a clinician's standpoint attention to mask complaints and changing masks when problems arise can improve adherence in select patients. The provider should be aware that changing the mask type from nasal to oronasal or vice versa might change the necessary effective treatment pressure that was initially identified during an in-laboratory titration.[113] Thus, for patients on fixed PAP therapy, the clinician should consider the need to adjust the pressure or to have the patient perform an in-laboratory PAP titration if problems that could result in reduced adherence with therapy persist after a mask change has been instituted.

Because many patients may complain of sleep disruption or difficulty initiating sleep during the first few days to weeks of CPAP therapy, several studies have evaluated the use of prescription hypnotics to improve adherence to CPAP treatment either in the sleep laboratory during a PAP titration study or during the first few weeks of therapy. Although some studies have demonstrated that, in newly diagnosed patients with severe OSA, treatment with eszopiclone 3 mg before an overnight titration study or during the first 14 days of PAP therapy has been associated with improved quality of CPAP titrations (greater proportion of patients with optimal or good titrations) or improved adherence to CPAP therapy over the first 6 months of treatment, respectively, these results are not typical of most of the literature regarding the use of hypnotics as adjunctive therapies to improve adherence to PAP therapy.[13,114] When compared with placebo or usual care, other randomized controlled studies have demonstrated no significant benefits, but no significant adverse effects of other hypnotic therapies (zaleplon or zolpidem), on CPAP adherence.[115,116] As with most studies, the data evaluating the effects of hypnotics on CPAP adherence have looked at relatively short-term adherence in specialized centers of care. The ability to generalize these data to a typical clinical population and office setting is uncertain based on the current literature, and care should be used when applying this approach to a given patient or population. Given the limited data in patients with OSA, the use of short-term or chronic hypnotics should generally be avoided in patients with OSA.

Role of the Sleep Specialist in Improving Adherence

Several retrospective and observational studies have shown that sleep specialist consultation, before an in-laboratory sleep study or during the initiation and follow-up of CPAP therapy, has been associated with improved CPAP adherence and other important outcomes, such as patient satisfaction and timeliness of care.[91,117] Alternatively, three randomized controlled trials in symptomatic patients with a high clinical suspicion of uncomplicated moderate to severe OSA demonstrated that management by either a specially trained nurse, nurse–primary

care physician team, or primary care physician resulted in outcomes (CPAP adherence and improvements in daytime sleepiness) that were similar to management by sleep specialists.[118-120] In addition to similar CPAP adherence, all of these studies demonstrated a significant cost savings in the non–sleep specialist group. Thus the data supporting the role of the sleep specialist in the treatment and overall management of all patients with uncomplicated moderate to severe OSA is not well defined based on the current literature. More research is necessary to better determine which groups of patients with OSA may receive the most benefit from sleep specialist management of CPAP therapy.

Current recommendations, based predominantly on expert opinion, suggest that patients should have initial office follow-up during the first few weeks of prescribed CPAP therapy. Thereafter patients using CPAP should be followed on an annual basis and as needed to troubleshoot problems as they arise.[9,23] Centers for Medicare and Medicaid Services has defined its own rules regulating how and when patients on CPAP should have office follow-up, and commercial payers have also adopted their own policies on CPAP follow-up and adherence monitoring. Based on the current outcomes literature, the optimal method or schedule for short- or long-term follow-up is not clear. Clinicians must determine appropriate follow-up based on a given patient's response to therapy as well as payer policies that may guide requirements to continue treatment.

Technology to Improve Adherence

In addition to technologic advancements in the delivery of PAP therapy that are discussed later in this chapter, several applications of technology have been employed in an attempt to improve PAP adherence. Interventions include the use of online PAP adherence monitoring software as described earlier, telemedicine, and patient interactive technologies. As noted previously, most patients overestimate their compliance with therapy and thus objective monitoring of CPAP therapy has been recommended by the AASM.[3,9] Although the literature supports the concept that CPAP use can be reliably determined by CPAP tracking systems, the role of objectively measuring PAP adherence and its effect on improving adherence in all patients are uncertain.[121] Limited data suggest that online monitoring of PAP adherence and the use of a telemedicine management strategy may be associated with improved adherence, although more data are required to better define the role of this approach.[122] Finally, although several PAP device manufacturers have developed software (smartphone and computer-based applications) aimed at improving patient involvement with their CPAP therapy, there are currently no randomized trials that have objectively evaluated the effect of this approach on adherence or, in fact, any outcomes.

TECHNOLOGIC ADVANCEMENTS IN THE DELIVERY OF POSITIVE AIRWAY PRESSURE THERAPY FOR OBSTRUCTIVE SLEEP APNEA

Although CPAP remains the mainstay of therapy for OSA, there are several other methods of delivering PAP therapy. This section of the chapter focuses on technologic advancements in the delivery of positive pressure therapy, including bilevel PAP, EPR devices, and APAP.

Bilevel Positive Airway Pressure Therapy

The potential benefits of bilevel PAP in treating patients with OSA were first described in 1990.[123] As opposed to CPAP, which delivers a fixed pressure throughout the respiratory cycle, bilevel PAP therapy allows the independent adjustment of the expiratory positive airway pressure (EPAP) and the inspiratory positive airway pressure (IPAP). In its initial description, bilevel PAP therapy demonstrated that obstructive events could be eliminated at a lower EPAP compared with conventional CPAP pressures.[123] Bilevel PAP is typically titrated during an attended in-laboratory sleep study. As is the case for CPAP titrations, the current guideline recommendations for bilevel PAP titration strategies are based on consensus opinion.[8] Although intuitively one would predict that bilevel PAP would increase adherence by reducing expiratory pressure–related discomfort and side effects, there are in fact no objective outcomes studies that show that bilevel PAP improves adherence and daytime sleepiness compared with CPAP in patients with uncomplicated OSA.[9,23,124]

Newer bilevel PAP systems have been introduced by several companies. The BiFlex device (Respironics, Murrysville, Pa.) differs from conventional bilevel systems in two major respects. First, the inspiratory pressure is reduced slightly near the end of inspiration, and the expiratory pressure is slightly reduced near the beginning of expiration. Second, the amount of pressure relief change of the EPAP during expiration is proportional to patient effort. Although the data regarding the use of traditional bilevel and BiFlex therapies do not demonstrate any advantages over CPAP therapy in patients with newly diagnosed OSA, one study has demonstrated a potential role for BiFlex therapy in patients who are noncompliant with CPAP therapy.[125] Ballard and colleagues studied a large group of OSA patient who were noncompliant with CPAP therapy despite significant education, attention to proper mask fitting, and the addition of heated humidification. After 3 months of therapy, those patients randomized to BiFlex therapy demonstrated significantly better nightly adherence (P =.03) compared with those who were randomized to continuing on standard CPAP therapy. Importantly, because BiFlex technology provides PAP through its own unique algorithm, these findings are specific to the BiFlex devices and cannot be generalized to other non-bilevel PAP therapies.

Overall, bilevel PAP therapy remains a reasonable option for CPAP-intolerant patients, patients with OSA with concurrent respiratory disease (e.g., COPD), and patients with obesity hypoventilation syndrome.[2,9,23] The role of bilevel PAP therapy, and its variants, in otherwise uncomplicated OSA remains unclear.[3,126]

Expiratory Pressure Relief Systems

A common complaint in many patients with OSA using CPAP is the uncomfortable feeling of exhaling against positive pressure. This consequence is one potential barrier to the long-term acceptance of CPAP therapy. Several PAP manufacturers have developed EPR systems in an attempt to remedy this potential problem. EPR device technologies allow pressure relief during exhalation with the goal of making CPAP therapy more comfortable. EPR technologies briefly reduce the CPAP pressure, between 1 and 3 cm H_2O, during exhalation and then return the pressure to its set CPAP setting before the initiation of inspiration. Certain EPR technologies monitor the patient's airflow during exhalation and reduce the

expiratory pressure in response to the airflow and patient effort. The amount of pressure relief varies on a breath-by-breath basis, depending on the actual patient's airflow, and is also dictated by the patient's preference setting on the device.

Although several PAP manufacturers have developed EPR devices for the marketplace, only the Philips Respironics (Respironics, Murrysville, PA) technology (C-Flex) has been evaluated in the peer-reviewed literature.[120-128] Several randomized controlled trials have evaluated the role of C-Flex technology compared with standard CPAP therapy in patients with uncomplicated, predominantly moderate to severe OSA. Overall, the use of such C-Flex technology at fixed pressure relief settings between 1 and 3 cm H_2O has not been associated with improved adherence in either parallel or crossover trials.[129] In addition, improvements in other commonly measured outcomes (subjective sleepiness, objective alertness, vigilance, or residual OSA) were similar to, but not better than, standard CPAP therapy. C-Flex therapy has not been shown to offer significant benefits in that subgroup of patients who require CPAP pressures of 9 cm H_2O or greater. Based on these data, the routine use of C-Flex technology is not recommended as a method to improve compliance or other major outcomes compared with fixed CPAP therapy. Further randomized controlled trials are necessary to determine whether this technology offers any objective advantages over fixed CPAP therapy in select groups of patients.

Autotitrating Positive Airway Pressure

APAP (also known as auto-, automated, autoadjusting, or automatic positive airway pressure) incorporates the ability of the PAP device to detect and respond to changes in upper airway flow and resistance in real time.[130] This section focuses on the literature related to APAP in the treatment of patients with previously diagnosed OSA because there is currently little evidence to support the use of APAP technology for the diagnosis of OSA.[131]

Currently available APAP devices use proprietary algorithms to noninvasively detect and respond to variations in patterns of upper airway inspiratory flow or resistance. Most APAP machines monitor a combination of changes in inspiratory flow patterns, including inspiratory flow limitation, snoring (indirectly measured through mask pressure vibration), reductions of airflow (hypopnea), and absence of flow (apneas), using a pneumotachograph, nasal pressure monitors, or alterations in compressor speed. Most units detect flow limitation through proprietary algorithms using flow-versus-time profiles to determine a flattening index. The other less commonly used technology uses the forced oscillation technique method, which is an alternative method that detects changes in patterns of upper airway resistance or impedance.[132-134] Because the forced oscillation technique method measures changes in upper airway resistance that are independent of patient activity and ventilatory effort, this technology has the potential advantage of better differentiating central apneas from obstructive apneas or mask leak. There are currently no peer-reviewed data to substantiate efficacy of such detection, or clinical outcomes, with such technology.

When upper airway flow or impedance changes have been detected, the APAP devices use proprietary algorithms to automatically increase the pressure until the flow or resistance has been normalized. After a therapeutic pressure has been achieved, the APAP devices typically reduce pressure until flow limitation or increases in airway resistance resume. Most devices have a therapeutic pressure range between 4 and 20 cm H_2O, giving the clinician the ability to adjust the upper and lower pressure limits based on the clinical conditions and the patient's response to therapy. This should be differentiated from bilevel PAP or autobilevel PAP (discussed later), in which a separate IPAP and EPAP are set with changes in pressure across each respiratory cycle. Similar to CPAP, expiratory relief and other pressure delivery modifications are available for APAP technologies, although these additional pressure modifications have not be shown to consistently improve several APAP-related outcomes, including in-laboratory titration success, PAP adherence, or daytime sleepiness.[135-137] Because pressures changes occur throughout the sleep period, some have postulated that APAP devices may actually increase sleep fragmentation.[138] This concern has not been substantiated in studies evaluating changes in sleep structure or in clinical trials that have measured subjective sleepiness as a main outcome. Specifically, the frequency of microarousals and sleep fragmentation induced by APAP devices appears to be small,[139] and clinical outcomes related to subjective sleepiness also show no significant differences compared with conventional CPAP therapy.[18,140-143]

Currently available APAP machines have several potential limitations. Most flow- and pressure-based APAP devices are somewhat limited in their ability to distinguish between central and obstructive apneas as well as large mask leaks.[144-147] These flow patterns are "interpreted" by these types of devices as an absence of flow, which in the cases of central apneas and leaks may erroneously lead to increases in pressure and worsening of the central events or leaks. Newer APAP algorithms appear to be better at differentiating obstructive from central events as well as compensating for large mask leaks. Also, the ability of the APAP devices to respond to sustained hypoventilation in the absence of upper airway obstruction is unclear because most APAP studies have excluded patients at high risk for hypoventilation, including patients with obesity hypoventilation syndrome or chronic respiratory diseases. Given these potential limitations in technology, as well as the exclusion of patients with many comorbid diseases from the randomized trials comparing APAP to in-laboratory titrated CPAP therapy, the current AASM Practice Parameters regarding the use of APAP recommend that APAP devices only be used for patients with uncomplicated moderate to severe OSA.[131,148,149] APAP devices typically should *not* be used in the patients with comorbid medical conditions that could potentially affect their respiratory patterns (complicated OSA), including patients with CHF, patients with lung diseases such as COPD; and patients expected to have nocturnal arterial oxyhemoglobin desaturation due to conditions other than OSA (e.g., obesity hypoventilation syndrome and other hypoventilation syndromes). Patients who do not snore (either because of palatal surgery or naturally) should not be titrated with an APAP device that relies on vibration or sound in the device's algorithm.[131,148,149] Finally, APAP devices are not recommended for split-night titrations given the lack of data to support such a practice.

There have been several randomized controlled trials that have compared APAP technology to conventionally titrated CPAP therapy for the treatment of uncomplicated OSA.[18,19,126,134,140-143,150-159] Compared with standard fixed CPAP therapy, APAP devices as a group are almost always

associated with a reduction in mean pressure across a night of therapy in the range of 2 to 2.5 cm H_2O, although peak pressures through the night tend to be higher than fixed CPAP therapy. Aside from these differences, APAP and standard CPAP are similar with regard to improvements in several outcomes, including objective adherence, ability to eliminate respiratory events, and subjective daytime sleepiness as measured by the Epworth Sleepiness Scale.[160] There are few data regarding improvements in blood pressure with APAP therapy and no long-term data regarding any cardiovascular outcomes. These findings have been consistently demonstrated for APAP therapy used as a primary chronic therapy and for APAP used for a short therapeutic trial to determine a fixed CPAP setting for ongoing CPAP therapy.

Most of the literature concerning APAP technology as a treatment for OSA has evaluated patients with uncomplicated predominantly moderate to severe OSA (AHI ≥15 events/hour), and therefore the results and recommendations that have been reviewed predominantly apply to this group of patients. The data comparing efficacy of APAP versus attended in-laboratory titrated CPAP in patients with mild OSA (AHI = 5 to 14 events/hour) are more limited.[159,160] Based on the available information, there appear to be similar improvements in important outcomes, including resolution of sleep-disordered breathing and daytime sleepiness and adherence with therapy between APAP and CPAP, even in patients with more mild disease, although it is difficult to make reliable recommendations concerning the use of APAP for this subgroup of patients.

Although the use of APAP as a therapy with or without changing the patient to a fixed CPAP device has also been well described, the optimal method for determining treatment success is controversial. Most of the newer PAP devices calculate several parameters, including device use time, an AHI, and leak data. Compliance with PAP therapy can be reliably determined using the various PAP tracking systems, but the validity of the PAP-calculated AHI data are not as easy to interpret because the various PAP manufacturers define respiratory events differently from each other and differently from the standard scoring definitions used by the AASM. In general, most studies comparing the APAP-calculated AHI to a PSG-determined AHI show that the PAP-calculated AHI tends to overestimate the AHI, especially at the lower end of the AHI spectrum.[161,162] In general, PAP-calculated AHIs of less than 10 events/hour tend to correlate with adequately treated sleep-disordered breathing events and have been associated with improved outcomes in randomized controlled trials.[121] This is especially true when these findings are associated with the resolution of nighttime snoring and daytime symptoms. Because the various proprietary APAP algorithms are far from perfect for detecting and resolving all sleep-disordered breathing events, the clinician should consider an in-laboratory attended PAP titration study when a patient is having difficulty with unattended APAP therapy or when residual daytime symptoms persist even if the APAP-calculated parameters suggest adequately treated OSA syndrome.[163]

Autobilevel Therapy for Obstructive Sleep Apnea

Autobilevel therapy has also been developed that, using proprietary algorithms, automatically adjusts both the EPAP and IPAP in response to sleep-disordered breathing events. Limited data indicate that, compared with CPAP, autobilevel therapy results in similar compliance and other important outcomes in patients who have had poor initial experiences with CPAP therapy.[164,165] There is currently no peer-reviewed literature evaluating outcomes with autobilevel therapy for OSA in PAP-naïve patients. Thus, unlike non-autobilevel PAP therapy, no recommendations can be made for autobilevel PAP therapy for treating patients with OSA.

In summary, APAP technologies appear to be as effective as conventional fixed CPAP therapy when used for treatment in attended and unattended settings in patients with moderate to severe uncomplicated OSA.[131] Although APAP technologies as a group reduce the mean treatment pressure across the night, they appear to result in similar objective adherence and improvements in other important clinical outcomes compared with in-laboratory titrated CPAP therapy. Although APAP therapy has demonstrated some shortcomings in the peer-reviewed literature, the technology is rapidly advancing. The main benefits of APAP technology in the future will likely be the ability to provide more rapid treatment to patients with uncomplicated OSA and possibly the saving of health care dollars by eliminating some attended in-laboratory sleep studies that are typically required for CPAP titrations.[18,19,141]

CLINICAL PEARLS

- CPAP is the first-line therapy for patients with moderate to severe OSA, especially for those with daytime symptoms.
- CPAP therapy consistently resolves sleep-disordered breathing events and improves symptoms of daytime sleepiness in symptomatic patients, especially for patients with moderate to severe disease. There are inconsistent data concerning the benefits of CPAP therapy with regard to neurocognitive function, mood, quality of life, and cardiovascular outcomes across the spectrum of disease severity. The data regarding the benefits of CPAP therapy in patients with more mild disease are even more controversial, especially in those without daytime symptoms or underlying cardiovascular disease.
- The role of CPAP therapy for patients without associated daytime symptoms across the spectrum of OSA severity is unclear based on the current data. Most randomized controlled trials in this patient group have failed to demonstrate improvements in important outcomes, including blood pressure control, cardiovascular morbidity and mortality, neurocognitive function, and quality of life.
- Adherence with CPAP therapy is suboptimal for many patients, although improvements in adherence have been consistently associated with systematic education with and without behavioral therapy. The roles of other interventions, including heated humidification, hypnotics, and telemedicine, to improve adherence to CPAP therapy are unclear based on limited or inconsistent outcomes data from observational and randomized controlled trials.
- The roles of advanced PAP technologies, including EPR and bilevel PAP pressure devices, are not clear because they have typically not been associated with improved adherence, daytime sleepiness, or quality of life in patients with OSA.
- APAP used in an unattended setting, either to determine a fixed CPAP setting or as a primary treatment, is reasonable therapy for patients with moderate to severe OSA without underlying comorbidities. Most of the data on APAP therapy have been limited to patients with daytime sleepiness; thus the role of APAP therapy in patients without associated daytime sleepiness is not clear.

SUMMARY

CPAP therapy remains the mainstay of treatment of patients with moderate to severe OSA, especially those patients with daytime sleepiness. The role of PAP therapy in patients with OSA in the absence of daytime sleepiness is not clear. Despite its potential to improve several clinical outcomes including daytime sleepiness, neurocognitive dysfunction, quality of life, and blood pressure, long-term adherence with therapy remains suboptimum. Newer technologies such as APAP have the potential to improve the treatment of OSA, with most data demonstrating that this technology is as effective as in-laboratory titrated CPAP in patients with uncomplicated moderate to severe OSA. Although the role of APAP in the treatment of OSA is still not well defined, it has the potential to improve the delivery of PAP therapy by replacing laboratory-based PAP titrations in patients with uncomplicated OSA, thus reducing the current sleep laboratory waiting times and potentially reducing health care spending on in-laboratory studies. Other technologic advancements such as EPR and bilevel PAP devices are supported by limited data and appear to offer no advantages over conventionally titrated CPAP therapy in most patients with OSA.

Selected Readings

Barbe F, Duran-Cantolla J, Sanchez-de-la-Torre M, et al. Effect of continuous positive airway pressure on the incidence of hypertension and cardiovascular events in nonsleepy patients with obstructive sleep apnea: a randomized controlled trial. *JAMA* 2012;**307**(20):2161–8.

Billings ME, Auckley D, Benca R, et al. Race and residential socioeconomics as predictors of CPAP adherence. *Sleep* 2011;**34**(12):1653–8.

Chai-Coetzer CL, Antic NA, Rowland LS, et al. Primary care vs specialist sleep center management of obstructive sleep apnea and daytime sleepiness and quality of life: a randomized trial. *JAMA* 2013;**309**(10):997–1004.

Fava C, Dorigoni S, Dalle Vedove F, et al. Effect of CPAP on blood pressure in patients with OSA/hypopnea a systematic review and meta-analysis. *Chest* 2014;**145**(4):762–71.

Holmqvist F, Guan N, Zhu Z, et al. Impact of obstructive sleep apnea and continuous positive airway pressure therapy on outcomes in patients with atrial fibrillation: results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). *Am Heart J* 2015;**169**(5):647–54.

Ip S, D'Ambrosio C, Patel K, et al. Auto-titrating versus fixed continuous positive airway pressure for the treatment of obstructive sleep apnea: a systematic review with meta-analyses. *Syst Rev* 2012;**1**:20.

Johnson KG, Johnson DC. Treatment of sleep-disordered breathing with positive airway pressure devices: technology update. *Med Devices (Auckl)* 2015;**8**:425–37.

Lettieri CJ, Shah AA, Holley AB, et al. Effects of a short course of eszopiclone on continuous positive airway pressure adherence: a randomized trial. *Ann Intern Med* 2009;**151**(10):696–702.

Marin J, Carrizo S, Vincente E, Agusti A. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* 2005;**365**(9464):1046–53.

Marin JM, Soriano JB, Carrizo SJ, et al. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. *Am J Respir Crit Care Med* 2010;**182**(3):325–31.

Parthasarathy S, Subramanian S, Quan SF. A multicenter prospective comparative effectiveness study of the effect of physician certification and center accreditation on patient-centered outcomes in obstructive sleep apnea. *J Clin Sleep Med* 2014;**10**(3):243–9.

Phillips CL, Grunstein RR, Darendeliler MA, et al. Health outcomes of continuous positive airway pressure versus oral appliance treatment for obstructive sleep apnea: a randomized controlled trial. *Am J Respir Crit Care Med* 2013;**187**(8):879–87.

Sanchez-de-la-Torre M, Nadal N, Cortijo A, et al. Role of primary care in the follow-up of patients with obstructive sleep apnoea undergoing CPAP treatment: a randomised controlled trial. *Thorax* 2015;**70**(4):346–52.

Schwab RJ, Badr SM, Epstein LJ, et al. An official American Thoracic Society statement: continuous positive airway pressure adherence tracking systems. The optimal monitoring strategies and outcome measures in adults. *Am J Respir Crit Care Med* 2013;**188**(5):613–20.

Sun H, Shi J, Li M, Chen X. Impact of continuous positive airway pressure treatment on left ventricular ejection fraction in patients with obstructive sleep apnea: a meta-analysis of randomized controlled trials. *PLoS ONE* 2013;**8**(5):e62298.

Weaver TE, Mancini C, Maislin G, et al. Continuous positive airway pressure treatment of sleepy patients with milder obstructive sleep apnea: results of the CPAP Apnea Trial North American Program (CATNAP) randomized clinical trial. *Am J Respir Crit Care Med* 2012;**186**(7):677–83.

Wozniak DR, Lasserson TJ, Smith I. Educational, supportive and behavioural interventions to improve usage of continuous positive airway pressure machines in adults with obstructive sleep apnoea. *Cochrane Database Syst Rev* 2014;(1):CD007736.

A complete reference list can be found online at ExpertConsult.com.

Chapter
116

Medical and Device Treatment for Obstructive Sleep Apnea: Alternative, Adjunctive, and Complementary Therapies

Susheel P. Patil; Ephraim Winocur; Luis Buenaver; Michael T. Smith

Chapter Highlights

- Medical treatments for obstructive sleep apnea (OSA) can be categorized on the basis of the pathophysiologic mechanisms (i.e., anatomic, neuromuscular, and neuroventilatory control) that the interventions generally target. Stratifying treatments by pathophysiologic target may be particularly useful in personalizing therapy for patients with OSA.
- The current and emerging medical and device therapies available for the treatment of OSA generally are indicated as alternative therapies when traditional therapies for OSA are poorly tolerated, or as adjunctive treatment to more standard therapy.
- Therapeutic interventions and strategies such as weight loss, positional therapy, hypoglossal nerve stimulation, use of expiratory nasal resistors, and oral pressure therapy can successfully treat OSA in the appropriate patient. Other approaches such as myofunctional therapy, nasopharyngeal stenting, high nasal flow therapy, application of compression stockings, and pharmacotherapy are not proven as efficacious treatment for OSA and currently should be considered as experimental, possibly alternative, or at times adjunctive therapies.
- Oxygen therapy alone has not been shown to improve outcomes in OSA. In fact, use of supplemental oxygen has been associated with increased duration of apneic episodes and the development of hypercapnia.
- A limited number of complementary and alternative medicine approaches also have been studied for the treatment of OSA.
- Even with otherwise effective treatment for OSA, some patients may experience persistent sleepiness despite adequate sleep time and may be appropriate candidates for adjunctive stimulant pharmacotherapy.

Obstructive sleep apnea (OSA) is a highly prevalent disorder that increases cardiovascular and metabolic disease–related morbidity and mortality, contributes to the risk of motor vehicle and occupational accidents, and reduces occupational productivity. The recurrent episodes of upper airway obstruction that characterize OSA have been attributed to both anatomic loads (e.g., retrognathia, micrognathia, excess pharyngeal muscosal tissues, large parapharyngeal fat pads) on the upper airway and impairments in neuromuscular responses.[1] Potential treatments for OSA can be considered in light of the mechanisms identified to be important in its pathogenesis (Figure 116-1).

Therapy for OSA traditionally has targeted reductions in anatomic loads on the upper airway using continuous positive airway pressure (CPAP) therapy, oral appliances, upper airway surgery, and weight reduction. However, therapeutic approaches for OSA targeting neuromuscular (e.g., muscle responsiveness) and neuroventilatory (e.g., arousal threshold, apnea threshold, and loop gain) mechanisms have also been investigated; these include electrical stimulation of the hypoglossal nerve, myofunctional therapy, and pharmacotherapy. Given known difficulties with adherence to conventional therapies, particularly CPAP and use of oral appliances, for OSA,[2] active investigation of alternative and adjunctive therapies continues.

This chapter presents an overview of medical and device treatments for OSA based on the pathophysiologic mechanisms (i.e., anatomic, neuromuscular, and neuroventilatory control) that the interventions target. Stratifying treatments based on pathophysiologic targets may be useful in personalizing optimal therapy for patients with OSA. Patient preference also plays an important role in therapy decisions.[3] Such treatments may be considered in three categories: primary, alternative, and adjunctive. In this chapter, *primary treatment* or *therapy* is defined as a treatment that should be considered as a first-line therapy. *Alternative treatment* or *therapy* refers to a therapy that should be considered when a primary therapy is poorly tolerated or ineffective. *Adjunctive therapy* is defined as a treatment that should be used in conjunction with a primary or alternative therapy. *Investigational treatment* or

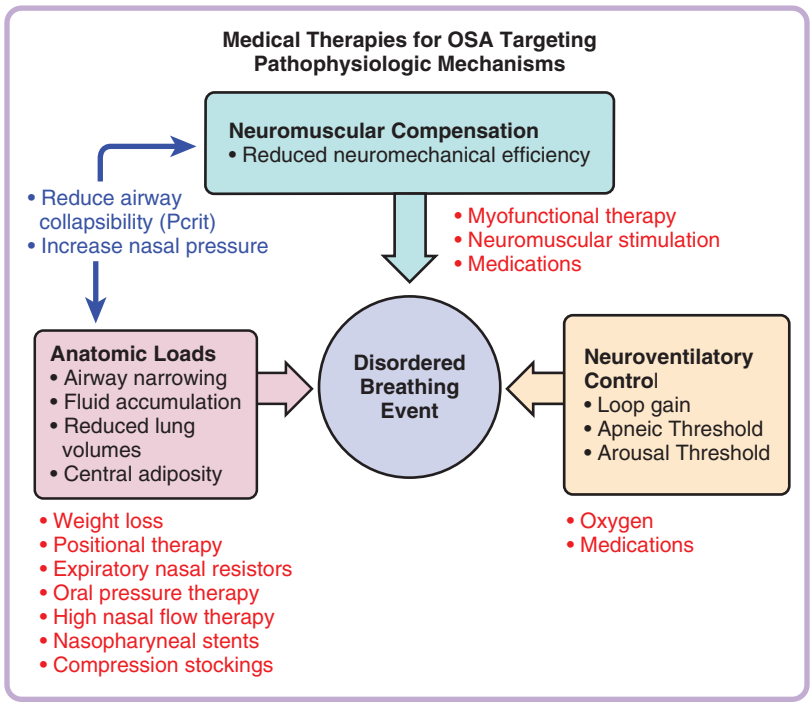


Figure 116-1 Medical Therapies for Obstructive Sleep Apnea (OSA), Stratified by Mechanisms Targeted. Certain therapies can be considered on the basis of the pathophysiologic mechanisms targeted. OSA is thought to occur as a consequence of increases in anatomic loads on the upper airway, impairments in neuromuscular compensation, or alterations in neuroventilatory control. Traditional therapies such as continuous positive airway pressure (CPAP) use nasal pressure to overcome anatomic loads. By contrast, upper airway surgery or weight loss result in reduced airway collapsibility (Pcrit). Other therapies that may relieve anatomic loads on the upper airway include positional therapy, use of expiratory nasal resistors, oral pressure therapy, nasopharyngeal stenting, and application of compression stockings. Therapies that address impairments in neuromuscular function include myofunctional therapies, use of certain medications, and neuromuscular stimulation. Therapies that may affect neuroventilatory control include use of medications to increase the arousal threshold and supplemental oxygen, which can affect loop gain.

therapy is defined as a treatment that cannot be currently recommended except in the setting of clinical research. Another important term, *complementary and alternative medicine* (CAM) *therapy*, is used in the current literature to refer to treatments that are not part of allopathic medical treatment for OSA. Recommendations regarding whether a specific current or emerging alternative medical or device therapy for OSA should be considered primary, alternative, adjunctive, or investigational are provided, with the recognition that this is not without controversy for some treatments. Furthermore, many of the treatments discussed may be classified in more than one category, depending on the specific clinical context. Concluding the chapter is a section on management of residual excessive sleepiness in patients with otherwise adequately treated OSA and good adherence to therapy, with a focus on use of pharmacologic stimulants.

THERAPIES PRIMARILY TARGETING UPPER AIRWAY ANATOMIC LOADS

Increases in upper airway anatomic loads may be incurred through several mechanisms, including (1) airway narrowing due to complex interactions between pharyngeal soft tissues and the bony enclosure within which the upper airway resides, (2) central adiposity–mediated increases in airway collapsibility (Pcrit) through reductions in lung volume and tracheal stiffness, and (3) fluid accumulation within pharyngeal soft tissues. Primary treatment options targeting reduction in anatomic loads have included CPAP therapy (see Chapters 115 and 116), upper airway surgeries (Chapter 149), and oral appliance therapy (see Chapter 147). However, a number of other medical and device therapies aimed at relieving anatomic loads imposed on the upper airway have been explored, including weight loss, positional therapy, expiratory nasal resistance therapy, oral positive-pressure therapy, and use of compression stockings. These therapeutic interventions have been studied as primary, adjunctive, and alternative OSA treatments.

Medical and Surgical Weight Loss

Excess weight has long been recognized as a major risk factor for the development of OSA. The attributable risk of OSA in overweight individuals (BMI ≥ 25 kg/m^2) is estimated to be 41%.[4] Obesity-related impairments in upper airway function appear to be mediated through several mechanisms that affect upper airway anatomy. First, obesity may alter pharyngeal airspace geometry. Data from imaging studies of the human upper airway demonstrate that increases in the lateral pharyngeal fat pads are seen in patients with OSA compared with weight-matched control subjects.[5-7] Enlarged lateral pharyngeal fat pads alter the airway geometry from a horizontal elliptical orientation to an anterior-posterior orientation, which can increase susceptibility of the airway to collapse.[8,9] Second, external mass loads imposed on the upper airway

increase airway collapsibility. In early studies using isolated animal upper airway preparations, application of external loads to the anterior neck and submandibular space led to elevations in Pcrit.[10] In more recent studies, investigators have demonstrated that lateral pharyngeal fat pad pressure fluctuations correlate with cyclic pharyngeal pressure changes, supporting the role of cervical fat depositions in increasing airway collapsibility.[11,12] Peripharyngeal fat deposits are clinically evident as enlarged neck circumference, with studies demonstrating a correlation between increasing neck circumference and increasing airway collapsibility.[13,14] Obesity also may indirectly impose upper airway anatomic loads through mechanical modulation of lung volumes. Central adiposity decreases functional residual capacity (FRC), reducing tracheal traction and thereby increasing upper airway collapsibility. This pathomechanism has been shown experimentally through manipulation of end-expiratory lung volumes in human volunteers, in whom reductions in lung volumes resulted in increases in Pcrit.[15,16]

Medical and surgical weight reduction has long been studied and implemented as a treatment for OSA[17-24]; however, randomized clinical trials have been published only in the past decade.[25-27] Early nonrandomized intervention studies of medical weight loss demonstrated that modest weight loss in the range of 10 to 20 kg in moderately to severely obese men with severe OSA resulted in an apnea-hypopnea index (AHI) reduction of 47% to 50%. A small subset of patients reduced their AHI below 20 (i.e., nighttime occurrence of fewer than 20/hour).[17,19]

More recent randomized, controlled studies in different patient populations have confirmed that medical weight loss decreases OSA severity. Collectively, these studies demonstrate that reductions in OSA severity with weight loss interventions are dose-dependent and sustained over a 1- to 4-year period despite a 30% to 50% weight regain. Furthermore, patients with more severe OSA at baseline demonstrate the greatest improvements in AHI,[25,26] with men tending to experience the greatest benefits.[26]

Tuomilehto and associates[27] studied obese patients with predominantly mild, positional OSA who undertook a 3-month program of a very-low-calorie diet (VLCD) and subsequent lifestyle modification. Patients in the intervention group, with a mean weight reduction of 10.7 kg at 12 months, exhibited a reduction in mean OSA severity (evidenced as occurrence of 4.0 fewer nighttime events/hour), whereas the control group subjects, with a mean weight reduction of 2.4 kg, showed no significant mean change in OSA severity (0.3/hour). Similar reductions in supine AHI were seen in the intervention and control groups (apopnea-hypopnea rate reduced by 6.5/hour and 5.9/hour, respectively). Despite the modest reduction in AHI, 61% of participants in the intervention group experienced resolution (AHI less than 5) of their OSA, compared with 32% in the control group. Improvement as evidenced by a decrease in mean OSA severity was sustained for an additional year despite a mean weight regain of 32% after termination of the supervised program.[28]

Johannson and colleagues[25] randomly assigned obese men with moderate to severe OSA either to a group managed with dietary intervention—VLCD combined with lifestyle counseling—or to a control group with no weight intervention. With a mean 18-kg weight reduction achieved after 9 weeks of intervention in the VLCD group, mean OSA severity decreased, with improved AHI as evidenced by 21 fewer events/hour during nighttime sleep, with 17% demonstrating resolution of their OSA. OSA-related symptomatic improvement was sustained (mean AHI reduction by 17 events/hour) at 1 year despite a 31% weight regain.[29]

Additionally, the Sleep AHEAD (Sleep Apnea in Look AHEAD [Action for Health in Diabetes]) study investigators randomly assigned overweight and obese patients with type 2 diabetes mellitus and OSA to either an intensive lifestyle intervention (ILI) for weight loss or to diabetes support and education (DSE). At 1 year, the ILI group achieved a mean 10.8-kg weight reduction with a concomitant mean decrease in AHI of 5.4; the DSE group demonstrated an increased mean AHI of 4.2, despite no significant change in weight. OSA resolved (for an AHI of less than 5) in 36.3% of those receiving ILI, compared with 10.7% in the DSE group. The improvement in AHI persisted at 4 years despite a 50% weight regain. At year 4, 44% of participants demonstrated an improvement in OSA severity category, compared with only 18% of participants given DSE. Moreover, nearly 21% of ILI participants exhibited complete remission of OSA, to achieve an AHI below 5, compared with only 3.6% of DSE participants.[30]

Surgical weight loss, in contrast with medical weight loss, can result in more dramatic weight reduction that is more likely to be sustained over time (see also Chapter 121). Current National Institutes of Health (NIH) consensus guidelines for surgical weight loss recommend that patients with a BMI of 40 kg/m^2 or greater, or with a BMI of 35 kg/m^2 or greater associated with an obesity-related comorbid condition, including OSA, with previous unsuccessful attempts at medical weight loss, can be considered as potential candidates for surgical weight loss.[31] Bariatric surgeries can include restriction-based techniques such as laparoscopic adjustable gastric banding (LAGB) or vertical sleeve gastrectomy. More dramatic weight loss can be achieved when such techniques are combined with malabsorptive interventions such as the Roux-en-Y bypass or biliopancreatic diversion surgery (see Chapter 121).

Data regarding the effects of bariatric surgery on OSA are predominantly from uncontrolled case series or nonrandomized studies. A large meta-analysis of data for more than 20,000 patients from 134 studies that were predominantly uncontrolled case series reported outcomes from various bariatric surgical procedures. The analysis found that patients on average lost 61.2% of their excess weight, with 85.7% achieving resolution of their OSA.[32] However, resolution of OSA was adjudicated on the basis of patient self-report, rather than postoperative polysomnography, in most of these studies. A meta-analysis that examined only studies in which polysomnography was used before and after surgery reported reductions in mean OSA severity as evidenced by a decrease in AHI (apnea-hypopnea rate of 15.8 events/hour down from 54.7 events/hour) in association with a 17.9 kg/m^2 reduction in BMI.[33] This observation of reduction in severity but incomplete resolution of OSA was confirmed by Dixon and coworkers,[34] who conducted a randomized, controlled trial comparing LAGB with medical weight loss therapy in patients with severe OSA over a 2-year period. Patients in the LAGB group lost more weight compared with the medical weight loss group (mean weight loss of 27.8 kg versus 5.1 kg, respectively). With this weight reduction, the LAGB group

had a tendency to greater reduction in OSA severity as reflected in mean AHI, although the difference was not statistically significant (apnea-hypopnea rate reduction by 25.5 events/hour versus 14.0 events/hour, respectively). In contrast with the medical weight loss studies, which demonstrated dose-dependent improvements in OSA severity with weight loss, this study demonstrated a nonlinear reduction in AHI with weight loss, with a plateau in AHI reduction after 10 kg of weight loss. Thus, although some patients will experience resolution of their OSA to achieve an AHI below 5/h with surgical weight loss, the vast majority will continue to have some level of OSA that may necessitate continued treatment other than weight loss after surgery.

For several reasons, some overweight and obese patients may not experience symptomatic improvement or resolution of OSA with weight loss. Responsible factors include insufficient weight loss, persistent anatomic defects from craniofacial morphology or nasopharyngeal obstruction, or continued disturbances in neuromuscular or neuroventilatory control that contribute to increased airway collapsibility (high Pcrit). For example, obese patients who lost approximately 20% of their baseline weight demonstrated a reduction in Pcrit. However, resolution of OSA occurred only in participants for whom Pcrit fell below a threshold of –4 cm H_2O.[35]

In summary, weight loss, whether by medical therapy or surgical intervention, generally reduces OSA severity and may be curative in certain patients. Unfortunately, those subsets of patients with OSA who are most likely to benefit from these interventions have yet to be defined. Furthermore, the potential success of weight loss in mitigating OSA severity is tempered by the substantial weight regain that occurs in many if not most patients over time, particularly in the setting of medical weight loss. In addition, the risk of a major adverse outcome (e.g., perioperative death, abdominal operation, venous thromboembolism, endoscopy, extended hospitalization) with surgical weight loss interventions is approximately 4% in the first 30 days after surgery.[36] Careful consideration of the risks and benefits of surgical weight loss interventions for OSA must be individualized.

Nevertheless, given the role of obesity in the pathogenesis of OSA, medical providers should advocate weight loss in all patients with OSA who are overweight or obese. Lifestyle interventions for weight loss not only have the potential to reduce OSA severity but may decrease morbidity and mortality from other obesity-related diseases such as metabolic syndrome, hypertension, cardiovascular disease, and diabetes mellitus.[37] Whether a weight loss program should be a primary, adjunctive, or alternative therapy depends on the patient's circumstances. For example, any overweight or obese patient treated for OSA with a primary therapy (e.g., CPAP, use of an oral appliance, upper airway surgery) should be prescribed weight loss as an adjunctive therapy. In patients with mild to moderate OSA associated with minimal daytime symptoms, weight loss could be recommended as a potentially primary therapy, provided that the patient is monitored for success with weight loss over a limited time frame. If a patient is unsuccessful with weight loss, then other primary therapies should be recommended. Patients with symptomatic OSA, however, should not be prescribed weight loss as a sole primary therapy; for example, symptoms of sleepiness may pose a safety risk and must be addressed using other traditional, primary OSA therapies. In patients who are motivated to pursue weight loss, referral to weight management programs that involve a multidisciplinary team, when available, may lead to sustained benefits over those achievable with a traditional weight loss program.[38]

Positional Therapy for Obstructive Sleep Apnea

Positional OSA typically is defined as that associated with an overall AHI less than 5, with a supine AHI that is at least twice the nonsupine AHI. The prevalence of positional OSA in affected patients overall is estimated to be approximately 56%, and it is more common among less obese patients and in those with mild to moderate OSA.[39-41] In view of the marked differences in AHI seen in these patients, positional therapy has been evaluated as a primary or alternative therapy for OSA. Many case series have demonstrated marked improvements in AHI with positional therapy in patients with positional OSA; however, randomized, controlled trial data on use of positional therapy as primary therapy are limited.[42] In one 4-week study, patients with positional OSA (mean AHI of 20.9) were randomly assigned to either a control group (lifestyle education for one session discussing exercise, weight loss, and sleep in the lateral position) or an active group (lifestyle education and the use of a tennis ball position modification device). There was a 46% versus 23% reduction in AHI in the active group versus the control group, respectively, which correlated with reduced supine sleep time in the active group. However, no difference between the groups was found for improvements in sleepiness, mood, or quality of life. Additional small randomized, controlled crossover trials have been performed comparing positional therapy with CPAP over 3 nights to 9 weeks.[43-45] In aggregate,[46] positional therapy has been less effective than CPAP in normalizing the AHI (mean posttreatment apnea-hypopnea rate of 6.2 events/hour versus 2.0 events/hour, respectively). The higher posttreatment AHI in the positional therapy groups in the data noted earlier was due to residual, nonsupine OSA, rather than to an inability to maintain the nonsupine position. Nevertheless, positional therapy improved sleep quality and quality of life measures similarly to CPAP in these studies, despite the higher posttreatment AHI.[43-45]

With most positional therapy techniques, an object is strapped to the back (tennis balls, squash balls, special vests), preventing the patient from sleeping in the supine position. This interventional strategy may disturb sleep architecture and sleep quality owing to arousals precipitated on turning from the right lateral position[47] to the left, resulting in poor long-term treatment adherence.[48] Alternative forms of positional therapy recently have been developed in attempts to improve longer-term adherence. For example, a new neck-worn device delivers a vibration when the patient moves supine, to provide feedback to the patient to shift to a non-supine position without significantly reducing total sleep time.[47,49] Other devices include sleep position trainers worn as a strap around the chest, which similarly vibrates when the supine position is detected,[50-52] and specially designed pillows to improve cervical positioning[53] The limited data available suggest that long-term adherence to positional therapy over 6 months may be comparable to or better than that reported for CPAP[51] but may be specific to the type of positional device used. Additional randomized, controlled studies are needed to demonstrate longer-term improvements in OSA severity status and OSA-related outcomes, as well as long-term adherence, before

positional therapy is considered as a primary treatment for positional OSA. Positional therapy should, however, be considered as adjunctive therapy added to primary therapies in patients with positional OSA. For example, positional therapy could be used in combination with an oral appliance to minimize the extent of mandibular advancement needed to normalize the AHI, or to improve the AHI when the appliance is already at maximum advancement and residual disordered breathing persists. Positional therapy also could be used as an adjunct to CPAP, to increase adherence by lowering the CPAP level in patients intolerant of such settings. Finally, use of positional therapy as an alternative modality in patients with positional OSA, particularly if the nonsupine AHI is near normal, could be considered when other primary therapies are not tolerated or if the patient is traveling without primary therapy appliances or devices. These decisions ideally should be guided by objective data from a sleep study, the patient's preferences, and the clinical response.

Expiratory Nasal Resistors

An expiratory nasal resistor (ENR) is a device containing a one-way valve that is superficially placed in the nares and is secured to the skin of the nose with an adhesive. The valve allows inspiration to occur unimpeded but partially closes during expiration (Figure 116-2). Closure of the valve creates expiratory nasal resistance that results in expiratory positive airway pressure (EPAP) within the pharynx, which is thought to stabilize the upper airway during subsequent inspirations.

Several mechanisms of action have been postulated for EPAP production with use of an ENR, aimed primarily at relieving mechanical loads on the upper airway.[54] First, expiratory pharyngeal airway dilation could reduce subsequent inspiratory airway narrowing. Dynamic imaging of the upper airway suggests that airway caliber is most narrow at end-expiration, when pharyngeal muscle tone is dependent on tonic muscle activity. Using fiberoptic endoscopy, investigators have reported significant expiratory narrowing of the pharyngeal airspace before the development of an apnea.[55] Thus, an increase in end-expiratory airway size has been hypothesized to prevent pharyngeal collapse during the subsequent inspiration. Second, similar to CPAP, ENR may increase lung volumes, thereby increasing tracheal traction and reducing upper airway collapsibility. Investigators, using MRI-based techniques to measure FRC, demonstrated that FRC increases by 47% during wakefulness in association with increases in nasal EPAP between 4 to 17 cm H_2O. The effect of ENRs on lung volumes, however, was mitigated when subjects breathed through the mouth, thus bypassing the ENR.[56] Finally, use of an ENR may improve OSA through indirect chemoresponsive mechanisms by inducing hypercapnia by means of ENR-induced hypoventilation. End-tidal CO_2 measurements have been shown to increase by approximately 2 to 6 mm Hg during ENR application with sleep.[56,57] Increases in CO_2 may then reduce airway collapsibility through recruitment of genioglossal muscle activity.[58,59]

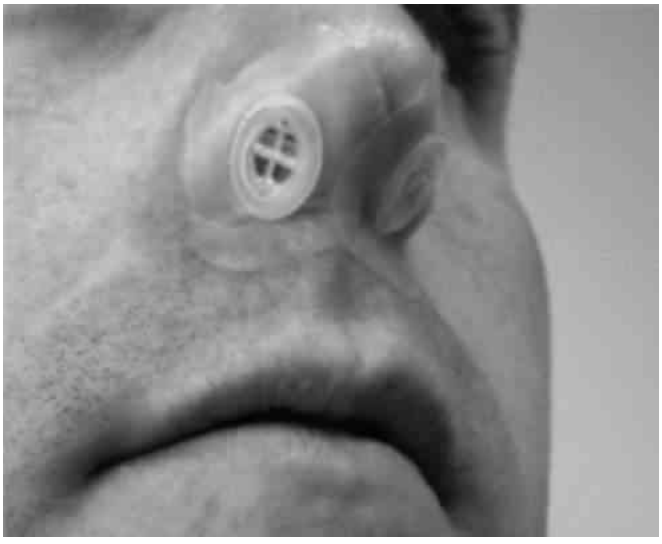


Figure 116-2 An Expiratory Nasal Resistor. This device generates end-expiratory positive airway pressure within the pharynx. (From Walsh JK, Griffin KS, Forst EH, et al. A convenient expiratory positive airway pressure nasal device for the treatment of sleep apnea in patients non-adherent with continuous positive airway pressure. *Sleep Med* 2011;12:147-52.)

Clinical trials of the effectiveness of ENRs have been performed[60-63] since the initial observation by Mahadevia and associates[64] that selective EPAP reduced apnea severity. Initial, uncontrolled studies demonstrated an approximately 50% reduction in AHI in patients with mild to moderate OSA, and 32% reduction in those with severe OSA.[60,61] Subsequently, a 3-month randomized, double-blinded clinical trial was conducted in patients with a new OSA diagnosis.[62] Participants were predominantly obese male patients with mild positional OSA. At 3 months, application of ENRs resulted in a 61% reduction in AHI, compared with 19% in the sham ENR group. The subgroup of patients with severe OSA also demonstrated a 61% reduction in AHI. Although improvements in sleep architecture were not observed with ENR use, decreases in subjective sleepiness based on the Epworth Sleepiness Scale (ESS) score were statistically significant. These effects appeared to be sustained with continued ENR use at 12 months in a follow-up study of a subset of adherent participants in whom ENR therapy reduced AHI by 50%, with decrease in number of events/hour to less than 10.[63] In contrast, in a study of ENR versus sham ENR therapy for 2 weeks in patients with severe OSA undergoing CPAP withdrawal, investigators demonstrated that ENR use did not lessen OSA severity.[65] Side effects reported with use of ENRs include headache, dry mouth, breathing discomfort, nasal itching, sleep maintenance insomnia, and vertigo and resulted in 7% discontinuing therapy.[62]

As indicated by the available evidence, ENR therapy should be considered an alternative therapy for the treatment of OSA in patients who are intolerant of traditional therapies such as CPAP or use of oral appliances. Patients with mild to moderate OSA, particularly positional OSA, appear to be most likely to respond, although in some situations patients with severe OSA also may respond. Patients with symptoms of nasal obstruction are less likely to tolerate ENR therapy, owing to increased nasal resistance, and are not good candidates for this therapy. Efficacy of ENR use in reducing OSA severity should be determined through objective sleep testing before long-term prescription of this therapy. For patients in whom use of ENRs has demonstrated efficacy, the ENR technique could be used as an alternative therapy during travel—for example, if the available electrical source is unreliable or if the equipment for CPAP therapy is perceived as too cumbersome. Additional research is necessary to confirm effectiveness

of this therapy, including whether use of ENRs reduces excessive sleepiness and cardiovascular risk similar to that seen with CPAP or oral appliances, before ENR should be considered a primary therapy.

Oral Pressure Therapy

Oral pressure therapy (OPT) represents another alternative therapy for OSA that aims at reducing upper airway mechanical loads. The OPT system consists of a customized mouthpiece, which is worn nightly, connected by flexible tubing to a console that generates a vacuum of approximately −50 cm H_2O that is applied to the oral cavity[66-68] (Figure 116-3). The negative pressure is isolated to the oral cavity as a consequence of the natural seal created between the soft palate and the tongue.[66] MRI studies performed during wakefulness indicate that OPT increases retropalatal airspace in the lateral and anterior-posterior dimensions through movement of the soft palate (anterior and superior) and the tongue (anterior-superior segment only). The retroglossal airspace is reduced, however, primarily owing to the resting position of the device. OPT may potentially reduce OSA severity through other mechanisms, including vacuum-mediated attenuation of airway collapse during inspiration, and through activation of upper airway negative pressure reflexes, which stabilize the upper airway through increased pharyngeal muscle activity during inspiration.[66]

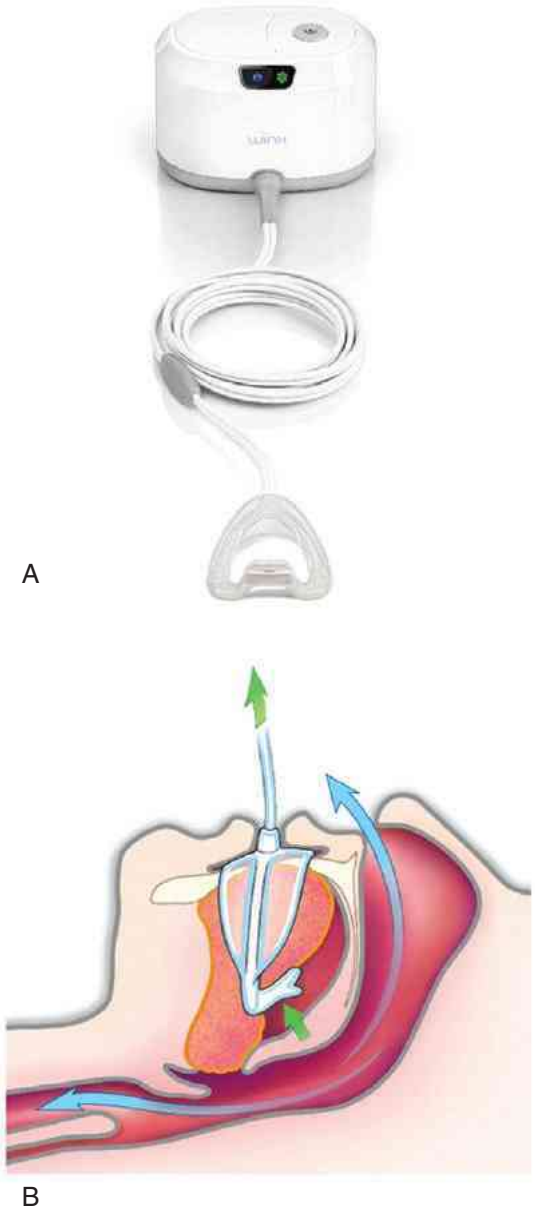


Figure 116-3 Oral Positive-Pressure Therapy Device. **A,** The device sits in the mouth and applies a gentle suction to keep the tongue and soft palate forward during breathing, while the patient lies supine. **B,** The effect of the suction pressure on tongue and palate position. (Images provided and reproduced with permission from Apnicure, Inc.)

Several clinical trials have been performed to assess the efficacy of OPT. Initial studies were uncontrolled and tested OPT in patients with OSA for one night compared with a baseline sleep study.[67,68] These preliminary studies demonstrated that in patients with an AHI of approximately 35, OPT reduced their AHI by 36% to 40%. AHI was reduced to less than 10 in 38% to 48% of the patients with OSA. A subsequent study performed a randomized, cross-over, first-night order study of OPT compared with control conditions, which was followed by an open-label, 4-week trial period.[69] Participants were naive to any OSA treatment, intolerant of CPAP, or actively using CPAP but electing to participate in the study. The pre-OPT mean AHI was 27.5 and was reduced by 51%, to 13.4 on OPT, which was sustained with use of OPT during sleep at the end of 4 weeks, with a mean AHI of 14.8. Thirty-two percent of patients were considered responders, defined as achieving an AHI below 10 and at least a 50% decrease in AHI from baseline. Baseline OSA severity, however, did not predict response to therapy. As in the initial studies, improvements in sleep architecture were demonstrated, characterized by reductions in stage N1 sleep, stage N1 shifts, overall sleep stage shifts, awakenings, arousal index, and by increases in stage R (rapid eye movement [REM]) sleep. In participants who were naive to any OSA treatment, improvements in subjective sleepiness and sleep-related quality of life were reported. However, participants who were actively using CPAP before the study did not demonstrate a reduction in sleepiness, presumably owing to the efficacy of CPAP in improving sleep status. Common side effects reported with OPT included oral tissue discomfort or irritation, dental discomfort, and dry mouth, with 5% of patients ($n = 3$) discontinuing therapy during the trial.

OPT is approved by the U.S. Food and Drug Administration (FDA) for primary treatment of OSA; in view of the limited available data, however, it should be considered an alternative therapy for OSA in patients intolerant of traditional therapies such as CPAP or use of oral appliances. Efficacy of the therapy in reducing OSA severity should be determined through sleep testing before this modality is prescribed. Randomized, controlled trials and comparative effectiveness studies need to be performed to demonstrate whether OPT decreases OSA-related morbidity similar to traditional OSA therapies.

High Flow Nasal Therapy

Several small studies have investigated the effects of high-flow-rate (20 to 30 L/minute), humidified air administered by nasal cannula for treating OSA (Figure 116-4). Such *high flow*

Figure 116-4 An Example of a High-Flow Nasal Therapy Device. This device administers warm, humidified air by nasal cannula at high flow rates. Although the system is "open" compared with the "closed" system inherent in continuous positive airway pressure (CPAP), a positive airway pressure (PAP) of approximately 2 cm H_2O can be generated. (Images provided and reproduced with permission from TNI-Medical.)

nasal therapy (HFN) is hypothesized to improve mechanical loads on the upper airway in a manner similar to that for CPAP, by increasing end-expiratory pharyngeal pressure by approximately 2 cm H_2O.[70] In addition, HFN may prevent pharyngeal collapse through neurally mediated mechanisms. HFN reduces ventilatory drive through reductions in the inspiratory duty cycle and respiratory rate, thereby improving mean inspiratory airflow with each breath.[70,71]

In a study of 11 adult participants with mild to moderate OSA, the mean AHI was reduced by 64% (from mean of 28/h to 10/h) with HFN.[70] In 8 participants, the AHI fell below 10. HFN also has been studied in 10 patients who had experienced a recent acute ischemic stroke (mean, 4.8 days).[72] HFN in this patient population with severe OSA had more modest effects, with an AHI reduction by 24% (mean of 30.8 down from 40.4). In the largest case series of 56 patients, modest reduction in OSA severity were seen with a mean AHI reduction from 22.6 to 17.2, with similar reduction seen for separate hypopnea and apnea indices. A therapeutic response, defined by an AHI of less than 10, with a 50% reduction in AHI from baseline, was seen in 27% of patients. In a separate study using the respiratory disturbance index, patients with predominantly obstructive hypopneas, respiratory effort–related arousals, or REM-related events appeared to be most likely to respond to HFN. By contrast, the nightly occurrence of more than 10% central apneas or more than 90% obstructive apneas predicted a poor response to HFN.[73] Although these initial results are promising, additional clinical trials are needed before HFN can be recommended as a primary or alternative therapy.

Nasopharyngeal Stents

Use of nasopharyneal stents, sometimes referred to as nasal trumpets, has been studied as a potential treatment for OSA since the 1970s.[74-76] Traditionally, such devices are used in the emergency setting to maintain an airway until intubation or tracheostomy can be performed. The stents are inserted into the nose and extend into the nasopharynx, protecting the airway from obstruction. If long enough, the stent may prevent obstruction of the oropharynx. The stents also prevent obstruction of the internal and external nasal valves and may decrease nasal resistance in some patients.[76] To date, no randomized, controlled trials have been performed demonstrating efficacy of nasopharyngeal stenting in improving OSA severity. A systematic review of nasopharyngeal stents identified five noncontrolled studies testing the efficacy of these devices in reducing the AHI in settings that included the sleep laboratory, the postoperative setting, and the home.[74] Overall, the AHI decreased by 49%, from a mean of 44.1to 22.7, and the minimum oxygen saturation improved from 66.5% to 75.5%. The tolerability of nasopharyngeal stenting on a night-to-night basis, however, remains to be established. Until controlled studies are performed, use of nasopharyngeal stents cannot be recommended as a primary or alternative therapy for OSA, except on an emergent basis in the hospital setting to mitigate active airway obstruction in patients with known OSA or suspected OSA based on snoring, witnessed apneas, and oxygen desaturation in an attempt to prevent respiratory failure and the need for emergency intubation.

Venous Compression Stockings

Investigations have examined the use of compression stockings in the treatment of OSA based on the observation that fluid displacement from the legs in awake, healthy subjects without OSA results in an increased neck circumference, pharyngeal narrowing with increased pharyngeal resistance, and increased airway collapsibility.[77-79] Furthermore, increases in OSA severity correlate with the amount of fluid displaced from the legs to the neck when patients were sleeping supine.[80] Subsequently, several studies have examined the effects of venous compression stockings on OSA severity.[81,82] One randomized, crossover study recruited 12 nonobese patients with chronic venous insufficiency and OSA and assigned them in random order to 1 week of wearing compression stockings and 1 week of no stockings. Compared with control conditions, the use of stockings resulted in a 62% reduction in leg fluid volume and a 60% increase in neck circumference that was associated with a 36% reduction in AHI (from 48.4 events/hour to 31.3 events/hour).[81] More recently, the effects of compression stockings were studied in a sample of 57 patients with OSA, who were randomly assigned to a control condition or the use of compression stockings for 2 weeks. Subjects who wore compression stockings demonstrated greater reduction in overnight decrease in leg fluid volume, which correlated with a higher morning upper airway cross-sectional area and a greater AHI reduction compared with

control data (reduction of 8.6 events/hour versus 0.9 event/hour, respectively).[83]

These results, although intriguing, do not support the use of compression stockings as primary therapy for the treatment of OSA. However, the use of compression stockings may have an adjunctive role in combination with primary therapies such as CPAP in treating OSA, particularly in patients who are in states of volume overload such as congestive heart failure, chronic venous stasis, or chronic kidney disease.

THERAPIES TARGETING PRIMARILY NEURAL AND NEUROMUSCULAR MECHANISMS

Pharyngeal obstruction has long been postulated to occur in part by disturbances in neuromuscular function during sleep.[84] OSA subjects have impaired dynamic responses to upper airway obstruction and have reduced tonic genioglossal muscle activity compared with age-, sex-, and BMI-matched healthy control subjects.[1,85] Furthermore, studies of motor unit potential morphology suggest signs of neurogenic remodeling of the genioglossus in persons with OSA.[86] Treatments for OSA thus aimed at relieving impairments in upper airway neuromuscular function or augmenting upper airway neuromuscular responses have been and continue to be investigated. Examples of such interventions are electrical stimulation of the hypoglossal nerve, myofunctional therapy, and pharmacotherapy, as discussed next.

Hypoglossal Nerve Stimulation

Hypoglossal nerve stimulation (HGNS) has been developed as a potential therapy for OSA, with at least one proprietary HGNS system approved by the FDA (see also Chapter 149).[87] HGNS stimulates the hypoglossal nerve, which in turn stimulates the genioglossus muscle, an upper airway dilator muscle. Early studies in animal models demonstrated the potential success of extrinsic electric stimulation of the genioglossus muscle in maintaining upper airway patency during sleep.[88-90] This initial success led to pilot human studies using submental stimulation of the genioglossus[91-93] or direct fine-wire stimulation of the genioglossus or hypoglossal nerve.[94-97] These studies, in addition to demonstrating increases in airflow during sleep in the obstructed airway (Figure 116-5), provided important lessons for the design of the most recent generation of HGNS systems. First, distal placement of electrodes along the hypoglossal nerve to provide selective stimulation of tongue protrusors, or in combination with tongue retractors, improved airway patency.[98] However, proximal nerve stimulation of tongue retractors alone led to airway obstruction.[99] Second, nerve stimulation synchronized with inspiration demonstrated the maximal benefit in airflow improvements and provided secondary benefits of extending battery life and minimizing neuromuscular fatigue.[97,99]

Several HGNS systems have now been developed using either synchronous, closed-loop stimulation or continuous, open-loop stimulation. These systems contain an implanted pulse generator similar to a cardiac pacemaker, which is connected to a respiratory sensing lead (Figure 116-6). A stimulus burst output is delivered and synchronized with inspiration to a cuff that is implanted around the hypoglossal nerve, immediately increasing inspiratory airflow. The continuous, open loop system uses an array of electrodes arranged within the electrode cuff. Stimulation is performed at a set duration, irrespective of the respiratory cycle, in a manner targeting different nerve fibers and minimizing stimulation of the duty cycle.[100,101]

Studies testing the current generation of HGNS systems have predominantly enrolled obese, male patients with moderate to severe OSA that had difficulties "tolerating" CPAP. The studies generally have excluded patients with moderate to severe obesity and some degree of central sleep apnea (greater than 5% or 25% of the AHI events are central or mixed apneas). One system, Inspire (Inspire Medical Systems, Maple Grove, Minnesota), uses drug-induced sleep endoscopy (DISE) as an exclusion criterion based on earlier studies suggesting that concentric airway collapse during DISE predicted greater chance of therapeutic failure.[102] Despite differing eligibility criteria, the studies in aggregate have demonstrated a mean AHI reduction of 50% to 70%, from a mean baseline rate of 32 to 45 events/hour.[100,103,104] Furthermore, in the Stimulation Therapy for Apnea Reduction (STAR) trial, patients who received the HGNS system demonstrated improvements in subjective sleepiness based on ESS scores and in sleep-related quality of life as assessed by the Functional Outcomes of Sleep Questionnaire (FOSQ).[103]

Adverse events related to HGNS have been reported. Short-term surgical risks include wound infections requiring removal of hardware, hematomas, and nerve palsy. Longer-term risks associated with repetitive tongue stimulation include soft tissue abrasions, discomfort with electrical stimulation, and dry mouth.[87,103] In addition, the need for DISE may carry additional risks, such as oxygen desaturation or hypoventilation, that may necessitate bag-mask ventilation or prolonged recovery from sedation due to enhanced drug sensitivity, which warrants careful monitoring by trained professionals.

The Inspire device recently has been FDA-approved. Given that roughly one third of patients in trials of this device were considered to be nonresponders (response was defined as an AHI reduction of at least 50%, with a residual AHI below 20) and had persistent OSA, even after correct application of inclusion and exclusion criteria, additional research is needed to refine optimal patient selection criteria. Further work also is needed to determine how HGNS stimulus settings can be optimized and managed over the long term. Finally, long-term effects on OSA severity and patient safety beyond 18 months remain to be reported. At this time, HGNS should not be considered a primary therapy for OSA until results of additional studies are available. HGNS is available only through specialized centers whose personnel have undergone the appropriate training with the device manufacturer. HGNS, however, should be considered as an alternative therapy in patients with significant difficulties in adhering to or benefiting from CPAP or other primary therapies who meet the inclusion-exclusion criteria defined in the study.

Myofunctional Therapy

The term *myofunctional therapy* is used to describe oropharyngeal exercises used to improve nasal breathing, facial appearance, and mandibular growth.[105] More recently, myofunctional therapy has been studied in the treatment of OSA, using specific voice lessons, musical instruments, and oropharyngeal exercises. An early, uncontrolled study of self-reported chronic snorers found that after a single voice lesson and practice of singing daily for 20 minutes at maximum volume and control

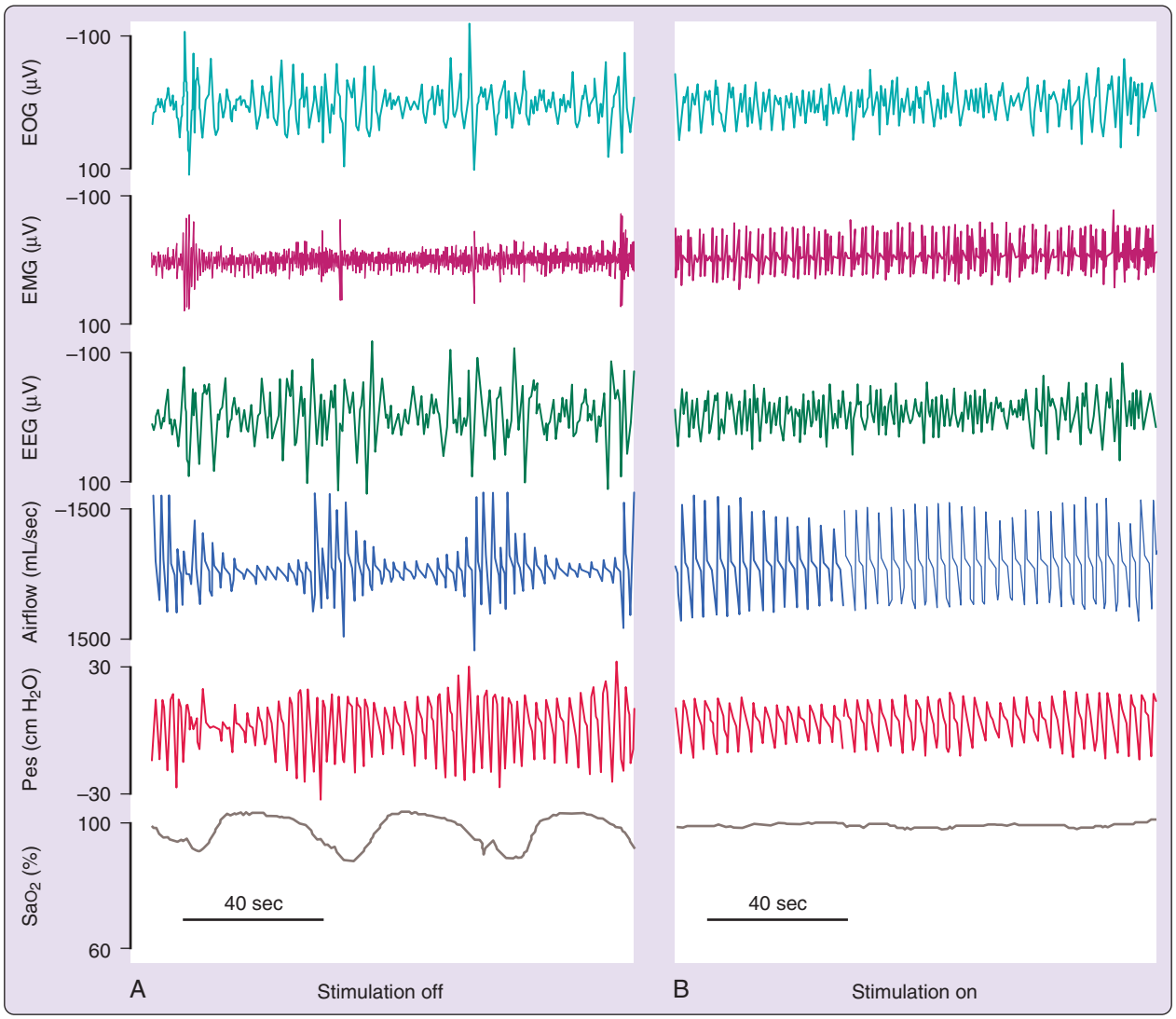


Figure 116-5 Effects of Hypoglossal Nerve Stimulation. **A,** Breathing pattern during non–rapid eye movement (NREM) sleep with hypoglossal stimulation off. **B,** Breathing pattern during NREM sleep with hypoglossal stimulation on. EEG, C3-A2 electroencephalogram; EMG, electromyogram; EOG, electrooculogram; Pes, esophageal pressure; Sa_{O_2}, oxyhemoglobin saturation. (From Schwartz AR, Bennett ML, Smith PL, et al. Therapeutic electrical stimulation of the hypoglossal nerve in obstructive sleep apnea. *Arch Otolaryngol Head Neck Surg* 2001;127: 1216-23.)

reduced the duration of loud snoring after 3 months. Training included instruction in the proper use of the diaphragm and the production of sounds and scales, which cause the soft palate to rise and fall.[106] Subsequently, a cross-sectional study of orchestra musicians completing a Internet-based survey found that double-reed woodwind musicians, in comparison with those who play other wind or non-wind instruments, had a reduced prevalence of OSA risk based on the Berlin questionnaire. Although the proportion of women in the double-reed instrument group was higher, which might have explained this association, the relation persisted after adjustments for sex. Practice duration with a double-reed instrument also was associated with a reduced OSA risk score based on the Berlin questionnaire[107] In another web-based survey study of orchestra members also using the Berlin questionnaire to assess OSA risk, wind instrument players were found to have a greater odds of OSA risk; however, this was no longer significant after adjustments for BMI.[108] Potential explanations for the discrepant findings could include that the latter study did not distinguish between wind instrument type (i.e., single-reed versus double-reed), and that double-reed instruments require both a relatively unique lip placement and a high degree of air resistance. Furthermore, statistical overadjustment for obesity by including this as a covariate may have occurred, because BMI is part of the Berlin questionnaire.

Overcoming such limitations of correlational design and the use of questionnaires to assess OSA risk, findings in at least three preliminary randomized, controlled trials provide promising support for the efficacy of oropharyngeal exercise training for OSA in selected patients. In the first of these studies, investigators performed a controlled trial and randomly assigned 25 participants with moderate OSA (with an AHI between 15 and 30)[109] either to receive digeridoo lessons,

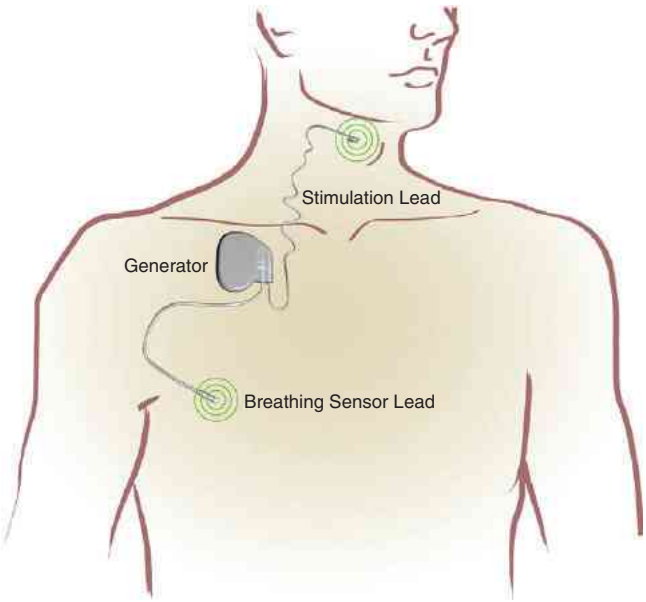


Figure 116-6 General Design of a Closed-System Hypoglossal Nerve Stimulator. (Image provided and reproduced with permission of Inspire Medical Systems, Inc.)

a traditional instrument of the aboriginal Australian culture, with daily practice or to a waitlist for lessons and followed the participants for 4 months. After 4 months of playing the digeridoo for at least 5 days/week for 20 minutes per day, patients in the digeridoo group achieved an AHI reduction of 10.7, compared with 4.5 in the control group participants. Decreases in subjective daytime sleepiness as measured by the ESS also were observed in the digeridoo group.

A more recent randomized, controlled trial of speech therapy–derived oropharyngeal exercises in patients with recently diagnosed moderate OSA studied the effects of isometric and isotonic exercises of the tongue, soft palate, and lateral pharyngeal wall compared with sham exercises performed over a 3-month period. Exercises were observed weekly by a speech therapist, and participants performed exercises for 30 minutes daily for 3 months. Adherence to exercises was monitored using a diary. The investigators found that the oropharyngeal exercise group demonstrated a mean AHI reduction of 38%, compared with 6% in the sham exercise group, for a mean reduction in AHI of 8.7 events/hour versus a mean increase of 1.5 events/hour, respectively; the pretreatment mean AHI was 22.4 for both groups.[110] Significant reductions in neck circumference were found in the oropharyngeal exercise group compared with the sham group (a decrease of 1.1 cm versus a gain of 0.2 cm, respectively) and correlated with the AHI reduction. As in the didgeridoo study, significant reductions were noted in ESS score, a measure of subjective daytime sleepiness.

Myofunctional therapy also has been successfully applied to treat residual symptoms of OSA after adenotonsillectomy (AT) in children.[111] Thirty children with an AHI greater than 1 after AT were randomly assigned to either an exercise regimen targeting nasal breathing, labial seal, lip tone, and tongue posture or no treatment (control group). Children practiced three times per day for 3 months. Oropharyngeal training significantly reduced the post-AT AHI by 58%, compared with 7% for the control group. With the recognition that a majority of children who have undergone AT continue to display at least mild residual OSA symptoms, oropharyngeal training has the potential to become an important adjunctive postsurgical therapy.

These preliminary studies of oropharyngeal muscle training are promising and support continued research in this area. The studies reported to date, however, generally are small and highly selective, with inclusion of only subjects with mild to moderate OSA; thus whether similar reductions in AHI would be observed in patients with severe OSA is unknown. Furthermore, the longevity of the effects is unknown beyond the initial 3-month training period. Further elucidation of the specific types of training exercises that are most strongly associated with decreased airway collapsibility during sleep is needed. Although myofunctional therapy cannot be currently recommended as a primary therapy for OSA, it may be a useful adjunctive modality for use with other OSA therapies or as an alternative therapy in patients who refuse or otherwise cannot benefit from primary treatments for OSA.

Pharmacotherapy for Obstructive Sleep Apnea Targeting Neuromuscular Control

Selecting and studying pharmacologic targets that might successfully treat OSA in humans constitute a challenging endeavor owing to the complexity of respiratory control, the multiple neurochemical pathways that drive respiration, interactions with the sleep state, and limitations in animal models of OSA.[112,113] An ideal pharmacologic agent for the treatment of OSA would need to possess multiple properties including the ability to achieve (1) maintainance of normal airway patency and respiratory drive during both non-REM (NREM) and REM sleep and (2) mitigation of the effects of intermittent arousals and hypoxemia.[114] No such pharmacologic intervention currently exists, although various drugs have been studied in this context. Current pharmacologic approaches for the management of OSA might best be described as alternative therapy in patients for whom other primary or even alternative OSA therapies have not been of benefit. In view of the current evidence, however, such an alternative therapy approach should be considered investigational, with perhaps a few exceptions, until appropriate clinical trials have been completed. Pharmacotherapies used as adjuncts to primary treatment modalities for OSA, including PAP, oral appliance therapy (OPT), surgical treatments, and weight loss, however, also have received attention. Agents that have been evaluated include REM-suppressing agents and drugs that improve airway patency (e.g., serotoninergic, cholinergic, and cannabinoid agents). Overall, the results of these approaches for OSA treatment to date generally have been disappointing, as reviewed next (see Table 116-1 for a summary of the findings).

Serotoninergic Agents

Serotoninergic neurons are known to regulate upper airway motor output, and several studies have investigated the possible beneficial effects of serotoninergic agents in patients with OSA. Serotoninergic control of respiration, however, is complex and remains poorly understood. Whereas some serotoninergic inputs are excitatory and facilitate respiration,[115] others inhibit upper airway motor neuron function.[116] Systemically administered agents that augment or attenuate

Table 116-1 Pharmacotherapies for Obstructive Sleep Apnea

Study*	Study Design	Generic Name	Influence on Osas	Comments
Lin et al, 2012[114]	Critical review	Ventilatory stimulants Serotoninergic drugs	↔	Expert opinion
Veasey, 2003[165]	Review	Serotonin agonists and antagonists	↔	Potential future expectations
Espinoza et al, 1987	Randomized crossover, placebo-controlled study	Aminophylline	↓	Effective only for central and mixed apneas
Carley et al, 1999[117]	Animal study	Mirtazapine	↓	Not recommended owing to adverse side effects
Carley et al, 2007[118]	Randomized, double-blind, placebo-controlled, three-way crossover study			
Castillo et al, 2004	Case report			
Guilleminault and Hayes, 1983[148]	Clinical trial	Naloxone, theophylline, bromocriptine	↔	

**Complete sources for unreferenced studies follow*: Castillo JL, Menendez P, Segovia L, Guilleminault C. Effectiveness of mirtazapine in the treatment of sleep apnea/hypopnea syndrome (SAHS). *Sleep Med* 2004;5(5):507-8; Espinoza H, Antic R, Thornton AT, McEvoy RD. The effects of aminophylline on sleep and sleep-disordered breathing in patients with obstructive sleep apnea syndrome. *Rev Respir Dis* 1987;136(1):80-4.
OSAS, Obstructive sleep apnea syndrome; ↓, ameliorate; ↔, no effect; ↑, exacerbate.

serotonin levels, therefore, might be expected to either decrease OSA severity or exacerbate the condition, respectively.[117]

Several clinical trials have tested serotoninergic medication–based regimens for OSA. Mirtazapine, an antidepressant with both 5-HT_1 agonist and 5-HT_3 antagonist effects, has been more widely investigated as a medication for OSA treatment. On the basis of an animal study in which mirtazapine was found to be effective in reducing central apneas in rats,[117] investigators conducted a randomized, double-blind, placebo-controlled, threeway crossover study of mirtazapine in 12 patients with OSA.[118] The results were positive in that the daily administration of 4.5 to 15 mg of mirtazapine for 1 week reduced the AHI by approximately 50% in adult patients with OSA, from a pre-treatment mean AHI of 22.3 to an on-treatment mean of 11.4. A subsequent randomized, controlled trial, however, demonstrated no significant benefits of mirtazapine (in a dose of 7.5 to 45 mg for 2 weeks) compared with placebo in moderating OSA severity: Compared with the pre-treatment mean AHI of 24.1, on-treatment mean AHI increased to 26.7 to 39.2, depending on the dose of mirtazapine.[119] Use of this medication was associated with a mean weight gain of approximately 1 kg. Relatively common side effects of mirtazapine include both sedation and weight gain, two problems linked with OSA itself.

Subsequently, investigators compared placebo, fluoxetine (a central 5-HT_2 agonist), odansetron (a peripheral 5-HT_3 antagonist), and combined fluoxetine and ondansetron in a randomized, controlled 4-week trial in 35 adults with mild to severe OSA. Combined high-dose therapy with fluoxetine and ondansetron showed some efficacy: This regimen reduced the mean AHI significantly compared with baseline at days 14 and 28 for the second half of full-night polysomnography, in contrast with no significant changes in AHI with placebo.[120] However, no subsequent confirmatory trials have been performed. Common side effects reported with use of these medications include headache, constipation, dry mouth, and hypersomnolence, although such effects were not seen during this 4-week clinical trial.

Protriptyline, a nonsedating tricyclic antidepressant, acts as a serotonin and norepinephrine reuptake inhibitor and also has been shown to have partial treatment effects in OSA at doses up to 30 mg. Mechanisms which contribute to improvements in OSA severity include reduction in REM sleep duration,[121] and increased hypoglossal and recurrent laryngeal nerve activity with increased upper airway motor tone.[122] In at least two small clinical trials, use of protriptyline for treatment of severe OSA (mean AHI range, 71 to 75) was associated with reductions in AHI by 21% to 33%,[121,123] predominantly as a consequence of reductions in apnea frequency and duration. Furthermore, subjective improvements with respect to sleepiness were reported by most patients despite significant residual disordered breathing, suggesting that protriptyline may have independent alerting effects.[121] Documentation of significant residual disordered breathing and hypoxemia, however, has diminished enthusiasm for this treatment when more effective therapies are available. Side effects of protriptyline include dry mouth, urinary hesitancy, constipation, confusion, and ataxia, all of which also may limit the use of medications of this class.[124]

Thus, as a group, serotoninergic medications have modest effects on OSA severity status. In light of the availability of more effective treatments for OSA and potentially significant side effects for some of these drugs, these medications should not be used for primary therapy for OSA but rather are best considered as adjunctive therapy with other OSA treatments. Use of serotoninergic medications also could be considered an alternative therapy in patients with OSA intolerant of other forms of OSA treatment, particularly in patients in whom these medications are already planned to be used for comorbid disorders such as depression (mirtazapine or protriptyline), anorexia (mirtazapine), migraine (protriptyline), and cataplexy (protriptyline or fluoxetine). Reductions in OSA severity with

these medications should not be assumed on the basis of decreased symptoms, and the patient's OSA status should be monitored by sleep testing.

Cholinergic Agents

Acetylcholine, a cholinergic neurotransmitter active primarily during REM sleep, is involved in the modulation of upper airway motor tone. Preliminary investigations of acetylcholinesterase inhibitors have been performed on the basis of findings in preclinical studies, which demonstrated that injection of physostigmine into cholinergic neurons located in the rostral ventrolateral medulla of anesthetized and vagotomized cats was followed by increased hypoglossal and phrenic nerve activity. This increased activity resulted in prolonged hypoglossal to phrenic nerve firing interval with consequent improvement in respiratory drive.[125] On the basis of such data, a double-blind, placebo-controlled trial was conducted in 12 men with moderate to severe OSA.[125] Physostigmine or placebo was injected intravenously on separate nights followed by an overnight sleep study. Comparison of sleep recordings for the physostigmine night and the placebo night demonstrated a slightly lower mean AHI (apnea-hypopnea rate of 41 events/hour and 54 events/hour, respectively), with the greatest reduction occurring during REM sleep (mean AHI of 54 versus 30, respectively).

Donepezil, a reversible inhibitor of the acetylcholinesterase enzyme often used to treat memory impairment in Alzheimer disease (AD), also has been tested as a potential agent for treatment of OSA in patients with and without AD. An initial study was performed in 23 patients with AD and mild to moderate OSA. This randomized, double-blind, placebo-controlled trial demonstrated that at 3 months, donepezil improved the mean AHI from 20.0 to 9.9 compared with placebo, for which the mean AHI did not change (23.2 versus 22.9). As would be expected with a cholinergic medication, increased REM sleep was observed in the donepezil group at 3 months.[127] Another double-blind, placebo-controlled trial of donepezil was conducted in 21 male patients with OSA but no AD. This study also found donepezil to improve mean AHI at 1 month, although the effects were more modest, with a mean AHI reduction of 23% (pretreatment mean of 42.2 versus posttreatment mean of 32.8) versus a mean AHI increase of 14% (pretreatment mean of 26.4 versus posttreatment mean of 31.0) in the placebo group.[128] This study found no differences in REM sleep between the groups. Side effects reported with donepezil included dizziness, nausea, headaches, vivid dreams, and nightmares. Although these results are promising, further confirmatory studies are needed in larger samples to determine if donepezil has a role in the treatment of OSA. For now, there may be a role for donepezil in patients with AD with comorbid OSA, when donepezil is already being considered for memory-related conditions and other primary or alternative OSA therapies are not tolerated.

Cannabinoids

Cannabinoid agonists have recently been investigated as a candidate target for OSA therapy. Preclinical data suggest that dronabinol, a nonselective cannabinoid type 1 (CB1) and type 2 (CB2) receptor agonist, increases phasic genioglossal activity and attenuates serotonin-induced apneas in rats when injected in the nodose ganglion.[129] Dronabinol is hypothesized to inhibit afferent vagal nerve activity, which may thereby result in disinhibition of upper airway motor neurons.[130] In an initial proof-of-concept human clinical trial,[131] participants with moderate OSA treated with CPAP were withdrawn from their CPAP regimen for 1 week. Participants were then given dronabinol in an escalated dose over 3 weeks, up to 10 mg. Pretreatment AHI was reduced by 29% from a pretreatment mean AHI of 48.8. Side effects noted during the study included somnolence and increased appetite without weight increase. Although these findings are of interest, controlled studies are needed on the potential effects of dronabinol in OSA, and this treatment should be considered investigational at present.

THERAPIES TARGETING PRIMARILY NEUROVENTILATORY MECHANISMS

Neuroventilatory mechanisms play an influential role in the expression of OSA severity. Physiologic parameters such as the arousal threshold, apnea threshold, CO_2 reserve, and circulatory time determine the subject's response to reduced ventilation from any cause whether mediated centrally (central hypoventilation) or peripherally (obstruction). Globally, these measures determine the degree of ventilatory instability (loop gain) and whether disordered breathing will be mitigated or perpetuated. Neuroventilatory mechanisms, therefore, present a potential therapeutic target for OSA treatment. Examples of such interventions are oxygen therapy and pharmacotherapies and are discussed next.

Supplemental Oxygen

Many of the consequences of OSA are attributable to nocturnal hypoxemia. Studies done before the widespread use of CPAP reported that supplemental oxygen administration during sleep in patients with OSA could significantly increase oxygen saturation of hemoglobin (Sao_2) but could also lengthen apneic spells, potentially leading to hypercapnia and respiratory acidosis.[132-134] These early studies found no improvement in subjective or objective measures of daytime sleepiness with nocturnal oxygen treatment.[135] However, reduction in OSA severity with oxygen therapy alone may depend on whether the patient has stable or unstable ventilatory control (low or high loop gain, respectively). In a study of subjects with severe OSA, oxygen reduced the AHI by 53% in the high loop gain group, compared with 8% in the low loop gain group.[136] A randomized clinical trial comparing oxygen administration alone with CPAP in patients with OSA and comorbid cardiovascular disease or multiple cardiovascular risk factors found that CPAP, but not nocturnal oxygen therapy, significantly reduced 24-hour mean blood pressure.[137] Thus oxygen therapy alone during sleep is not recommended as a primary or alternative therapy for most patients with OSA.

However, subgroups of patients with OSA who might benefit from oxygen therapy are recognized. Patients with significant cardiovascular disease (e.g., coronary artery disease or cerebrovascular disease) and an only marginally elevated frequency of abnormal breathing events during sleep, but who experience severe oxyhemoglobin desaturation during those events, might benefit in terms of reduced risk of myocardial ischemia with oxygen supplementation.[138,139] The use of supplemental oxygen also could be considered as alternative therapy in patients with OSA and significant intermittent

hypoxemia who are intolerant of a primary therapy such as CPAP, to minimize potential cardiovascular and metabolic risks. In the absence of any high-grade evidence, such use should be considered controversial, particularly because oxygen administration is not without risks (e.g., hypercapnia, fire risk). The clinician in this situation should consider titration of oxygen in an attended sleep study setting to document the optimal minimum oxygen dose for efficacy in patients with OSA in preventing hypoxemia and minimizing hypercapnia.

Oxygen also may be added as adjunctive therapy to PAP in patients in whom CPAP or bilevel PAP regimens are effective in treating their OSA, but in whom hypoxemia persists owing to ventilation-perfusion mismatching or hypoventilation.[140] This may occur in patients with severe obesity or a so-called overlap syndrome (see Chapter 119). Patients with OSA who require supplemental oxygen therapy during wakefulness almost always will require supplemental oxygen during sleep, even if PAP therapy maintains a patent upper airway.[141] However, it should be determined if persistent oxygen desaturation in patients on CPAP is related to hypoventilation,[10] because use of bilevel positive pressure in patients who are hyperventilating despite adequate control of OSA may obviate the need for added oxygen.

Transtracheal Oxygen Delivery

Several reports have described the use of transtracheal oxygen administration in patients with OSA who are intolerant of CPAP.[142-144] One study described the use of this modality as salvage therapy in a patient with overlap syndrome and in a patient with persistent hypoxemia despite CPAP and in-line oxygen.[145] These data are too limited to recommend this mode of oxygen delivery; accordingly, transtracheal oxygen use in OSA as either an alternative or adjunctive therapy should be considered investigational.

Pharmacotherapy for Obstructive Sleep Apnea Targeting Neuroventilatory Control

Several drugs have been studied targeting neuroventilatory control mechanisms such as a high loop gain state and a low arousal threshold. Although more often considered as potential treatments for central sleep apnea (see Chapter 110), some of these pharmacologic approaches have been examined in the setting of OSA. With the possible exception of acetazolamide, these approaches should be considered either as alternative therapies for use in patients who have demonstrated intolerance to primary therapies or as investigational therapies.

Carbonic Anhydrase Inhibitors

Acetazolamide, a carbonic anhydrase inhibitor that induces a metabolic acidosis, thereby increasing ventilation, has been studied primarily in patients with central sleep apnea secondary to high altitude exposure or heart failure, who often have unstable ventilatory control (i.e., high loop gain). Similarly, some patients with OSA also have been shown to have an elevated loop gain.[146] In one study of patients with OSA treated with CPAP, administration of acetazolamide for 7 days reduced the mean loop gain by 41% and the mean AHI by 41%.[147] However, difficulties with tolerability of acetazolamide may preclude its long-term use, because patient-reported side effects have included paresthesias, altered taste, nocturia, and hypokalemia (when it is used in combination with a diuretic), with the need to monitor serum bicarbonate levels over time. Currently, use of acetazolamide should be considered only as alternative or adjunctive therapy for OSA.

Methylxanthines, Opioid Antagonists, and Dopamine Agonists

A randomized, crossover, placebo-controlled trial in 1987 assessed the efficacy of infusion of aminophylline, a methylxanthine derivative known to have respiratory stimulant properties, in male subjects with moderate OSA.[53] Aminophylline decreased the frequency of central and mixed apneas but did not affect the frequency or duration of obstructive apneas. Mean and minimum arterial oxygen saturation values in sleep also were unchanged, and sleep architecture was markedly disturbed.

In recognition of the effects of opiate agonists on respiratory depression, another early investigation tested multiple medications including naloxone (an opioid antagonist), theophylline, and bromocriptine mesylate (a dopamine agonist). None of these agents had any significant beneficial effects on the frequency or duration of obstructive apneic and hypopneic spells or on oxygen desaturation indices.[148] Thus none of these medications are appropriate for use as as primary agents in the treatment of OSA.

Sedatives and Hypnotics

The use of sedatives and hypnotics in the treatment of OSA appears to be counterintuitive, on the basis of concerns regarding worsening of OSA secondary to the myorelaxant and central nervous system sedative effects of many of these medications.[149] However, a low arousal threshold in a patient with OSA may result in a premature arousal before compensatory neuromuscular mechanisms have sufficient time to restore complete upper airway patency. A premature arousal when combined with an underlying state of ventilatory instability may result in persistent disordered breathing during sleep. One study[146] demonstrated that 37% of patients with OSA have a low arousal threshold, raising the possibility that increases in the arousal threshold, as may be achieved with sedative-hypnotics, may represent a therapeutic target for this population.

Initial clinical studies testing this possibility have looked at the use of eszopiclone, a nonbenzodiazepine sedative, or trazodone, an antidepressant medication with serotoninergic, antihistaminic, and antiadrenergic effects. One study randomly assigned 17 subjects with OSA with a nadir Sao_2 greater than 70% to receive one night of eszopiclone 3 mg or one night of placebo.[150] The arousal threshold, quantified by degree of nadir epiglottic pressure level associated with electroencephalogram arousal, was observed to increase by 18% in stage N2 (NREM stage 2) sleep , whereas the mean AHI was reduced by 23% (rate decrease from 31 events/hour to 24 events/hour) in the eszopiclone group compared with the placebo group. No significant difference in hypoxemia severity was seen between the two groups. In another, more rigorous study of double-blind, placebo-controlled cross-over design, however, participants with mild to moderate OSA were randomly assigned to receive either eszopiclone or placebo for two consecutive nights. No significant difference in AHI was seen between the groups.[151] Several studies have examined the use of trazodone to treat OSA.[152-154] Trazodone given at

100 mg for one night was reported to increase the arousal threshold by 32% but did not reduce the AHI[153] A subsequent study of 15 patients with severe OSA given trazodone 100 mg or placebo had conflicting results, demonstrating no significant changes in arousal threshold but mild improvements in AHI with trazodone compared with placebo (reported rate of 28.5 events/hour vs. 38.7 events/hour, respectively).[154] These data suggest that at the very least, certain sedatives and hypnotics may not exacerbate OSA and could be used in patients with comorbid insomnia when indicated. In view of the noted conflicting results, however, the use of sedatives and/or hypnotics to treat OSA, whether as adjunctive or alternative therapy, currently should be considered investigational.

Complementary and Alternative Medicine Therapy for Obstructive Sleep Apnea

The term *complementary and alternative medicine* (CAM) typically is used to refer to treatments that are not part of allopathic medical training.[155] Such treatments are not restricted to medications and typically lack a clear and compelling mechanism of action that targets known pathophysiology. In the following discussion, the concept of *CAM therapy* is differentiated from *adjunctive* and *alternative therapy* as described earlier in the chapter. CAM therapies may potentially be used as adjunctive or alternative treatments for OSA. In view of the well-established cardiovascular, cerebrovascular, and metabolic consequences of OSA and the neuromechanical nature of the disease, the role for CAM treatments, if not specifically contraindicated, is expected to be minimal. As noted earlier, however, PAP, the "gold standard" treatment for OSA, is difficult to adhere to for many patients.[2] Recent surveys suggest that a majority of patients with OSA are actively interested in CAM approaches.[156] Setting aside the general absence of well-controlled studies evaluating complementary interventions for OSA, we present a brief review of preliminary studies of acupuncture therapies, considered a form of CAM.[157,158] In general, however, other CAM-designated treatments such as use of herbal and dietary supplements and manipulative therapies (e.g., tai chi) have been studied in a very limited and uncontrolled fashion and consequently are not further discussed here. With the possible exception of acupuncture, reports of CAM treatments are of insufficient scientific quality to support their use. Placebo effects and manipulation of presleep expectancies,[159,160] which represent potential mechanisms by which some of these interventions could improve sleep parameters, have been shown to persist into and alter sleep-related physiologic measures. Indeed, some work demonstrates that REM sleep may play a role in the persistence of next-day placebo analgesia effects.[159]

Acupuncture

Most studies of CAM therapies for OSA have not used randomized, placebo-controlled experimental designs. Acupuncture is a notable exception. One single-blind study of acupuncture randomly assigned 36 patients with previously untreated OSA for 10 weeks to either an acupuncture group, a sham acupuncture group, or a control group (in which sleep hygiene and weight loss counseling was provided).[161] Weekly acupuncture significantly improved the mean AHI by approximately 50%, from a recorded rate of 19.9 events/hour to 10.1 events/hour, compared with the sham acupuncture or control group, for which mean AHI was unchanged or significantly worsened, respectively.[161] The acupuncture group also demonstrated improvement in subjective measures of sleepiness based on the ESS and in some quality of life domains based on the short form 36 health survey questionnaire (SF-36). A more recent randomized, controlled study by the same group of investigators compared manual acupuncture and electroacupuncture for one session with no treatment (control condition) in 40 patients with previously untreated, moderate OSA. Both manual acupuncture and 10-Hz electroacupuncture administered just before sleep significantly reduced the mean AHI by approximately 50% (apnea-hypopnea rate for manual acupuncture: 21.9 events/hour reduced to 11.2 events/hour; rate for 10-Hz electroacupuncture: 20.6 events/hour reduced to 10.0 events/hour), compared with 2-Hz electroacupuncture and no treatment (control condition) (both groups: no significant change in mean AHI).[162]

A randomized, controlled trial of auricular plaster therapy, a form of acupuncture administered only to the ear, compared this treatment against vitamin C supplementation in the control group.[163] Forty-five participants with severe OSA were randomly assigned to receive either auricular acupuncture three to five times per day or vitamin C three times daily for 10 days. Patients in the auricular acupuncture group demonstrated modest, but statistically significant, improvement in the mean AHI (pretreatment apnea-hypopnea rate of 72.4 events/hour versus posttreatment rate of 59.2 events/hour), in contrast with the control group (pretreatment rate of 73.5 events/hour versus posttreatment rate of 72.0 events/hour).

Although the mechanisms by which acupuncture may modulate OSA severity, and the longevity of such effects, are unknown, these preliminary studies are promising and merit further research. Acupuncture should not replace established primary treatments in current use, but if this modality is added, it should be performed using published protocols by professionals trained in this discipline to promote standardization of effects.

PHARMACOLOGIC MANAGEMENT OF "RESIDUAL" EXCESSIVE DAYTIME SLEEPINESS WITH ADEQUATELY TREATED OBSTRUCTIVE SLEEP APNEA

Even with objectively documented successful treatment of OSA, including acceptable adherence to therapy, it has been estimated that as many as 10% of patients with OSA continue to report significant excessive daytime sleepiness (EDS).[164] The significance of EDS cannot be overstated in view of its role in contributing to motor vehicle accidents, impaired psychological functioning, and reduced work performance.[45] The cause of such "residual" EDS, however, can be difficult to definitively ascertain. Data from mouse models suggest that intermittent hypoxia may result in irreversible oxidative injury to brain centers associated with sleep.[165]

As part of the clinical management of patients with appropriately treated OSA but persistent EDS, other contributing factors and conditions such as insufficient sleep time, insomnia, medication-related side effects, or other comorbid sleep disorders should be carefully ruled out. Successful treatment of OSA should be documented by objective measures of adherence to therapy and a normal AHI with the prescribed

CPAP regimen as confirmed by a sleep study. Objective documentation of EDS with a Multiple Sleep Latency Test (MSLT) or Maintenance of Wakefulness Test (MWT) could be considered, although this is not a requirement for most third party payers, and such testing typically is performed when the possibility of a primary CNS hypersomnolence disorder such as narcolepsy is a concern. If EDS persists despite adequate OSA treatment, nonsympathomimetic stimulants (e.g., caffeine) can be used, as well as psychostimulant medications (i.e., nonamphetamine or amphetamine derivatives). Such regimens should be considered as part of an overall management strategy to improve daytime alertness.[166]

Traditionally, amphetamine-class medications have been used to treat residual EDS in patients with OSA, on the basis of data for subjects with narcolepsy and sleep-restricted persons (Chapter 90). The potential for harmful cardiovascular consequences and potential negative mood- and sleep-related effects with this class of medications has led to investigational use of pharmacologic agents other than amphetamines, for which these effects seem less likely to occur. Comparative effectiveness studies of different stimulants in the treatment of residual EDS in patients with OSA are needed because many of the amphetamine derivatives are considerably less expensive than the nonamphetamine medications. The use of these medications is not discussed further here owing to the absence of data regarding their use in the treatment of residual EDS in patients with OSA.

Modafinil and armodafinil, the *R*-isomer of modafinil, are nonamphetamines currently FDA-approved for the treatment of residual hypersomnolence in patients with OSA considered to be otherwise adequately treated with PAP.[167] The wake-promoting effects of the medication are incompletely understood but are reported to be due primarily to dopaminergic-mediated pathways.[168] Several relatively large randomized, placebo-controlled clinical trials have demonstrated that modafinil[148,149,169,170] and armodafinil[167,171-174] can safely reduce EDS, as indicated by subjective and objective measures,[175] and can improve quality of life in patients with OSA adequately treated with CPAP.[176,177] For example, an initial 4-week randomized, double-blind, placebo-controlled, parallel group study[169] of modafinil versus placebo reported normalization of the ESS score in 51% of the modafinil group versus 27% in the placebo group. Although the mean sleep latency on an MSLT was decreased in the modafinil group compared with the placebo group, normalization of the mean sleep latency (to less than 10 minutes) was similar in both groups (29% versus 25%, respectively). A subsequent randomized, controlled trial[170] performed over 12 weeks evaluated the effects of placebo versus modafinil at a 200-mg or 400-mg dose. Improvements in ESS scores and mean sleep latency on the MWT were reported, with similar improvements for the 200-mg and 400-mg modafinil dose groups. Improvements in sleep-related quality of life and global function as assessed by the FOSQ and Clinical Global Impression of Change questionnaires were observed to a similar degree in both modafinil groups compared with placebo. No change in CPAP adherence was reported in these two studies.[169,170] However, in a separate 12-week open label continuation of modafinil from the initial study,[169] mean CPAP use was observed to decline from 6.3 hours/night to 5.9 hours/night,[178] suggesting that such agents may in fact reduce adherence to CPAP. Accordingly, clinicians should continually remind their patients to be adherent to CPAP, to maximize the wake-promoting effects of both therapies.

A similar literature base is available regarding the efficacy of armodafinil for residual EDS in patients with treated OSA.[167,171-174] Armodafinil has a duration of action that is 10% to 15% longer than that of modafinil. Data from a pooled analysis[172] of two 12-week multicenter, double-blind, placebo-controlled, parallel group clinical studies[176,179] found that adjunctive treatment with armodafinil in CPAP-adherent participants with OSA coupled with residual EDS significantly improved wakefulness, long-term memory, and ability to engage in activities of daily living. Armodafinil also reduced patient-reported fatigue, evaluated separately from sleepiness, and was well tolerated in terms of side effects.[172] Treatment with armodafinil showed no effect on subsequent CPAP adherence, but active monitoring of CPAP adherence by the clinician is advocated in this setting.[174] A multicenter, flexible-dose, open-label study found that armodafinil remained effective for more than 12 months in patients with residual EDS and treated OSA.[173]

The most commonly reported adverse events in the studies that were associated with both medications included headache (occurring in approximately 15% to 20% of subjects), nausea (in 10% to 20%), insomnia (in 5% to 10%), and anxiety (in 5% to 15%). Up to 15% of patients in one study discontinued medications owing to such adverse events. A rare but serious adverse event that clinicians should be aware of is the occurrence of serious rashes, including Stevens- Johnson syndrome and toxic epidermal necrolysis, which typically occur within 5 weeks of the initiation of therapy but in rare cases may appear later than that. Rare cases of multiorgan hypersensitivity manifesting as fever, rash, and organ system dysfunction have been reported with modafinil and armodafinil, as well as anaphylactoid reaction with armodafinil. Modafinil and armodafinil may decrease the effectiveness of hormonal birth control systems, so women should be advised of this possibility, with consideration given to use of nonhormonal contraceptive approaches.

In summary, the use of pharmacologic stimulants in patients with OSA should be considered as adjunctive therapy for the management of residual EDS in patients with OSA adequately treated with CPAP, with documentation of acceptable adherence to CPAP, and after exclusion of other causes of EDS. More controversial is whether stimulants should be considered as an alternative therapy for patients with EDS due to OSA who are intolerant of treatments for their OSA. At least two studies suggest that use of modafinil during a 2-day CPAP withdrawal period or for 2 weeks in patients with mild to moderate untreated OSA resulted in significant improvements in driving performance in a driving simulator, decrease in subjective sleepiness, and better scores for attention and vigilance on the psychomotor vigilance test.[180,181] In patients in high-risk situations in which alertness and performance are critical (e.g., professional drivers, military personnel) and those for whom use of primary PAP therapy is interrupted (such as with an unreliable electrical source during travel), there may be a role for stimulants as sole treatment for brief periods. However, continued reinforcement of adherence to the patient's prescribed OSA treatment, as well as monitoring for medication side effects, is necessary to avoid potential long-term adverse effects of stimulants.

CLINICAL PEARLS

- Despite the efficacy of traditional primary treatments such as CPAP, oral appliance therapy, and upper airway surgery for OSA, patient factors such as suboptimal adherence or poor tolerance of a therapy may lead to incomplete treatment.
- In such situations, alternative therapies should be considered, either alone or as adjunctive treatment with other OSA treatment options (e.g., use of positional therapy with supplemental oxygen in a patient with positional OSA and persistent hypoxemia).
- In patients with OSA and residual sleepiness, adjunctive pharmacologic stimulant therapy can be considered, but only after ensuring that OSA treatment and sleep time are adequate, and that no other sleep disorders are present that may explain the residual sleepiness.

SUMMARY

In light of the fact that traditional primary treatment modalities for OSA, including CPAP, use of oral mandibular advancement devices, and upper airway surgery, often have poor adherence rates and/or insufficient long-term outcomes (e.g., upper airway surgery for OSA), adjunctive and alternative treatment options for OSA should be considered in nonadherent patients and those who are otherwise not fully benefiting from primary and adjunctive therapies. Furthermore, OSA treatments can be individualized in accordance with the known pathophysiologic basis for OSA in a particular patient, and with the patient's personal preferences. Treatment measures such as weight loss, positional therapy, HGNS, ENRs, and OPT have been studied and shown to successfully treat OSA in appropriate, sometimes selected patients. Other potential therapies such as myofunctional therapy, nasopharyngeal stenting, highnasal-flow therapy, use of compression stockings, and pharmacotherapies do not have proven efficacy and should currently be considered as adjunctive or investigational therapy, rather than primary therapy, for OSA. Oxygen therapy alone has not been shown to improve outcomes in OSA, although treatment of severe hypoxemia with supplemental O_2 alone or in combination with PAP can be considered a reasonable adjunctive or alternative treatment option in some patients, including those who cannot maintain adherence to or benefit from primary therapies, or who need supplemental O_2 with PAP to adequately ameliorate sleep-related hypoxemia. Promising preliminary data are available for the use of acupuncture, a form of CAM therapy for OSA, as an adjunctive or alternative therapy, but this modality requires further investigation before it can be recommended for use in OSA. Even with effective treatment for OSA, some patients may experience persistent sleepiness despite adequate sleep time and may be appropriate candidates for adjunctive stimulant therapy. In view of the difficulties that some patients experience with primary therapies for OSA, continued investigation into alterative and adjunctive OSA treatment options can be expected.

Selected Readings

Berry RB, Kryger MH, Massie CA. A novel nasal expiratory positive airway pressure (EPAP) device for the treatment of obstructive sleep apnea: a randomized controlled trial. *Sleep* 2011;**34**:479–85.

Billings KR, Maddalozzo J. Complementary and integrative treatments: managing obstructive sleep apnea. *Otolaryngol Clin North Am* 2013;**46**: 383–8.

Campbell T, Pengo MF, Steier J. Patients' preference of established and emerging treatment options for obstructive sleep apnoea. *J Thorac Dis* 2015;**7**(5):938–42.

Colrain IM, Black J, Siegel LC, et al. A multicenter evaluation of oral pressure therapy for the treatment of obstructive sleep apnea. *Sleep Med* 2013; **14**:830–7.

Foster GD, Borradaile KE, Sanders MH, et al. A randomized study on the effect of weight loss on obstructive sleep apnea among obese patients with type 2 diabetes: the Sleep AHEAD study. *Arch Intern Med* 2009;**169**: 1619–26.

Freire AO, Sugai GC, Chrispin FS, et al. Treatment of moderate obstructive sleep apnea syndrome with acupuncture: a randomised, placebo-controlled pilot trial. *Sleep Med* 2007;**8**:43–50.

Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med* 2014;**370**(24):2276–85.

Greenburg DL, Lettieri CJ, Eliasson AH. Effects of surgical weight loss on measures of obstructive sleep apnea: a meta-analysis. *Am J Med* 2009;**122**: 535–42.

Guimaraes KC, Drager LF, Genta PR, et al. Effects of oropharyngeal exercises on patients with moderate obstructive sleep apnea syndrome. *Am J Respir Crit Care Med* 2009;**179**:962–6.

Lin CM, Huang YS, Guilleminault C. Pharmacotherapy of obstructive sleep apnea. *Expert Opin Pharmacother* 2012;**13**:841–57.

McGinley BM, Patil SP, Kirkness JP, et al. A nasal cannula can be used to treat obstructive sleep apnea. *Am J Respir Crit Care Med* 2007;**176**: 194–200.

Puhan MA, Suarez A, Lo CC, et al. Didgeridoo playing as alternative treatment for obstructive sleep apnoea syndrome: randomised controlled trial. *BMJ* 2006;**332**:266–70.

Sukhal S, Khalid M, Tulaimat A. Effect of wakefulness-promoting agents on sleepiness in patients with sleep apnea treated with CPAP: a meta-analysis. *J Clin Sleep Med* 2015;[Epub ahead of print].

Tuomilehto HP, Seppa JM, Partinen MM, et al. Lifestyle intervention with weight reduction: first-line treatment in mild obstructive sleep apnea. *Am J Respir Crit Care Med* 2009;**179**:320–7.

White LH, Lyons OD, Yadollahi A, et al. Effect of below-the-knee compression stockings on severity of obstructive sleep apnea. *Sleep Med* 2015; **16**:258–64.

Woodson BT, Soose RJ, Gillespie MB, et al; STAR Trial Investigators. Three-year outcomes of hypoglossal cranial nerve stimulation for obstructive sleep apnea: the STAR trial. *Otolaryngol Head Neck Surg* 2016. [In press].

A complete reference list can be found online at ExpertConsult.com.

Chapter
117

Obstructive Sleep Apnea and the Central Nervous System: Neural Adaptive Processes, Cognition, and Performance

Ivana Rosenzweig; Terri E. Weaver; Mary J. Morrell

Chapter Highlights

- Patients with obstructive sleep apnea (OSA) demonstrate variable degrees of cognitive, emotional, and performance deficits.
- OSA is increasingly recognized as one of the potentially modifiable risk factors for dementia; its multiple effects on the central nervous system are acknowledged, albeit their nature and prognosis are yet to be fully understood.
- During nocturnal apnea-hypopnea episodes and sleep fragmentation, both maladaptive and adaptive pathways are likely initiated in the brain of patients; the net result likely depends on the chronicity of process and idiosyncratic characteristics of each patient.
- Treatment of OSA with continuous positive airway pressure results in consistent improvement in cognition and performance, although the magnitude of improvement is variable.
- The role of continuous positive airway pressure and its long-term effectiveness with regard to cognitive and performance deficits need further study.

Obstructive sleep apnea (OSA) is one of the potentially modifiable risk factors for dementia,[1-4] and it is commonly associated with serious cardiovascular and metabolic comorbidities.[5-7] Nocturnal episodes of complete or partial pharyngeal obstruction in patients with OSA result in intermittent hypoxia, reoxygenation, hypercapnia, and sleep fragmentation.[8,9] An increase in respiratory effort, in association with hypoxemia or hypercapnia, triggers the frequent sleep arousals, which usually terminate the apneic episodes but also contribute to abnormal sleep architecture and lighter and less restorative sleep.[10] Progressive changes in sleep quality and structure, changes in cerebral blood flow, neurovascular and neurotransmitter changes, and the cellular redox status and neural regulation in OSA patients all may constitute contributing factors to cognitive decline.[8,11-13]

Increased road traffic accidents, reduced quality of life, excessive daytime sleepiness, labile interpersonal relationships, and decreased work and school efficiency have all been documented in OSA patients.[13] These impairments and deficits are not always reversed with treatment.[14] Beneficial effects of treatment on cognitive performance, sleepiness, and neural injury in OSA (Figure 117-1) are, however, documented in recent meta-analyses[15,16] and a meta review.[17] Two studies also suggest beneficial effects of continuous positive airway pressure (CPAP) therapy in minimally symptomatic, and older OSA patients, respectively.[18,19] In a recent study of a well-characterized longitudinal cohort (the Alzheimer's Disease Neuroimaging Initiative cohort), the self-reported presence of untreated sleep-disordered breathing, including "obstructive sleep apnea" and "sleep apnea," was associated with an earlier age at cognitive decline, up to a decade.[4] This association was found to be significant even when accounting for possible confounding factors such as sex, apolipoprotein ε4 status, diabetes, depression, body mass index, cardiovascular disease, hypertension, age at baseline, and education of participants. Moreover, this link appeared significantly attenuated in patients who used CPAP, suggesting that use of CPAP may delay progression, or onset, of cognitive impairment.[4] However, the effect of CPAP on delay in age at Alzheimer disease dementia onset was not demonstrated in this study.[4]

The current dearth of fully effective treatments for the central nervous system (CNS) sequelae of OSA is likely to be a reflection of an as yet poorly understood intricate interplay of both adaptive and maladaptive processes with the hypoxemia, reoxygenation, hypercapnia or hypocapnia, and sleep fragmentation that occur in the CNS of OSA patients.[13] The overall net result of ongoing neuroinflammatory processes and ischemic preconditioning for each particular patient depends on the stage of this OSA-induced dynamic process, effects on other body systems, cognitive reserve, and idiosyncratic susceptibility.[11,13,20,21] Thus, different therapeutic approaches might benefit different stages and conversely might aggravate damage in some patients.[11,13,20] This chapter addresses recent clinical and translational findings regarding the effects of intermittent hypoxia and sleep fragmentation on the CNS, describes the known cognitive and psychological deficits

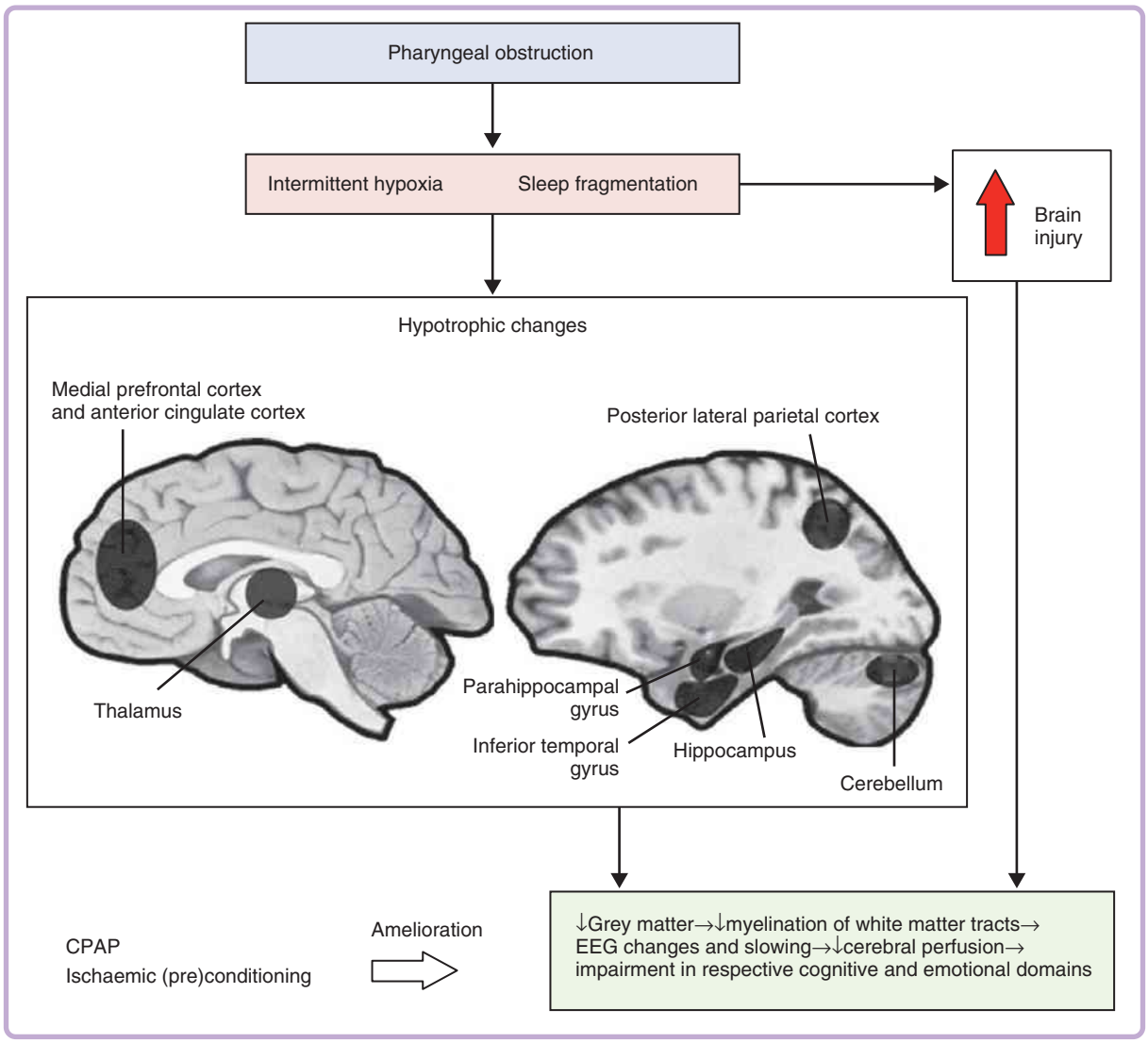


Figure 117-1 Brain regions and mechanisms involved in sleep apnea injury. The nocturnal episodes of complete or partial pharyngeal obstruction result in intermittent hypoxia and sleep fragmentation. Both intermittent hypoxia and sleep fragmentation can aggravate brain injury (*red arrow*) and cause hypotrophic changes in several brain regions shown.[128] Ensuing neurophysiologic and neurochemical changes can also manifest in cognitive and emotional deficits that can be ameliorated (*white arrow*) with continuous positive airway pressure therapy (CPAP) and/or ischemic preconditioning. EEG, Electroencephalography. (From Rosenzweig I, Glasser M, Polsek D, et al. Sleep apnoea and the brain: a complex relationship. *Lancet Respir Med* 2015;3:404–14.)

in patients with OSA, and proposes etiologic mechanisms behind the complex relationship between OSA and the CNS.

NEUROPATHOLOGY OF OBSTRUCTIVE SLEEP APNEA

Changes in cerebral blood flow that occur during obstructive apneas[22] and apnea-induced hypoxemia, combined with reduced cerebral perfusion, likely predispose patients to nocturnal cerebral ischemia.[23,24] In addition, an altered resting cerebral blood flow pattern in several CNS regions has been shown in OSA, along with hypoperfusion during the awake states.[25] Numerous clinical studies have demonstrated changes in the electroencephalogram of OSA patients compared with healthy individuals, including aberrant cortical excitability[26-28] and an associated array of neurocognitive deficits.[10] Taken collectively, such studies have also delineated a putative neurocircuitry "fingerprint" of OSA-induced brain injury and have suggested a disconnection of the frontal regions (Figure 117-2) and a disruption of the (cerebellar)-thalamocortical oscillator, with involvement of the hippocampal formation.[9,10] It has been previously suggested that the constellation of symptoms frequently encountered in OSA patients, such as depression, disturbances in attention, dysmetria of thought and affect, and executive and verbal memory deficits,[29-31] point to similarities with two other recognized neurologic clinical syndromes, frontal lobe syndrome and the cerebellar cognitive affective syndrome.[10,32]

The prefrontal model posits that the sleep disruption, intermittent hypoxemia, and hypercapnia experienced by OSA

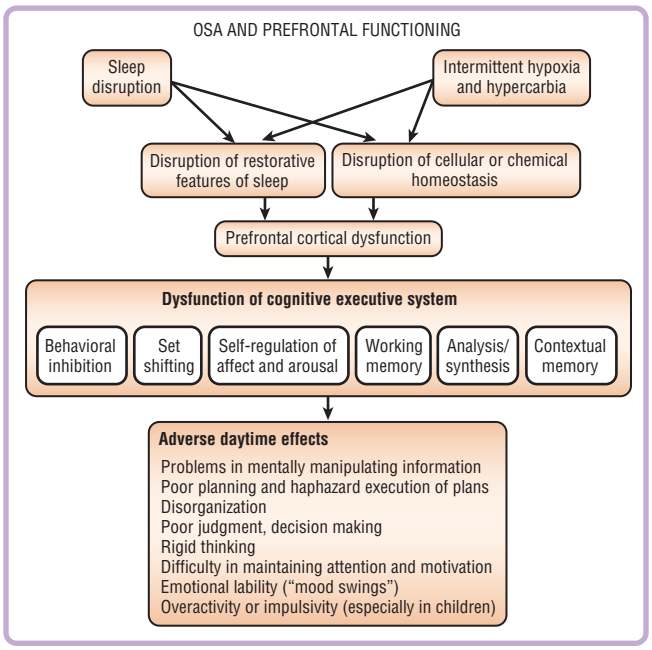


Figure 117-2 The proposed prefrontal model. In this model, obstructive sleep apnea–related sleep disruption and intermittent hypoxemia and hypercarbia alter the efficacy of restorative processes occurring during sleep and disrupt the functional homeostasis and neuronal and glial viability within particular brain regions, particularly the prefrontal regions of the brain cortex. (From Beebe DW, Gozal D. Obstructive sleep apnea and the prefrontal cortex: towards a comprehensive model linking nocturnal upper airway obstruction to daytime cognitive and behavioral deficits. *J Sleep Res* 2002;11:1–16.)

patients alter the normal restorative process that occurs during sleep, generating cellular and biochemical stresses that result in disruption of functional homeostasis and altered neuronal and glial viability within certain brain regions, primarily the prefrontal regions of the brain cortex.[33,34] This model has been proposed as a theoretical framework for the relationship between sleep fragmentation and nocturnal hypoxemia and predominantly frontal deficits (see Figure 117-2).[34] OSA-induced neuropathologic alterations can lead to destabilization of the executive system, causing behavioral disturbance in inhibition, maintenance of performance, self-regulation of affect and arousal, working memory, analysis and synthesis, and contextual memory.[33,34] Alterations in the executive system can adversely affect cognitive abilities, resulting in maladaptive types of behavior as depicted in Figure 117-2.[33,34] Nonetheless, unlike some other neurologic disorders, the impairments associated with OSA are more likely to produce inefficient performance rather than inability to perform.[34] For example, when memory- or divided attention–related neuronal circuitry is incapacitated, other CNS systems and circuitries likely get recruited in an effort to compensate.[33,34] However, if such systems are themselves affected by sleep fragmentation or hypoxemia, their compensatory contributions might be suboptimal. This may account for the increased activation of the prefrontal cortex under conditions of sleep deprivation documented by functional magnetic resonance imaging.[33,34] Impairments in performance of OSA patients can be further explained by deficits in elementary cognitive functions, specifically, sensory transduction, feature integration, and motor preparation and execution, which are required, even in simple response-time tasks.[34,35] Corresponding to the listed deficits in OSA patients, the neuroanatomic regions that have most commonly been reported in clinical and animal studies as affected in OSA suggest that both the cerebellar modulation of neural circuits and the normal state-dependent flow of information between thalamus (and basal ganglia) and frontoparietal cortex are likely to be affected in susceptible patients (see Figure 117-1).[10,36-41]

Some clinicians have argued against such a reductionist approach to OSA-induced brain injury and point out that emerging research indicates that the relationship between OSA disease severity and cognitive dysfunction is the product of a multitude of susceptibility and protective factors and that sleep fragmentation, hypoxemia, and cognitive reserve are only three such aspects.[9,10,16] Other commonly overlooked factors are duration of the disease, role of the blood-brain barrier, presence of hypertension, metabolic dysfunction, and systemic inflammation, levels of cerebral blood flow, and genetic vulnerability.[16] Further research is necessary to provide a clear understanding of the risk for neurocognitive dysfunction and the benefit and optimization of treatments.

Affected Neurocognitive Domains

Despite contradictory results and ongoing polemics in the field, most studies to date agree that patients with OSA can have significant deficits in attention and vigilance, long-term visual and verbal memory, visuospatial and constructional abilities, and executive function.[13,17,29] Several associations have been recognized, including the association between worsening global cognitive functioning and the severity of hypoxemia as well as the association between attention and vigilance dysfunction and the degree of sleep fragmentation.[13,17] Consensus is less strong on the effects of OSA on working memory and short-term memory.[17] In some studies, language ability and psychomotor functioning have been shown to be largely unaffected by OSA,[17] whereas others have pointed to psychomotor slowing as the most vulnerable cognitive domains and also the least responsive to treatment with CPAP.[42] Similarly, several studies showing impairments in language abilities in patients with severe OSA have not shown agreement on whether phonemic or semantic domains have the greatest effect.[43] Neurodevelopmental stages of adolescents and children with OSA appear to dictate a higher risk for this deficit.[44]

In children with OSA, the results of studies assessing cognitive performance and effects of treatment are similarly divergent.[45,46] In a recent study of children 7 to 12 years of age with sleep-disordered breathing (SDB), who were followed for 4 years, treatment of the SDB led to improvements in several aspects of neurocognition, collectively categorized as performance IQ.[45] Performance IQ represents fluid intelligence that is reflective of incidental learning, and it describes one's ability to adapt to new situations.[47] In this study, improvements were recorded in tasks associated with spatial visualization, visuomotor coordination, abstract thought, and nonverbal fluid reasoning.[45] However, overall improvements in academic ability or behavior were less clear. Furthermore, tendency to worsening of verbal IQ, which, unlike performance IQ, is more likely to be affected by formal education and learning experiences, was noted in a treated group.[45] A definitive explanation for this finding was not provided, and no statistically significant association between the reduction

in verbal IQ performance and treatment was demonstrated.[45] Conversely, in another influential study, younger children with SDB followed for 12 months of treatment showed significant improvements in academic performance.[46] The different neurodevelopmental ages of children and different test parameters used provide for a complex clinical data set, against which no finite conclusions can be drawn. Nonetheless, particular patterns and associations seem to be emerging from this and earlier work, among which the association between performance IQ and slow wave activity (SWA) during non–rapid eye movement (NREM) sleep is perhaps the strongest one.[45,48] It has been argued that cognitive improvements in treated OSA patients may reflect increased stability of brain activity during sleep, allowing crucial synaptic repair and maintenance to occur and counteracting toxic effects of arousal and hypoxic effects of OSA.[45,49] This argument is concordant with findings showing that the neurochemical and gene environments of sleep and sleep activity patterns present crucial window periods during which the brain can restore cellular homeostasis, increase signal-to-noise ratio, and reinforce neuronal circuitry for subsequent cognitive processing demands.[12,50,51]

PROPOSED MECHANISTIC ROLE FOR PERTURBED SLEEP IN PATIENTS WITH OBSTRUCTIVE SLEEP APNEA

Sleep and sleep deprivation alter molecular signaling pathways that regulate synaptic strength, plasticity-related gene expression, and protein translation in a bidirectional manner.[51] Moreover, sleep deprivation can impair neuronal excitability, decrease myelination, and lead to cellular oxidative stress and misfolding of cellular proteins.[51,52] Frequent brief awakenings lead to fragmented sleep that negatively affects the next day's cognitive and emotional functioning, in a manner similar to that of total sleep deprivation.[2] Several studies have attempted to assess whether OSA patients are more vulnerable to sleep-loss-induced performance deficits, with special emphasis on driving performance variables, with varied results.[53-56] From the practical point of view it is of major interest to develop reliable and practical bedside tests to help clinicians advise patients on their individual risk for traffic accidents.[13] Preclinical animal studies suggest that sleep fragmentation independently affects similar brain regions to those affected by intermittent hypoxia, as occurs in OSA.[8] Also, clinical studies of the effects of sleep deprivation on cognition in the general population suggest comparable cognitive impairments to those seen in OSA.[57] Frequent partial arousals during sleep in OSA patients contribute to abnormal sleep architecture and symptoms of excessive daytime somnolence (i.e., sleepiness).[8,9] An independent association between excessive daytime somnolence and cognitive impairment has been demonstrated, and several prospective studies have shown that excessive daytime somnolence is associated with an increased risk for cognitive decline and dementia.[1] Further, in a prospective cohort study of Japanese American men in the Honolulu-Asia Aging Study, lower nocturnal oxygenation and reduction in stage 3 (slow wave) NREM sleep were associated with the development of microinfarcts and brain atrophy.[58] Conversely, men with longer slow wave sleep time showed slower cognitive decline.[58]

The relationship between OSA and its effect on selected sleep stages merits particular attention, given that each of the sleep stages, with its attendant alterations in neurophysiology, is associated with facilitation of important functional learning and memory processes[12] (also see Chapter 22). In OSA patients, the proportion of stage 2 NREM sleep (N2) has been shown to be increased, whereas proportions of stages 1 and 3 NREM sleep (N1, N3) and rapid eye movement (REM) sleep are decreased.[43] Limited experimental studies conducted to date have shown specific impairments of sleep-dependent consolidation of verbal declarative information in patients with OSA.[59] Furthermore, several recent clinical studies suggest disturbed spatiotemporal evolution of sleep spindles in patients with OSA during the night.[60,61]

However, dynamic analysis of sleep architecture is required to fully gauge the neurophysiologic effect of sleep fragmentation on sleep in OSA patients.[13] For example, in one study of mild OSA the exponential decay function of SWA was demonstrated to be significantly slower in OSA patients compared with controls.[62] This was due to the more even distribution of SWA throughout the night, without significant decrease in total slow wave and REM sleep time. These results show that mild sleep fragmentation can alter the dynamics of SWA, without significantly decreasing the amounts of slow wave and REM sleep, and emphasize the need to perform SWA decay analysis in sleep fragmentation disorders.[62] In the same study, a decrease in spindle activity was observed in N2 and N3 sleep that was not attributed to an increase of SWA.[60,62] Such a reduction in total spindle density has also been reported in sleep maintenance insomnia and is likely to be related to sleep fragmentation.[60-62]

The model proposed by Landmann and colleagues[63] suggests an integrative framework for the qualitative reorganization of memory during sleep.[63] It further builds on studies that have shown that sleep facilitates the abstraction of rules and the integration of knowledge into existing schemas during slow wave sleep.[50,51,63] REM sleep, on the other hand, has been shown to benefit creativity that requires the disintegration of existing patterns.[63] Both respective sleep stages have been commonly reported as reduced or fragmented in patients with OSA, and their dysregulation could underlie some of the frequently reported cognitive and performance deficits in OSA patients.[27,43] In line with this argument, one study that investigated the neurocognitive deficits in OSA found that the number of microarousals during the night was the best predictor of episodic memory deficit.[64] Traditionally, obstructive events during NREM sleep have been viewed as associated with greater cognitive deficits or impaired quality of life, whereas REM sleep events have been shown to be associated with greater sympathetic activity, arterial hypertension, and cardiovascular instability in patients with OSA.[65,66] Recently, the role for fragmented REM sleep in spatial navigational memory in OSA patients has been addressed with a physiologically relevant stimulus.[67] During this study, patients spent two different nights in the laboratory, during which they performed timed trials, before and after sleep, on one of two unique three-dimensional spatial mazes.[67] Normal consolidation of sleep was achieved with use of therapeutic CPAP throughout the first night, whereas during the second night CPAP was reduced only during the REM stages. Patients showed improvements in maze performance after a night of normal sleep, but those improvements were significantly reduced following a night of isolated REM disruption, without changes in psychomotor vigilance. Noted cognitive

improvements were positively correlated with the mean REM run duration across both sleep conditions.[67]

It has been argued that the sense of excessive daytime sleepiness and of feeling unrefreshed in the morning in some OSA patients could be due to the inability to augment NREM SWA or REM sleep. Moreover, in some OSA patients, reduction of REM sleep can lead to dissociation of REM traits with other sleep stages, further affecting critical sleep windows for memory formation and consolidation.[12] Equally, it has been shown that when high homeostatic demands are not fully met during sleep, in the subsequent wake period microsleeps can occur in highly active regions of the brain[68] and can lead to concomitant disability for the function subserved by that region.[50,68] To what degree this takes place in OSA patients and whether this also contributes to attention-vigilance dysfunction and the higher frequency of traffic accidents noted for this patient group are yet to be fully defined.[13] Previously reported retarded SWA decay throughout the night, even in patients with mild OSA, further supports the notion of nonrestorative sleep in OSA.[62]

Several recent studies have aimed to discern the role of sleep in cognition and cognitive decline, with potential effect on the way we consider sleep in patients with OSA. For example, as depicted in Figure 117-3, *A*, it has been suggested that the amount of atrophy in the medial prefrontal cortex (mPFC), predicts the extent of disrupted slow wave (N3) sleep in older people, and consequent impaired overnight episodic hippocampal memory consolidation.[51,69] The mPFC area has been shown to be independently affected by OSA (see Figure 117-1) and has been known to be involved in the generation of slow waves.[13,69] It has been proposed that improving slow wave sleep in older adults (irrespective of their OSA status) may represent a novel treatment for minimizing cognitive decline in later life.[69]

The importance of sleep spindles in cognition has also gained interest over the past several years.[13,70] It has been shown that, during the night, OSA patients, unlike healthy controls, display a significant proportion of slow spindles in the frontal, central, and parietal regions.[60] One recent study has shown that older adults who express fewer prefrontal fast sleep spindles also exhibit a proportional impairment in hippocampal functioning during the subsequent wake periods and, with that impairment, a deficit in the ability to form new episodic memories.[71] Fast sleep spindles represent part of a coordinated NREM sleep–dependent memory mechanism, and it is thought that hippocampal sharp-wave ripples provide feedback excitation, which initiates neuroplasticity in spindle-activated cortical neurons.[51] Relative to slow sleep spindles, fast sleep spindle activity is associated with greater hippocampal activation and greater hippocampal-cortical functional connectivity.[2,71] The sleep architecture of even mild OSA patients shows a high degree of sleep fragmentation, which results in a different time course of SWA and a decreased sleep spindle index compared with controls.[62] Whether this deregulated spindle formation and activity present another contributory facet to cognitive complaints in patients in OSA, however, remains a conjecture at this point. Nonetheless, taken together, these studies suggest a possible role for OSA-induced brain injury in the acceleration, or even initiation, of cognitive decline in older adults (see Figure 117-3, *A*).[1-3,31,69] The exact pathophysiology of such an association remains elusive.[13]

Mental Health and Sleep Associations in Obstructive Sleep Apnea Patients

A bidirectional relationship between sleep and the function of the brain circuitry involved in emotions is increasingly supported by studies that further build on long-standing clinical observations of co-occurring mood and sleep disorders.[13,72] Unsurprisingly, then, a variety of mental health issues, such as affective disorders, emotional lability, and depression, have been reported as highly prevalent in individuals with OSA, with some studies reporting that up to 63% of OSA individuals are so affected[73] despite considerable heterogeneity and a high risk for bias in these studies.[13] Evidence from various studies is particularly suggestive of a role for REM sleep in selective emotional memory processing and sleep-dependent emotional memory depotentiation (see Figure 117-3, *B*).[72] Moreover, REM sleep is suggested to play a role in recalibrating the sensitivity and specificity of the brain's response to emotional events, both positive and negative.[72] This recalibration effect likely occurs, at least in part, as a result of modulation of noradrenergic brainstem activity and the responsive profiles of the amygdala and mPFC, two regions critically involved in detecting emotional salience.[13,72]

Of the psychiatric disorders, the evidence for increased prevalence of OSA is particularly strong in major depressive disorder and posttraumatic stress disorder (PTSD),[74,75] both independently associated with REM sleep disturbance.[13] Even though the causal relationship between these affective disorders and OSA is unclear and is likely to be multifactorial, the potential sleep mechanics of their interaction is worthy of further consideration.[13] PTSD is independently associated with decreases in the total time spent in REM sleep. It is also associated with marked fragmentation of REM sleep, indicative of arousal-related awakenings from REM sleep linked to adrenergic surges.[72] CPAP adherence has been shown to be reduced in veterans with PTSD and comorbid OSA.[74] Based on the current knowledge of OSA-induced sleep deficits, it can be argued that in PTSD patients with comorbid OSA, the additive effect of sleep disturbances associated with the OSA can further impair the quantity and quality of REM sleep. This would likely also affect the REM noradrenergic "housekeeping" function because it has been shown that REM sleep reduces, and thus likely restores, concentrations of CNS noradrenaline to baseline, allowing for optimal awake state functioning.[13,72] More specifically, several studies suggest that quiescence of locus coeruleus activity, a brainstem structure that is a source of noradrenergic input, during REM sleep throughout the night restores the appropriate next-day tonic-phasic response specificity within the emotional salience network (e.g., locus coeruleus, amygdala, mPFC).[72] It is hence feasible that OSA-induced REM fragmentation could further aggravate the hyperadrenergic state of some PTSD patients and lead to decreased connectivity between the PFC and amygdala and thus exaggerated amygdala reactivity.[72] The functional outcome may be an aggravated disease course and worse prognosis.[13,72] Of note, in the prospective Honolulu-Asia Aging Study, in which men (n = 3801) aged 71 to 93 years at baseline (1991) were followed until their death, higher nocturnal oxygenation during REM sleep was associated with less gliosis and neuronal loss in the locus coeruleus.[58]

Major depression, on the other hand, is associated with exaggerated REM sleep qualities and deficiency in

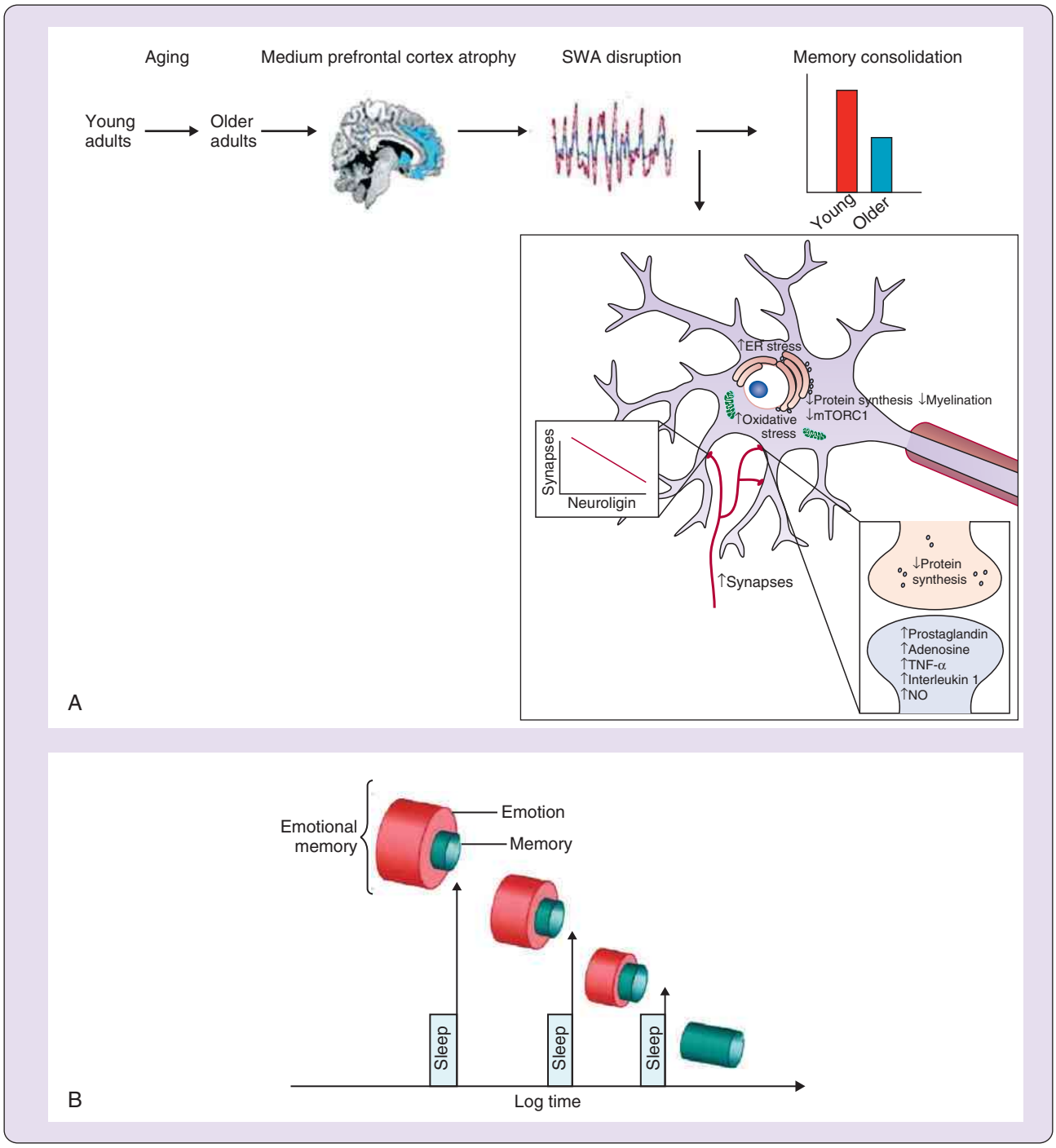


Figure 117-3 The proposed role for sleep in cognition and emotions. **A,** *Cognitive sleep:* Sleep apnea and aging can independently cause gray matter atrophy in the prefrontal cortex. Atrophy can mediate the degree of slow wave activity (SWA) disruption, whereas SWA in turn can mediate the degree of impaired memory retention.[69] SWA activity disruption likely also leads to cellular stress.[52] **B,** *Emotional sleep:* Conceptual schematics of "the sleep to forget and sleep to remember" model are shown, as described by Goldstein and Walker.[72] Over one or several nights and numerous repetitions of this REM mechanism, sleep transforms an emotional memory into a memory of an emotional event that is no longer emotional.[72] ER, Endoplasmic reticulum; IL-1, interleukin-1; mTORC1, mammalian target of rapamycin contact 1; NO, nitric oxide; TNF-α, tumor necrosis factor-α. (From Rosenzweig I, Glasser M, Polsek D, et al. Sleep apnoea and the brain: a complex relationship. *Lancet Respir Med* 2015;3:404–14.)

monoamine activity.[72] The bidirectional-dual relationship between major depression and OSA has been suggested by findings of several studies.[31] In some OSA patients, fragmented REM sleep can precipitate a vicious cycle of impaired REM regulation and rebound REM augmentation.[13] This, along with concomitant changes in neurotransmitter systems caused by hypoxemia, could further lead to reduced monoamine activity, with associated increased negative rumination and ensuing depression, in genetically predisposed individuals.[13] Through its effects on REM sleep, comorbid OSA might

also lead to dysfunctional consolidation and depotentiation of emotional memory from prior affective experiences.[13,72] It has been proposed that this may result in a condition of chronic anxiety within autobiographic memory networks (see Figure 117-3, *B*).[72] In support of this recent meta-analysis of randomized controlled trials of treatment of OSA, a significant improvement in depressive symptoms was reported.[76]

Even though the previously argued theoretical constructs of a bidirectional relationship between fragmented or disturbed sleep in OSA and psychiatric disorders are indirectly supported by animal and neuroimaging studies of sleep,[72] the underlying mechanics are likely to be more complex and, as such, require further well-designed studies.[13]

NEUROINFLAMMATION AND ISCHEMIC PRECONDITIONING

Cognitive and emotional complaints of OSA patients may also be explained by oxidative and neuroinflammatory effects of OSA on the CNS emotional salience network.[10,13,31] In OSA, repetitive occlusions of the upper airway lead to intermittent hypoxia and recurrent hypoxemia, typically characterized by short cycles of hypoxemia and reoxygenation.[20] However, the patterns vary greatly among patients, and, depending on the idiosyncratic characteristics of each individual, the end results might be either adaptive or maladaptive.[20] The outcome will likely depend on the dynamic interplay between the specific type and amount of reactive oxygen-nitrogen species produced, duration and frequency of such production, the intracellular localization, and microenvironmental antioxidant activity.[11] Additional interplay depends on factors such as genetic makeup, nutrition, and other lifestyle-related variables, all of which affect the redox status.[11,20] A variety of studies to date suggest that the severity of hypoxia, its duration, and its cycle frequency are fundamental determinants of outcomes (Figure 117-4, *A*).[77,78] For example, it has been generally acknowledged that short, mild, and lower cycle frequencies of intermittent hypoxia may generate beneficial

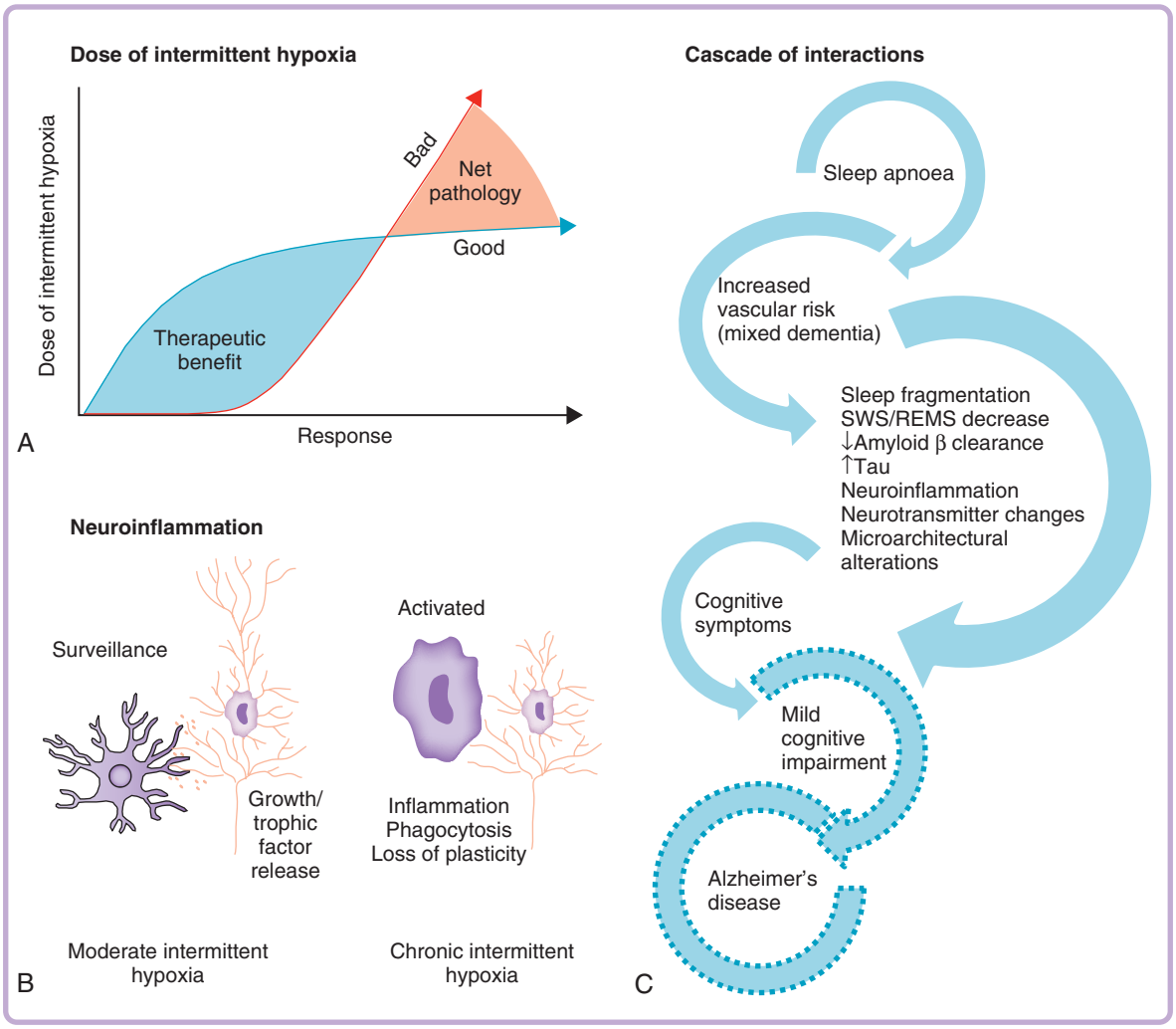


Figure 117-4 Adaptive and maladaptive processes induced by intermittent hypoxia. **A,** Conceptual presentation of the net effect of cycles of intermittent hypoxia, of varied length and frequency, over a period of time (minutes to days to weeks), as described by Dale and colleagues.[78] High doses still elicit neuroadaptive mechanisms, but the balance is shifted and maladaptive processes such as neuroinflammation **(B)** are likely to be instigated. Finding an optimal dose is key to developing effective treatment.[78] **C,** Possible cascade of interactions between sleep apnea and Alzheimer disease. (From Rosenzweig I, Glasser M, Polsek D, et al. Sleep apnoea and the brain: a complex relationship. *Lancet Respir Med* 2015;3:404–14.)

and adaptive responses in the brain, such as ischemic preconditioning.[20] Conversely, chronic, moderate to severe, and high-frequency intermittent hypoxia can induce maladaptive disruption of homeostatic mechanisms, leading to dysfunction and sterile neuroinflammation.[11,20]

Ischemic preconditioning represents a generalized adaptation to ischemia by a variety of cells.[21,79] In OSA, induction of ischemic preconditioning is thought to be due to the activation of several gene programs, including the hypoxia inducible factor-1, vascular endothelial growth factor, erythropoietin, atrial natriuretic peptide, and brain-derived neurotrophic factor.[80,81] Various end mechanisms and pathways have been shown to play a role in preconditioning, including those of long-term facilitation of phrenic motor output, chemoreflex activation, vascular remodeling, neo-angiogenesis, productive autophagy, reactive gliosis, various synaptic alterations, and modulation of adult hippocampal neurogenesis.[11,82,83] CPAP treatment of OSA has been shown to partially reverse structural imaging changes in gray matter of hippocampal regions and to ameliorate some of the associated cognitive deficits, possibly also by modulating adult neurogenesis.[84] In a recent neuroimaging study, coexistence of hypotrophic and hypertrophic changes in the brain of OSA patients was taken to reflect the evolving nature of OSA-associated brain injury.[36] It has been proposed that at any given time ongoing maladaptive neuroinflammatory processes likely exist alongside adaptive mechanisms of increased brain plasticity and ischemic preconditioning.[13] As a corollary to these findings, in a recent study that compared the cognitive performance of patients with high and low levels of OSA-related hypoxemia, controlling for demographic factors and other aspects of OSA severity, an unexpected advantage of higher levels of hypoxemia on memory was demonstrated in a carefully matched clinical cohort.[85]

Several studies also suggest that, under certain conditions, intermittent hypoxia can increase immune defenses without exacerbating inflammation.[11,20] Moreover, in animals, short-lasting hypoxic exposures mimicking OSA have been associated with recruitment of bone marrow–derived pluripotent stem cells, which exhibited upregulation of stem cell differentiation pathways, particularly involving CNS development and angiogenesis.[20]

Another powerful central neuroprotective adaptive mechanism for ischemic events has been demonstrated following the activation of the intrinsic neurons of the cerebellar fastigial nucleus.[86] Neurostimulation of these nuclei appears to provide "protective" reduced excitability of cortical neurons during subsequent ischemic episodes and to lead to reduced immunoreactivity of cerebral microvessels.[10] Also, a "compensatory" entraining of cerebellum by hypertrophic hippocampi has been proposed to occur in some younger patients with mild OSA.[36] Although there are no direct monosynaptic anatomic connections between hippocampi and cerebellum, their connectivity is thought to be important for the control of movement under states of heightened emotion and novel conditions and for associative learning.[10,13] Failed adaptation of cerebellar networks to injury, of any etiology, has been shown to lead to cognitive deficits and hyperactivity, distractibility, ruminative behaviors, dysphoria, and depression in some patients.[10,32]

Neuroinflammation in Obstructive Sleep Apnea

There are, however, relevant maladaptive effects of intermittent hypoxia.[13] These include neuroinflammation, and although the exact neurocellular sources for associated processes are still incompletely defined, activation of astroglia is likely to be important.[11,13,78] In addition, oligodendrocytes, myelin-producing cells of the CNS, have been shown to be selectively sensitive to hypoxia and sleep fragmentation.[87,88] The subsequent loss of buffering functions can ultimately contribute to pathologic processes, such as increased glial proliferation and microglial activation (see Figure 117-4, *B*).[13,78] Astroglial and microglial cells play critical roles in regional blood flow regulation and inflammatory processes in the brain, as well as critical coordination of bioenergetics through lactate transport.[78] Under normal conditions, microglia in the healthy CNS exhibit a surveillance phenotype that synthesizes and releases neuroprotective growth and trophic factors.[78] However, severe and prolonged hypoxia can activate microglia toward a toxic, proinflammatory phenotype that triggers pathology, including hippocampal apoptosis, impaired synaptic plasticity, and cognitive impairment.[78] Neuroinflammation has been shown to independently increase the brain's sensitivity to stress, resulting in stress-related neuropsychiatric disorders, such as anxiety and depression.[13,89] Dynamic changes in transcription of inflammatory genes have been demonstrated following exposure to intermittent hypoxia.[13,78] Increased prostaglandin E_2 neural tissue concentrations have also been demonstrated in hippocampal and cortical regions accompanied by lipid peroxidation of polyunsaturated fatty acids.[78] Similarly, it has been shown that increased carbonylation- and nitrosylation-induced oxidative injury emerges in susceptible brain regions following exposure to intermittent hypoxia and promotes excessive daytime somnolence.[11,78] Recently, toll-like receptor 4 (TLR4) expression and activity have been demonstrated to be increased on monocytes of patients with OSA.[90] Similarly, ligands for TLR4 have been shown to be increased in the serum of children with OSA.[13,90] The microglia of the cortex and brainstem exhibit TLR4 expression after chronic intermittent hypoxia, when it is postulated to play a region-specific and differential (adaptive or maladaptive) role.[13,90] This finding is of particular interest because TLR4 has also been strongly implicated in several inflammatory and neurodegenerative disorders, including vascular dementia and Alzheimer disease.[90] In cognitively healthy adults, intermittent hypoxia has been correlated with increases in phosphorylated and total tau and amyloid β^{42} concentrations in cerebral spinal fluid, key components of Alzheimer pathology.[1,13] Similarly, cerebral amyloidogenesis and tau phosphorylation, along with neuronal degeneration and axonal dysfunction, have been demonstrated in the cortex and brainstem of animals exposed to intermittent hypoxia.[2] Taken together, these findings support the role for neuroinflammatory processes in cognitive and emotional deficits of OSA patients. They further suggest a close association between hypoxemia-induced maladaptive processes and dementia (see Figure 117-4, *C*).[13]

NEUROLOGIC DISORDERS AND COGNITIVE AND PERFORMANCE DEFICITS ASSOCIATED WITH OBSTRUCTIVE SLEEP APNEA

Several neurologic disorders have been associated with OSA.[10] For example, adults with epilepsy appear at increased risk for OSA.[91] Conversely, OSA is a recognized independent risk factor for stroke.[10] OSA has been associated with seizure

exacerbations in older adults with epilepsy, and treatment with CPAP may represent an important avenue for improving seizure control in this population.[10,92,93] OSA-induced brain injury is believed to exacerbate neural damage during incident stroke as well as to increase the risk for a subsequent stroke.[94,95]

Additionally, an increasing body of evidence from animal studies suggests that cerebral amyloidogenesis and tau phosphorylation, two cardinal features of Alzheimer disease, can be triggered by intermittent hypoxia.[2] Intermittent hypoxia and associated generation of reactive oxygen species, known to occur during nocturnal apneic episodes, have been shown to initiate neuronal degeneration and axonal dysfunction in the cortex and brainstem of animals.[2,10] Also, oligodendrocytes, myelin-producing cells of the CNS, are selectively sensitive to hypoxia and sleep fragmentation.[10,88] However, it is not clear to what extent this particular vulnerability contributes to the widely reported hypotrophic white matter changes in the brains of some OSA patients, including the fornices and corpus callosum.[10,96,97] Impaired learning capabilities have been documented in children with OSA, along with increased hyperactivity and incidence of attention deficit disorders.[8] On the other end of the age spectrum, as noted earlier, several clinical studies have suggested that older patients with OSA might suffer accelerated brain atrophy, cognitive decline, and the onset and severity of dementia.[77,98,99]

It has been estimated that approximately 80% of OSA patients complain of both excessive daytime sleepiness and cognitive impairment, and half also report personality changes.[34] However, the exact prevalence of neurocognitive deficits in patients with OSA remains unknown. One in four patients with newly diagnosed OSA has appreciable neuropsychological impairments.[34,100] Various studies suggest that memory impairments can be found in up to 9% of OSA subjects; 2% to 25% have problems with sustained attention, and 15% to 42% demonstrate difficulties with executive functioning.[34,101] Moreover, the increased frequency of work-related and traffic accidents in OSA patients may be taken as a surrogate indicator of neurobehavioral performance deficits.[34,102,103] Patients with OSA are 37 times more likely to complain of sleepiness compared with nonsnoring healthy controls. Work limitation in terms of difficulties with time management, mental tasks, interpersonal relationships, and work output have all been associated with excessive daytime sleepiness.[34] OSA patients are 7.5 times more likely to have difficulties with concentration at work, have a ninefold increase in difficulty learning new tasks, and are 20 times more likely to have problems performing monotonous tasks.[34,104] In addition, occupational accidents have been reported to occur in 50% of male OSA patients, whereas the risk for occupational accidents in women with OSA has been reported as six times greater than in controls.[34,102,105]

Of particular note is the finding that motor vehicle drivers, regardless of OSA status, do not always perceive their impairment and continue to drive while sleepy.[34,106] Overall, compared with normal controls, OSA patients are 2 to 13 times more likely to experience a driving-related traffic accident.[107] Such accidents are more likely to occur in those who manifest greater daytime sleepiness.[34,107] However, OSA has also been associated with motor vehicle crashes independent of daytime sleepiness.[108] Sleepiness due to work schedules and sleepiness due to OSA are independent risk factors for accidents.[34]

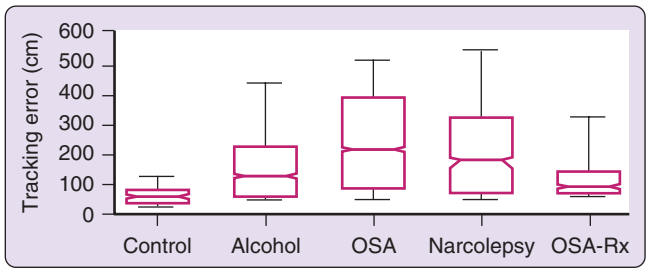


Figure 117-5 Summary of tracking errors in different groups on the Divided Attention Driving Task. OSA, Obstructive sleep apnea. (From George CF. Vigilance impairment: assessment by driving simulators. *Sleep* 2000; 23[Suppl. 4]:S115–18.)

For example, in commercial vehicle drivers, in whom both of these sleepiness-promoting conditions coexist, those with the highest level of sleepiness have a twofold increase in multiple accidents.[109] The data for OSA and automobile crashes are numerous and consistent: as a group, OSA patients' risk for motor vehicle collisions is increased twofold to fourfold (Videos 117-1 and 117-2).[34] On driving simulators, OSA patients hit more obstacles, have increased error in tracking and visual search, have increased response time to secondary stimuli, and drive out of bounds more times compared with non-OSA control subjects (Figure 117-5).[110] Still, not all OSA patients who drive have accidents, and as many as two thirds never have a collision.[106,110] A means to identify OSA patients at greatest risk for motor vehicle collisions is still not clear based on available literature, and this complicates decision making from a medical and legal perspective.[34]

ASSESSMENT OF COGNITIVE AND NEUROBEHAVIORAL PERFORMANCE DEFICITS IN OBSTRUCTIVE SLEEP APNEA

To understand the cognitive and neurobehavioral performance deficits that affect patients with OSA, it is helpful to consider these from a categorical perspective.[34] The effects of sleep loss on performance include changes in cognitive performance, difficulty with working memory, slowing of response or inability to sustain attention across the duration of the task, declines in best effort or fastest response, lapses, and false responses.[34,111] As noted earlier, in OSA, hypoxemia-reoxygenation cycles with attendant biochemical and cellular alterations cause dysfunction of the prefrontal cortex, among other CNS regions.[34] This results in impaired executive function manifesting as false responses, problems with working memory and contextual memory, problems with cognitive processing in addition to deficits in the pattern of responses, and self-regulation of affect and arousal.[33,34] A description of the performance deficits and commonly used assessment techniques in OSA patients is given in Box 117-1.[34] Tests that can readily be performed in the clinical setting include the Digit Symbol Substitution Task (90-second test) to assess cognitive processing and the Psychomotor Vigilance Task (10-minute task) to evaluate the ability to sustain attention.[34] Summary information regarding the neurobehavioral tests may be found elsewhere.[112] The effects of OSA on cognitive processing, memory, sustained attention, and executive and motor functioning are further shown in Figure 117-6, which reports the effects of OSA in patients relative to healthy adults.[34]

Box 117-1 DEFINITION AND ASSESSMENT OF COGNITIVE AND NEUROBEHAVIORAL DEFICITS ASSOCIATED WITH OBSTRUCTIVE SLEEP APNEA

Cognitive Processing

Behavior

Decreased ability to digest information
- Slowing on task
- Increased errors
- Decline in total number correct and/or completed per unit of time

Measures Commonly Used to Assess Deficit

Self-paced tasks of short duration (1 to 5 minutes), including arithmetic calculations, communication, or concept attainment
- Paced Auditory Serial Addition Task (PASAT)
- Trail Making Test Parts A and B: sequencing numbers (A) or letters and numbers (B)
- Category Test: six sets of items organized around different principles with a seventh set comprising previously shown items
- Digit Symbol Substitution Test: supplying matching symbol given the corresponding number
- Digit Backward: stating verbally provided numbers in reverse order
- Letter Cancellation: cancellation of target alphabets from presentation of randomized alphabets

Memory

Behavior

Decreased ability to register, store, retain, and retrieve information

Measures Commonly Used to Assess Deficit

Short-term memory: timed tasks of up to 10 minutes that require free recall of words, numbers, paragraphs, or figures
- Probed, Recall Memory Task (words)
- Digit Span Forward (numbers)
- Wechsler Memory Scale Story Task (paragraph)
- Rey Auditory-Verbal Learning Test (figure)

Long-term memory: presenting the subject with lists of items that are longer than the seven-item memory capacity
- California Verbal Learning Test

Procedural memory: gradual acquisition and maintenance of motor skills and procedures
- Mirror Tracing Task
- Rotary Pursuit Task

Sustained Attention or Vigilance

Behavior

Inability to maintain attention over time.
- Slowing of response time (time on task)
- Increased errors
- Reduction in the fastest optimal response times
- Periods of delayed or no response (lapses)
- Response to stimuli when none presented (false responses)

Measures Commonly Used to Assess Deficit

Short-duration tasks (30 minutes)
- Psychomotor Vigilance Task (PVT)
- Four Choice Reaction Time Test
- Steer Clear
- Continuous performance tests

Divided Attention

Behavior

Inability to respond to more than one task or stimuli, such as with driving

Measures Commonly Used to Assess Deficit

Divided Attention Driving Test (DADT): mimics vigilant-related behavior essential to driving
- Tracking (the ability to stay within the driving lane)
- Visual search (looking for and avoiding obstacles, traffic lights, etc.)

Executive Functioning

Behavior

Problems with manipulating and processing information
Inadequate planning and execution of plans
Disorganization: poor judgment, decision making
Inflexible: emotional liability
Impulsivity
Difficulty maintaining motivation

Measures Commonly Used to Assess Deficit

Volition component or intentional behavior
- Assessed by asking the patients' preferences, what they like to do, or what makes them angry

Planning component
- Porteus Maze Test
- Tower Tests: Tower of London, Tower of Toronto, Tower of Hanoi
- Wisconsin Card Sorting Test

Purposive action
- Tinkertoy Test

Effective performance
- Random Generation Task

Modified from Dinges D. Probing the limits of functional capability: the effects of sleep loss on short-duration tasks. In: Broughton R, Ogilvie R, editors. *Sleep, arousal, and performance*. Boston: Birkhauser; 1992. p. 177–188.

An additional issue to consider when assessing patients with OSA is their subjective cognitive and emotional complaints.[13] A detailed analysis of important studies in the field has suggested only a weak correlation between (subjective) cognitive complaints in patients with OSA and their objective cognitive functioning.[13,113] Divergent results of subjective versus objective complaints have been recognized in other medical populations, and several possible explanations for this in OSA patients have been suggested.[13] For example, an insufficient specificity of current tests for deficits documented in OSA is evident and largely acknowledged. Currently used and validated objective tests for cognition are frequently designed to assess deficits found in patients with traumatic brain injury and as such do not specifically assess impairments in OSA-induced brain injury.[13,113] Cognitive domains are not unitary constructs, and only the carefully deconstructed analysis of

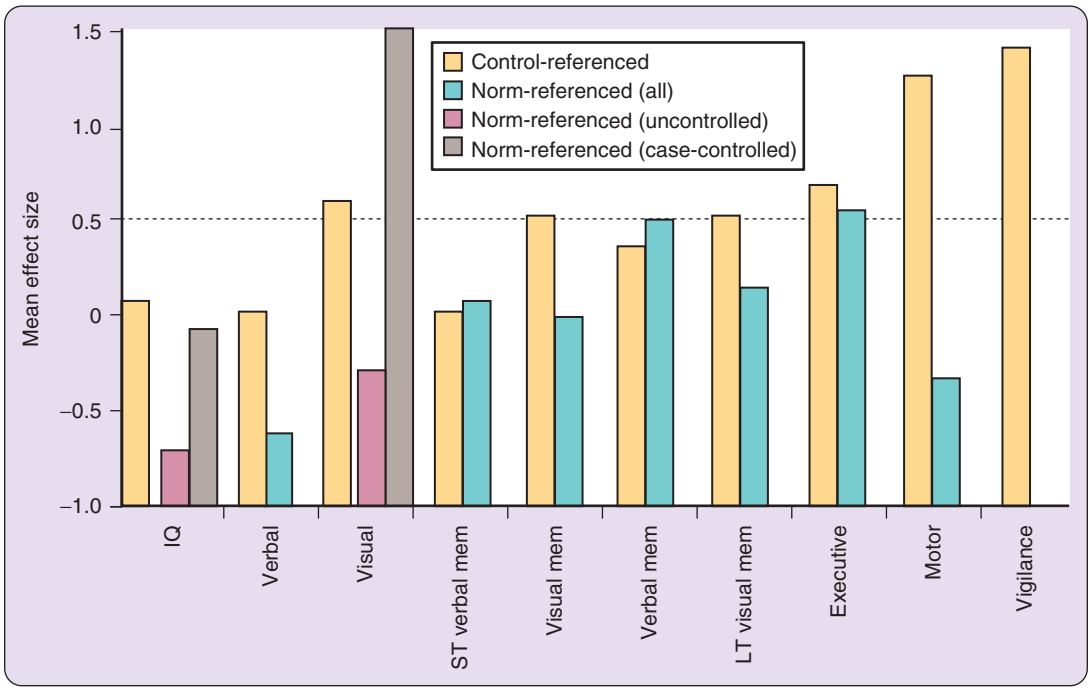


Figure 117-6 **Summary of mean effect sizes across domains and data sets.** Positive values indicate deficits relative to healthy adults, and negative values indicate strengths relative to healthy adults. The data set for moderate intelligence and visual functioning is split into case-controlled and uncontrolled samples for domains where study design (case-controlled versus uncontrolled) moderated the data. ST Mem, Short-term memory; LT Mem, long-term memory. (Modified from Beebe DW, Groesz L, Wells C, et al. The neuropsychological effects of obstructive sleep apnea: a meta-analysis of norm-referenced and case-controlled data. *Sleep* 2003;26:298–307, p. 302.)

their different subcapacities and their vulnerabilities to a range of risks and protective factors specific to OSA can provide a more realistic assessment of an individual's disability.[13,16] Similarly, a number of impairments may be secondary to other symptoms of OSA, such as sleepiness itself, or they can be a sign of psychological distress.[113,114]

To date, subjective cognitive complaints have been largely ignored in randomized controlled trials of treatments for OSA patients. However, given that subjective cognitive complaints are linked to the quality of life, work productivity, and health care utilization of patients, it is important that future studies account for these.[113]

EFFECT OF OBSTRUCTIVE SLEEP APNEA TREATMENT ON ASSOCIATED NEUROCOGNITIVE DEFICITS AND DISORDERS

Nonpharmacologic and pharmacologic treatments for OSA have been shown to improve cognitive outcomes in OSA patient subpopulations, as described in Chapter 114. The results of several meta-analyses suggest that CPAP treatment reduces sleepiness complaints and mood problems and that it improves objective cognitive functioning in OSA patients.[14,15,113,115,116] However, many questions regarding treatment with CPAP, the most pivotal of which are to whom, when, and for how long should CPAP treatment be administered, remain to be clarified.[13] The optimal treatment protocols, likely in combination with other lifestyle or pharmacologic approaches, may only be achieved once the full spectrum of the neuropathology of OSA and its dynamic fingerprinting are understood.[9,10,13] For example, it has been shown that in some treatment-compliant patients, the beneficial effect of CPAP on symptoms of sleepiness and sleep quality can be obtained after only few days of treatment. On the other hand, the effects on other subjective and objective cognitive symptoms are less well defined, and in to provide similar therapeutic effects, much longer duration of treatment may be required.[110,117] Two recent studies suggest that prolonged treatment might in fact be required in patients with severe OSA.[118,119] In one of these, an almost complete recovery of white matter tract pathology in patients with severe OSA was demonstrated in association with significant improvement in memory, attention, and executive functioning, only when 1 year of CPAP adherence was achieved.[118] The functional neuroanatomy of OSA has been highlighted in a study that documented that 3 months of treatment with CPAP improved cognitive function in several domains that corresponded to gray matter volume increases in frontal and hippocampal regions.[84] Most studies investigating treatments for OSA, however, fail to account for incomplete reversal of tissue damage or deficits in cognition, suggesting that early initiation of a prolonged treatment regimen might be necessary to optimize improvements in the neurocognitive disease process associated with OSA.[120-122]

The need for a longer duration of treatment with CPAP in elderly patients compared with younger patients has been suggested by the findings of a small pilot study that found that treatment of severe OSA in Alzheimer disease patients of mild to moderate severity was associated with significantly slower cognitive decline over 3 years.[119] Further, 1 year of CPAP treatment has been shown to improve sleepiness and quality of life in older people with OSA.[19]

Although less striking, limited evidence with drugs such as donepezil, physostigmine, and fluticasone also points to better cognitive outcomes in treated patients, likely necessitating longer treatments.[123,124] Pharmacologic treatment may also be required in patients with OSA who, despite adequate CPAP use, continue to complain of residual sleepiness.[125] It has been suggested that, among the most common explanations for persistent sleepiness in OSA patients, low CPAP compliance, inadequate CPAP titration leading to residual respiratory events and sleep fragmentation, mask or mouth leaks, treatment emergent central sleep apnea, behaviorally induced insufficient sleep syndrome, comorbid psychiatric disorders, sedative medication use, and undiagnosed coexisting sleep disorders predominate.[126] However, it has been recognized that some compliant CPAP users can still experience excessive daytime somnolence even after sleep hygiene improvement, optimization of CPAP treatment, and comorbid disorders management, and those patients are then considered as suffering from true residual sleepiness.[126] Although its pathophysiologic mechanisms remain unclear based on retrospective studies, the prevalence of such residual sleepiness can be estimated at approximately 10%.[126]

In clinical cases in which sleepiness is deemed severe enough to require treatment with an alerting drug, an objective evaluation at baseline should be done. This will allow for proper assessment of vigilance on treatment.[126] Of alerting drugs commonly used in other sleep disorders, modafinil and armodafinil have been shown to have some effects on CPAP-resistant sleepiness[126] (see also Chapter 43). A recent meta-analysis of the effect of modafinil and armodafinil in patients with residual sleepiness suggested improved objective and subjective measures of sleepiness, wakefulness, and patients' perception of disease severity, with overall good tolerance and minimal side effects.[125] Moreover, a trend toward decreased CPAP after treatment with these agents was also observed.[125] Methylphenidate,[126] dexamphetamine, venlafaxine, and atomoxetine are yet to be tested specifically for this indication. Clinical trials of residual sleepiness treatment with histamine-3 receptor agonists are underway and may provide a useful alternative in countries where modafinil and armodafinil are not approved for the treatment of residual sleepiness.[127] Future large prospective studies are required to better define predictive baseline characteristics and possible causal mechanisms for residual sleepiness as well as to inform and guide clinicians in choosing the most appropriate pharmacologic treatments.

CLINICAL PEARLS

OSA is increasingly recognized as one of the potentially modifiable risk factors for cognitive and performance deficits and dementia in adults. During untreated apnea-hypopnea episodes, intermittent hypoxemia, reoxygenation, and hypercapnia or hypocapnia occur, along with sleep fragmentation and changes in cerebral blood flow. These may, independently and in combination, result in cognitive deficits and reduced daytime performance, with functional consequences for work and school efficiency. Clinician awareness of these impairments and their prompt treatment will reduce the burden of illness on the individual patient with OSA as well as the public health risk.

SUMMARY

Patients with OSA demonstrate variable degrees of cognitive and performance deficits. Such deficits are more easily identified in those with more severe OSA.[34] The long-term effectiveness of CPAP with regard to reducing cognitive and performance deficits in patients with mild OSA remains undetermined and needs further exploration.[34] The disruption of normal sleep physiology by OSA has been increasingly recognized as an underappreciated factor regarding such deficits, which, together with hypoxemia and other already recognized factors, may further aggravate age-related memory deficits in patients with OSA.[2,3,69,71] Clinically, this dynamic interplay underscores numerous subjective and objective cognitive and emotional complaints in some patients.[13,31,72] An understanding of the proportional effect of these factors in each individual OSA patient is a major challenge because they typically occur simultaneously and, in all likelihood, target similar neurocircuitry.[13] Persistent deficits, even after prolonged treatment with CPAP in some patients, suggest that early detection of the CNS sequelae in OSA is vital so that appropriate treatment can be administered before irreversible atrophic and metabolic changes occur. However, the optimal timing and duration of treatment and the optimal treatment population are still unclear and must be addressed in future prospective randomized controlled trials.[13] Studies discussed in this chapter strongly suggest that tapping into the therapeutic potential of ischemic preconditioning, while working on ameliorating the acute and chronic effects of neuroinflammation, may offer legitimate therapeutic targets in OSA.[2,11,78] Similarly, although they are in their infancy, studies of clinical approaches that target the sleep disturbance factors of this intricate equation advocate a significant future treatment intervention potential.[13]

Despite the need for more evidence regarding cognitive and performance deficits in community-acquired samples in sham CPAP–controlled studies, there is significant documentation that untreated and CPAP nonadherent OSA patients are at risk for traffic and occupational accidents.[34] Recent findings also raise valid questions about the mechanics of associations between OSA and dementia and further highlight the public health importance of detecting and targeting patients with OSA at highest risk for cognitive decline.

ACKNOWLEDGMENTS

Supported by the Wellcome Trust [103952/Z/14/Z] and the NIHR Respiratory Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust, Imperial College, London.

Selected Readings

Dalmases M, Solé-Padullés C, Torres M, et al. Effect of CPAP on cognition, brain function and structure among elderly patients with obstructive sleep apnea: a randomized pilot study. *Chest* 2015;**148**(5):1214–23.

Ferini-Strambi L, Marelli S, Galbiati A, et al. Effects of continuous positive airway pressure on cognition and neuroimaging data in sleep apnea. *Int J Psychophysiol* 2013;**89**(2):203–12.

Gagnon K, Baril AA, Gagnon JF, et al. Cognitive impairment in obstructive sleep apnea. *Pathol Biol (Paris)* 2014;**62**(5):233–40.

Harper RM, Kumar R, Ogren JA, et al. Sleep-disordered breathing: effects on brain structure and function. *Respir Physiol Neurobiol* 2013;**188**(3):383–91.

Kheirandish-Gozal L, Yoder K, Kulkarni R, et al. Preliminary functional MRI neural correlates of executive functioning and empathy in children with obstructive sleep apnea. *Sleep* 2014;**37**(3):587–92.
Kilpinen R, Saunamäki T, Jehkonen M. Information processing speed in obstructive sleep apnea syndrome: a review. *Acta Neurol Scand* 2014;**129**(4):209–18.
Lal C, Siddiqi N, Kumbhare S, Strange C. Impact of medications on cognitive function in obstructive sleep apnea syndrome. *Sleep Breath* 2015;**19**(3):939–45.
Lim DC, Pack AI. Obstructive sleep apnea and cognitive impairment: addressing the blood-brain barrier. *Sleep Med Rev* 2014;**18**(1):35–48.
Olaithe M, Bucks RS. Executive dysfunction in OSA before and after treatment: a meta-analysis. *Sleep* 2013;**36**(9):1297–305.
Zhu Y, Fenik P, Zhan G, et al. Degeneration in arousal neurons in chronic sleep disruption modeling sleep apnea. *Front Neurol* 2015;**6**:109.

A complete reference list can be found online at ExpertConsult.com.

Obstructive Sleep Apnea and Metabolic Disorders

Mary Sau-Man Ip

Chapter 118

Chapter Highlights

- With the common risk factor of obesity, obstructive sleep apnea (OSA) and metabolic disorders often coexist. Obesity itself is considered a metabolic disease. With global escalation of obesity trends, the current and future health care burdens of these conditions are of immense concern.
- OSA produces intermittent hypoxia and sleep disruption, with evidence for downstream cascades of sympathetic activation, oxidative stress, and inflammation—pathways that align with the pathogenetic mechanisms in metabolic disorders.
- Growing epidemiologic and clinical evidence suggests that OSA may modulate metabolic outcomes. The confounding effects of obesity on metabolic disorders have, however, been difficult to dissect. Of greater clinical relevance may be potential synergistic effects between OSA and obesity, mediated partly through exacerbation of adipose tissue dysfunction.
- Animal and cell-based studies, mostly using intermittent hypoxia regimens as a surrogate model of OSA in humans, have provided evidence for deleterious effects on various tissues and cells in the pathogenesis of metabolic dysfunction and have elucidated relevant molecular pathways.
- Despite suggestive data from human observational studies, no definitive evidence has yet emerged to indicate that controlling OSA would result in improvement in metabolic function of significant clinical impact. Future studies need to address the challenges of small heterogeneous samples, diverse methodology for metabolic evaluation, and issues regarding withholding treatment for symptomatic OSA for substantial periods in longitudinal cohort follow-up or in randomized controlled studies.

Sleep modulates body metabolism, and sleep restriction or disturbance can have negative metabolic effects. Although each metabolic disorder has specific pathogenetic pathways, all share common grounds of engagement of hormones, oxidative stress, and inflammation, with obesity as a prevalent phenotypic feature.[1,2] Recurrent obstructed breathing events in obstructive sleep apnea (OSA) characteristically result in repeated cycles of hypoxia-reoxygenation with consequent disruption of sleep architecture, which may trigger downstream cascades that align with the mediating mechanisms for cardiometabolic dysfunction.[3,4] Hence, beyond the common link of obesity in the close partnership between OSA and metabolic disorders, great interest has focused on the potential role of OSA in the causation or aggravation of metabolic dysfunction per se, or in concert with other factors.

The metabolic network is one of intricate cross-talk among various organs, tissues, and cells and their respective signaling pathways. Thus the relationship between OSA and metabolic disorders is unlikely to comprise a set of discrete, one-to-one unidirectional connections but rather can be depicted as a complex network with interplay of various organs and tissues, along with multiple positive or negative feedback mechanisms.[1,2,5] Furthermore, metabolic function is subject to genetic as well as behavioral influences such as dietary intake and physical exercise, and these factors contribute to individual metabolic outcomes in persons with and without OSA.

OBSTRUCTIVE SLEEP APNEA AND METABOLIC DYSREGULATION: PATHOGENESIS AND MECHANISMS

Obesity and Adiposity

With advances in the understanding of adipose tissue biology, fat tissue, which traditionally has been viewed as a storage depot of energy, is now known as an active system with autocrine, paracrine, and endocrine functions, propagating signals to entrain metabolic cooperation of other organs and tissues.[2,5] Under conditions of positive energy balance, fat accumulates, and its distribution is crucial to health outcomes. Visceral fat in particular becomes dysfunctional, with altered nonesterified free fatty acid metabolism which may contribute to hepatic insulin resistance, dyslipidemia, and altered release of adipocytokines which are mediators of dysmetabolism.[2,6] The expansion of adipose tissue with hypertrophied adipocytes may cause cellular hypoxia, and oxidative stress and inflammation, initiating adipose tissue dysfunction.[6] Obesity is thus characterized by a state of chronic low-grade systemic and adipose tissue inflammation, fueling cardiometabolic dysfunction.[5,6,7]

Adiposity holds a unique role in the consideration of links between OSA and metabolic disorders. It is well established that obesity is a major risk factor for various metabolic disorders and a key component of the metabolic syndrome.[2,7]

Furthermore, obesity is now considered a metabolic disease in itself, with many systemic manifestations.[8] Obesity carries major public health impact, and its prevalence is escalating globally.[9] Obesity is the most common risk factor for OSA, although BMI accounts for only a small part of the variability of OSA severity as reflected in the apnea-hypopnea index (AHI), implying a multifactorial nature of OSA pathogenesis.[10] OSA and its severity have been reported to be associated with central obesity (involving the neck, trunk, and abdomen) and abdominal visceral fat, more so than with BMI, particularly in men.[10-12] Apart from affecting breathing mechanics predisposing to upper airway collapse, abdominal fat is a source of the adipokine leptin, which may modulate ventilatory control and upper airway function (see Chapter 120). In parallel, obesity also is the predominant risk factor for metabolic dysfunction. In clinical practice, abdominal obesity, as measured by waist circumference, is recommended as a useful screening tool for metabolic disorders.[13] The accumulation of visceral fat may result in lipid overflow, with further ectopic fat deposition in sites such as skeletal muscle and liver, and promote insulin resistance, whereas in the pancreas, lipid excess may impair insulin secretion.[2]

In OSA, neck fat deposition is considered to be of importance in upper airway dimensions and function, promoting structural narrowing and functional collapse.[10] Neck circumference, an established predictor of OSA, also has been shown to be a novel measure of cardiometabolic risk in the Framingham data.[14] Open to speculation, however, is whether neck circumference was a surrogate marker for OSA in those data in terms of cardiometabolic risk.

Inasmuch as obesity is a common and strong risk factor for both OSA and metabolic disorders, it is not surprising that patients with OSA not uncommonly have metabolic comorbid diseases. The ongoing enigma is whether and to what extent OSA per se is involved in the causation and/or aggravation of various metabolic disorders, including obesity. It is hypothesized that OSA exerts systemic effects on different end organs and tissues, with adipose tissue as one of the targets. With its strategic position in the metabolic network, dysfunctional adipose tissue is likely to play an important role in further metabolic dysfunction in OSA.[15] Additional evidence points to a more adverse metabolic profile even in lean subjects with OSA, compared with those without OSA.[16-18] It is not known, however, if OSA may act as a factor to convert presumably metabolically healthy adipose tissue in nonobese subjects to metabolically unhealthy tissue—a mechanism that has been proposed in metabolic dysgenesis.

It has been speculated that patients with OSA are inherently predisposed to weight gain and may experience difficulty in losing weight compared with subjects without OSA,[19] possibly relating to selective differences in body metabolism regulated by hormones such as insulin, leptin, or ghrelin. Repeated apneic spells and arousals during sleep are associated with excessive daytime sleepiness, which may reduce motivation to engage in physical activity and thereby predispose affected persons to weight gain over time. Furthermore, OSA may promote abdominal obesity through increasing insulin resistance, and/or disturbing sleep quality and quantity. The presence of sleep apnea in men with central obesity was found to attenuate metabolic improvement in response to a lifestyle intervention program, compared with men without OSA.[20]

Insulin Resistance and Glucose Metabolism

The body maintains glucose homeostasis mainly through the action of insulin on various tissues. In type 1 diabetes mellitus (DM), pancreatic beta cell failure of insulin secretion is the key defect, whereas in type 2 DM, which accounts for more than 90% of cases of DM globally, insulin resistance in muscle and liver is the primary pathophysiologic defect. Insulin is secreted in response to an increase in glucose concentration and reduces glucose levels by suppressing hepatic gluconeogenesis and promoting glucose uptake in skeletal muscle and fat. With increasing insulin resistance, pancreatic beta cells respond with a compensatory increase in insulin secretion such that glucose homeostasis and a constant blood glucose level can be maintained. When this compensatory mechanism is deficient or overwhelmed, impaired glucose tolerance and overt DM ensue.[21] Although the origins of insulin resistance can be traced to genetic background, the epidemic of DM is related to the parallel epidemic of obesity and physical inactivity. Insulin resistance is closely but not exclusively linked to visceral obesity, and worsening insulin resistance may further stimulate fat accumulation and encourage ectopic fat deposition.[2]

Insulin also regulates glycogenesis, lipogenesis, and protein synthesis, and insulin resistance may occur in many cells and tissues, within which there may be selective hormone resistance for different pathways.[22] In concert with other mechanisms, insulin resistance predisposes affected patients to endothelial dysfunction, which underlies many cardiometabolic diseases.[23]

In keeping with the common factor of obesity, it is not surprising that OSA is strongly associated with the spectrum of insulin-glucose dysmetabolism. OSA may contribute independently to insulin resistance and glucose dysmetabolism through its pathophysiologic profile of intermittent hypoxia, sympathetic activation, oxidative stress, and inflammation.[3] Such data, are, however, subject to accuracy and variability of research methodology. A variety of methods for the evaluation of glucose metabolism have been deployed in sleep research[24,25] (Table 118-1). To accurately measure insulin sensitivity, it is necessary to use a method that observes in some fashion the metabolic effect of insulin given intravenously. Other methods are simpler to perform, but the results are affected by any degree of beta cell failure.[26,27] Furthermore, findings from epidemiologic or clinical studies are subject to the adequate control of the confounding factors. that are not easy to accurately identify or quantify.

Lipid Metabolism and Dyslipidemia

Lipids, including cholesterol, triglycerides, and others, are transported in the body as lipoprotein complexes in body fluids (plasma, interstitial fluid, and lymph) passing into and out of tissues, and metabolized through exogenous and endogenous pathways. The exogenous pathway operates for dietary lipids absorbed through the gastrointestinal tract, transporting them to the liver and other peripheral tissues, especially fat and muscles. The endogenous pathway refers to hepatic secretion and metabolism of lipoproteins, and their transport to peripheral tissues. As noted earlier, under conditions of positive energy balance, excess fat accumulates in adipose tissue and other organs as ectopic fat.[2,15] Lipid metabolism is regulated by both genetic and nongenetic factors. Secondary changes in plasma levels of lipids occur in a variety of diseases

Table 118-1 Assessment Tools for Glucose Metabolism in Clinical Practice and Research

Test	Brief Methodology	Parameter(s) Measured	Comments
Blood glucose	Fasting venous blood sample for plasma glucose level	Fasting glucose level	Conventional test for diagnosis of DM/impaired fasting glucose
Hemoglobin A_{1c} (HbA_{1c})	Spot venous blood sample for glycated hemoglobin level	Glycemic status over past 2–3 months	Used in clinical practice to assess glycemic control in past 2–3 months in DM HbA_{1c} ≥6.5% is used for diagnosis of DM (ADA/WHO); HbA_{1c} of 5.7%–6.4% is used for diagnosis of prediabetes (ADA) Higher levels predict worse diabetic complications
Oral glucose tolerance test (OGTT)	Oral glucose loading (75 g) followed by evaluation of 2-hour post–blood glucose loading	Impaired glucose tolerance (IGT)	2-hour glucose ≥11.1 mmol/L for diagnosis of DM 7.8–11 mmol/L for diagnosis of IGT
	Oral glucose loading followed by evaluation of glucose every 30 minutes; simultaneous insulin levels measured	Insulin sensitivity	May be reflecting insulin secretion in response to glucose loading rather than insulin sensitivity Poor test reproducibility due to variability of gastrointestinal absorption and other factors
Hyperinsulinemic euglycemic clamp	A dose-response curve for data on exogenous insulin is generated by measuring the variable infusion rate of glucose required to maintain euglycemia	Insulin sensitivity	Gold standard for assessing insulin sensitivity The steady-state rate of peripheral glucose utilization (M value) is measured as milligrams of glucose used per kilogram of body weight per minute Labor-intensive investigation
Homeostasis model assessment (HOMA)	Fasting venous blood sample with glucose and insulin measurements		First derived from epidemiologic studies
	HOMA-IR: insulin (μU/mL) × glucose (mmol/L)/22.5	Insulin resistance: HOMA-IR	Measures basal insulin resistance and insulin secretion
	HOMA-β: [20 × insulin (μU/mL)]/[glucose (mmol/L) − 3.5]	Insulin secretion: HOMA-β	Reflects mainly hepatic insulin resistance
Frequently sampled intravenous glucose tolerance test (FSIGT, FSIVGTT)	Fasting baseline blood glucose (and insulin), followed by frequent sampling after glucose injection (for insulin sensitivity, insulin is injected 20 minutes later) for 3 hours. A computer model describing plasma dynamics (minimal model) is applied for deriving metabolic parameters	Assesses both pancreatic beta cell secretory capacity and peripheral glucose uptake in response to the bolus IV glucose Additional information on insulin sensitivity is gained by administration of insulin 20 minutes after the glucose load	Validated for insulin sensitivity against hyperglycemic euglycemic clamp No need for on-line measurements or external control of infusion Reflects whole-body insulin sensitivity
Short insulin tolerance test (SITT)	Administration of exogenous insulin followed by monitoring of fall in blood glucose over the next 30 minutes, to derive the glucose disappearance rate	Insulin sensitivity	Validated for insulin sensitivity against hyperglycemic euglycemic clamp No need for on-line measurements or external control of infusion

ADA/WHO, American Diabetes Association/World Health Organization; DM, diabetes mellitus; IV, intravenous.
From Lam DC, Lam KS, Ip MS. Obstructive sleep apnoea, insulin resistance and adipocytokines. *Clinic Endocrinol (Oxf)* 2015;82(2):165–77.

and clinical conditions, including obesity, smoking, insulin resistance, DM, and liver disorders. Obesity is frequently, although not invariably, accompanied by hyperlipidemia. An increase in fat mass is associated with increased release of free fatty acids to the liver, where they are reesterified in hepatocytes to form triglycerides, which are then packaged for secretion back into the circulation.[2] Obesity is a predisposing factor for development of fatty liver, in which disruption of hepatic biosynthesis of lipids occurs.[15]

Dyslipidemia is a major risk factor for atherosclerosis. In the clinical setting, lipids usually are classified as total cholesterol, high-density lipoprotein cholesterol (HDL cholesterol) and low-density lipoprotein cholesterol (LDL cholesterol), and triglycerides. The levels of these parameters have different implications regarding vasculopathy and cardiovascular disease (CVD). Multiple epidemiologic studies have demonstrated a strong relationship between serum cholesterol levels and coronary heart disease, and randomized controlled trials (RCTs) have unequivocally demonstrated that lowering cholesterol reduces clinical events due to atherosclerosis. LDL cholesterol is considered deleterious, whereas HDL cholesterol is believed to be protective, although in fact little evidence exists regarding the benefit of raising HDL cholesterol.[28] Elevated fasting triglyceride levels have not been correlated with significant CVD risk, but it has been suggested that postprandial hypertriglyceridemia may be a bigger CVD risk factor than fasting triglyceride levels.[29]

In the context of these biologic pathways and clinical outcomes, it is clear that the pathophysiology of OSA holds biologic plausibility regarding causation and/or promotion of dyslipidemia through modulation of fat or liver metabolism.

Liver Injury and Related Metabolic Dysregulation

As discussed previously, the liver plays a pivotal role in the regulation of both lipid and glucose metabolism. Obesity is known to result in liver injury as nonalcoholic fatty liver disease (NAFLD), which is another recently proposed addition to the list of disorders for inclusion in the metabolic syndrome.[30] Obesity causes intracellular accumulation of lipids in the liver, designated *hepatic steatosis* in light of the associated histopathologic changes observed. NAFLD ranges in severity from hepatic steatosis (presence of fat in more than 5% of hepatocytes), to steatohepatitis (i.e., nonalcoholic steatohepatitis [NASH]), to liver fibrosis and cirrhosis, and it may be a risk factor for hepatocellular cancer.

Obesity and insulin resistance are major risk factors for NAFLD; conversely, fat accumulation in the liver may cause hepatic insulin resistance with enhanced hepatic glucose production.[30] The reported prevalence of NAFLD varies, ranging between 30% and 100% in obesity and between 10% and 75% in type 2 DM.[31]

Obesity alone, however, does not appear to account for all cases of NAFLD. Although weight loss can significantly improve NASH histologic activity scores, the potential triggers for progression of NAFLD are not fully understood. OSA, possibly through the pathomechanism of intermittent hypoxia may promote the progression of simple steatosis to the more severe forms of NAFLD.[31]

Neurohumoral Activation

The hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic nervous system play important roles in energy balance, body metabolism, and the pathogenesis of obesity.[32] Other than exerting a prominent regulatory function on blood pressure, cortisol is an anabolic hormone that promotes insulin resistance and dyslipidemia, whereas catecholamines upregulate hormone-sensitive lipase and increase circulating free fatty acid levels, induce beta cell apoptosis, and adversely affect adipokine profile.

Recurrent asphyxia from obstructed breathing, as occurs in OSA, poses a potent stress to the body, resulting in the so-called "fight or flight" phenomenon. Stress-related neurohumoral activation is a potential mechanistic pathway for metabolic dysregulation in OSA. Animal models of chronic intermittent hypoxia (CIH) demonstrate sympathetic activation contributing to CIH-induced hypertension.[33] Experimental studies in healthy subjects subjected to short-term sleep deprivation, sleep fragmentation, or intermittent hypoxia have demonstrated alterations in hormonal profiles in association with altered glucose metabolism.[34] Sleep fragmentation induced by acoustic stimuli without hypoxia in healthy volunteers resulted in alterations in insulin-glucose metabolism, which were accompanied by alterations in daytime heart rate variability as a marker of sympathetic activation.[35,36] CIH may potentially sensitize the carotid body, contributing to increase in sympathetic nerve activity and increase in blood pressure.[37]

Studies of patients with OSA consistently demonstrate increased muscle sympathetic nerve activity even in the awake state, as well as increased output of urinary catecholamines, although the extent to which obesity itself contributes to such sympathetic activation remains unclear.[38,39] Robust data show that CPAP treatment of OSA can rapidly reduce sympathetic activity, indicating that OSA itself induces sympathetic activation.[40] In a study of subjects undergoing sleep study to investigate suspected OSA, the independent determinants of serum adiponectin levels included insulin resistance and urinary catecholamine levels but not the presence or severity of OSA, suggesting a complex relationship among these pathophysiologic parameters.[41]

The associations between OSA and other hormones from the HPA axis, such as cortisol, are not clear. In a study of obese subjects, despite a lack of difference in baseline cortisol levels between those with and those without OSA, the subjects in the OSA group showed a reduction in heart rate and greater cortisol suppression with dexamethasone after 3 months of CPAP treatment, suggesting that untreated OSA may lead to abnormally high activation of the sympathetic nervous system and HPA axis.[42] Overall, the literature on cortisol status and associated dysmetabolism in OSA has not been abundant or consistent.[43]

Intermittent Hypoxia

Recurrent apneas and hypopneas in OSA generate intermittent hypoxia, which appears to hold a pivotal position in the pathogenesis of metabolic dysfunction in OSA. The chronic if intermittent oxygenation deficit in OSA results in systemic tissue and cellular hypoxia, with a plethora of downstream effects. Mounting evidence suggests that intermittent hypoxia simulates ischemia-reperfusion and can activate oxidative stress and inflammation, which are key pathogenetic pathways in cardiometabolic dysfunction.[1,4]

Intermittent hypoxic exposure in animals in vivo or in cell cultures in vitro allows the controlled interrogation of cellular and molecular mechanisms that may occur in various tissues,

organs, and cells under different conditions. Exposure to intermittent hypoxia for 6 to 8 hours in the 24-hour time clock over days or weeks, termed *chronic intermittent hypoxia* (CIH), as noted earlier in association with OSA in humans, often is used in experimental settings. A wide array of intermittent hypoxia regimens is used in different laboratories, which may partly explain the sometimes discrepant results. Different tissues and organs in murine models demonstrate individualized oxygenation profiles[44] and oxidative stress responses[45,46] to intermittent hypoxia challenge. For example, intermittent hypoxic exposure has been shown to cause oxygen partial pressure swings in the liver, whereas such fluctuations were found to be attenuated in muscle and markedly so in fat, which instead showed steady hypoxia.[44] Compared with lean mice, obese mice had lower baseline liver oxygen tension but similar fat and muscle tissue oxygen tensions, whereas both obese and lean mice exhibited similar tissue partial pressure changes with intermittent hypoxic exposure.[44] Another caveat regarding this mechanism as a pathogenetic trigger is the likely presence of adaptive mechanisms, and the balance or imbalance of these various otherwise poorly delineated factors probably determines eventual health outcomes.[47,48]

Intermittent hypoxia–induced upregulation of nuclear factor kappa B (NF-κB), the master transcriptional switch of inflammation, has been demonstrated in a variety of cells and tissues—leukocytes, vascular cells, fat cells, cardiovascular tissue, and liver tissue, with increased production of inflammatory gene products downstream of activation of NF-κB, such as tumor necrosis factor alpha (TNF-α), interleukin 6 (IL-6), and C-reactive protein (CRP).[44,45,49-53] Data have been controversial regarding intermittent hypoxia induction of hypoxia-inducible factor-1 (HIF-1), a transcription factor critical to physiologic responses to hypoxia, including erythropoiesis, angiogeneis, and glucose metabolism.[48] Severe, but not moderate, intermittent hypoxia has been found to elicit HIF-1 activation in PC12 pheochromocytoma cells,[54] whereas moderate intermittent hypoxia has been found to elicit preferential activation of NF-κB, but not HIF-1, in HeLa cells.[49]

In a lean mouse model of CIH (at 12 weeks), liver histologic examination showed marked accumulation of glycogen in hepatocytes, with evidence of increased hepatic levels of oxidative stress biomarkers and activation of NF-κB, and these changes were followed by sensitization to acetaminophen-induced liver toxicity.[52] Exposure to CIH in a mouse model of high-fat, high-cholesterol diet–induced obesity, compared with similar dietary manipulation alone, caused liver oxidative stress and hepatic inflammation in addition to hepatic steatosis.[53]

Taken collectively, animal and cell data indicate that CIH may act independently, or in concert with obesity, alcohol, or drugs, to promote oxidative stress, inflammation, and dysmetabolism. In rigorously controlled experimental settings, human volunteers have been exposed to intermittent hypoxic regimens to investigate specific metabolic responses. In some studies involving patients with OSA, dysmetabolism arising from OSA has been inferred to be due specifically to sleep hypoxemia, when the metabolic parameter studied showed significant correlations with deoxygenation parameters such as oxygen nadir, oxygen desaturation index, or duration of oxygen desaturation.

In a mouse model of intermittent hypoxia, insulin resistance increased in lean and genetically obese mice, as well as those with dietary obesity, and was dependent on the disruption of the leptin pathways.[55,56] With exposure to intermittent hypoxic regimens of increasing severity, progressively elevated insulin resistance and leptin levels were found in lean mice, whereas increase in leptin levels plateaued in obese mice, suggesting that adiposity may have overwhelmed the effect of the hypoxia on adipokine production.[44] Insulin resistance persisted in intermittent hypoxia–exposed mice that were subjected to pharmacologic denervation of the sympathetic and parasympathetic nervous systems,[56] which did not support the hypothesized mechanistic role of sympathetic activation on glucose dysmetabolism in OSA. Lean rats exposed to different regimens of CIH demonstrated "dose-related" increases in serum insulin, probably reflecting augmented beta cell secretion in response to increased insulin resistance, alongside a commensurate circulating adipocytokine profile with increases in leptin, IL-6, and TNF-α and a decrease in adiponectin.[50] Lean or obese mice exposed to CIH demonstrated evidence of oxidative stress in liver tissue, with upregulation of inflammatory markers; this line of evidence has given rise to the speculation that intermittent hypoxia may contribute to hepatic insulin resistance through the pathophysiologic changes of NASH.[52,53] However, one study of CIH exposure for 4 weeks demonstrated exacerbation of insulin resistance and induction of steatohepatitis in mice with diet-induced obesity, but not in lean mice.[57]

Intermittent hypoxia–induced alterations in adipose tissue and cell metabolism pave the way for adverse downstream effects on insulin resistance and glucose metabolism. In vitro studies of 3T3-L1 adipocytes have demonstrated hypoxic exposure dose–dependent upregulation of proinflammatory activities, represented by the profile of NF-κB, HIF-1, glucose transport factor-1, TNF, IL-6, leptin, and adiponectin.[50] Synthetic sympathomimetics also could suppress adiponectin gene expression in preadipocyte cell lines independently of the intermittent hypoxia mechanism.[58]

Intermittent hypoxia may modulate pancreatic beta cell function. Lean mice exposed to CIH and glucose infusion demonstrated pancreatic beta cell replication with increased insulin secretion,[59] whereas another study of CIH reported associated beta cell proliferation and enhanced cell death, which was mediated by oxidative stress.[60] Other than induction of insulin resistance and hepatocyte glucose output, intermittent hypoxic exposure for 14 days in lean mice increased oxidative stress in the pancreas and impaired beta cell function, and cessation of hypoxic exposure could not fully reverse observed changes in glucose metabolism.[61]

Intermittent hypoxia also has been shown to be a key factor in the upregulation of genes responsible for hepatic lipid biosynthesis, promotion of oxidation of serum lipids, and modulation of the neurohormonal axes that influence signaling pathways in lipid transport and synthesis.[62] Exposure to CIH has led to increase in circulating levels of triglycerides, total cholesterol, and LDL and VLDL cholesterol in a hypoxic dose-dependent manner in both lean and obese murine models.[63] CIH exposure in a mouse model of dietary obesity exacerbated diet-induced dyslipidemia, with evidence of atherosclerosis in the aorta.[64] At a molecular level, insulin's effects on lipogenesis were mediated predominantly by the transcription factor sterol regulatory element–binding protein (SREBP)-1c, which controls the expression of genes required for cholesterol, fatty acid, triglyceride, and phosopholipid

synthesis.[65] Mechanistic studies focusing on the liver have further demonstrated that intermittent hypoxia induces increased lipolysis with free fatty acid flux into the liver, with sequential upregulation of the master transcriptional factors HIF-1 and NF-κB, SREBP-1, and stearoyl coenzyme A desaturase-1 (SDC-1).[64,65] The outcomes are hepatic steatosis, aggravation of hepatic insulin resistance, and disruption of hepatic lipoprotein biosynthesis and secretion. A murine model of severe CIH was shown to result in poor clearance of triglyceride-rich lipoproteins, attributed to reduced lipoprotein lipase activity, which was decreased by 80% in adipose tissue.[66] These results resonate with the finding of postprandial hyperlipemia in subjects with OSA, which could be decreased with CPAP treatment.[67]

In human studies, intermittent hypoxia simulating that seen in OSA has been induced by altering inspired oxygen concentrations. Young healthy men exposed to 8 hours of intermittent hypoxia (through inhalation of "air" with 5% oxygen alternating with 21% oxygen) demonstrated impairment of insulin sensitivity, glucose disposal, and pancreatic islet cell function.[68]

Derangement of Sleep Quality and Quantity

The impact of sleep quality and sleep derangement on metabolism is discussed in Chapter 20. Recurrent obstructed breathing gives rise to cerebral arousals and disturbs sleep architecture, posing another trigger for metabolic dysregulation in OSA. Several epidemiologic studies have found a curvilinear relationship between sleep duration and obesity or glucose dysmetabolism, and subjects who reported shorter sleep durations, compared with those averaging more than 7 to 8 hours of sleep per night, exhibited greater degrees of obesity and glucose dysmetabolism.[69] Alterations in energy-regulating hormones including leptin and grhelin have been implicated.[70] Healthy subjects who were subjected to experimental sleep restriction demonstrated abnormal glucose metabolism[71] and impairment of insulin signaling in their subcutaneous adipocytes,[72] compared with after normal sleep. Studies using acoustic stimuli to induce sleep fragmentation,[36] with reduction of slow wave and REM sleep and preserved total sleep time,[35] showed that sleep disruption without hypoxia could lead to a decrease in insulin sensitivity, as well as impaired non–insulin-dependent glucose disposal and inadequate compensatory increase in insulin secretion.

In a study of 226 children, both obesity and OSA, and to a greater extent the combination of the two, were associated with reduced circulating levels of G protein–coupled receptor 120 (GPR120), a long-chain free fatty acid receptor protective against insulin resistance and systemic inflammation.[73] GPR120 levels correlated with insulin resistance, but not with dyslipidemia or CRP levels. Among the sleep parameters, GPR120 levels showed the strongest independent association with respiratory arousal index, which provoked the query of whether sleep curtailment or disruption, rather than the other pathophysiologic disturbance of OSA, was the etiologic mechanism.

Oxidative Stress

Oxidative stress is a state caused by imbalance between the production of reactive oxygen species (ROS) highly damaging to cells and the antioxidant activity that counteracts ROS. Oxidative stress activates redox-sensitive transcription factors that regulate inflammatory processes and downregulate nitric oxide synthase, with consequent reduced nitric oxide in endothelial cells, leading to microvascular and macrovascular endothelial dysfunction. These cellular mechanisms also may be operative in other tissues and organs and may underlie the development of insulin resistance, dyslipidemia, arterial hypertension, and other cardiometabolic derangements.[6,74] The recurrent hypoxia-reoxygenation cycles in OSA (i.e., intermittent hypoxia and particularly CIH, as described previously) are thought to be analogous to ischemia-reperfusion injury known to produce ROS in the reperfusion phase. Evidence for occurrence of increased oxidative stress in OSA is not consistent, however. A majority of studies have found that subjects with OSA have increased biomarkers of oxidative stress compared with control subjects without OSA, which decreased after CPAP treatment of OSA, whereas a few studies reported reduced antioxidant activity.[4,75,76] For example, subjects with OSA were found to have an increase in lipid peroxidation, glycated end products of oxidation, serum or urinary reactive oxygen metabolites, ROS production from leukocytes or monocytes, and reduction of circulating nitric oxide, although there have also been negative studies.[4] Discrepant findings for the presence of oxidative stress in OSA may be related to the difficulty of assessing oxidative status accurately.[77] Furthermore, obesity itself has been associated with enhanced oxidative stress systemically and locally in adipose tissues, constituting a confounding factor in the study of oxidative stress potentially due to OSA.[6,15,75]

Inflammation and Alteration of the Adipocytokine Profile

Obesity and metabolic syndrome are known as proinflammatory states, with elevated circulating levels of proinflammatory mediators, often referred to as adipocytokines (see earlier under Obesity and Adiposity).[7] Adiopokines are produced predominantly from adipose tissues, whereas cytokines are released by a variety of cells and tissues. Collectively, adipocytokines function as key mediators linking obesity and metabolic disorders. [7] It is hypothesized that OSA may modulate the expression and release of adipokines to favor adverse metabolism.[11] Intermittent hypoxia, oxidative stress, and sympathetic activity are capable of modulating release of proinflammatory adipocytokines from a variety of cells and tissues, including adipose tissues.[15,44,49-54] Investigation of alterations of adipocytokine regulation in subjects with OSA, however, have yielded highly variable results, which may be attributed to the presence of confounding factors, in particular obesity.[24]

OBSTRUCTIVE SLEEP APNEA AND METABOLIC DYSREGULATION: CLINICAL ASSOCIATIONS AND TREATMENT

Obesity

It has been estimated that a range of 15% to 90% of obese subjects have OSA, depending on age, gender, and BMI.[78] Generally, OSA prevalence is higher among persons with morbid obesity, and OSA severity is greater in obese than in leaner subjects.[10,78] Longitudinal data of the Wisconsin cohort indicated that a 10% gain in body weight increased the chance of developing moderate to severe OSA by a factor of 6, and that every 1% increase in body weight was associated with a 3% increase in AHI.[79] Conversely, weight loss reduced OSA

severity but to a less substantial degree than that seen with exacerbation with weight gain. If OSA alters energy balance to promote obesity, it follows that treatment of OSA should mitigate obesity. However, studies on the impact of treatment of OSA on body weight or adiposity have not been supportive of a beneficial effect on body composition. In an RCT of CPAP treatment for OSA, therapeutic CPAP decreased daytime sleepiness and promoted physical activity over a 3-month period, but no change in body weight was seen.[80] Studies on the impact of CPAP treatment on abdominal fat have yielded conflicting data,[15] with randomized, sham-controlled trials for 8 and 12 weeks, respectively, showing no change in abdominal fat quantified by imaging.[81,82] In fact, a recent meta-analysis of data from randomized trials suggests that CPAP treatment of OSA may promote increase in BMI and body weight.[83] On the other hand, interventions including lifestyle modifications, antiobesity medications, and bariatric surgery (also termed metabolic surgery) can produce weight loss and have undoubted metabolic benefits as well as resulting in symptomatic improvement in obese subjects with OSA, although the relative impact of various weight reduction regimens in this regard remains underinvestigated.[84] In a controlled trial of surgical versus conventional therapy in severely obese subjects with OSA, greater weight loss in the surgical group did not translate to significantly greater reduction in AHI than that achieved with conventional weight loss approaches.[85] A daunting challenge is to find the optimal weight control measure that not only is effective in improving anthropometric aspects and various health outcomes but also is acceptable to the individual patient and sustainable in real-life situations.

Insulin and Glucose Metabolism

The clinical relationship between OSA and a range of glucose dysmetabolism disorders ranging from insulin resistance to overt diabetes has been extensively investigated and regularly reviewed.[3,86,87] Several population-based studies from the United States, Hong Kong, Korea, Brazil, and Europe consistently found that OSA was associated with increased insulin resistance or evidence of impaired glycemic status including DM, despite adjustment for obesity and other confounders.[17,88-93] The Sleep Heart Health Study (SHHS) cohort of 5874 subjects in the United States did not show an independent relationship between OSA and DM,[94] although in a subgroup of overweight, middle-aged men in the SHHS, an independent association was found between OSA and insulin resistance/glucose intolerance.[89] In 2014, the European Sleep Apnea Database (ESADA) reported its cross-sectional analysis of data on 6616 participants and found that increasing OSA severity was associated with increased likelihood of type 2 DM and worse glycemic control in the diabetic subjects despite adjustment for confounding variables.[93] The Wisconsin Sleep Cohort, which found an independent association between severity of untreated OSA and prevalence of DM at baseline, did not find any increase in incident DM at a 4-year follow-up evaluation.[88] Other longitudinal follow-up cohort studies from the United States[95] and Japan[96] have found associations between baseline OSA and incident diabetes over 3 to 4 years, as did another study from Australia, although the incidence of DM was very low in that cohort.[97] A historical cohort of 8678 patients undergoing diagnostic study for OSA in a single clinical center in Toronto showed that DM developed in 1017 subjects with OSA (11.7%) over a median follow-up period of approximately 67 months.[98] With full adjustment for confounders, subjects with AHI greater than 30 had a 30% higher hazard of developing DM than those with AHI below 5. Adjusted odds ratios for the incidence of DM in moderate to severe OSA compared with those without OSA in different studies have ranged from 1.31[98] to 13.45.[87]

Conversely, a high prevalence of OSA has been reported among diabetic populations[94,99-104] (Figure 118-1). Several studies in diabetic subjects suggest that severity of OSA is associated with worse glycemic control,[93,102] although this correlation is not universally seen.[103]

A majority of studies analyzing nondiabetic subjects have reported independent associations between OSA and insulin resistance or sensitivity and/or other measures of glycemic health, with dose-dependent effect of OSA on such measures of metabolic impairment.[86] However, some studies found that such a relationship was confounded by obesity, such that the association was abolished after adjustments for BMI and/or other measures of adiposity.[86] The ESADA cohort study found that OSA severity independently predicted glycemic health assessed by HbA_{1c} in nondiabetic subjects.[92] Nonobese subjects with OSA, compared with BMI-matched or BMI-adjusted counterparts without OSA, also had more insulin resistance or glucose dysmetabolism.[16-18,105,106]

It has been proposed that excessive daytime sleepiness may be a phenotypic marker for insulin resistance in OSA. Associated abdominal or visceral obesity in OSA could contribute to sleepiness in OSA through hypercytokinemia.[107] Waist circumference and visceral fat have demonstrated high correlations with insulin resistance[13] and also with OSA prevalence[10,15] or severity.[12,15] Excessive daytime sleepiness was reported to be a useful indicator of moderate or severe OSA in white diabetic subjects,[108] but this was not the case in a Chinese diabetic population.[103] In a case-control study of nondiabetic subjects with similar BMI and AHI, those with excessive daytime sleepiness (as defined by a mean Epworth Sleepiness Scale [ESS] score of 16) had higher scores on homeostasis model assessment for insulin resistance (HOMA-IR) than those without sleepiness (mean ESS score of 4), and indices of insulin resistance were improved with CPAP treatment only in the group with baseline excessive sleepiness.[109]

Despite abundant positive data supporting an independent association of OSA and disturbance of glucose homeostasis, cross-sectional studies cannot be considered definitive for a causal link. Reported data on the effect of treatment of OSA, usually with CPAP, on insulin-glucose metabolism remain highly controversial.[86,87,110] A number of observational studies in either diabetics or non-diabetics with OSA suggested improvements in insulin resistance or glycemic status with CPAP treatment, but such results are by no means consistent, and most of the RCTs did not provide definitive evidence for an improvement in insulin-glucose metabolism in response to CPAP.[40] Several RCTs of CPAP treatment for OSA without DM ranging from 1 to 12 weeks in duration did not find consistent improvement in insulin resistance measured by HOMA-IR, or insulin sensitivity using OGTT or a hyperinsulinemic clamp.[111-113] Data from subset analysis or open continuation phase of these RCTS, and from observational studies, suggest that severity of OSA,[112] BMI,[111,114] CPAP adherence,[115] sample size, and/or treatment duration[113] may contribute to the determination of metabolic effects. Reported

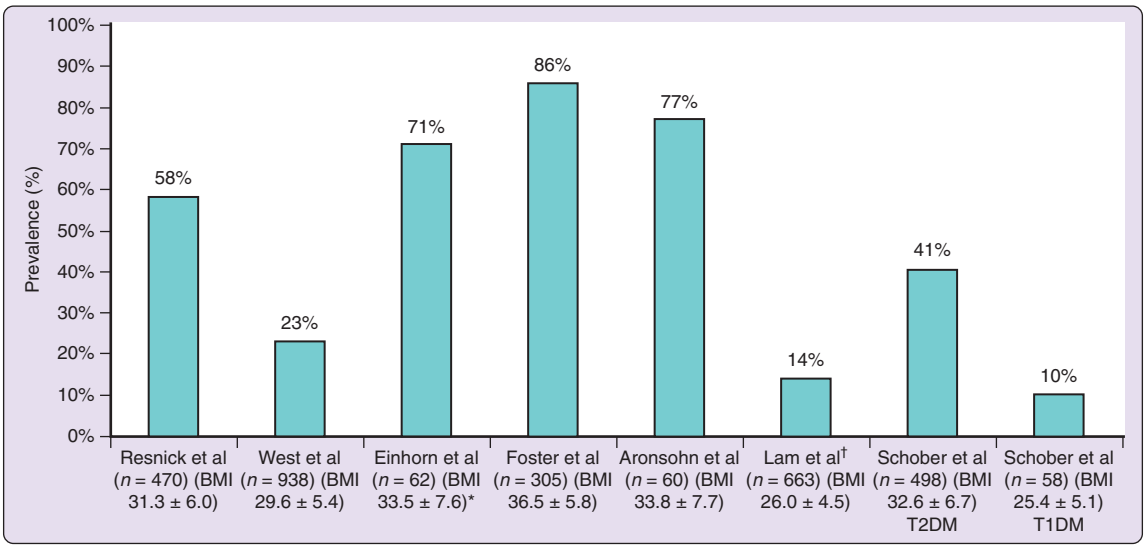


Figure 118-1 Prevalence of obstructive sleep apnea (OSA) in subjects with diabetes mellitus.*BMI available only in a bigger cohort of *n* = 279. †Chinese (Asian criteria for obesity: BMI greater than 25 kg/m^2). T1DM, Type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. (Data from referenced studies: Resnick et al., 2003[94]; West et al., 2006[99]; Einhorn et al., 2007[100]; Foster et al., 2009[101]; Aronsohn et al., 2010[102]; Lam et al., 2010[103]; Schober et al., 2011—both cohorts.[104])

data on the confounding influence of obesity are controversial, with conflicting evidence that either obese (Chinese)[111] or nonobese (non-Asian white) subjects[114] can show more improvement in insulin sensitivity in response to CPAP. A rigorously conducted RCT that investigated the effects of CPAP alone versus weight reduction alone versus both interventions in a cohort of obese subjects with severe OSA showed that weight reduction but not CPAP alone over 24 weeks improved insulin sensitivity.[116] The beneficial effect was not further enhanced in the group receiving both interventions, suggesting that obesity has a more dominant impact than OSA on insulin resistance in persons with OSA. Because weight loss also may lead to changes in OSA, however, the absence of information on sleep-disordered breathing after the treatment period in those data disallows definitive analysis of the complicit interactions of multiple risk factors and conditions that occur in this patient population.[117]

In diabetic subjects with OSA, data from observational studies on treatment of OSA tend to be favorable regarding improvement in indicators of glycemic status.[86] In the only randomized sham-CPAP controlled study reported to date, neither HbA_{1c} nor insulin sensitivity measured with the hyperinsulinemic euglycemic clamp showed any improvement with CPAP over a 3-month period,[118] but the low average CPAP use of 2.5 hours per night in the "therapeutic" CPAP group is considered inadequate to produce any change in OSA-related sequelae. Rapid eye movement (REM)-related event frequency, as indicated by REM-AHI, but not NREM-related event frequency, was found to be associated with worse glycemic control in type 2 DM, suggesting potential treatment implications regarding the need for adequate inclusion of REM sleep periods in overnight CPAP usage.[119]

Recently, an RCT of CPAP in 39 subjects with prediabetes and OSA found that 8 hours of CPAP use every night for 2 weeks, documented by nightly in-laboratory sleep monitoring, could reduce glucose response to oral glucose tolerance testing and improve insulin sensitivity.[120] As with a previous 1-week RCT in nondiabetic men,[111] these findings can be regarded as proof-of-concept for an adverse effect of untreated OSA on glucose metabolism, which is reversible with control of OSA, but the relevant clinical impact remains elusive.

In the face of increased insulin resistance, compensatory increase in insulin secretion occurs, but this homeostatic mechanism of beta cell function may be lost with chronic insult, ageing or pre-diabetic state. In clinical subjects or experimental human models of OSA, studies have shown conflicting data: pancreatic insulin secretion may be enhanced or impaired. In a study of 118 nondiabetic subjects with OSA evaluated with an intravenous glucose tolerance test (FSIVGGT), despite "OSA dose–dependent" impairment of insulin sensitivity, no increase in pancreatic beta cell insulin output was seen.[121] By contrast, in a study of 26 lean young men free of cardiometabolic disease, the presence of mild to moderate OSA was associated with insulin resistance and an increase in insulin secretion,[106] and a study of 45 severely obese adults also found that OSA was associated with increased beta cell function in those with normal glucose metabolism.[122]

Additional data indicate that coexistence of obesity, pregnancy, and OSA may well pose a conglomerate predisposition to gestational DM. Despite adjustment for pre-pregnancy BMI, a diagnosis of gestational DM has been strongly associated with a diagnosis of OSA.[123]

The diverse results regarding the effect of OSA or its treatment on glucose metabolism as illustrated above may be due to the use of different investigative tools that are not directly comparable, and to different sample characteristics. Host factors, intrinsic or extrinsic, play important roles in the determination of insulin-glucose metabolism in OSA, including age, BMI and adiposity, duration of OSA, prevailing glycemic status, genetic susceptibility, and variable exposures to external factors such as diet and exercise. Adherence to and duration of treatment for OSA also must be considered in the appraisal of any potential metabolic response. An overview of the

literature indicates that assessment tools, sample size, patient characteristics, and treatment duration and adherence are factors that need to be carefully addressed in the design of future clinical studies.

Although current investigative data are dominated by studies of the action of OSA towards glucose homeostasis and DM, the possibility of diabetic autonomic neuropathy as a predisposing factor toward sleep-related pharyngeal collapse and OSA has been raised.[124] In subjects with type 1 DM in particular, who usually are not overweight or obese, a higher prevalence of OSA than in the general population raises the possibility of contribution by autonomic neuropathy,[104] although data are very limited in this regard.[124]

Dyslipidemia

Clinical data regarding the relationships of OSA and lipids and the effect of treatment of OSA on lipids are mostly gleaned from studies that include several metabolic parameters as end-points, and relatively few which specifically investigated dyslipidemia as the primary measure of interest. Epidemiologic studies comprising relatively large numbers of subjects have identified an association between OSA and dyslipidemia independent of confounding variables. In the American SHHS cohort with mean age of 62 years, stepwise regression models identified that the respiratory disturbance index was independently determined by higher total cholesterol levels in men, and lower HDL cholesterol levels in women.[125] The European SYNAPSE study of 846 participants with mean age of 68 years showed that severe OSA was independently associated with low HDL cholesterol.[126] Oxygen desaturation index and AHI were independent predictors of HDL cholesterol levels, and the relationships were more pronounced in those not receiving lipid-lowering agents.[126] The Brazilian San Paola sleep cohort found that an AHI of 15 or less and a longer duration of oxygen saturation below 90% were independently associated with elevated fasting glucose and triglyceride levels and HOMA-IR.[91] By contrast, no independent association was identified between OSA (defined by AHI of 5 or less) and cholesterol or triglyceride levels in community-dwelling middle-aged Chinese residents of Hong Kong.[90]

Clinical studies of patients with OSA have reported various adverse lipid profiles, including elevations of total cholesterol, triglycerides, LDL cholesterol or lower HDL cholesterol levels.[62] Apart from promoting a dyslipidemic profile of cholesterol or triglycerides, subjects with OSA have been shown to have higher levels of oxidized or dysfunctional lipids which are more atherogenic, ascribed to increased oxidative stress.[127,128]

Observational intervention studies and RCTs regarding the effect of OSA treatment on dyslipidemia profiles have demonstrated variable changes in lipid parameters,[15,40] although some studies with short treatment durations may not have allowed adequate time for changes in circulating lipid levels. In a single-center longitudinal follow-up study of 127 patients with OSA, positive airway pressure treatment for 6 months significantly increased HDL cholesterol levels.[129] Pooled data from two RCTs on metabolic profile in OSA suggested that CPAP treatment of OSA results in a lowering of serum total cholesterol.[130] In a randomized controlled cross-over trial with dyslipidemia as the primary end point, therapeutic CPAP treatment for 2 months compared with placebo CPAP in 30 subjects with severe OSA (defined as mean AHI of 41) reduced postprandial hypertriglyceridemia and also lowered fasting and postprandial total cholesterol levels.[67] However, in the Icelandic Sleep Apnea Cohort, a 2-years follow-up assessment of CPAP treatment outcomes in 199 subjects with newly diagnosed OSA compared with 118 nonusers did not show any change in fasting lipid levels with CPAP treatment.[131] A recent systematic review of randomized controlled studies did not show any alteration in lipid levels with CPAP treatment of OSA.[40]

Hepatic Dysfunction and Nonalcoholic Fatty Liver Disease

A number of reports of liver enzyme elevations in adults and children with OSA have been published.[31,132,133] The National Health and Nutrition Examination Survey (NHANES) data between 2005 and 2010 for 10,541 adults reported that 15% had NAFLD and 7.2% had sleep disorders, identified as sleep apnea in 64.7% of those affected, and sleep apnea was independently associated with NAFLD with an odds ratio of 1.39 (95% confidence interval, 0.98 to 1.97).[134] Definitive evidence of liver injury in OSA was mostly derived from findings in morbidly obese subjects who underwent liver biopsy during bariatric surgery.[15] As with other metabolic dysfunction, the association of fatty liver with OSA is substantially confounded by obesity, and an independent relationship has not been firmly established.[31]

In studies of obese subjects, those with moderate or severe OSA and severe sleep-related hypoxemia exhibited more significant changes in indices of hepatic inflammation than those with mild OSA,[135,136] and in a series of subjects undergoing bariatric surgery, the absence of OSA was found to be an independent predictor of normal findings on liver histologic analysis.[137] Patients with severe OSA (AHI greater than 50) were found to have more insulin resistance and to have higher percentage of steatosis, as well as higher prevalence of necrosis and fibrosis, than patients with milder OSA with similar BMI.[138] In a study of 65 consecutive children with biopsy-proven NAFLD, 60% were shown to have OSA on polysomnography, and the presence and severity of OSA were associated with features of NASH and fibrosis, independently of BMI, abdominal adiposity, metabolic syndrome, and insulin resistance.[139] This relationship held in the nonobese children with NAFLD. Of note, the duration of oxyhemoglobin saturation below 90% correlated with increased hepatocyte apoptosis and fibrogenesis.

Despite the concern that OSA may worsen NAFLD, the major causes of premature death in subjects with NAFLD have been identified as type 2 DM and CVD, rather than the liver disease itself.[140] This correlation serves, however, as a reminder of the potential role of additional liver injury from OSA as a pathway to greater cardiometabolic burden.

Inflammation and Alteration of Adipocytokines

Multiple studies have addressed the profile of proinflammatory mediators in OSA, and results have been diverse and confounded by obesity.[24] A meta-analysis of 51 studies found higher levels of CRP, TNF-α, IL-6, and other molecules in patients with OSA than in control group subjects.[141] The Icelandic cohort study of 454 subjects with untreated OSA showed that OSA severity, as reflected by the degree of nocturnal oxygen desaturation, but not AHI, correlated significantly with levels of IL-6 and CRP.[142] An association of

BMI with IL-6 was found only in obese participants, and an independent association of OSA severity and CRP levels was found for minimum oxygen saturation only. Although a meta-analysis found that treatment of OSA with CPAP improved levels of CRP, TNF-α, and IL-6, the studies pooled for this analysis generally were small, nonrandomized trials,[143] and well-designed studies have failed to demonstrate that CPAP alters inflammation markers in OSA.[24,40]

Leptin regulates appetite and energy intake, and hyperleptinemia in obesity reflects leptin resistance. Hyperleptinemia is associated with increased insulin resistance and cardiometabolic morbidities. Subjects with OSA have consistently been found to have elevated plasma leptin levels compared with healthy subjects, although whether the increase is related independently to OSA or is simply due to the concomitant adiposity remains controversial.[24] Some studies have suggested that nocturnal hypoxemia, rather than AHI itself, is a better indicator of the effect of OSA on leptin levels. The Icelandic Sleep Apnea Cohort study of 452 patients with untreated OSA (mean BMI, 32.7 kg/m^2) showed that the dominant determinants of leptin levels were still obesity and gender, although OSA severity as measured by AHI explained a significant variance (3.2%) in leptin levels in the nonhypertensive group, with the relationship strongest in nonobese, nonhypertensive subjects.[144] Adiponectin has insulin-sensitizing, antiinflammatory, and antiatherogenic properties, and hypoadiponectinemia is associated with reduced insulin sensitivity, type 2 DM, and the metabolic syndrome. A majority of relevant studies in OSA have found that hypoadiponectinemia strongly correlates with obesity and insulin resistance as in general populations, but the relationship of adiponectin levels and OSA is heavily confounded by obesity.[24] One study suggested that the degree of sympathetic activation, rather than sleep-disordered breathing indices, contributed to the determination of adiponectin levels.[41] Another study found that nonobese men with severe OSA, compared with nonobese control subjects who did not have OSA, demonstrated more impaired insulin resistance (higher HOMA-IR) and a profile of higher 24-hour levels of leptin, CRP, IL-6, and TNF-α but similar levels of adiponectin.[18]

Obstructive Sleep Apnea and Metabolic Syndrome

Although the considerable disagreement in the medical community over terminology and diagnostic criteria has yet to be resolved, *metabolic syndrome* is conceptually accepted as a clustering of multiple metabolic risk factors for CVD and DM.[13] The key features of metabolic syndrome are abdominal obesity, insulin resistance, atherogenic dyslipidemia, a prothrombotic state, and an inflammatory profile. This constellation of metabolic aberrations often is accompanied by arterial hypertension and/or type 2 DM, in keeping with the relevant genetic or exogenous predispositions.[2] Different sets of clinical criteria for definition of metabolic syndrome have been reached by various expert panels,[13] and one widely used set of criteria from the National Cholesterol Education Program Adult Treatment Panel III is presented in Table 118-2[145,146] for reference. Besides the core components, an increasing number of conditions are proposed to be included in the metabolic syndrome "family," and OSA is one such condition because of its strong associations with other core factors and its potential role in causing CVD and glucose dysmetabolism.[146,147]

Table 118-2 Definition of Metabolic Syndrome of the National Cholesterol Education Program Adult Treatment Panel III (NECP-ATIII)

Risk Factor	Defining Level
Abdominal obesity (waist circumference)	
Men	>102 cm
Women	>88 cm
Triglycerides	≥150 mg/dL
High-density lipoprotein cholesterol (HDL cholesterol)	
Men	<40 mg/dL
Women	<50 mg/dL
Blood pressure	≥130/≥85 mm Hg
Fasting glucose	≥110 mg/dL
Asian criteria for abdominal obesity*	
Men	≥90 cm
Women	≥80 cm

*Asian abdominal obesity criteria: data from *The IDF consensus worldwide definition of the metabolic syndrome.* Brussels: International Diabetes Federation; 2006. <http://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf>.
Data from National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002; 106:3113–421.

As illustrated in previous sections of this chapter, strong and probably independent associations of OSA with various individual clinical components of the metabolic syndrome are likely. Studies have demonstrated a 5- to 7-fold increase in the association of OSA with this phenotypic entity, compared with those without OSA, and the association was independent of BMI and age; a similar association has been found in subjects with versus those without metabolic syndrome.[3,11] Of a total of 228 patients referred for OSA evaluation, 146 patients proved to have OSA, of whom 60% had metabolic syndrome, whereas of those without significant OSA, only 40% had metabolic syndrome.[148] The presence of OSA in subjects with classical metabolic syndrome as defined by current international guidelines may add to the associated inflammatory and cardiometabolic burden.[149] To date, definitive evidence that treating OSA reduces the occurrence of conventionally defined metabolic syndrome is lacking. With respect to clinical impact, this may be just a matter of threshold—if indeed treatment of OSA can produce adequate improvement in individual components of the metabolic syndrome, it is reasonable to anticipate improvement in the adverse cardiometabolic outcomes predicted by metabolic syndrome. However, despite increasing research focus in this area, treatment of OSA has not yet been associated with consistent and clinically important improvement in the metabolic components of obesity, insulin-glucose dysmetabolism, dyslipidemia, or systemic inflammation.[40,150]

CONCLUSIONS AND PERSPECTIVES

The strong association between OSA and metabolic disorders has received intense interest, against the background of a

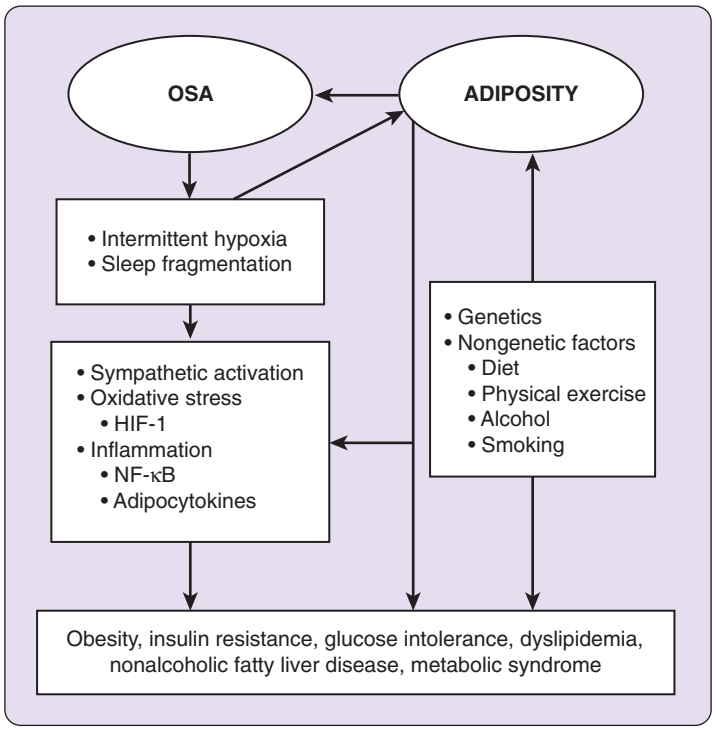


Figure 118-2 Proposed mechanistic links of obstructive sleep apnea (OSA) and metabolic disorders. HIF-1, Hypoxia-inducible factor-1; NK-κB, nuclear factor kappa B.

global epidemic of obesity and obesity-related diseases, and the modern lifestyle of sleep curtailment. In human disease, despite growing evidence for an independent or additive adverse effect of OSA on metabolic function, the relationship remains controversial owing to the strong potential confounding effect of obesity and the unaccounted influence of numerous intrinsic or exogenous factors that may affect body metabolism. Furthermore, the influence of OSA and metabolic dysfunction may be bidirectional, and the putative alterations of metabolic function by OSA, such as increased insulin resistance with promotion of visceral adiposity, may in turn aggravate OSA, allowing a vicious circle to be established (see Figure 118-2). The demonstration of a beneficial metabolic effect of treatment of OSA, moreover, has been elusive. Overall, many studies are of limited sample size and/or heterogeneous sample characteristics, and because of the need for symptomatic treatment in this condition, conducting RCTs or longitudinal follow-up studies of untreated disease with adequate duration is inherently difficult. Treatment results for children, young adults, and a range of middle-aged to elderly adults can hardly be directly compared; obese subjects may behave differently from lean subjects metabolically, and the difference may not be adequately resolved with adjustment for BMI. Furthermore, factors such as the duration of OSA before sleep study diagnosis are not possible to define accurately, owing to the nonspecific and insidious onset of symptoms in most cases, and the prevailing metabolic status in the individual patient influence whether "reversibility" could still be attained with treatment of OSA even if it is the culprit for dysmetabolism. Finally, any metabolic benefits that are demonstrated in the study setting, especially in short-term studies, need to be reproducible in real life, where lifestyle factors and treatment adherence interpose.

Animal and cellular studies using intermittent hypoxia or other surrogate models for OSA allow manipulation of experimental conditions to provide insights into the physiologic or molecular mechanisms that may be at play in the human disease. Accordingly, such studies can be expected to pave the way for extension into translational work in humans.

SUMMARY

In view of the many factors that may influence metabolic health and disease, it is unrealistic to expect that one unifying path applies to all. Although any demonstrated adverse impact of OSA on metabolic function may be of limited clinical effect, the economy of scale would translate such effect into an enormous health care burden when coupled with the sweeping epidemic of obesity and anticipated escalation of rates of related morbidity in the coming decades. Meanwhile, holistic care mandates a multidimensional approach, in both clinical practice and research, to the management of patients presenting with either OSA or a metabolic disorder. Strong evidence suggests that controlling body weight is of pivotal importance in improving metabolic health, including in subjects with OSA. OSA is gaining professional recognition as a serious comorbid condition to be accorded appropriate attention in the management of cardiometabolic disorders.[151,152] Recent clinical practice guidelines note that OSA should be identified and appropriately treated in the comprehensive care of patients with DM.[152] It is of pivotal importance that both the medical profession and the public be aware of the clustering of OSA and metabolic disorders, with the consequent need for determination of optimal methods of prevention, early recognition, and relevant clinical management.

CLINICAL PEARL

Strong associations are recognized between OSA and various metabolic disorders, notably obesity, type 2 DM, dyslipidemia, and NAFLD. Despite abundant suggestive evidence for independent associations between OSA and metabolic dysfunction, a causal or aggravating role for OSA in dysmetabolism is not yet delineated, and treatment of OSA has not been definitively shown to prevent or improve metabolic dysfunction. A reasonable holistic clinical approach mandates high vigilance regarding the clustering of these conditions, with relevant screening and specific management considered accordingly. The need for appropriate body weight control measures is particularly relevant in this regard.

Selected Readings

Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on epidemiology and prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;**120**(16):1640–5.

Arnardottir ES, Lim DC, Keenan BT, et al. Effects of obesity on the association between long-term sleep apnea treatment and changes in interleukin-6 levels: the Icelandic Sleep Apnea Cohort. *J Sleep Res* 2015;**24**(2):148–59.

Bonsignore MR, McNicholas WT, Montserrat JM, et al. Adipose tissue in obesity and obstructive sleep apnoea. *Eur Respir J* 2012;**39**(3):746–67.

Chirinos JA, Gurubhagavatula I, Teff K, et al. CPAP, weight loss, or both for obstructive sleep apnea. *N Engl J Med* 2014;**370**(24):2265–75.

Kent BD, Grote L, Bonsignore M, et al. Sleep apnoea severity independently predicts glycaemic health in nondiabetic subjects: the ESADA study. *Eur Respir J* 2014;**44**(1):130–9.

Lam JC, Mak JC, Ip MS. Obesity, obstructive sleep apnoea and metabolic syndrome. *Respirology* 2012;**17**(2):223–36.

Lavie L. Oxidative stress—a unifying paradigm in obstructive sleep apnea and comorbidities. *Prog Cardiovasc Dis* 2009;**51**(4):303–12.

Pamidi S, Wroblewski K, Stepien M, et al. Eight hours of nightly CPAP treatment of obstructive sleep apnea improves glucose metabolism in prediabetes: a randomized controlled trial. *Am J Respir Crit Care Med* 2015;**192**(1):96–105.

Preis SR, Massaro JM, Hoffmann U, et al. Neck circumference as a novel measure of cardiometabolic risk: the Framingham Heart study. *J Clin Endocrinol Metab* 2010;**95**:3701–10.

Strand LB, Carnethon M, Biggs ML, et al. Sleep disturbances and glucose metabolism in older adults: the cardiovascular health study. *Diabetes Care* 2015;**38**(11):2050–8.

A complete reference list can be found online at ExpertConsult.com.

Overlap Syndromes of Sleep and Breathing Disorders

Chapter 119

Jose M. Marin; Santiago J. Carrizo

Chapter Highlights

- Sleep depresses the central control of breathing and muscle tone. These changes have no health effects in healthy subjects. However, in patients with pulmonary diseases such as chronic obstructive pulmonary disease (COPD), asthma, and interstitial lung diseases, these changes may aggravate gas exchange abnormalities and induce significant hypoxemia and hypercapnia, especially during rapid eye movement sleep.
- As COPD, asthma, and obstructive sleep apnea (OSA) are prevalent disorders in adults, overlap of OSA and either COPD or asthma is frequent in clinical practice. Such overlap, termed overlap syndrome, carries an excessive risk for worsened sleep- and awake-related outcomes than any one of these conditions alone, including increased risk of COPD exacerbations and mortality.
- Practitioners should identify the coexistence and severity of OSA in patients with COPD or asthma and the presence and severity of the overlap syndrome, and they should establish a personalized treatment in each case. A sleep study should be considered in any patient with COPD or asthma with signs and symptoms of OSA, such as snoring and excessive daytime sleepiness, or with inappropriate awake hypoxemia.
- Noninvasive ventilation (continuous positive airway pressure, bilevel positive airway pressure), with supplemental oxygen if necessary, should be prescribed after an appropriate titration process.

OVERVIEW

Sleep is associated with adaptive changes of the airways, lungs, and chest wall mechanics. In patients with chronic pulmonary diseases, such physiologic changes as well as the pathophysiologic changes of sleep breathing disorders, such as obstructive sleep apnea (OSA), may result in acute and chronic adverse effects, including precipitation or worsening of hypoxemia, hypercapnia, and bronchoconstriction, which in the long term can contribute to worsened outcomes in these chronic pulmonary disorders. This chapter reviews the clinically imperative concepts of overlap syndromes, defined as the coexistence in the same patient of OSA and one or more of the following chronic respiratory conditions: chronic obstructive pulmonary disease (COPD), asthma, and pulmonary hypertension (PH). The clinical relevance of identifying the coexistence of a primary sleep disorder such as OSA in patients with these chronic respiratory disorders lies not only in the diagnosis of an overlap syndrome; it involves a worse prognosis for these coexistent respiratory diseases and the need for specific treatment of the concomitant sleep-disordered breathing (SDB). Physicians who care for patients with such respiratory disorders should recognize these associations.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

According to the latest version of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) strategy document,[1] COPD is defined as a preventable and treatable disease, characterized by persistent airflow limitation that is usually progressive and associated with an enhanced chronic inflammatory response in the airways and the lung to noxious particles or gases. A clinical diagnosis of COPD should be considered in any patient with dyspnea, chronic cough, or sputum production and a history of exposure to risk factors for the disease. Spirometry is required to make the diagnosis in this clinical context; a postbronchodilator forced expiratory volume in the first second of expiration to forced vital capacity (FEV_1/FVC) ratio of less than 0.7 confirms the presence of persistent airflow limitation and thus of COPD.

Sleep disturbance is common in COPD without other coexistent primary sleep disorders. In a large survey conducted in North America and Europe, 40% of patients reported problems with their sleep.[2] In a European survey, 78.1% of patients with COPD reported some degree of nighttime symptoms, including one or more of the following: dyspnea, cough with increased sputum production, wheezing, and difficulty with maintenance of sleep. The prevalence of such nighttime symptoms was positively correlated with the severity of spirometrically measured airflow obstruction.[3] Polysomnography (PSG) studies show that these patients have problems initiating or maintaining sleep, reduced rapid eye movement (REM) sleep, and frequent microarousals. This poor sleep quality increases in parallel with the frequency of nocturnal respiratory symptoms, like cough and wheezing,[4] and COPD severity.[5]

In COPD patients, sleep is associated with reduced rib cage contribution to breathing, diaphragmatic inefficiency, and increased accessory muscle contribution to breathing.[6] The result is a reduction in functional residual capacity, which may augment ventilation-perfusion mismatching and hypoxemia. More than 50% of COPD patients with daytime arterial oxygen saturation of hemoglobin (Sao_2) above 90% on breathing room air, and without concomitant OSA, experience significant oxygen desaturation during sleep, defined as spending at least 30% of the night with Sao_2 below 90%.[7] Daytime gas exchange abnormalities are, however, somewhat predictive of sleep oxygen desaturation among COPD patients.[8] Given the shape of the oxyhemoglobin dissociation curve, patients on the steep portion of the curve (e.g., Pao_2 < 60 mm Hg on breathing room air during the daytime) would be expected to have a greater fall in Sao_2 during sleep, particularly during REM sleep. Accentuated physiologic hypoventilation in COPD is also the consequence of decreased central respiratory drive response to chemical and mechanical inputs,[9] increased upper airway resistance due to a loss of tone in the upper pharyngeal muscles,[10] and reduced efficiency of diaphragmatic contraction due to lung hyperinflation.[11] The consequences of nocturnal hypoxemia and hypercapnia are well known and include arrhythmias and PH. In addition, recent data suggest that disturbed sleep is an independent risk factor of COPD exacerbations and mortality.[12]

Chronic Obstructive Lung Disease and Obstructive Sleep Apnea Overlap Syndrome

Definitions and Classifications

The GOLD definition of COPD also underlines that "exacerbations and comorbidities contribute to the overall severity in individual patients."[1] OSA is recognized as one of these comorbidities. OSA is characterized by sleep-related pathologically increased upper airway resistance, repetitive decrease or absence of inspiratory and expiratory airflow, sympathetic activation, and intermittent oxyhemoglobin desaturation and hypercapnia.[13] The diagnosis of OSA requires a PSG study, with five or more apneas or hypopneas per hour of sleep (i.e., the apnea-hypopnea index [AHI]) consistent with OSA.[14] The consensus definitions of severity in OSA, mild (AHI ≥5 and <15 episodes/hour), moderate (AHI ≥15 and <30 episodes/hour), and severe (≥30 episodes/hour), are in part based on published literature that has found an association between such severity definitions of OSA and risk of excess mortality in this sleep-related breathing disorder.[15-19]

The coexistence of COPD and OSA was first described as the overlap syndrome by David Flenley 30 years ago.[20] He pointed out that PSG should be considered in COPD patients with obesity, snoring, or morning headache associated with nocturnal oxygen therapy to assess for the presence of associated OSA. He believed that the clinical course and prognosis of such "overlap patients" were worse than for patients suffering from COPD or untreated OSA alone. These opinions remain valid today. Nevertheless, at present, the term *overlap syndrome* is not a formal diagnostic designation for patients suffering from OSA and COPD. In the individual patient, it is better to describe the underlying lung disease and the associated abnormality of the sleep disorder. A classification of severity for this entity is not available, and health outcomes appear to depend on the severity of OSA and COPD independently.

Epidemiology of COPD/OSA Overlap Syndrome

In pulmonary clinics, OSA and COPD are two of the most prevalent chronic respiratory disorders. It is estimated that 10% of the general population has moderate to severe COPD as defined by an FEV_1/FVC ratio of less than 0.7 plus an FEV_1 of less than 80% predicted.[21] The prevalence of COPD increases with age and is directly related to the prevalence of tobacco smoking, but outdoor and indoor air pollution are also major COPD risk factors. The prevalence and burden of COPD are projected to increase in the coming decades because of continued exposure to COPD risk factors and the aging of the world's population.

Among men and women between the ages of 30 and 60 years, 20% and 9%, respectively, had an AHI of at least 5 events/hour in the Wisconsin Sleep Cohort Study.[22] Since this report was published 20 years ago, data from the same ongoing cohort provide prevalence estimates of moderate to severe SDB of the sleep apnea type (AHI ≥ 15 events/hour), thus showing a substantial increase during the last 2 decades.[23] The sex disparity of OSA ends at around the age of 55 years, with a sharp rise among postmenopausal women.[23-25]

There are, however, no studies that directly assess the prevalence of the OSA/COPD overlap syndrome. Because COPD and OSA are each increasing throughout the world in association with an aging population, presumably the overlap syndrome is becoming more prevalent. In clinical series, it has been noted that approximately 11% of patients with at least moderate OSA, as defined by an AHI of more than 20 events/hour, have airflow limitation on spirometry.[26] In a European population study of patients with predominantly mild COPD, the coincidence of OSA syndrome (AHI > 5 events/hour accompanied by excessive daytime sleepiness) occurred in 1% of the total population.[27] The Sleep Heart Health Study, a community-based cohort study that included 5954 participants who had PSG and spirometry at baseline, found that 19% had airway obstruction (defined as FEV_1/FVC < 0.7) that was predominantly mild. The prevalence of OSA, defined as a respiratory disturbance index of more than 10 events/hour, was not higher in subjects with airway obstruction (defined as FEV_1/FVC < 0.7) compared with the nonobstructed population.[28] There were 254 participants (4.3%) who had both characteristics: obstructive airways disease and sleep apnea. As expected, respiratory disturbance index increased with higher body mass index (BMI) in participants with and without airway obstruction. Age effect was not specifically addressed in this study. In short, the few available population studies of the association between COPD and OSA (i.e., overlap syndrome) show great variability in the prevalence of this association. It does appear that the world's adult population is affected in a range between 1% and 4%. This range likely reflects, at least to some extent, differences in the criteria used to define OSA and the age and weight of the subjects studied.

Sleep in Patients with COPD/OSA Overlap Syndrome

In the Sleep Heart Health Study, patients with OSA/COPD overlap syndrome had a lower total sleep time, lower sleep efficiency, and higher daytime sleepiness as assessed by the Epworth Sleepiness Scale[29] than did patients with COPD alone. They were also more likely to have greater sleep-related

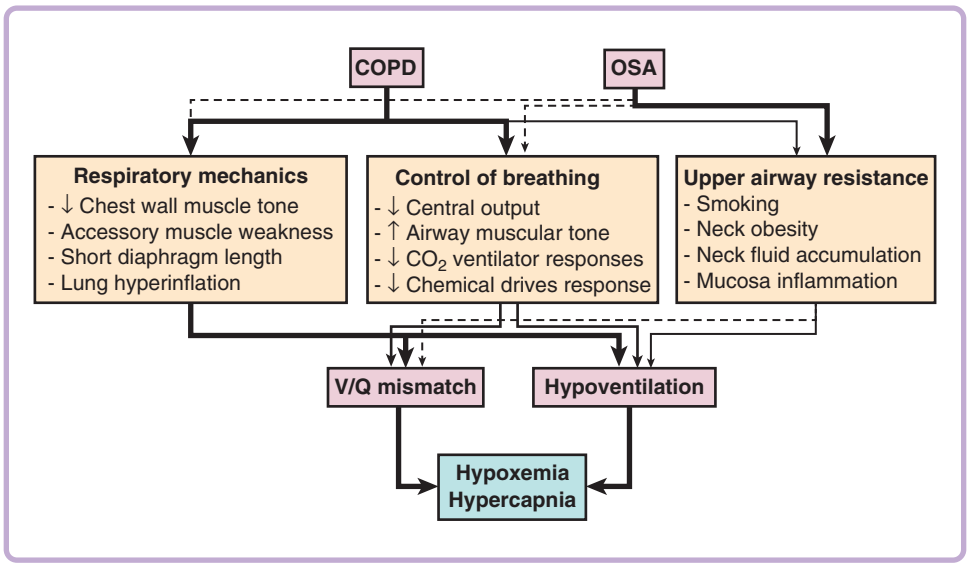


Figure 119-1 Pathways involved in producing sleep-related hypoxemia and hypercapnia in chronic obstructive pulmonary disease (COPD) and obstructive sleep apnea (OSA) overlap syndrome.

oxygen desaturation compared with participants with OSA or airway obstruction alone.[28] Most important, patients with overlap syndrome, compared with patients with either COPD or OSA alone, display more profound oxygen desaturation during sleep as well as worse daytime hypoxemia and hypercapnia.[26]

Risk Factors for COPD/OSA Overlap Syndrome

Patients with COPD can incur specific OSA risks, including obesity irrespective of airflow obstruction severity,[30] active smoking,[31,32] and both pharyngeal and lower extremity edema associated with episodic use of oral corticosteroids and impaired cardiac output.[33] There is also evidence that patients with advanced COPD who lose weight may show reduced diathesis for upper airway obstruction.

Sleep and Breathing Pathophysiology of COPD/OSA Overlap Syndrome

Because obesity also reduces functional residual capacity during sleep, overweight and obese patients with COPD/OSA overlap syndrome are particularly subject to a reduction of alveolar volume and greater gas exchange abnormalities during sleep apneas and hypopneas (Figure 119-1). Further, respiratory control center output is reduced during sleep, especially during REM sleep,[34] including blunted ventilatory responses and mouth occlusion pressure responses to CO_2.[35] During obstructive apneic episodes, to overcome the upper airway resistance and to maintain adequate airflow to the lung, increased diaphragmatic and abdominal muscle effort is required. This can be particularly difficult in COPD patients who already have increased intrathoracic airway resistance and lung hyperinflation at baseline. When COPD patients develop such obstructive apnea episodes, the compensatory response of the respiratory center is slower, apneas are longer, and changes in Pao_2 and Pco_2 are more intense compared with non-COPD subjects. Patients with COPD/OSA overlap syndrome who have awake hypoxemia are especially prone to nocturnal oxygen desaturation by being on the steep portion of the oxyhemoglobin dissociation curve.

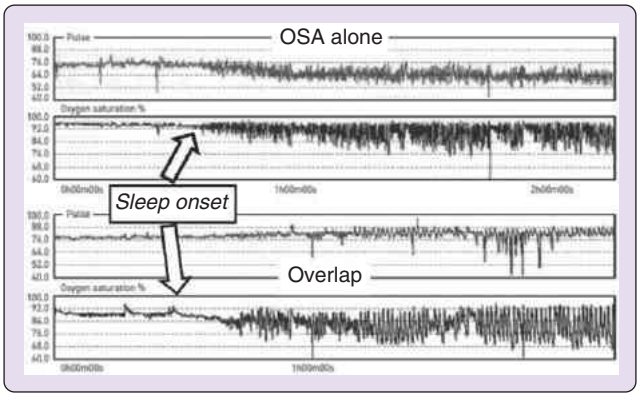


Figure 119-2 Typical pattern during sleep of a patient with obstructive sleep apnea (OSA) alone (*upper panel*) and chronic obstructive pulmonary disease (COPD)/OSA overlap syndrome (*lower panel*). Note the pattern of persistent O_2 desaturation in overlap patients; in contrast to the OSA patients, O_2 saturation does not return to baseline between apnea episodes.

Clinical Features of COPD/OSA Overlap Syndrome

Compared with patients with COPD alone or OSA alone, overlap patients of similar ages tend to be more obese and to have more comorbid conditions.[36] They also report more daytime sleepiness[28] and poorer quality of life[37] than either COPD or OSA patients without overlap syndrome. Sleep recordings of patients with COPD/OSA overlap show a lower total sleep time, lower sleep efficiency, and greater sleep fragmentation than those with COPD or OSA alone. More severe nocturnal O_2 desaturation is also a characteristic feature in COPD/OSA overlap patients compared with either condition alone. Subjects with OSA alone return to a normal O_2 saturation (Sao_2) in sleep between obstructive events (i.e., intermittent hypoxemia), whereas in COPD alone, as a result of the diathesis to sleep-related hypoventilation and ventilation-perfusion mismatch as noted before, nocturnal O_2 saturation characteristically decreases more evenly throughout sleep and at the termination of an apnea or hypopnea episode tends not to return to the initial baseline level (Figure 119-2).

A typical patient with COPD/OSA overlap syndrome has a reduced awake and asleep baseline Sao_2, a lower mean sleep-related Sao_2, and a longer time in hypoxemia than patients with OSA or COPD alone.

The majority of patients with OSA alone do not develop significant sleep-related hypercapnia because of interapnea hyperventilation. However, if the patient also has COPD, the abnormal mechanical and chemical ventilatory responses as noted before may result in postapnea CO_2 levels that do not return to baseline. Over time, a progressive desensitization of the respiratory center in response to OSA-related hypoxic-hypercapnic episodes develops, such that patients with COPD/OSA overlap syndrome can remain hypercapnic during sleep.[38] Of note, continuous positive airway pressure (CPAP) treatment for the OSA (see Diagnosis and Management of COPD/OSA Overlap Syndrome) can partially reverse this phenomenon.[39] Although daytime hypercapnia can develop in OSA without COPD, awake hypercapnia is much more frequent in the patient with overlap syndrome.[40] Both daytime hypoxemia and hypercapnia have been found to be predictors of right-sided heart failure in COPD patients,[41] and therefore these should be considered potentially treatable markers of otherwise poorer prognosis in COPD/OSA overlap.

Excessive sleepiness in patients with OSA alone is associated with decrements in school and work performance.[42] Further, there is also a strong association between OSA severity, as measured by the AHI, and the risk of traffic accidents.[43] It is reasonable to expect that in patients with COPD/OSA overlap syndrome, such performance decrements and risks reflect the sum of the severity of the sleep disorders of both entities, but such consequences of the COPD/OSA overlap syndrome have not been evaluated specifically. Similarly, whereas OSA is considered an independent risk factor for insulin resistance, with OSA severity predicting risk for incident diabetes,[44] neither COPD alone nor COPD/OSA overlap has been specifically linked with risk of metabolic disorders.

Both OSA and COPD alone are associated with an increased risk of cardiovascular morbidity and mortality. For example, epidemiologic data show a strong association between OSA and incident arterial hypertension,[45] particularly refractory hypertension. In COPD alone, however, arterial hypertension prevalence is similar to that of the general population, and patients with COPD/OSA overlap appear to have the same prevalence rates as patients with OSA alone.[36] Untreated OSA patients are also particularly susceptible to development of atrial fibrillation,[46] as are patients with COPD alone, likely related to nocturnal O_2 desaturation.[47,48] A community-based retrospective cohort analysis, including data collected on 2873 patients older than 65 years, confirmed an increased risk of new-onset atrial fibrillation in COPD/OSA overlap syndrome compared with OSA or COPD alone.[49]

Epidemiologic data indicate that incidence of coronary artery disease, stroke, and heart failure is increased in OSA[15,16,50] and COPD[51]; no such incidence data are available for COPD/OSA overlap. However, Chaouat et al[26] demonstrated that patients with COPD/OSA overlap syndrome have increased daytime pulmonary vascular resistance compared with patients with OSA alone, whereas Sharma et al[52] recently documented a higher right ventricular mass and

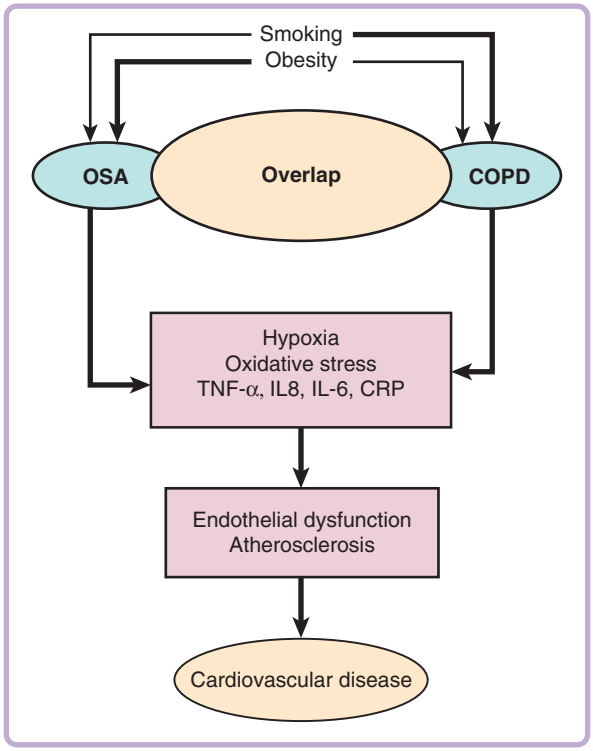


Figure 119-3 Schematic illustrating potential pathways involved in producing accelerated cardiovascular disease as a result of obstructive sleep apnea (OSA), chronic obstructive pulmonary disease (COPD), and COPD/OSA overlap syndrome. COPD, Chronic obstructive pulmonary disease; CRP, c-reactive protein; IL, interleukin; TNF-α, tumor necrosis factor-alpha.

remodeling indices in overlap syndrome compared with patients with COPD alone. In addition, arterial stiffness, a surrogate marker of subclinical atherosclerosis, has also been found to be significantly higher in subjects with COPD/OSA overlap than in those with OSA alone.[53] Finally, whereas increased oxidative stress is associated with both COPD and OSA, with evidence of increased circulating proinflammatory cytokines and leukocytes in both disorders, no specific data exist regarding COPD/OSA overlap syndrome and risk and prevalence of such oxidative stress compared with COPD or OSA alone. Potential key risk factors for endothelial dysfunction, atherosclerosis, and ultimately cardiovascular diseases are depicted in Figure 119-3.

In both COPD alone and OSA alone, the risk of excess all-cause mortality increases in association with increasing severity of these disorders. The excess of mortality is most marked in younger individuals with OSA[54] and in more elderly patients with COPD.[55] Overall, evidence indicates that mortality is increased in COPD/OSA overlap patients. For example, in OSA patients studied at sleep clinics, the coexistence of COPD has been found to increase the risk of death compared with patients with OSA alone.[56] We have recently confirmed this in a large cohort of patients with an average age of 57 years, referred with suspected SDB. In addition to PSG, all patients underwent spirometry as a routine procedure.[36] During a median follow-up period of more than 9 years, all-cause mortality was higher in the overlap group untreated for OSA (42.2%) than in the COPD-only group (24.2%) (Figure 119-4). In the COPD patients, comorbid untreated OSA remained a risk factor for death even after

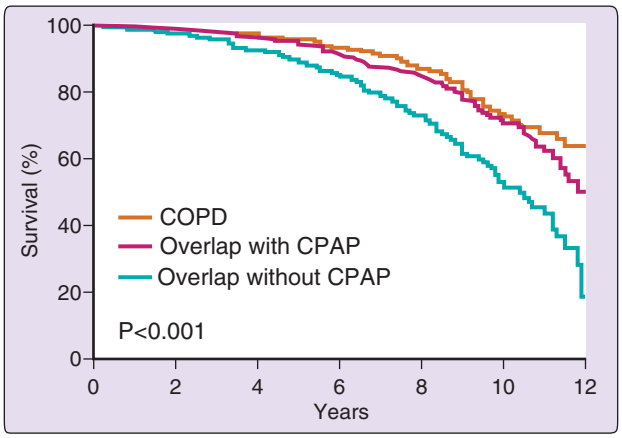


Figure 119-4 Kaplan-Meier survival curves of chronic obstructive pulmonary disease (COPD) patients without obstructive sleep apnea (OSA), patients with COPD and coexisting untreated OSA (overlap group), and patients with overlap syndrome treated with continuous positive airway pressure (CPAP). The differences in survival for COPD alone and COPD/OSA overlap syndrome treated with CPAP are statistically different compared with patients with untreated overlap syndrome ($P < .001$). (Reprinted with permission of the American Thoracic Society. Copyright © 2014 American Thoracic Society. Marin JM, Soriano JB, Carrizo SJ, et al. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. *Am J Respir Crit Care Med* 2010;182:325–31. Official Journal of the American Thoracic Society.)

adjustment for FEV_1 percentage predicted as a surrogate of COPD severity. There was a significantly higher number of cardiovascular deaths in patients with COPD only and untreated overlap syndrome compared with overlap patients treated appropriately for their OSA with CPAP. Interestingly, the second most frequent cause of death was cancer in patients with both OSA and COPD alone.[57,58]

Nocturnal death risk appears to be increased in COPD compared with the general population, mainly during COPD exacerbations.[59] Nocturnal hypoxemia, an important pathophysiologic feature of OSA, is associated with sudden cardiac death (SCD). Gami et al[60] reported on 10,701 consecutive adults undergoing diagnostic PSG and sought to identify the risk of SCD associated with OSA. During an average follow-up of 5.3 years, 142 patients had resuscitated or fatal SCD. Independently of well-established risk factors, SCD was best predicted by age older than 60 years, AHI above 20, mean nocturnal Sao_2 below 93%, and nadir nocturnal Sao_2 below 78%. No data are available in this study regarding the risk of nocturnal death in patients with COPD/OSA overlap versus COPD or OSA alone. Nevertheless, the report by McNicholas and FitzGerald[59] documented that nocturnal death was higher among patients admitted for acute exacerbation of chronic bronchitis or emphysema than in patients admitted for other causes. It is possible that an increased sympathetic activity along with a reduction in the perfusion of oxygen to the myocardium can increase the risk of arrhythmias and mortality during nighttime hours in COPD patients. Whether the coexistence of OSA (i.e., COPD/OSA overlap) increases this risk remains unknown.

Diagnosis and Management of COPD/OSA Overlap Syndrome

There are no specific guidelines for the diagnosis or treatment of COPD/OSA overlap syndrome. In the appropriate clinical context, PSG and spirometry should be performed to confirm the existence of the syndrome and to establish its severity. It has been appropriately stated that PSG should be considered in patients with COPD "when OSA is suspected because of either symptoms or the development of hypoxemic complications—cor pulmonale and polycythemia—with daytime Pao_2 greater than 60 mm Hg."[61]

The therapeutic management of identified COPD/OSA overlap syndrome patients should, in general, be based on optimizing treatment for both conditions (COPD and OSA) following corresponding clinical recommendations.[1,43] The goal of such therapy includes improvement in subjective outcomes, such as sleep fragmentation, sleep quality, and daytime sleepiness, as well as optimization of more objective data regarding daytime alertness and function and COPD- and OSA-specific cardiopulmonary outcomes, such as frequency of COPD exacerbation. Correction of hypoxemia and hypercapnia during sleep is considered especially important to reduce cardiovascular complications and to increase survival.

Noninvasive ventilation (NIV), currently typically applied as positive airway pressure (PAP) delivery thorough a nasal or face mask, is the most effective treatment for OSA. Continuous PAP (CPAP) is the optimal PAP therapy for most patients with OSA; bilevel PAP, which delivers a higher pressure during inspiration than during expiration, may also be used if a pressure gradient that increases alveolar ventilation is necessary, effective, and tolerated.

In COPD, NIV in a specifically ventilatory mode (usually bilevel PAP) is consistently shown to be highly effective in the setting of acute and acute-on-chronic hypercapnic respiratory insufficiency. In contrast, data regarding the effects of NIV on quality of life, lung function, gas exchange, and long-term survival have been contradictory when it is used in the chronic setting in COPD patients, in part because of the absence of studies of sufficient power and duration.[62] In the United States, NIV is reimbursed for patients with severe COPD and all the following criteria: (1) OSA has been ruled out; (2) awake $Paco_2$ is 52 mm Hg or higher; and (3) sleep oximetry shows Sao_2 of 88% or less for 5 minutes or more while breathing supplemental O_2 at 2 liters/minute or at the patient's prescribed FIo_2.[63]

Data have now accrued specific to OSA/COPD overlap syndrome regarding nocturnal NIV, specifically CPAP. In a long-term cohort study, overlap syndrome patients not treated with CPAP demonstrated both an increased risk of death from any cause and an increased risk of hospitalization for COPD exacerbation compared with overlap patients who were treated with and adhered to CPAP.[36] In another observational study, the use of CPAP added to long-term oxygen therapy improved survival among overlap patients with chronic respiratory failure.[64] Finally, a retrospective analysis of 227 patients with COPD/OSA overlap syndrome treated with CPAP revealed that a greater time on CPAP was associated with a reduced risk of death after controlling for common risk factors.[65]

The choice between CPAP and bilevel PAP can be determined during the titration session, based on the pattern of SDB. In cases in which OSA predominates and there is no coexistent consistent sleep-related hypoventilation, CPAP may be most appropriate to treat the OSA component. In cases in which there is evidence of any degree of nocturnal hypoventilation in addition to the apneic episodes, bilevel

PAP may be more appropriate. Nevertheless, there is no specific evidence in the literature of the superiority of CPAP or bilevel PAP for the treatment of patients with OSA/COPD overlap syndrome regarding long-term outcomes. Supplemental oxygen should be added to the mask or the PAP circuit if the otherwise optimal-appearing PAP regimen (whether CPAP or bilevel PAP) alone fails to provide satisfactory oxygenation. The ideal setting in which to adjust these parameters is the sleep laboratory, and such "titrations" should be conducted by well-trained technicians with the design, guidance, and interpretation of clinicians with sleep breathing expertise.

In most patients with COPD alone, nocturnal hypoxemia, when present, is corrected with supplemental O_2 through a nasal cannula. Nevertheless, alveolar ventilation of such patients is particularly dependent on the peripheral stimulant effect of hypoxemia. Therefore, to minimize the tendency toward CO_2 retention, particularly during sleep hours, such O_2 supplementation should be titrated carefully. The emergence of morning headache after O_2 initiation in patients with COPD is an indication to perform a PSG study to exclude the coexistence of OSA or to investigate the development of CO_2 retention. In OSA, supplemental oxygen treatment without PAP can eliminate or reduce nocturnal hypoxemia, but it does not reduce the AHI, daytime hypersomnolence,[66] or nocturnal blood pressure.[67] The role of oxygen supplementation as a solo nocturnal therapy in COPD/OSA overlap syndrome has not been sufficiently explored, and at present it is recommended that nocturnal O_2 be used as a complement to NIV in patients with COPD/OSA overlap syndrome.

No specific studies have been conducted on sleep quality, SDB, or long-term clinical outcomes to evaluate the effects of pharmacologic treatment in patients with COPD/OSA overlap syndrome. Potential use of pharmacologic therapy in overlap syndrome can therefore only be extrapolated from limited existing data about such treatment in OSA and COPD alone. There is in fact currently no established role for pharmacologic treatment of OSA alone, whereas patients with COPD alone receive pharmacologic treatment according to current recommendations.[1] The most common drugs currently prescribed in stable COPD, such as long-acting anticholinergics and long-acting beta agonists, have been shown to improve nocturnal O_2 saturation but not quality of sleep.[68,69] Theophylline, potentially useful for patients with COPD and SDB as a central respiratory stimulant with enhancement of the activity of the respiratory muscles,[70] is currently not clearly shown to be efficacious in improving COPD-related sleep breathing disorders or perturbed quality of sleep. Inhaled corticosteroids used in patients with stable COPD have not been specifically linked with either enhanced or decreased sleep continuity. Benzodiazepine sleep aids are typically avoided in patients with COPD and with OSA because of concerns that they may decrease the arousal response to hypercapnia, induce hypoventilation, and decrease upper airway muscle tone. There is evidence that nonbenzodiazepine hypnotics do not decrease respiratory drive and do not cause daytime drowsiness[71]; however, the indications for and contraindications to any type of sleep aid in these conditions, whether OSA or COPD alone or COPD/OSA overlap syndrome, remain to be better established.

The role of surgery in the treatment of COPD/OSA overlap as well as the need for special precautions regarding preoperative and postoperative evaluation and care in such patients undergoing surgery for treatment of their COPD or OSA, including lung transplantation, lung volume reduction, upper airway surgery, and bariatric surgery, also remains to be established.

ASTHMA

According to the Global Strategy for Asthma Management and Prevention of the Global Initiative for Asthma (GINA), asthma is defined by a history of respiratory symptoms such as wheeze, shortness of breath, chest tightness, and cough that vary over time and in intensity, together with variable expiratory airflow limitation.[72] It is typically recognized as a heterogeneous disease, usually characterized by chronic airway inflammation. Current asthma prevalence across all ages in the United States is 8.2%.[73]

Sleep has a deep impact on the morbidity and mortality of patients with asthma. Classical studies linked nocturnal asthma to an increased risk of mortality, with 70% of deaths and 80% of respiratory arrests caused by asthma occurring during nocturnal hours.[74] The normal physiologic changes that affect the lung during sleep and how those changes may contribute to nocturnal asthma have recently been reviewed in depth.[75]

Nocturnal asthma is characterized by coughing, wheezing, or dyspnea that interrupts and disturbs sleep, with such patients complaining of frequent arousals and poor sleep quality.[76] Nocturnal asthma generally indicates poor control of asthma and the need to modify overall asthmatic treatment.[72] Sleep studies done when current asthma treatment was not available showed a lower sleep efficiency, more awakenings, and less stage 3–4 sleep in asthmatics compared with nonasthmatic subjects.[77] Cognitive performance, as tested by psychometric testing, has also been shown to be impaired in patients with nocturnal asthma.[78] Circadian peak expiratory flow variation of 20% or more, a surrogate parameter of asthma instability, has been associated with poorer daytime cognitive performance compared with healthy control subjects.[79] Effective asthma treatment resulted in the recovery of cognitive impairment to a level of performance comparable to that of the healthy control subjects, paralleled by a reduction of circadian peak expiratory flow variation below 10% and by the resolution of nocturnal asthma symptoms.

Asthma and OSA Overlap Syndrome

Epidemiology

In adults, diseases with a known association with asthma include gastroesophageal reflux disease, rhinosinusitis, obesity, mental disorders, and OSA. The GINA initiative recommended investigation for the coexistence of OSA (i.e., asthma/OSA overlap syndrome) in all patients with asthma, especially in those with severe asthma, difficult to control asthma, and asthma with associated obesity.[72]

There are, however, few population-based data to identify the prevalence or severity of OSA in adult asthmatics. In a U.S. academic institution, patients in the asthma clinic and internal medicine clinic were surveyed for OSA risk with the Berlin Questionnaire, a validated instrument with a positive predictive value of 0.89.[80] OSA risk, as determined by the Berlin Questionnaire, was higher in the asthma group (39.5%) than in the internal medicine group (27.2%; $P = .004$). In a

Canadian cohort of patients with asthma, whose severity was established in accordance with the American Thoracic Society criteria,[81] OSA as defined by an AHI of 15 or more events per hour of sleep was present in 88% of patients with severe asthma, 58% of patients with moderate asthma, and 31% of controls without asthma.[82] From these limited data and in the absence of robust population-based studies, it appears that the prevalence of asthma/OSA overlap syndrome, defined as the coexistence of both entities in the same patient, likely is high in asthmatics, especially those with the most severe forms of asthma.

In patients with nocturnal asthma whose quality of sleep does not improve with proper antiasthma treatment, the coexistence of OSA (i.e., asthma/OSA overlap syndrome) should be excluded. Whereas there are no specific studies regarding the effect of coexistent OSA on the quality of sleep and daytime function in asthmatics, it is rational to expect an additive adverse effect of OSA on these outcomes.

Pathophysiology and Risk Factors for Asthma/OSA Overlap Syndrome

Numerous pathophysiologic and clinical factors contribute to an increased diathesis for OSA in patients with treated and untreated asthma; similarly, there are many pathophysiologic and clinical factors that increase the diathesis for asthma, including nocturnal asthma, in patients with OSA. Such factors, therefore, not only constitute risk factors for the presence and severity of asthma/OSA overlap syndrome but also thus represent the adverse clinical manifestations of asthma/COPD overlap. Obesity is a risk factor for both asthma and OSA and therefore for asthma/OSA overlap syndrome. There is a "dose response" effect of increasing BMI on increasing risk of incident asthma, especially in women.[83]

Many patients with nonatopic asthma and most atopic asthmatics suffer from nasal obstruction due to rhinitis and chronic sinusitis, which cause nasal congestion and airflow resistance, and nasopharyngeal polyps, which reduce airway caliber. These lead to increasing intrathoracic and pharyngeal negative pressure, which promotes upper airway collapse during inspiration, snoring, and obstructive apnea.[84] Similarly, in patients with chronic asthma, persistent mucosal inflammation affects the upper airway by decreasing cross-sectional area of the pharynx, promoting upper airway collapse.[85]

Inhaled corticosteroids are the most effective and most widely used drugs in asthma. Their long-term effects on the collapsibility of the pharynx remain unknown. However, the effects of oral corticosteroids on the upper airway are well known and are generally adverse, including myopathy of the muscles of the pharynx, fatty infiltration of the pharyngeal wall, and accumulation of liquid in the neck. In asthma clinics, asthmatics requiring frequent bursts or consistent use of oral corticosteroids were found to have a high prevalence of OSA (>90%) after adjustment for BMI and neck circumference.[86]

Factors potentially effective in reducing the risk and severity of OSA in patients with asthma include the same factors as in the case of patients with COPD/OSA overlap: weight loss, sleep in the lateral decubitus position, and smoking cessation. The effect of adjusting asthma medications to improve concomitant OSA has not been studied. In nonasthmatic patients with OSA, there is both molecular and clinical evidence of the ability of inhaled corticosteroids to reduce upper airway inflammation and to improve AHI in a subgroup of patients with concomitant allergic rhinitis.[87] In clinical practice, nasal inhaled corticosteroids and oral antileukotrienes may be beneficial for reducing snoring and obstructive apneas in children with asthma and OSA, but such an effect has not been proved in adults. These noted factors, both predisposing to and protective against OSA in patients with asthma, are shown in schematic form in Figure 119-5.

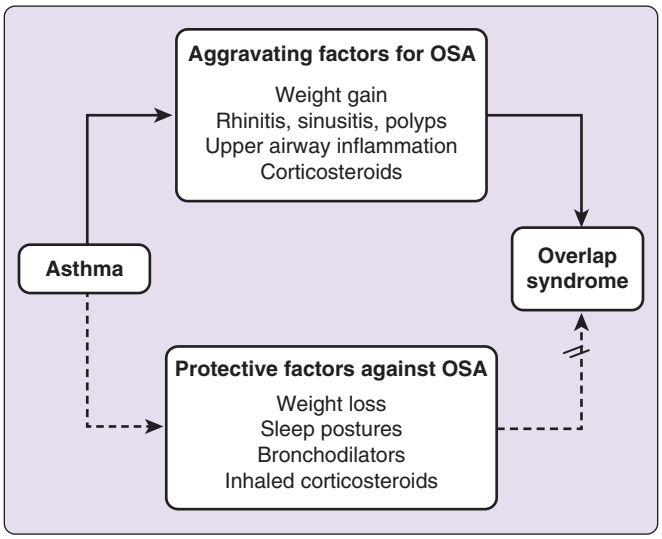


Figure 119-5 Interactions between asthma and obstructive sleep apnea (OSA) contributing to asthma/OSA overlap syndrome.

The potential mechanisms by which OSA may worsen asthma are also multifactorial. Obstructive apneic episodes are associated with repetitive arousals from sleep, perturbations in autonomic activity, and intermittent hypoxemia.[43] Increased vagal tone during obstructive apnea episodes can contribute to nocturnal asthma through stimulation of muscarinic receptors of the central and upper airways. Negative intrathoracic pressure during obstructive events leads to intermittent loss of lower esophageal sphincter tone; associated gastroesophageal reflux is associated with bronchial microaspiration of gastric acid, potentially promoting nocturnal asthma.[88] By stimulation of carotid body receptors, intermittent hypoxia can enhance bronchial responsiveness through vagal pathways.[89] Chronic intermittent hypoxia in OSA may also induce a low-grade systemic inflammation characterized by the elevation of serum proinflammatory cytokines and chemokines. Local inflammatory changes of the upper airways similar to those noted in asthma are also prominent in OSA. Such inflammatory changes may reduce airway caliber and at the same time increase underlying bronchial hyperresponsiveness, thus representing a potential asthma trigger.

Clinical Outcomes and Treatment in Asthma/Overlap Syndrome

In contrast to COPD/OSA overlap syndrome, there are no long-term studies that have evaluated asthma outcomes, either nocturnal or awake, among patients with comorbid untreated OSA or OSA outcomes in OSA patients with comorbid asthma. Consequently, there are currently no guidelines specific to the management of asthma/OSA overlap syndrome. However, in asthma/COPD overlap patients, OSA treatment with CPAP has important potential pathophysiologic

beneficial effects for asthmatics, including reducing gastroesophageal reflux, airway and systemic inflammation, and airway smooth muscle contractility.[89] Therefore, CPAP appears to have significant potential clinical benefit in the treatment of the asthma/OSA overlap syndrome, and data do exist documenting that CPAP treatment for comorbid OSA improves asthma symptoms, decreases use of rescue medication, and improves asthma-specific quality of life.[90-93] Further, in a short-term randomized trial, CPAP use decreased airway reactivity in asthmatics without OSA, possibly through reducing bronchial inflammation.[94] Longer term studies are needed to determine the optimal application of CPAP in patients with and without asthma/OSA overlap in improving asthma symptoms, medication need, and overall cardiorespiratory and quality of life outcomes.

Second-line treatments for OSA, such as mandibular advancement devices and upper airway surgery, have not been prospectively evaluated in patients with asthma/OSA overlap. However, bariatric surgery for patients with OSA and morbid obesity may be effective not only for OSA resolution but also for improving asthma.[95] There are no studies assessing clinical outcomes related to the use of asthma medications in asthma/OSA overlap syndrome. At this time, therefore, it appears that asthma in patients with OSA should be treated according to current asthma treatment guidelines[72] in addition to optimizing treatment of the comorbid OSA.

INTERSTITIAL LUNG DISEASE

Diffuse parenchymal lung disease, also known as interstitial lung disease (ILD), represents more than 200 nonmalignant, noninfectious entities characterized by inflammatory and fibrotic changes affecting alveolar and air spaces. ILD is characterized by progressive dyspnea, hypoxemia, and restrictive-ventilatory limitation. It should be suspected in the appropriate clinical context when there is evidence of impairment of gas exchange or restrictive lung function deficit. Confirmation comes when there is a pattern of usual interstitial pneumonia on lung biopsy or on computed tomography scan.[96] The incidence of ILD appears to be increasing mainly because of improvements in the ability to diagnose the condition due to advances in chest imaging.[97]

Sleep is often disturbed among patients with ILD, which contributes to daytime fatigue in this population.[98] Compared with control subjects, Perez-Padilla et al[99] reported worse sleep quality in patients with ILD, with more time in stage N1 (33.7% of total sleep time versus 13.5%), less time in REM sleep (11.8% versus 19.9% of total sleep time), and more fragmentation of sleep. In this study, patients with awake hypoxemia (Sao_2 < 90%) had greater abnormalities in sleep structure than did those with Sao_2 above 90%. Potential mechanisms that contribute to sleep fragmentation include hypoxemia, hypercapnia, and cough. Most patients report nocturnal cough as an important cause of nocturnal awakenings. Esophageal dysmotility and reflux, also prevalent in ILD, and the pulmonary fibrotic process itself are the main intermediate mechanisms that explain nocturnal cough.[100]

Hypoxemia during sleep is also common and tends to be worse in those with more severe daytime hypoxemia.[99] During REM sleep, O_2 desaturation is often more severe than that occurring during exercise.[101] The role of nocturnal desaturation on health outcomes in patients with ILD has been evaluated retrospectively in a large cohort of patients with ILD.[102] In this study, desaturation index was defined as the number of desaturation events above 4%/hour. Desaturation was present in 37% of patients, and 31% of them had PH on echocardiography. Increased desaturation index was associated with higher mortality independent of age, gender, BMI, and PH. These data indicate the need to conduct sleep studies in patients with ILD. If nocturnal hypoxemia is detected, it is reasonable to treat these patients with oxygen therapy, at least until randomized trials are available.

ILD and Obstructive Sleep Apnea Overlap Syndrome

OSA is prevalent in ILD but clearly underrecognized. In a sample of 50 patients with stable ILD, OSA was confirmed with PSG in 88%.[103] Of those, 68% were moderate to severe (AHI >15 events per hour of sleep). It appears that severity of OSA, as indicated by AHI, inversely correlates with total lung capacity and, interestingly, poorly correlates with BMI.[103] The mechanistic relationship between OSA and ILD and the impact of comorbid OSA on the natural history of ILD remain unknown. Nevertheless, clinicians should evaluate the potential coexistence of SDB in patients with ILD as the appropriate treatment can improve the patient's quality of life and may improve survival.

PULMONARY HYPERTENSION

PH is defined as a mean pulmonary artery pressure higher than 25 mm Hg.[104] The current classification of PH consists of five categories: (1) primary pulmonary arterial hypertension (PAH); (2) PH due to left-sided heart disease; (3) PH associated with chronic pulmonary diseases, such as COPD and OSA; (4) chronic thromboembolic PH; and (5) PH due to various disorders, such as sarcoidosis or systemic vasculitis.[104] The mechanisms by which OSA can lead to the development of PH as a potential long-term complication are reviewed in Chapter 127.

Pulmonary Hypertension and Sleep-Disordered Breathing Overlap Syndrome

SDB overall can be considered a spectrum of ventilatory disorders during sleep that include OSA, central sleep apnea, and sleep-related hypoventilation. The prevalence of SDB/PH overlap is not known. One study conducted to determine the prevalence and significance of nocturnal oxygen desaturation in patients with PH, using home oximetry studies, showed that 69.7% of patients spent more than 10% of sleep time with Sao_2 below 90%.[105] Nocturnal hypoxemia correlates with advanced PH and right ventricular dysfunction. Interestingly, 60% of this subgroup with nocturnal hypoxemia had no exertional hypoxemia. In a small study of patients with idiopathic PH who had full PSG, Schulz et al[106] found that 30% of these patients had periodic breathing, defined as a crescendo-decrescendo pattern of hyperventilatory phases alternating with central apneas or hypopneas of at least three consecutive cycles. Most of these patients, however, had normal nocturnal oximetry. In another study of 38 PH patients who had ambulatory cardiorespiratory sleep studies, 45% had 10 or more apnea-hypopnea events per hour.[107] A subgroup of 22 patients also had in-laboratory PSG. Among patients who underwent both studies, home sleep studies accurately predicted an AHI of 10 events or more during PSG (area under the

receiver operating characteristic curve, 0.93; $P = .002$). The corresponding value for pulse oximetry was 0.63 (P = not significant). Therefore, when SDB is suspected among patients with PH, evaluation should include full PSG or modified cardiorespiratory sleep studies rather than pulse oximetry alone.

In the largest series of patients with confirmed PH by right-sided heart catheterization,[108] home cardiorespiratory sleep studies demonstrated that among 169 patients, 26.6% had an AHI above 10 events. Of these, 27 patients (16%) had OSA and 18 patients (10.6%) had central sleep apnea. Despite these limited data, it seems that the prevalence of SDB/PH overlap appears to be higher in PH patients than in the general population.

The extent to which SDB contributes to the PH patient's symptoms and disease progression is unclear. In a retrospective review of 52 consecutive patients with PAH referred for assessment of possible SDB, 71% had SDB (56% of these had primarily OSA and 44% primarily central sleep apnea).[109] There were no differences in cardiopulmonary hemodynamics at baseline assessed by right-sided heart catheterization between patients with PAH only and those with SDB/PH overlap. After a median follow-up of 4.7 years, no differences in survival between those with and without SDB were observed. In this study and in the other studies commented on before,[105,107] there was a lack of subjective daytime sleepiness as assessed by the Epworth Sleepiness Scale in the PAH population with or without coexistent SDB, similar to that found in patients with heart failure with and without SDB; such a phenomenon could be explained by elevated sympathetic nervous activity in both heart failure and PH[110,111] that can act as an adrenergic cortical alerting mechanism. Predictors of SDB in patients with PH do not differ from those of the general population, being mainly older age and BMI.[105-109]

Coexistent OSA may contribute to worsening of underlying PAH. During obstructive events, negative intrathoracic pressures result in right ventricular overload.[112] This is aggravated by intermittent hypoxia and elevated sympathetic nervous activity. Together, these mechanisms contribute to right ventricular hypertrophy and ultimately cardiac failure. Conversely, patients are more predisposed to development of OSA if they accumulate fluid in the neck during sleep.[113] Such rostral shift of fluid from the legs during the daytime to the neck at night has been demonstrated in patients with left ventricular failure[114] (Figure 119-6).

Because of the absence of long-term studies, the prognosis of SDB/PH overlap remains unknown. Similarly, no studies have systematically evaluated the effect of SDB treatment on SDB/PH overlap syndrome outcomes. However, the presence of PH may have prognostic importance in patients with OSA; for example, an observational study of 83 patients with OSA (AHI > 5) who underwent pulmonary artery catheterization for unspecified reasons documented 1-, 4-, and 8-year survival rates that were lower among patients with PH (mean pulmonary artery pressure > 25 mm Hg at rest) than among those without PH.[115]

We believe that at the present time, patients with PH should be evaluated by PSG when presenting with symptoms that suggest the coexistence of SDB, such as daytime hypoxemia or heart failure. Treatment with CPAP, supplemental oxygen, or both is ideally titrated in the sleep laboratory to customize the treatment as in the other overlap syndromes described before.

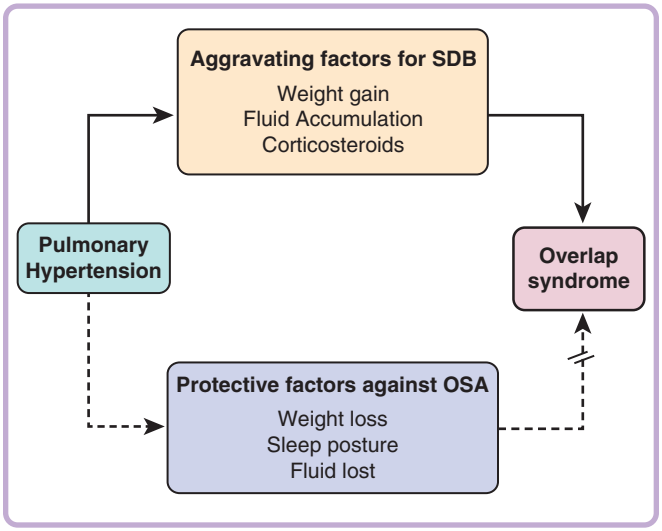


Figure 119-6 Interactions between pulmonary hypertension and sleep disordered breathing (SDB).

INSOMNIA IN PULMONARY DISEASES

Chronic insomnia is a major health problem that leads to worse quality of life and decreased productivity.[116] It is estimated that approximately 10% of the general population is insomniac.[117] There has been little interest in the study of insomnia in pulmonary disorders. Specifically, there are no data on the prevalence and burden of insomnia in patients with PH or ILD. In COPD, relatively older studies reported a high prevalence of self-reported insomnia compared with non-COPD subjects that appears related to the severity of the respiratory symptoms.[118] According to the American Academy of Sleep Medicine, insomnia is defined as history of frequent difficulty in initiating or maintaining sleep and significant disruption of daytime functioning for at least 1 month.[119] Recently, using the American Academy of Sleep Medicine criteria, Budhiraja et al[120] interviewed 183 patients with COPD about sleep complaints. Insomnia was present in 27.3% of participants. Severity of COPD as assessed by pulmonary function test (FEV_1 < 50% predicted) or by the Medical Research Council dyspnea scale was not different among participants with insomnia or without insomnia. Interestingly, the presence of insomnia was associated with increased daytime sleepiness and worse quality of life. There are no studies that have evaluated the causality of factors associated with insomnia in patients with COPD or that have evaluated the insomnia as a determinant of health outcomes in COPD.

Because of this high prevalence of insomnia, it is common that COPD patients request medicines to improve their quality of sleep. Benzodiazepines are prescribed for non-COPD insomniacs because they shorten sleep latency, improve sleep efficiency, and decrease arousal frequency. Nevertheless, these agents should be avoided if possible in COPD patients because they reduce alveolar ventilation, diminish arousal response, and increase apnea frequency, and therefore they can worsen hypoxemia and hypercapnia.[121]

Some nonbenzodiazepine hypnotics, such as zolpidem,[122] and melatonin receptor antagonists, such as ramelteon,[123] have been reported to have no adverse effects on gas exchange in patients with COPD. Recently, the safety profile of suvorexant, an orexin receptor antagonist approved for treatment of insomnia in the United States, was evaluated in COPD. Suvorexant, at up to twice the maximum recommended dose, did not cause SBDs in a multicenter, randomized, double-blind, placebo-controlled, crossover study in patients with mild to moderate COPD.[124] At the present time and in the absence of comparative studies, we consider that physicians caring for COPD patients with insomnia should preferably use nonbenzodiazepine hypnotics with which they are most familiar.

Sleep disturbances in patients with asthma relate to the occurrence of nocturnal asthmatic crisis. The prevalence of insomnia symptoms was significantly higher among asthmatics than among nonasthmatics (47.3% versus 37.2%) in a postal questionnaire sent to a random sample of 45,000 adults in Sweden.[125] In this study, the risk of insomnia increased with the severity of asthma. In another recent online survey of adolescents from the general community, it was reported that almost twice as many adolescents with severe asthma had clinically significant insomnia than adolescents with mild or no asthma.[126] Daytime sleepiness was frequent in this population, and 28% of its variance was accounted for by insomnia severity, whereas only 2% was accounted for by asthma severity. Insomnia remains a common problem among asthmatics that should be addressed in any patient as part of his or her comprehensive treatment.

CLINICAL PEARLS

- Overlap of OSA with COPD, the COPD/OSA overlap syndrome, affects more than 1% of adults. The prevalence of OSA overlap with asthma, the asthma/OSA overlap syndrome, and the prevalence of SDB overlap with PH, the SDB/PH overlap syndrome, are not well defined. Nevertheless, the coexistence of OSA in asthma and PH increases with increasing severity of both pulmonary disorders.
- Obesity increases the risk of OSA in both COPD and asthma populations.
- Untreated OSA in OSA/COPD overlap is associated with worsened clinical outcomes for both the OSA and the comorbid pulmonary disorder. Conversely, effective identification and treatment of OSA reduce diurnal and nocturnal symptoms and improve clinical outcomes in patients with OSA/COPD.

SUMMARY

OSA and COPD, each a prevalent and clinically important condition in adults, carry numerous common risk factors, including obesity and smoking. It is estimated that the coexistence of OSA and COPD, the COPD/OSA overlap syndrome, affects more than 1% of the general population. The presence of such overlap, when the OSA is untreated, carries a risk of more adverse diurnal and nocturnal physiologic and clinical outcomes, including greater sleep fragmentation, more severe nocturnal hypoxemia, and increased overall mortality, than is documented for COPD alone and OSA alone. Effective identification and treatment of the comorbid OSA and the other features of SDB in the COPD/OSA overlap syndrome improve overall clinical outcomes in the condition.

Asthma, ILD, and PH are also linked with OSA and, in the case of PH, other types of SDB, such as central sleep apnea and sleep-related hypoventilation, by common risk factors and mutually exacerbating pathophysiologic and clinical features. The prevalence of asthma overlap with OSA and PH overlap with SDB is not well defined but increases as the severity of both asthma and PH increases. As with COPD and OSA overlap, effective treatment of the comorbid OSA, using well-established therapy with CPAP, improves asthma-related and overall pathophysiologic and clinical outcomes of the asthma/OSA overlap syndrome, including airway and systemic inflammation, asthma control, and asthma-specific quality of life.

Selected Readings

Flenley DC. Sleep in chronic obstructive lung disease. *Clin Chest Med* 1985;**6**:651–61.

Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med* 2014;**370**:2276–85.

Greenberg H, Cohen RI. Nocturnal asthma. *Curr Opin Pulm Med* 2012;**18**:57–62.

Marin JM, Soriano JB, Carrizo SJ, et al. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. *Am J Respir Crit Care Med* 2010;**182**:325–31.

Mohsenin V. Obstructive sleep apnea: a new preventive and therapeutic target for stroke: a new kid on the block. *Am J Med* 2015;**128**:811–16.

Montplaisir J, Walsh J, Malo JL. Nocturnal asthma: features of attacks, sleep and breathing patterns. *Am Rev Respir Dis* 1982;**125**:18–22.

Mulloy E, McNicholas WT. Ventilation and gas exchange during sleep and exercise in severe COPD. *Chest* 1996;**109**:387–94.

Soler X, Gaio E, Powell FL, et al. High prevalence of obstructive sleep apnea in patients with moderate to severe chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2015;**12**(8):1219–25.

Teodorescu M, Broytman O, Curran-Everett D, et al; National Institutes of Health, National Heart, Lung and Blood Institute Severe Asthma Research Program (SARP) Investigators. Obstructive sleep apnea Risk, asthma burden, and lower airway inflammation in adults in the Severe Asthma Research Program (SARP) II. *J Allergy Clin Immunol Pract* 2015;**3**(4):566–75.

Vestbo J, Hurd SS, Agusti AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;**187**:347–65.

A complete reference list can be found online at ExpertConsult.com.

Chapter 120

Obesity-Hypoventilation Syndrome

Babak Mokhlesi

Chapter Highlights

- Obesity-hypoventilation syndrome (OHS) has been conventionally and to some extent arbitrarily defined by the combination of obesity and daytime hypercapnia during wakefulness occurring in the absence of an alternative neuromuscular, mechanical, or metabolic explanation for hypoventilation. This syndrome is also invariably accompanied by sleep disordered breathing (e.g., obstructive sleep apnea or sleep hypoventilation), and therefore sleep disordered breathing is included as one of the diagnostic criteria in some definitions of OHS.
- During the last 3 decades, the prevalence of extreme obesity has markedly increased in the United States and other countries. With such a global epidemic of obesity, the prevalence of OHS is likely to increase.
- Patients with OHS have a lower quality of life with increased health care expenses and are at higher risk for development of pulmonary hypertension and early mortality due to cardiopulmonary complications compared with eucapnic patients with obstructive sleep apnea.
- OHS often remains undiagnosed until late in the course of the disease. Early recognition is important as these patients have significant morbidity and mortality if they are left untreated. Effective treatment can lead to significant improvement in patient outcomes, underscoring the importance of early diagnosis.

HISTORICAL PERSPECTIVE

The association between obesity and hypersomnolence has long been recognized. Of historical interest, obesity-hypoventilation syndrome (OHS) was described well before obstructive sleep apnea (OSA) was recognized in 1969.[1-3] In 1955, Auchincloss et al[4] described in detail a case of obesity and hypersomnolence paired with alveolar hypoventilation. One year later, Bickelmann et al[5] described a similar patient who finally sought treatment after his symptoms caused him to fall asleep during a hand of poker, despite having been dealt a full house of aces over kings. Although other clinicians had made the comparison some 50 years earlier,[6] Bickelmann popularized the term *Pickwickian syndrome* in his case report by noting the similarities between his patient and the boy Joe (Figure 120-1), Mr. Wardle's servant in Charles Dickens' *The Posthumous Papers of the Pickwick Club*.

DEFINITION

OHS has been conventionally and to some extent arbitrarily defined by the combination of obesity (body mass index [BMI] ≥ 30 kg/m^2) and daytime hypercapnia (partial pressure of arterial CO_2 [$Paco_2$] ≥ 45 mm Hg at sea level) during wakefulness occurring in the absence of an alternative neuromuscular, mechanical, or metabolic explanation for hypoventilation. This syndrome is also invariably accompanied by a sleep breathing disorder (SBD), and therefore SBD is included as one of the diagnostic criteria in some expert definitions of OHS.[7] Approximately 90% of patients with OHS have OSA, defined by an apnea-hypopnea index (AHI) of 5 events/hour or more. The remaining patients have nonobstructive sleep hypoventilation. The American Academy of Sleep Medicine has arbitrarily defined sleep hypoventilation in adults by the following criteria: the $Paco_2$ (or surrogate, such as end-tidal CO_2 or transcutaneous CO_2) is above 55 mm Hg for more than 10 minutes or there is an increase in the $Paco_2$ (or surrogate) above 10 mm Hg (compared with an awake supine value) to a value exceeding 50 mm Hg for more than 10 minutes.[8] This point is relevant because although the definition suggests a diurnal pathologic process, overnight polysomnography is required to determine the pattern of nocturnal SBD including hypoventilation (obstructive or nonobstructive) and to individualize therapy, particularly the optimal mode of positive airway pressure (PAP).

OHS is a diagnosis of exclusion and should be distinguished from other conditions that are commonly associated with awake hypercapnia (Box 120-1).

EPIDEMIOLOGY

Nearly 1 of 3 adults in the world are overweight (BMI ≥ 25 kg/m^2), and almost 1 in 10 adults are obese (BMI ≥ 30 kg/m^2). This "obesity epidemic" is associated with myriad comorbidities including OHS. Between 1986 and 2005, the prevalence of morbid obesity (BMI ≥ 40 kg/m^2) increased by fivefold in the United States, affecting 1 in every 33 adults. Similarly, the prevalence of BMI of 50 kg/m^2 and higher has

Figure 120-1 Joe the "fat boy." (Detail from "Mr. Pickwick in Chase of His Hat." Illustration by Robert Seymour. In: Dickens C. *The posthumous papers of the Pickwick Club*. Published in serial form. London: Chapman and Hall; 1836. Courtesy The Beinecke Rare Book & Manuscript Library, Yale University.)

Box 120-1 DIAGNOSTIC FEATURES OF OBESITY-HYPOVENTILATION SYNDROME

Obesity

Body mass index ≥ 30 kg/m^2

Chronic Hypoventilation

Awake daytime hypercapnia (sea-level arterial Pco_2 ≥ 45 mm Hg)

Possible role of serum venous bicarbonate or calculated bicarbonate >27 mEq/L from capillary blood gas

Sleep Breathing Disorder

Obstructive sleep apnea (apnea-hypopnea index [AHI] ≥ 5 events/hour)

Nonobstructive sleep hypoventilation (AHI <5 events/hour, $Paco_2$ above 55 mm Hg for more than 10 minutes or an increase in the $Paco_2$ [or surrogate] above 10 mm Hg [compared with awake supine $Paco_2$] to a value >50 mm Hg for >10 minutes during sleep, or sustained hypoxemia with oxygen saturation ≤88% without obstructive respiratory events)

Exclusion of Other Causes of Hypoventilation

Severe obstructive airways disease (e.g., chronic obstructive pulmonary disease)

Severe interstitial lung disease

Severe chest wall disorders (e.g., kyphoscoliosis)

Severe hypothyroidism

Neuromuscular disease

Congenital hypoventilation syndromes

increased by 10-fold in the United States, affecting 1 in every 230 adults.[9] With such epidemic obesity, the prevalence of OHS is likely to increase.

Thirteen studies have reported a prevalence of OHS between 8% and 20% in patients referred to sleep centers for evaluation of SBD.[10-12] A meta-analysis of 4250 outpatients with obesity and OSA (mean BMI range between 30 and 44 kg/m^2 and mean AHI range between 40 and 60 events/hour) who did not have chronic obstructive pulmonary disease reported a 19% prevalence of awake hypercapnia.[13] On the basis of these data, approximately 19% of obese patients with OSA have OHS. East Asian populations are known to have OSA at a lower BMI compared with other populations, probably because of cephalometric differences.[14] Therefore, in these populations, OHS may be more prevalent at a lower BMI range than in non-Asian populations.[11,14-16] The prevalence of obesity-associated hypoventilation among consecutive patients with BMI higher than 35 kg/m^2 hospitalized on medical wards (excluding critical care units) has been reported to be 31%.[17] Although it remains unclear why the prevalence of obesity-associated hypoventilation in this hospitalized cohort was higher than the reported prevalence in outpatient obese patients with OSA, it may be related to the facts that the investigators enrolled subjects with a higher BMI (>35 kg/m^2 as opposed to >30 kg/m^2) and there was high prevalence of diuretic use (64% of the patients).

Prevalence estimates for OHS vary significantly across studies, owing partly to differences in sample characteristics, disease definitions, and assessment procedures.[10] In populations of patients with concomitant OSA, as the degree of obesity increases, the prevalence of OHS increases (Figure 120-2).[10] Laaban and Chailleux[18] reported an OHS prevalence of 11% in a cohort of 1141 patients with OSA with a mean BMI of 30 kg/m^2, whereas Mokhlesi et al[19] reported a prevalence of 24% in patients with OSA and a mean BMI of 44 kg/m^2. Among non-Asian populations, the prevalence of OHS is 8% to 11% among patients with OSA with BMI of 30 to 35 kg/m^2 and increases to 18% to 31% among patients with OSA with BMI of 40 kg/m^2 and higher.[18-21] The prevalence of OHS in the general population is unknown but can be estimated. The most recent report from the Centers for Disease Control and Prevention has estimated that approximately 6.4% of the general U.S. adult population has morbid or severe obesity (BMI ≥ 40 kg/m^2), and the prevalence is substantially higher at 12.2% amongst non-Hispanic blacks.[22] If we conservatively estimate that half of patients with this degree of obesity have OSA and that approximately 20% of these OSA patients have OHS, the prevalence of OHS can be estimated as roughly 0.6% (approximately 1 in 160 adults in the U.S. population). OHS may be more prevalent in the United States than in other nations because of its obesity epidemic. With such an epidemic, the prevalence of OHS is likely to increase, and therefore there is a need for a high index of suspicion on the part of clinicians to optimize early recognition and treatment of this syndrome.

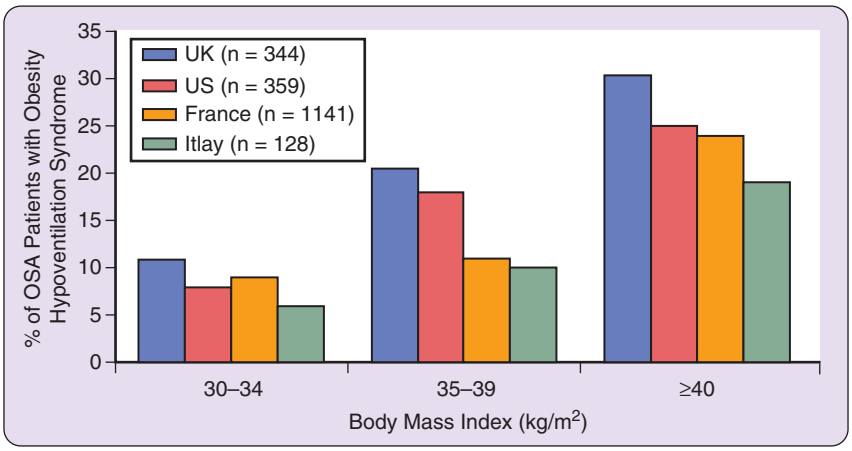


Figure 120-2 Prevalence of obesity-hypoventilation syndrome in patients with obstructive sleep apnea (OSA), sorted by body mass index (BMI). In the U.K. study,[21] the mean BMI was nearly 40 kg/m^2, and 38% of subjects had a BMI higher than 40 kg/m^2. Similarly, in the U.S. study,[19] the mean BMI was 43 kg/m^2, and 60% of subjects had a BMI higher than 40 kg/m^2. In contrast, the mean BMI in the French study[18] was 34 kg/m^2, and only 15% of subjects had a BMI higher than 40 kg/m^2. Italian data[30] were provided by Professor Onofrio Resta (personal communication).

CLINICAL PRESENTATION AND DIAGNOSIS

OHS is typically diagnosed either when an afflicted patient reaches a high state of acuity, in the form of acute-on-chronic hypercapnic respiratory failure,[23] or, alternatively, when ambulatory care is escalated to include evaluation by pulmonary or sleep specialists.[20] Unfortunately, a delay in diagnosis is common; the diagnosis typically occurs during the fifth and sixth decades of life, and during this delay, OHS patients use more health care resources than comparably obese normocapnic patients.[17,23-25] In one study, 8% of all admissions to a general intensive care unit met diagnostic criteria for obesity-associated hypoventilation (BMI > 40 kg/m^2; $Paco_2$ > 45 mm Hg; and no evidence of musculoskeletal disease, intrinsic lung disease, or smoking history). All of these patients presented with acute-on-chronic hypercapnic respiratory failure.[26] Of these patients, nearly 75% were misdiagnosed and treated for obstructive lung disease (most commonly chronic obstructive pulmonary disease) despite having no evidence of obstructive physiology on pulmonary function testing.

Patients with OHS tend to be morbidly obese (BMI ≥ 40 kg/m^2), have severe OSA (≥30 obstructive respiratory events/hour of sleep), and are typically hypersomnolent. Compared with patients with eucapnic OSA and similar BMI, patients with OHS are more likely to report dyspnea and to manifest cor pulmonale. Box 120-2 provides the typical portrait of an OHS patient based on the clinical features of a large combined cohort of OHS patients reported in the literature.[16-19,25,27-35] Whereas severe obesity (BMI ≥ 40 kg/m^2) is a predominant risk factor for OHS, not all patients with severe obesity develop OHS. There are significant physiologic differences between obese patients who have OHS and similarly obese patients without OHS as summarized in Box 120-3.[36]

Although the definitive test for alveolar hypoventilation is a room air arterial blood gas analysis, an elevated serum bicarbonate level due to metabolic compensation of respiratory acidosis is supportive of OHS.[19] Mokhlesi et al[19] first demonstrated that a venous serum bicarbonate threshold of 27 mEq/L, suggestive of chronic respiratory acidosis, could be used for OHS diagnosis in obese patients with diagnosed OSA. Their data demonstrated that among obese patients with OSA and normal renal function, serum bicarbonate level below 27 mEq/L had a 97% negative predictive value for excluding a diagnosis of OHS. Macavei et al[21] assessed ear

Box 120-2 CLINICAL FEATURES OF PATIENTS WITH OBESITY HYPOVENTILATION SYNDROME*

Clinical Features	Mean (Range)
Age (years)	52 (42–61)
Male (%)	60 (49–90)
Body mass index (kg/m^2)	44 (35–56)
Neck circumference (cm)	46.5 (45–47)
pH	7.38 (7.34–7.40)
Arterial PCO_2 (mm Hg)	53 (47–61)
Arterial PO_2 (mm Hg)	56 (46–74)
Serum bicarbonate (mEq/L)	32 (31–33)
Hemoglobin (g/dL)	15
Apnea-hypopnea index	66 (20–100)
SpO_2 nadir during sleep (%)	65 (59–76)
Percent sleep time SpO_2 < 90%	50 (46–56)
FVC (% predicted)	68 (57–102)
FEV_1 (% predicted)	64 (53–92)
FEV_1/FVC	0.77 (0.74–0.88)
Medical Research Council dyspnea class 3 or 4 (%)	69
Epworth sleepiness scale score	14 (12–16)

*Features are based on aggregated sample of 757 patients from 15 studies.[20]

Box 120-3 PHYSIOLOGIC DIFFERENCES BETWEEN EUCAPNIC MORBIDLY OBESE PATIENTS AND THOSE WITH OBESITY-HYPOVENTILATION SYNDROME

	Eucapnic Morbid Obesity	Obesity-Hypoventilation Syndrome
Waist:hip ratio	↑	↑↑
FEV_1/FVC	Normal	Normal/↓
Total lung capacity	Normal	Slight ↓
Functional residual capacity	↓	↓
Vital capacity	Normal or ↓	↓↓
Expiratory reserve volume	↓	↓↓
Work of breathing	↑	↑↑
Hypercapnic/hypoxic ventilatory drive	Normal	↓
Inspiratory muscle strength	Normal	↓

FEV_1, Forced expiratory volume in first second; FVC, forced vital capacity.

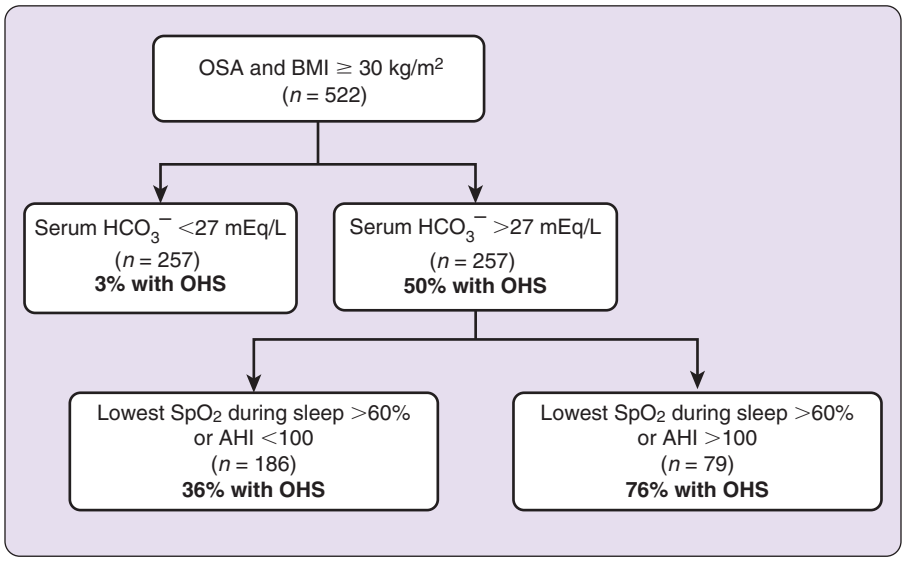


Figure 120-3 Decision tree to screen for obesity-hypoventilation syndrome (OHS) based on observation in 522 obese patients with OSA (BMI ≥ 30 kg/m² and AHI ≥ 5). Among those with a venous serum bicarbonate level above 27 mEq/L, OHS was present in 50% of patients. Very severe OSA (AHI > 100 events/hour or Spo_2 nadir during sleep < 60%) increased the prevalence of OHS to 76%.[19] OSA, Obstructive sleep apnea; AHI, apnea-hypopnea index; BMI, body mass index.

lobe capillary blood gas samples from patients referred to a sleep center and determined that bicarbonate values calculated from the Henderson-Hasselbalch formula have similar predictive values. A calculated serum bicarbonate level of 27 mEq/L and higher had a sensitivity of 85% and a specificity of 89% for the diagnosis of OHS among their patient sample. Two additional studies have confirmed serum bicarbonate to be an independent and reliable predictor of OHS.[12,37] Figure 120-3 shows the prevalence of OHS in obese patients with OSA (BMI ≥ 30 kg/m² and AHI ≥ 5) using a serum bicarbonate level combined with other readily available measures, such as severity of OSA.[19] Indeed, several investigators have suggested incorporating serum venous bicarbonate (HCO_3^-) levels into the definition of OHS, particularly because using a single measurement of arterial Pco_2 for OHS diagnosis is susceptible to a number of confounding factors, including the impact of the patient's periprocedural anxiety leading to hyperventilation.[38]

In addition to blood gas sampling and serum venous bicarbonate assessments, daytime finger pulse oximetry (Spo_2) may be a valuable tool for clinicians in screening for possible OHS.[39] Resting hypoxemia during wakefulness is not a typical feature of either patients with OSA or patients with obesity. Therefore, abnormal resting pulse oximetry during wakefulness should increase the suspicion for OHS among obese OSA patients.[19,40,41] Similarly, significant sleep-associated hypoxemia, defined as oxygen saturation below 85% for more than 10 continuous minutes, in an obese patient with OSA should raise suspicion for presence of sleep hypoventilation and possibly OHS.[42] In a meta-analysis, the mean difference of percentage of total sleep time with Spo_2 spent below 90% was 37.4% (56.2% for OHS, 18.8% for eucapnic obese OSA

patients) with very little overlap in the 95% confidence intervals.[13]

Ultimately, a rise in carbon dioxide levels (≥45 mm Hg) during wakefulness is necessary to define hypoventilation. There are a variety of techniques to measure carbon dioxide, such as daytime arterial blood gases, arterialized capillary blood gases, venous blood gases, and end-tidal carbon dioxide and transcutaneous carbon dioxide monitoring. Each of these techniques has its advantages and disadvantages.[43,44] The most reliable and practical method for identifying sleep hypoventilation is to measure carbon dioxide levels continuously during sleep by end-tidal or transcutaneous monitoring.[41] Improving technologies should greatly expand our ability to identify and to quantify nocturnal hypoventilation in sleep laboratories or even at home.

MORBIDITY AND MORTALITY

The majority of OHS patients are severely obese and have severe OSA.[18] Although severe obesity[45] and severe OSA are independently associated with increased risk of mortality,[46-50] OHS may contribute further.[51] A retrospective study reported that 7 of 15 patients with OHS (46%) who refused long-term noninvasive PAP therapy died during an average of 50 months of follow-up.[32] A prospective study by Nowbar et al[17] observed a group of 47 severely obese patients after hospital discharge. The 18-month mortality rate for patients with untreated OHS was higher than for the control cohort of 103 patients with obesity alone (23% versus 9%) despite the fact that the groups had similar BMI, age, and number of comorbid conditions. When adjusted for age, sex, BMI, and renal function, the hazard ratio of death in the OHS group was 4.0 in the 18-month period. Only 13% of the 47 patients were treated for OHS after hospital discharge. The difference in survival was evident as early as 3 months after hospital discharge. In contrast, Budweiser et al[34] conducted a retrospective analysis of 126 patients with OHS who were highly adherent to noninvasive ventilation (NIV) during sleep, with the NIV modality initiated in pressure support mode and after an adaptation period switched to pressure-cycled assist control mode, finding the 1-, 2-, and 5-year survival rates to be 97%, 92%, and 70%, respectively. Similarly, in a large retrospective study, 110 patients with OHS treated with NIV in the form of bilevel PAP (mean inspiratory PAP of 18.5 ± 2.5 cm H_2O and mean expiratory PAP of 8.4 ± 1.9 cm H_2O) were matched with 220 patients with OSA treated with continuous PAP (CPAP; mean pressure of 8.9 ± 1.7 cm H_2O).[51] Despite similar rates of adherence to PAP therapy (mean bilevel PAP use of 6.2 ± 3.0 hours/night vs. mean CPAP use of 5.8 ± 3.2 hours/night; $P = .29$), the 5-year mortality rates were 15.5% in the OHS cohort and 4.5% in the OSA cohort ($P < .05$). Patients with OHS had a twofold increase (odds ratio, 2; 95% confidence interval, 1.11-3.60) in the risk of mortality compared with those with OSA. Using bilevel PAP less than 4 hours/night emerged as the strongest independent predictor of mortality in patients with OHS.[51] Together, these studies suggest that treatment with NIV may lower the short-term mortality of patients with OHS (Figure 120-4).[7,34] Accumulating evidence from prospective cohort studies suggests that long-term survival may be better in OHS patients treated chronically with home NIV (most commonly in the form of bilevel PAP therapy) compared with CPAP therapy.[52]

The morbidity associated with a diagnosis of OHS can be varied, as illustrated by Jennum et al,[53] who evaluated 755 patients with a diagnosis of OHS (using *International Classification of Diseases, Tenth Revision* diagnostic codes) from a Danish national patient registry; in the 3 years before OHS diagnosis, these patients were more likely than age- and sex-matched controls to be diagnosed with a variety of medical conditions, including cellulitis, carpal tunnel syndrome, type

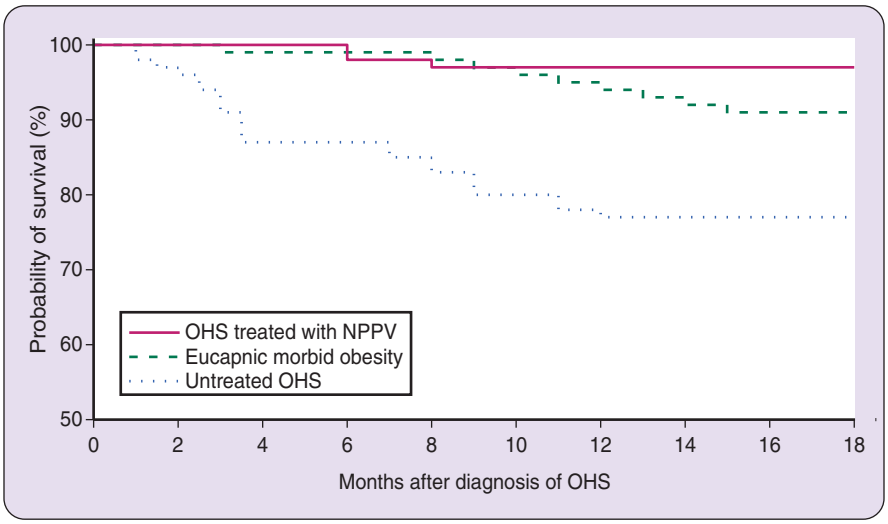


Figure 120-4 Survival curves for patients with untreated obesity-hypoventilation syndrome (OHS; $n = 47$; mean age, 55 ± 14 years; mean body mass index [BMI], 45 ± 9 kg/m^2; mean $PaCO_2$, 52 ± 7 mm Hg) and eucapnic obese patients ($n = 103$; mean age, 53 ± 13 years; mean BMI, 42 ± 8 kg/m^2) as reported by Nowbar et al[17] compared with patients with OHS treated with nocturnal positive pressure ventilation (NPPV) therapy ($n = 126$; mean age, 55.6 ± 10.6 years; mean BMI, 44.6 ± 7.8 kg/m^2; mean baseline $PaCO_2$, 55.5 ± 7.7 mm Hg; mean adherence with NPPV of 6.5 ± 2.3 hours/day). (Data for OHS patients treated with NPPV provided courtesy Stephan Budweiser and colleagues from the University of Regensburg, Germany.[34] Reprinted with permission of the American Thoracic Society. Copyright American Thoracic Society.)

2 diabetes, congestive heart failure, obstructive lung disease, and arthritis of the knee. It remains unclear if these conditions would be more prevalent than in an obese matched cohort with uncomplicated OSA. Furthermore, quality of life ratings among OHS patients appear to be lower than among those with hypoventilatory respiratory disorders such as obstructive lung disease.[54]

Cardiovascular morbidity is of particular concern in OHS.[51] Kessler et al[55] found a pulmonary hypertension prevalence of 58% among a cohort of 34 OHS patients compared with just 9% among a sample of similar OSA patients. Similarly, Berg et al[33] compared 20 OHS patients from a Canadian health registry with obese matched controls. OHS patients in their study were nine times more likely to have a diagnosis of cor pulmonale and nine times more likely to have a diagnosis of congestive heart failure. Moreover, hospitalized patients with obesity-associated hypoventilation are at increased risk of admission to the intensive care unit and need for invasive mechanical ventilation compared with hospitalized patients with eucapnic obesity.[17]

Accordingly, identifying patients with OHS in a timely manner is important, and treatment with PAP therapy should be initiated and monitored without delay to avoid adverse outcomes, such as readmission to the hospital, acute-on-chronic respiratory failure requiring intensive care monitoring, or death. More important, adherence to therapy should be emphasized and monitored objectively.[56]

PATHOPHYSIOLOGY

The partial pressure of CO_2 in the arterial blood ($Paco_2$) is determined by the balance between CO_2 production and elimination. Although the main reason for reduced CO_2 elimination is reduced alveolar ventilation due to an overall decreased level of ventilation (i.e., minute ventilation), maldistribution of ventilation with respect to pulmonary capillary perfusion (i.e., an increase in physiologic dead space) may contribute as well. Further, the rate of CO_2 production in OHS is of particular physiologic concern; severely obese patients, with or without OHS, have increased work of breathing, increased oxygen cost of breathing, and increased CO_2 production compared with lean individuals.[57-59] The majority of individuals with severe obesity maintain homeostasis by increasing alveolar ventilation and associated CO_2 elimination, thereby averting progression to OHS. This is achieved by tight compensatory mechanisms that require an intact integration between respiratory control and acid-base regulatory systems. Ultimately, inadequate elimination of CO_2 relative to CO_2 production leads to chronic hypercapnia in patients with OHS. In addition to the differences as illustrated in Box 120-3, there are a variety of physiologic differences between patients with OHS and those with eucapnic obesity with or without OSA, such as increased upper airway resistance,[60] decreased respiratory system compliance compared with similarly obese subjects without OHS,[61] ventilation-perfusion mismatching secondary to pulmonary edema[62] or low lung volumes/atelectasis,[63] and, most important, impaired central response to hypoxemia and hypercapnia. Although these mechanisms contribute in varying degrees to the gas exchange abnormality observed in patients with OHS, the combination of SBD, a blunted central response to hypercapnia and hypoxia, and renal buffering can explain the progression from sleep hypoventilation to chronic daytime hypoventilation.[64-67]

Severe obesity (BMI ≥ 40 kg/m^2) increases the work of breathing because of the excess weight on the thoracic wall and abdomen.[61,68] However, it is unclear what role, if any, these altered mechanics have in the pathogenesis of OHS. The lung compliance of OHS patients is less than that of equally obese controls (0.122 versus 0.157 L/cm H_2O). This can be explained by the lower functional residual capacity (1.71 versus 2.20 L). There is an even greater difference in chest wall compliance between the two groups (OHS, 0.079 L/cm H_2O; obese controls, 0.196 L/cm H_2O).[61] Patients with OHS also have a threefold increase in lung resistance that has also been attributed to a low functional residual capacity.[61,69] The changes in lung mechanics are frequently demonstrated on spirometry by a low forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV_1) and a normal FEV_1/FVC ratio. The spirometric abnormalities may be related to the combination of abnormal respiratory mechanics and weak respiratory muscles.[29,30,70,71] The abnormal respiratory system mechanics in subjects with severe obesity imposes a significant load on the respiratory muscles and leads to a significant increase in the work of breathing, particularly in the supine position.[61,68] As a result, morbidly obese patients dedicate 15% of their oxygen consumption to the work of breathing compared with 3% in nonobese individuals.[58]

The maximal inspiratory and expiratory pressures are normal in eucapnic morbidly obese patients but are typically reduced in patients with OHS.[72-74] Patients with mild OHS, however, may have normal inspiratory and expiratory pressures.[75] Further, the role of diaphragmatic weakness in the pathogenesis of this disorder remains uncertain because patients with OHS can generate similar transdiaphragmatic pressures at any level of diaphragmatic activation compared with eucapnic obese subjects.[73] In a study by Sampson and Grassino,[73] patients with OHS were able to generate equivalent transdiaphragmatic pressures as eucapnic obese patients during hypercapnia-induced hyperventilation, suggesting that respiratory muscle weakness may not play a role in the development of OHS. In addition, the OHS group showed no evidence of acute diaphragmatic fatigue (or neuromuscular uncoupling) throughout the hypercapnic trial when measured by the ratio of peak electrical activity of the diaphragm to peak transdiaphragmatic pressure, which theoretically eliminates the variable of the patient's cooperation. Potentially more accurate assessments of diaphragmatic strength (e.g., by cervical magnetic stimulation) have not been performed in patients with OHS.[76]

Patients with OHS are able to voluntarily hyperventilate to eucapnia,[77] evidence for a defective, "blunted" central respiratory drive. Further, patients with OHS do not hyperventilate to the same degree as eucapnic morbidly obese patients when rebreathing CO_2.[71,73,75] This deficit improves in most patients after PAP therapy.[75,78,79] In addition, patients with OHS do not augment their minute ventilation to the same degree as eucapnic obese OSA patients when breathing a hypoxic gas mixture.[75,79] This blunted hypoxic drive also improves with PAP therapy,[75,79] suggesting that such blunted drive is a secondary effect of the syndrome (and necessary for its persistence) but not the origin of it. Obesity, genetic predisposition, SBD, and leptin resistance have all been proposed as mechanisms for the blunted response to hypercapnia. Such

blunted respiratory response to hypercapnia is unlikely to be genetic because the ventilatory response to hypercapnia is similar between first-degree relatives of patients with OHS and control subjects.[80]

Leptin, a satiety hormone produced by adipocytes, stimulates ventilation.[81-84] Obesity leads to an increase in CO_2 production and load.[57,59,81] Therefore, with increasing obesity, the excess adipose tissue leads to increasing levels of leptin to increase ventilation to compensate for the additional CO_2 load. This is likely the reason that most severely obese individuals do not develop awake hypercapnia. Patients with OHS and OSA have significantly higher leptin levels compared with lean or BMI-matched subjects without OSA. Although the independent contribution of OSA or OHS to leptin production remains unclear, the data suggest that excess adiposity is a much more significant contributor to elevated serum leptin levels than the presence of OSA or OHS.[85-88] Patients with OHS, however, have a higher serum leptin level than eucapnic subjects with OSA matched for percentage body fat, and AHI and serum leptin levels each drop after treatment with PAP.[87,89,90] These observations suggest that patients with OHS might be resistant to leptin. For leptin to affect the respiratory center and increase minute ventilation, it has to penetrate the cerebrospinal fluid (CSF). The leptin CSF to serum ratio is fourfold higher in lean individuals compared with obese subjects (0.045 ± 0.01 vs. 0.011 ± 0.002; $P < .05$).[91] Individual differences in leptin CSF penetration may explain why some obese patients with severe OSA develop OHS and others do not.

OSA may well contribute to the ventilatory control defect because treatment with CPAP or bilevel PAP typically improves the response to hypercapnia.[75,78,79] The $P_{0.1}$ response to hypercapnia (a sensitive measure of respiratory drive) improves as early as 2 weeks and reaches normal levels after 6 weeks of therapy with PAP in patients with OHS who demonstrate an awake $Paco_2$ between 46 and 50 mm Hg. The response of minute ventilation to hypercapnia improves by the sixth week of PAP therapy but does not completely normalize,[75] and although such findings are not universal,[78,92,93] OSA appears well established in the pathophysiologic mechanism of OHS by the resolution of hypercapnia in most patients after treatment with either tracheostomy or PAP therapy.[11,27,32,56,75,94-96]

Norman et al[64] have proposed an elegant mathematical model that explains the transition from acute hypercapnia during OSA to chronic daytime hypercapnia. In most patients with OSA, the hyperventilation after an apnea eliminates all CO_2 accumulated during the apnea.[97] However, if the interapnea hyperventilation is inadequate or the ventilatory response to the accumulated CO_2 is blunted, it could lead to an increase in $Paco_2$ during sleep.[65] Even in this acute setting, during sleep the kidneys can retain small amounts of bicarbonate to buffer the decrease in pH. If the time constant for the excretion of the small amount of accumulated bicarbonate is slow, the patient will have a net gain of bicarbonate, which may blunt the respiratory drive and lead to CO_2 retention during wakefulness to compensate for the retained bicarbonate.[64] Further, the combination of a decreased response to CO_2 and a slow rate of bicarbonate excretion will lead to a blunted respiratory drive for the next sleep cycle. Indeed, obese eucapnic individuals with an elevated serum bicarbonate level exhibit a blunted response to hypercapnic and hypoxic stimulation tests compared with equally obese eucapnic individuals with normal serum bicarbonate level.[39] Further research is needed to elucidate whether these individuals represent a subgroup of "early OHS" and whether they are at increased risk of progressing to overt daytime hypercapnia over time.

Many studies have tried to identify risk factors associated with hypercapnia in patients with OSA, but the results have been mixed.[16,18,19,25,28-31,98] In a large meta-analysis of 15 studies, Kaw et al[13] identified three factors that were significantly associated with chronic hypercapnia in nonchronic obstructive pulmonary disease obese patients with OSA: (1) severity of obesity as measured by the BMI, (2) severity of OSA measured by either the AHI or hypoxemia during sleep, and (3) degree of restrictive chest physiology.

TREATMENT

Treatment modalities for patients with OHS are based on different aspects of the underlying pathophysiologic mechanism of the condition: reversal of SBD (OSA and nonobstructive sleep hypoventilation), weight reduction, and possibly pharmacotherapy. Nocturnal PAP therapies are considered first-line treatment and are effective in improving patient outcomes.[99,100] However, treatment strategies that include weight reduction and physical activity should also be offered to patients with OHS to improve their metabolic and cardiovascular risk profiles.[101-103]

Positive Airway Pressure Therapy

PAP, in the form of continuous PAP (CPAP), was first described in the treatment of OHS in 1982.[94] Whereas subsequent studies confirmed its efficacy, failure of CPAP in some cases has led to uncertainty as to whether CPAP should be attempted initially or if NIV (most commonly in the form of bilevel PAP) is a better modality.[18,56,92,94,104] In one prospective study of outpatients with severe OHS, 57% of patients were successfully titrated with CPAP alone. In these patients, CPAP was titrated to treat OSA, and the mean pressure required was 14 cm H_2O.[42] The remaining 43% of patients failed to respond to CPAP titration because of persistent hypoxemia at therapeutic or near-therapeutic pressures that had successfully treated OSA. In these patients, the oxygen saturation remained below 90% for more than 20% of total sleep time. Because this was a single-night titration study, the question of whether residual hypoxemia would resolve with long-term treatment was not evaluated systematically.[105] Even though several studies have described the efficacy of both NIV and CPAP, only one randomized controlled trial has directly compared the two modes.[106] In this study, 45 consecutive patients with OHS underwent a full night of CPAP titration. Nine patients (20%) had persistent hypoxemia (arbitrarily defined as 10 continuous minutes of $Spo_2 < 80\%$ without observed apneas) during the CPAP titration and were excluded from the study. The remaining 36 patients who had a successful CPAP titration night with resolution of OSA and hypoxemia were subsequently randomized to either CPAP or NIV (bilevel PAP in the spontaneous mode without a backup respiratory rate). The two groups were well balanced in terms of body habitus, severity of awake hypercapnia, OSA, and nocturnal hypoxemia at baseline. During titration polysomnography, CPAP was increased in increments of 1 cm H_2O with the aim of preventing obstruction, flow limitation,

desaturation, and arousal. During the bilevel PAP titration, the expiratory PAP (EPAP) was started at 2 cm H_2O below the pressure needed to abolish obstructive apneas during the CPAP titration or at 5 cm H_2O, whichever was higher. The EPAP was then increased in increments of 1 cm H_2O to resolve obstructive apneas. The inspiratory PAP (IPAP) was initially set 4 cm H_2O higher than EPAP and then increased to eliminate hypopneas and to improve oxygen saturation. After 3 months, there was no significant difference between the groups in terms of adherence to PAP therapy or improvement in daytime sleepiness, hypoxemia, or hypercapnia. However, this relatively small clinical trial found benefits in favor of NIV over CPAP, especially in sleep quality. This study confirms that CPAP can be successful in some patients with OHS as long as OSA and nocturnal hypoxemia are effectively treated. On the other hand, the exclusion of patients with severe nocturnal hypoxemia who were not responsive to CPAP therapy suggests that a subgroup of patients with OHS may need more advanced forms of ventilation. Therefore, NIV is not superior to CPAP a priori; rather, treatment should be individualized to each patient.

The American Academy of Sleep Medicine has proposed guidelines for the titration of NIV in patients with chronic alveolar hypoventilation syndromes, although not specifically for OHS.[107] The most common mode of NIV used in clinical practice is bilevel PAP. During in-laboratory titration, EPAP is increased until obstructive apneas are resolved.[108] If hypoxemia is persistent or estimated tidal volumes are lower than expected for the patient's ideal body weight, pressure support needs to be increased.[107] Pressure support is the difference between IPAP and EPAP. Most patients with OHS require a pressure support level of at least 8 to 12 cm H_2O (i.e., an IPAP pressure setting that is at least 8 to 12 cm H_2O above EPAP) to achieve effective ventilation.[27,32,109,110] Some patients with OHS may experience central apneas during NIV therapy. Central apneas could occur in OHS during CPAP or NIV titration because of decreased respiratory drive, heart failure, or unstable ventilatory control (high loop gain).[111] Advanced modes, such as bilevel PAP with a backup rate, can help alleviate central apneas in OHS. In the spontaneous/timed (S/T) or timed mode, a backup respiratory rate of 10 to 12 breaths/minute should be initiated and titrated upward by one or two increments generally not exceeding 16 breaths/minute. The backup respiratory rate should be initiated when a patient with hypoventilation syndrome manifests central apneas or inappropriately low respiratory rate and consequent low minute ventilation. Therefore, to perform an adequate NIV titration during sleep in patients with OHS, it is important to monitor several parameters during polysomnography, such as mask flow, delivered pressure, air leak, estimated exhaled tidal volume, and triggered backup mechanical breaths.[107] Transcutaneous CO_2 monitoring, if it is available, provides useful information about the effectiveness of NIV or CPAP titration. Scoring of respiratory events during NIV titration can be challenging, and a systematic description of these events has been proposed.[112]

In the minority of patients with OHS who do not have OSA, EPAP can be set at 5 cm H_2O and IPAP can be titrated to improve ventilation.[109,110] Switching to NIV should also be considered if the $Paco_2$ does not normalize after 3 months of CPAP therapy with objective evidence of adherence to prescribed therapy.

There is accumulating evidence suggesting that sleep hypoventilation can be better controlled by NIV settings and modes that optimize delivery of nocturnal ventilation with the use of either a higher mandatory backup respiratory rate of the ventilator[113] or pressure-volume hybrid modes.[114] Two of these hybrid modes are "average" volume-assured pressure support (AVAPS) and "intelligent" volume-assured pressure support. These pressure-volume hybrid modes of pressure support, volume-controlled ventilation deliver a more consistent tidal volume with the comfort of pressure support ventilation. With these hybrid modes, the pressure support or assistance delivered during the inspiratory phase aims to ensure a certain tidal volume that is calculated as a function of predicted body weight (usually 8 to 10 mL/kg ideal body weight or at 110% of the patient's tidal volume). The device assesses the preset tidal volume or minute ventilation during a variable time window of 1 to 5 minutes. The operating IPAP (or pressure assist) level is then allowed to fluctuate between a minimum and maximum pressure support level to ensure the target tidal volume. If a patient's tidal volume or minute ventilation decreases below a certain threshold, the device responds by increasing the IPAP and restores the tidal volume to approximately the preselected target volume. Such devices have an EPAP-minimum and EPAP-maximum range that needs to be preset as well. Although higher inspiratory pressures achieved with volume-targeted hybrid modes may also optimally relieve dyspnea, they may be more disruptive to sleep in some patients.[115] Additional settings may include spontaneous or timed respiratory rate settings, and newer technology has automated the respiratory rate selection on the basis of the patient's minute ventilation and proportion of breaths that are triggered versus spontaneous over a period of time. In a randomized controlled trial of 50 patients with OHS, volume-targeted pressure support mode was compared with fixed bilevel PAP S/T mode (with a backup respiratory rate).[116] In this study, there was no significant difference between the two advanced PAP modes after 3 months of therapy. Both PAP modalities significantly improved daytime $Paco_2$, sleep hypoventilation (as measured by transcutaneous CO_2), hypoxemia during sleep, and quality of life. The lack of such demonstrable difference between AVAPS and fixed bilevel PAP S/T in this trial may be due to the carefully "optimized" bilevel PAP S/T setting in these clinical research conditions. There was no significant difference in the levels of PAP delivered between the two groups. Patients randomized to AVAPS received mean pressures of IPAP 22 ± 5/EPAP 9 ± 1 cm H_2O with a backup rate of 14 breaths/minute versus bilevel PAP S/T mode with mean pressures of IPAP 23 ± 4/EPAP 10 ± 4 cm H_2O with a backup rate of 14 breaths/minute.[116] Despite these high-pressure settings, the mean adherence to therapy was reasonable and not different between the two groups (AVAPS, 4.2 hours/day; bilevel PAP S/T, 5.1 hours/day). Comparative post hoc analysis revealed that patients in whom more than 50% of the breaths were delivered as the backup respiratory rate experienced a greater control of nocturnal carbon dioxide by transcutaneous CO_2 monitoring, improved daytime $Paco_2$, and enhanced health-related quality of life at 3 months,[116] which supports the hypothesis that controlled NIV, which minimizes patient ventilatory effort in sleep, may help unload the respiratory muscles and provide optimal nocturnal ventilatory control and improved patient outcomes. In the largest

clinical trial to date, the Spanish Sleep Network investigators performed a randomized controlled trial comparing three treatment strategies in 221 patients with OHS. The treatment strategies consisted of NIV, CPAP, and lifestyle modification (control group). For NIV the 16 centers involved in the study were allowed to use a variety of ventilators all of which were set in the volume targeted pressure support mode (mean IPAP 20 ± 3.3 cm H_2O and mean EPAP 7.7 ± 1.8 cm H_2O; backup respiratory rate of 12–15 breaths per minute and tidal volumes of 550–660 ml). The average CPAP pressure was 11 ± 2.5 cm H_2O. At two months, NIV and CPAP were superior to control group in improving $PaCO_2$, clinical symptoms, and polysomnographic parameters. However, there were no significant differences in the degree of improvement in $PaCO_2$ between NIV and CPAP. Adherence to the PAP modalities was not significantly different (mean adherence to NIV was 5.3 ± 2.3 h/day and CPAP was 5.3 ± 2.1 h/day). Although some health-related quality-of-life assessments, FEV_1, and 6-minute-walk distance improved more with NIV than with CPAP, the long-term significance of these functional improvements requires further investigation. This clinical trial may be able to shed light on the long-term impact of different treatment modalities as the investigators plan to follow these patients for 36 months after randomization.[116a]

There are numerous technical challenges with applying NIV in OHS. Advanced PAP modalities such as volume-targeted pressure support technology rely heavily on and function properly when unintentional air leak from the noninvasive mask remains low.[117] Whereas bench studies have shown that most NIV devices underestimate air leak and exhaled tidal volume, there is significant variability among manufacturers.[118]

The most common reason for persistent hypercapnia and hypoxemia in patients with OHS treated with PAP is lack of adherence to the PAP therapy. In a retrospective study of 75 outpatients with stable OHS, patients who used CPAP or bilevel PAP therapy for more than 4.5 hours/day had a considerably greater improvement in blood gases than less adherent patients ($\Delta Paco_2$ 7.7 ± 5 vs. 2.4 ± 4 mm Hg, $P < .001$; ΔPao_2 9.2 ± 11 vs. 1.8 ± 9 mm Hg, $P < .001$).[56] The degree of improvement in ventilation and gas exchange, which can be seen as early as 2 to 4 weeks after therapy,[56,75,119] may allow discontinuation of daytime oxygen supplementation in many patients with OHS.[56] However, the improvement in chronic daytime gas exchange abnormalities (i.e., hypercapnia and hypoxemia) even in patients who are adherent to PAP therapy is neither universal nor complete.[52,102] Other possibilities behind persistent hypercapnia include inadequate CPAP pressure or insufficient NIV support; CPAP failure in OHS patients who do not have significant OSA; other causes of hypercapnia, such as chronic obstructive pulmonary disease; and metabolic alkalosis due to high doses of loop diuretics. In two studies,[56,106] the $Paco_2$ did not improve significantly in approximately a quarter of patients who had undergone successful PAP titration in the laboratory and were highly adherent (>6 hours/night) with either CPAP or bilevel PAP therapy. It is conceivable that volume-targeted pressure support or higher levels of pressure support with fixed bilevel PAP in the S/T mode would be more effective in normalizing ventilation and gas exchange. Reports of persistent hypoventilation after tracheostomy[27] highlight the need for aggressive nocturnal mechanical ventilation in addition to support of upper airway patency in at least a subset of OHS patients.

Early follow-up is imperative and should include assessment of adherence to PAP therapy; patients with OSA, although not specifically with OHS, frequently overestimate CPAP adherence.[120-122] Changes in serum bicarbonate level and improvements in resting room air pulse oximetry and end-tidal CO_2 measurements during wakefulness could be used as less invasive surrogates of ventilation if the patient is reluctant to undergo follow-up measurement of arterial blood gases.

Oxygen Therapy

In some patients with OHS, oxygen supplementation is necessary after the resolution of apneas and hypopneas during PAP titration, both CPAP and NIV, to keep Spo_2 above 88% to 90%. In two studies, the percentage of OHS patients requiring supplemental oxygen after adequate CPAP titration (i.e., resolution of obstructive apneas and hypopneas) was as high as 43%.[42,56] In contrast, in a relatively comparable group of patients with OHS, only 12% undergoing aggressive NIV titration with relatively high levels of pressure support (~13 cm H_2O above an average EPAP of 10 cm H_2O) required such oxygen supplementation.[116] This finding suggests that higher levels of pressure support during PAP titration must be considered to achieve adequate oxygenation and ventilation during sleep in a large proportion of patients with OHS. Oxygen supplementation as monotherapy without resolution of upper airway obstruction with CPAP or adequate ventilatory support with NIV is strongly discouraged. Two well-controlled clinical trials have reported that in a significant proportion of patients with OHS who were tested during wakefulness and in steady state, supplemental oxygen at high[123] and medium concentrations[124] worsened hypercapnia (because of a drop in tidal volume and minute ventilation). It is plausible that the risk of CO_2 retention is even higher during acute-on-chronic hypercapnic respiratory failure in OHS.[125]

Weight Reduction

Bariatric surgery has variable long-term efficacy in treating OSA.[126] A meta-analysis that included 12 studies with 342 patients who underwent polysomnography before bariatric surgery and after maximum weight loss reported a 71% reduction in the AHI, from baseline of 55 (95% confidence interval, 49-60) to 16 (95% confidence interval, 13-19).[127] Whereas only 38% achieved cure defined by AHI below 5, this drastic improvement in the severity of SBD would likely be enough to normalize daytime blood gases in most patients with OHS. It is also known that in the 6 to 8 years after weight reduction surgery, patients experience approximately 7% weight gain, which may lead to an increase in the AHI.[128,129] Only one study has examined the impact of bariatric surgery in patients with OHS. One year after surgery in 31 patients with OHS, the Pao_2 increased from an awake baseline on breathing room air of 53 to 73 mm Hg and $Paco_2$ decreased from 53 to 44 mm Hg after approximately 50 kg of weight loss (baseline BMI, 56 ± 13 kg/m^2; BMI at 1 year, 38 ± 9 kg/m^2). In the 12 patients in whom an arterial blood gas measurement was available 5 years after surgery, values had worsened, with the mean Pao_2 dropping to 68 mm Hg and $Paco_2$ increasing to 47 mm Hg.[130] The BMI in these 12 patients was 40 ± 10 kg/m^2.

The general perioperative mortality is between 0.5% and 1.0%. Untreated OHS may be associated with higher operative mortality.[131-133] The independent risk factors associated with mortality are intestinal leak, pulmonary embolism, preoperative weight, and hypertension. Depending on the type of surgery, intestinal leak occurs in 2% to 4% of patients and pulmonary embolism occurs in 1% of patients.[132] Ideally, patients with OHS should be treated with PAP therapy before undergoing surgical intervention to decrease perioperative morbidity and mortality. Moreover, PAP therapy using the patient's preoperative settings should be initiated immediately after extubation to avoid postoperative respiratory failure,[133-136] particularly because there is no evidence that PAP therapy initiated postoperatively leads to anastomotic disruption or leakage.[135,137] Such settings, however, may not be optimal for ventilation and/or oxygenation in the immediate postoperative period, particularly with the use of analgesic and/or sedative medication, and must be monitored and adjusted accordingly.

Tracheostomy

Tracheostomy was the first therapy described for the treatment of OHS.[138] In a retrospective study of 13 patients with OHS, tracheostomy was associated with significant improvement in concomitant OSA. However, in seven patients, the AHI remained above 20. Residual respiratory events were associated with persistent respiratory effort, suggesting that disordered breathing was caused by obstructive hypoventilation through an open tracheostomy rather than by central apneas. On occasion, excessive neck skin folds can intermittently obstruct the tracheostomy orifice. However, the overall improvement in the severity of SBD after tracheostomy leads to the resolution of hypercapnia in the majority of the patients with OHS.[139] Currently, tracheostomy is generally reserved for patients who are intolerant of or not adherent to PAP therapy. Patients with tracheostomy may require additional nocturnal ventilation as tracheostomy alone does not treat any central hypoventilation that may be present.[140] Polysomnography with the tracheostomy open is necessary to determine whether nocturnal ventilation is required and to specifically titrate the mode and levels of ventilation necessary.[27]

Respiratory Stimulation

Respiratory stimulants can theoretically increase respiratory drive and improve daytime hypercapnia, but such data in patients with OHS are extremely limited. Medroxyprogesterone acts as a respiratory stimulant at the hypothalamic level.[141] The results of treatment in patients with OHS have been contradictory. In a series of 10 men with OHS who were able to normalize their $Paco_2$ with 1 to 2 minutes of voluntary hyperventilation, treatment with 60 mg/day of oral medroxyprogesterone for 1 month resulted in normalization of the $Paco_2$ (from 51 mm Hg to 38 mm Hg) and improvement in the Pao_2 (49 mm Hg to 62 mm Hg).[142] In contrast, medroxyprogesterone did not improve $Paco_2$, minute ventilation, or ventilatory response to hypercapnia in three OHS patients who remained hypercapnic after tracheostomy.[92] Further, medroxyprogesterone may increase the risk of venous thromboembolism, particularly in this population whose mobility is limited.[143,144] In addition, high doses of medroxyprogesterone can lead to breakthrough uterine bleeding in women and to decreased libido in men.

Acetazolamide induces metabolic acidosis through carbonic anhydrase inhibition, which decreases serum bicarbonate, shifts the CO_2 response to the left, and increases minute ventilation.[92,145] Acetazolamide may also favorably affect OSA by improving loop gain.[146-148]

Most but not all patients with OHS can normalize their $Paco_2$ with 1 minute of voluntary hyperventilation.[77] The inability to eliminate CO_2 with voluntary hyperventilation may be due to mechanical impairment. In one study, the ability to decrease the $Paco_2$ by at least 5 mm Hg with voluntary hyperventilation was the main predictor of a favorable response to respiratory stimulants.[149] Ideally, however, respiratory stimulants should not be used in patients who cannot normalize their $Paco_2$ with voluntary hyperventilation (because of limited ventilation or mechanical impairment); it can lead to an increase in dyspnea or even worsening of acidosis with acetazolamide. Overall, pharmacotherapy with respiratory stimulants cannot be currently recommended as monotherapy in patients with OHS.

Hyperviscosity impairs oxygen delivery and can counteract the beneficial effects of erythrocytosis. Phlebotomy has not been systematically studied in patients with OHS who develop secondary erythrocytosis. In adult patients with congenital cyanotic heart disease, phlebotomy has been recommended if the hematocrit is above 65% only if symptoms of hyperviscosity are present.[150] However, it is difficult to extrapolate this recommendation to patients with OHS because many symptoms of hyperviscosity are similar to the symptoms of OHS. Reversal of hypoventilation and hypoxemia with PAP therapy eventually improves secondary erythrocytosis, and therefore phlebotomy is not needed in patients with OHS.[151]

CLINICAL PEARL

Clinicians should recognize that approximately 8% to 20% of obese patients referred for polysomnography for suspicion of OSA have OHS. The prevalence of OHS is even higher in severely obese patients with OSA. Unfortunately, OHS is typically underrecognized and undertreated. Delay in diagnosis and treatment leads to significant health care resource utilization and increased morbidity and mortality. Therefore, a high index of suspicion is required to diagnose OHS in a timely fashion to improve patient outcomes. Nocturnal PAP therapies are considered first-line treatment and are effective in improving patient outcomes. Whereas significant advances have been made in the delivery of nocturnal PAP therapy, adherence to such therapy remains suboptimal in many patients with OHS. Therefore, comprehensive management should include strategies to improve PAP adherence. Although PAP therapy improves nocturnal and daytime hypoventilation and quality of life, weight reduction and increase in physical activity should be included as part of comprehensive treatment strategies to improve the metabolic and cardiovascular risk profiles of patients with OHS.

SUMMARY

With the current global epidemic of obesity, the prevalence of OHS is likely to increase. Despite the significant morbidity and mortality associated with the syndrome, it is often unrecognized, and treatment is frequently delayed. A high index of suspicion can lead to early recognition of the syndrome and initiation of appropriate therapy. Significant advances have

been made in the delivery of PAP therapy and NIV. Clinicians should encourage adherence to PAP therapy to avert the serious adverse outcomes of untreated OHS.

Selected Readings

Berger KI, Rapoport DM, Ayappa I, Goldring RM. Pathophysiology of hypoventilation during sleep. *Sleep Med Clin* 2014;**11**:289–300.

Berry RB, Chediak A, Brown LK, et al. NPPV Titration Task Force of the American Academy of Sleep Medicine. Best clinical practices for the sleep center adjustment of noninvasive positive pressure ventilation (NPPV) in stable chronic alveolar hypoventilation syndromes. *J Clin Sleep Med* 2010;**6**:491–509.

Böing S, Randerath WJ. Chronic hypoventilation syndromes and sleep-related hypoventilation. *J Thorac Dis* 2015;**7**(8):1273–85.

Combs D, Shetty S, Parthasarathy S. Advances in PAP treatment modalities for hypoventilation syndromes. *Sleep Med Clin* 2014;**11**:315–26.

Gonzalez-Bermejo J, Perrin C, Janssens JP, et al. SomnoNIV Group. Proposal for a systematic analysis of polygraphy or polysomnography for identifying and scoring abnormal events occurring during non-invasive ventilation. *Thorax* 2012;**67**:546–52.

Hart N, Mandal S, Manuel A, et al. Obesity hypoventilation syndrome: does the current definition need revisiting? *Thorax* 2014;**69**:83–4.

Huttmann SE, Windisch W, Storre JH. Techniques for the measurement and monitoring of carbon dioxide in the blood. *Ann Am Thorac Soc* 2014;**11**:645–52.

Janssens JP, Borel JC, Pepin JL. Non-PAP treatment modalities in obesity-hypoventilation syndrome: role of exercise, nonsurgical and surgical weight reduction, tracheostomy, respiratory stimulants, and oxygen. *Sleep Med Clin* 2014;**11**:357–64.

Masa JF, Corral J, Alonso ML, et al. Efficacy of different treatment alternatives for obesity hypoventilation syndrome. Pickwick Study. *Am J Respir Crit Care Med* 2015;**192**:86–95.

Mokhlesi B, Tulaimat A, Faibussowitsch I, et al. Obesity hypoventilation syndrome: prevalence and predictors in patients with obstructive sleep apnea. *Sleep Breath* 2007;**11**:117–24.

Murphy PB, Davidson C, Hind MD, et al. Volume targeted versus pressure support non-invasive ventilation in patients with super obesity and chronic respiratory failure: a randomised controlled trial. *Thorax* 2012;**67**:727–34.

Murphy PB, Hart N. Outcomes for obese patients with chronic respiratory failure: results from the observational and randomized controlled trials. *Sleep Med Clin* 2014;**11**:349–56.

Pepin JL, Borel JC, Janssens JP. Obesity hypoventilation syndrome: an underdiagnosed and undertreated condition. *Am J Respir Crit Care Med* 2012;**186**:1205–7.

Piper AJ, Gonzalez-Bermejo J, Janssens JP. Sleep hypoventilation: diagnostic considerations and technological limitations. *Sleep Med Clin* 2014;**9**:301–14.

Piper AJ, Grunstein RR. Obesity hypoventilation syndrome: mechanisms and management. *Am J Respir Crit Care Med* 2011;**183**:292–8.

A complete reference list can be found online at ExpertConsult.com.

Chapter 121

Obstructive Sleep Apnea, Obesity, and Bariatric Surgery

Eric J. Olson; Anita P. Courcoulas

Chapter Highlights

- Excessive body weight is a growing global health issue. In the United States, two of every three adults weigh more than their ideal body weight. Obesity (defined as a body mass index of 30 kg/m^2 or greater) predicts increased morbidity and mortality. One of the health conditions that obesity has a significant impact on is obstructive sleep apnea.
- Inconsistent results from dietary, behavioral, and pharmacologic weight loss therapies have led to increasing interest in bariatric surgery, which encompasses a variety of abdominal operations that restrict caloric intake, absorption, or both. The global total number of bariatric procedures performed annually is estimated at more than 300,000.
- Familiarity with the principles and applications of bariatric surgery is emerging as an appropriate requirement for sleep medicine practitioners, in view of the frequency with which coexistent obesity and sleep-related breathing disorders, obstructive sleep apnea and obesity-hypoventilation syndrome, are encountered in clinical practice. The sleep specialist has an important role in a comprehensive perioperative bariatric care program.

DEFINITIONS AND OVERVIEW

The increasing proportion of people who weigh more than their ideal body weight is a worldwide health concern, with significant medical, psychological, and economic ramifications. In adults, overweight and obesity traditionally have been defined by the *body mass index* (BMI), which is the quotient of the weight in kilograms divided by the height in meters squared. Table 121-1 depicts the National Heart, Lung, and Blood Institute's weight classification system for adults, in which *overweight* is defined as a BMI of 25 to 29.9 kg/m^2 and *obese* is defined as a BMI of 30 kg/m^2 or more.[1] Excess abdominal fat, defined by a waist circumference of greater than 40 inches (102 cm) in men and greater than 35 inches (88 cm) in women, is an independent predictor of risk for type 2 diabetes mellitus, dyslipidemia, hypertension, and cardiovascular disease in adults with a BMI between 25 and 34.9 kg/m^2.[1] Overweight and obesity are a result of a complex interplay of genetic and sociocultural forces that lead to long-term positive energy balance.[2] Obesity is associated with myriad complications, including obstructive sleep apnea (OSA).

Obesity is one of the most important risk factors for OSA.[3] Excessive body weight may increase propensity for upper airway narrowing during sleep by altering the function and the geometry of the pharynx.[3] In addition, obesity may alter ventilatory control and respiratory muscle function, leading to obesity hypoventilation syndrome (OHS), which is characterized by the combination of obesity, chronic hypercapnia in the absence of another identifiable cause, and usually some component of sleep-related breathing disorder, most commonly OSA (see Chapter 120).[4] Treatment for OSA includes continuous positive airway pressure (CPAP), oral appliances, upper airway surgeries, and risk factor modifications, including weight loss.

For patients desiring to lose weight, initial interventions include dietary modifications to reduce energy intake, enhanced physical activity to increase energy expenditure, and behavioral therapies to overcome barriers to compliance.[1] Pharmacotherapy may be considered for patients with a BMI of 30 kg/m^2 or more, or for those with a BMI of 27 kg/m^2 or more and obesity-related disease who fail to achieve their weight loss targets after 6 months of diet and lifestyle changes.[1]

Surgical therapy for obesity, or *bariatric surgery,* is indicated for morbidly obese persons for whom other attempts at nonsurgical approaches to weight control have failed. Some bariatric surgery procedures restrict food intake, and others induce malabsorption or maldigestion.[5] The number of bariatric procedures being performed has increased dramatically as a result of the rise in prevalence of severe obesity and refinement of operative techniques.

This chapter considers the epidemiology of overweight and obesity, potential mechanisms linking overweight and obesity with OSA, indications for bariatric surgery, technical aspects of common bariatric procedures, perioperative management of patients with OSA, and outcomes of bariatric surgery, including its impact on OSA.

EPIDEMIOLOGY

Epidemiology of Overweight and Obesity

According to the latest National Health and Nutrition Examination Survey (NHANES), for the year 2011 to 2012, 68.5% of U.S. adults were either overweight or obese, 34.9% were obese, and 6.4% had class 3 obesity, which translates to a total of approximately 15 million adults with a BMI of 40 kg/m^2 or greater.[6] Among U.S. youth, 31.8% were either overweight or obese and 16.9% were obese.[6] Between 1980 and 2007, the prevalence of obesity doubled, and the prevalence of class 3

Table 121-1 Classification of Overweight and Obesity by Body Mass Index

Category	Body Mass Index (kg/m^2)
Underweight	<18.5
Normal	18.5–24.9
Overweight	25–29.9
Obesity	
Class 1	30–34.9
Class 2	35–39.9
Class 3 (extreme)	≥40

From North American Association for the Study of Obesity and the National Heart, Lung, and Blood Institute. *The practical guide: identification, evaluation, and treatment of overweight and obesity in adults.* NIH publication 00-4084. Bethesda (Md.): National Institutes of Health; 2000. <http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm>.

obesity nearly quadrupled.[7] Figure 121-1 shows the prevalence of adult obesity in 2014 per state and territory in the United States.[8] Obesity prevalence also has roughly doubled among Canadian adults, although the overall prevalence of adult obesity in Canada (approximately 25%) remains lower than in the United States.[9] Obesity rates in the United States are highest among non-Hispanic black adults (47.8%), followed by Hispanic Americans (42.5%) and then non-Hispanic whites (32.6%) and Asian Americans (10.8%).[6] Obesity prevalence in Native Americans is similar to that in non-Hispanic blacks.[10] Data from the Framingham Heart Study attribute a reduction in life expectancy of 7.1 years in nonsmoking women and 5.9 years in nonsmoking men at age 40 to obesity,[11] with the increased mortality resulting primarily from cardiovascular disease.[12]

Epidemiologic Association between Overweight/Obesity and Obstructive Sleep Apnea

Cross-sectional analyses of clinical and population samples have demonstrated notable colocalization of OSA and

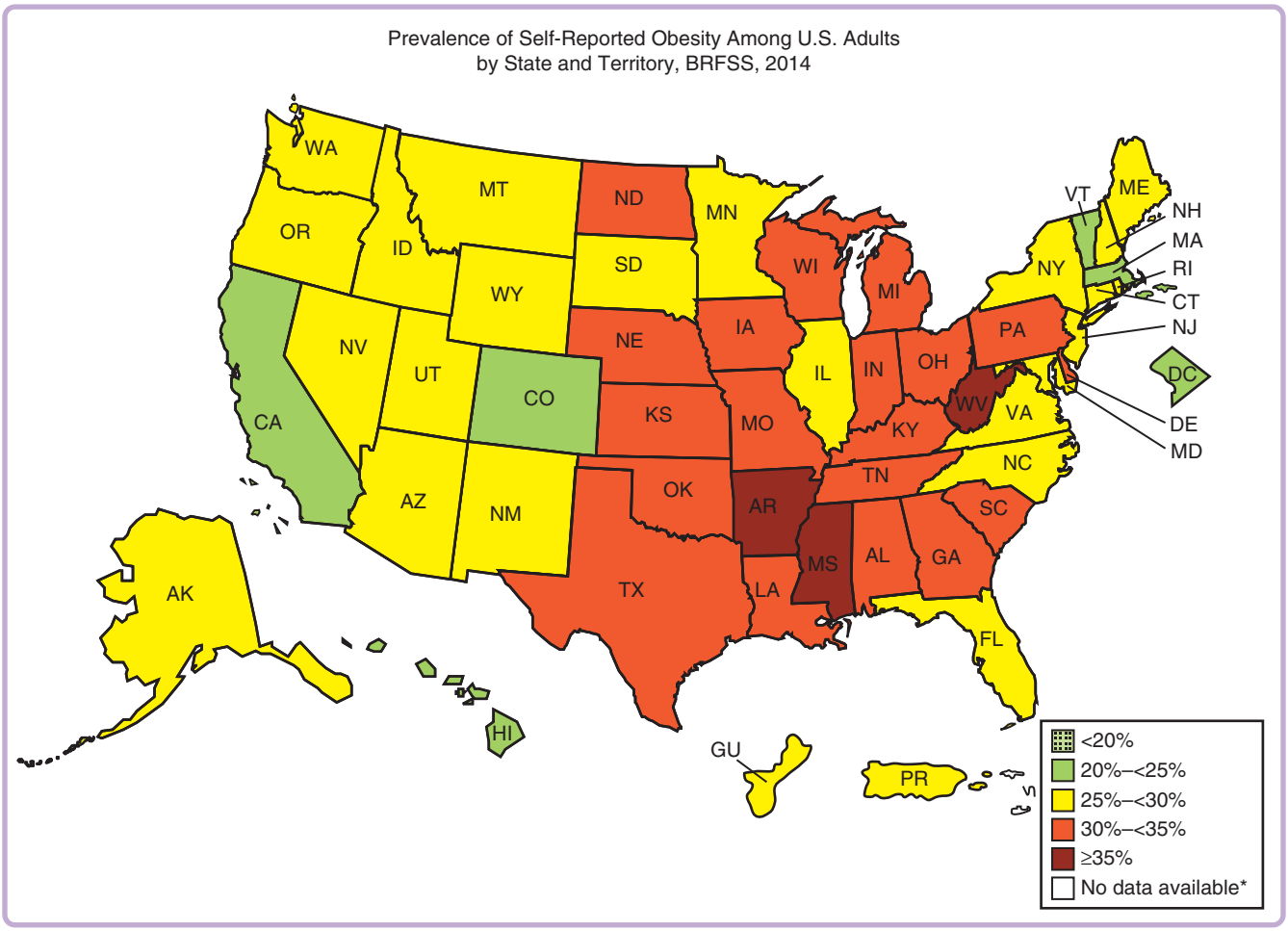


Figure 121-1 Prevalence of self-reported obesity among adults in the United States by state and territory per the Behavioral Risk Factor Surveillance System (BRFSS), 2014. No state has an obesity prevalence of less than 20%. Eighteen states have an obesity prevalence between 30% and less than 35%, and 2 states have an obesity prevalence of 35% or greater. The South has the highest prevalence of obesity, followed by the Midwest, the Northeast, and the West. (From Centers for Disease Control and Prevention. Obesity prevalence maps, 2013. <http://www.cdc.gov/obesity/data/prevalence-maps.html.>)

obesity.[3] OSA has been found in 50% to 80% of obese patients seen in clinical settings,[3] and from 60% to 90% of adults with OSA may be overweight.[13] In the Wisconsin Sleep Cohort, an increase of 1 standard deviation (5.7 kg/m^2) was associated with a fourfold risk of an apnea-hypopnea index (AHI) of 5 events/hour or greater.[14] Furthermore, the Sleep Heart Health Study reported a dose-dependent relationship between increasing BMI and OSA: The prevalence of an AHI of 15 or higher was 10% in the lowest BMI quartile (16 to 24 kg/m^2), as opposed to 32% in the highest quartile (32 to 59 kg/m^2).[15]

Longitudinal population and clinical samples also indicate that weight and AHI change congruently. In the Wisconsin Sleep Cohort, each 1% increase (or decrease) in weight was associated with a 3% increase (or decrease) in AHI, and in patients with mild OSA (AHI of 5 to 15) at baseline, a 10% weight gain led to a sixfold risk for developing moderate to severe OSA (AHI of 15 or higher).[16] In the Sleep Heart Health Study, parallel changes in weight and AHI were similarly found over a 5-year follow-up period, but AHI increased more with weight gain than it decreased with weight loss.[17] Men tend to suffer a greater increase in AHI with weight gain than women,[17] whereas BMI has a greater effect on AHI in postmenopausal women than in premenopausal women.[3] Increasing age may attenuate the association of BMI and AHI.[15]

PATHOGENESIS

Mechanism Linking Obesity to Obstructive Sleep Apnea Risk

Upper airway anatomic and neuromuscular factors may be influenced by parapharyngeal fat accumulation in several ways.[18] In obesity, upper airway size may be compressed by the deposition of adipose tissue, especially in the lateral pharyngeal fat pads, intraluminal structures (tongue, soft palate, uvula), and neck.[19-23] During wakefulness, increased pharyngeal dilator muscle activity provides compensation; the state-dependent attenuation of pharyngeal muscle activity with sleep, by contrast, leaves the upper airway vulnerable to collapse.[24] Accrual of fat around the upper airway may alter soft tissue properties, thereby heightening the propensity to collapse by increasing upper airway compliance,[25,26] or may change upper airway geometry,[3] with a consequent decrease in the ability of pharyngeal muscles to dilate the airway.[27]

Obesity, a chronic inflammatory state itself, may contribute to increasing upper airway tissue inflammation,[28] or upper airway neuropathic damage through the pathophysiologic changes of diabetes mellitus.[29] Abdominal viscera fat accumulation in patients with OSA also is of likely pathogenetic importance, as highlighted by population surveys reporting a twofold to threefold increase in prevalence of symptomatic OSA in men compared with women,[14,30] with a central distribution of fat involving the neck, trunk, and abdominal viscera typical in men, versus fat deposition in the lower body and extremities characteristically seen in women. The interaction of hormonal changes and accompanying increases in central fat deposition may contribute to the increased prevalence of OSA in postmenopausal women.[3] Central obesity–induced reduction in lung volume[31] decreases "tracheal tug," a caudally directed, pharyngeal-stabilizing, lung volume–dependent traction force directed along the trachea.[32] The reduction in lung volumes coupled with globally increased oxygen demand also may promote the oxygen desaturation that accompanies obstructive apneas and hypopneas.

An evolving discussion in sleep medicine concerns the existence of a bidirectional relationship between obesity and OSA. The intermittent hypoxia, sympathetic activation, and sleep fragmentation caused by repeated episodes of obstructive apnea and hypopnea produce metabolic alterations that provide biologically plausible means by which OSA also may exacerbate overweight and obesity (see Chapter 118). The pathogenetic interactions between obesity and OSA, however, are complex and not fully explored, leaving the subject open for active research.[13]

BARIATRIC SURGERY FOR MEDICALLY COMPLICATED OBESITY

Bariatric surgery has emerged in the context of the rising prevalence of severe obesity, heightened concern about associated comorbid medical conditions, and the limited success of traditional weight loss approaches of diet, exercise, and behavior modification. Although the number of bariatric procedures performed had been increasing for many years, only 1% of clinically eligible patients are being treated for medically complicated obesity in this manner.[33] Lack of access to bariatric surgery probably results from many as-yet poorly studied factors, including lack of insurance coverage, poor understanding of the procedures and their effects, and prohibitive costs—facility costs alone for bariatric surgery, for example, range from $10,000 to $15,000 per case.[34]

Patient Selection

Guidelines originally issued by the National Institutes of Health[35] and recently reaffirmed by the American Association of Clinical Endocrinologists, The Obesity Society, and the American Society for Metabolic and Bariatric Surgery (AACE-TOS-ASMBS)[36] state that bariatric surgery is a treatment option for severely obese patients who failed to achieve weight loss on a structured and monitored exercise and diet program and who have a BMI of 40 kg/m^2 or greater, or a BMI of 35 kg/m^2 or greater in conjunction with one or more obesity-related severe comorbid conditions.[1] Such obesity-associated comorbid conditions include diabetes mellitus, arterial hypertension, dyslipidemia, coronary artery disease, pseudotumor cerebri, asthma, venous stasis, severe urinary incontinence, debilitating arthritis, gastroesophageal reflux disease, nonalcoholic fatty liver disease, and OSA.[36] The most recent AACE-TOS-ASMBS clinical practice guidelines also state that bariatric surgery may be offered to patients with a lower BMI of 30 to 34.9 kg/m^2 along with diabetes mellitus or metabolic syndrome, with the acknowledgment that current evidence to support this expansion of potential bariatric surgery candidates is based on limited and short-term data demonstrating benefit and is not universally accepted.[36]

Patients preparing to undergo bariatric surgery must complete nutritional and psychological evaluation screening to rule out untreated depression, substance abuse, or a history of untreated eating disorders. The potential bariatric surgery recipient must demonstrate a complete understanding of the risks and benefits of the operation as a weight loss "tool," recognize the necessity after surgery to limit portion size and food types, and agree to follow a postoperative vitamin supplementation regimen. Patients should be deemed ineligible

Table 121-2 Patient Selection Criteria for Bariatric Surgery

Factor	Criteria
Weight (adults)	BMI of 40 kg/m^2 or greater, with no comorbid conditions BMI of 35 kg/m^2 or greater with obesity-associated comorbidity
Weight loss history	Failure of previous nonsurgical attempts at weight reduction, including nonprofessional programs (e.g., Weight Watchers)
Commitment	Expectation that patient will adhere to postoperative care Follow-up visits with physician(s) and team members Recommended medical management, including use of dietary supplements Instructions regarding any recommended procedures or tests
Exclusion	Reversible endocrine or other disorders that can cause obesity Current drug or alcohol use Uncontrolled, severe psychiatric illness Lack of comprehension of risks, benefits, expected outcomes, alternatives, and lifestyle changes required with bariatric surgery

From Mechanick JI, Youdim A, Jones DB, et al. Clinical practice guidelines for the perioperative nutritional, metabolic, and nonsurgical support of the bariatric surgery patient—2013 update: cosponsored by American Association of Clinical Endocrinologists, The Obesity Society, and American Society for Metabolic and Bariatric Surgery. *Obesity* 2013;21:S1–27.

for surgery if they cannot understand or will not commit to the dietary changes and lifestyle modifications necessary to complement the procedure. In addition, poor surgical or anesthetic risk status (as with advanced congestive heart failure or suboptimally controlled angina), older age (greater than 50 years), reversible endocrine or other disorders that might cause obesity, and active addiction behaviors are contraindications to bariatric surgery. Selection criteria for bariatric surgery are summarized in Table 121-2.[36]

Rationale for Patient Assessment for Obstructive Sleep Apnea before Bariatric Surgery

High OSA prevalence has been reported among patients being considered for bariatric surgery. In a large series of consecutive patients undergoing bariatric surgery ($n = 342$) in whom preoperative polysomnography was performed regardless of suspicion for OSA, the prevalence of OSA (defined as an AHI of 5 or higher) was 77%; 19% had moderate (AHI of 15 to 30) and 27% had severe OSA (AHI higher than 30).[37] In bariatric surgery candidates, many potential reasons to consider the possibility of OSA preoperatively have been documented: Concurrent OSA may complicate the intubation and/or increase the difficulty of mask ventilation in obese patients. Commonly used perioperative drugs have inhibitory influences on central ventilatory drive, protective upper airway reflexes, and arousal mechanisms, which may further jeopardize the airway of the severely obese patient with known OSA. Upper airway edema associated with endotracheal intubation and forced supine positioning also can acutely aggravate OSA risk after bariatric surgery. Spells of desaturation associated with postoperative obstructive apneic episodes or hypoventilation may be exaggerated by the interaction of obesity-related reduction in pulmonary functional residual capacity with factors in the postoperative milieu. OSA may increase risk for and/or destabilize comorbid conditions in the obese patient, such as hypertension, atrial fibrillation, heart failure, and diabetes mellitus, and these conditions may require attention preoperatively or adversely affect the postoperative course. Therefore close collaboration among the sleep specialist, the anesthesiologist, the bariatric surgeon, and the patient is crucial for proper planning to mitigate OSA-related complications perioperatively.

Preoperative Assessment for Bariatric Surgery in the Patient without Known Obstructive Sleep Apnea

The AACE-TOS-ASMBS clinical practice guidelines[36] stipulate that the possibility of OSA should be considered in *all* bariatric surgery candidates. However, uncertainties exist regarding the specifics of the extent of the preoperative OSA evaluation.

Discernment of OSA status begins with a sleep-focused history and physical exam by the bariatric surgery team. The diagnostic features of OSA are discussed in Chapters 113 and 114. No cardinal symptom of OSA, such as snoring or excessive daytime sleepiness, is singularly sufficient to predict the presence of OSA or its severity in bariatric surgery candidates. Furthermore, no a single best metric of body habitus for predicting OSA has been identified.[3] Instead, a combination of symptoms and signs is more discriminatory, so many prediction formulas combining clinical parameters have been created to hone clinicians' detection of OSA. An example of a screening tool for OSA extensively studied in preoperative patients is the STOP-Bang instrument.[38] This questionnaire poses "yes-or-no" questions about **s**noring, **t**iredness, **o**bserved apneas, blood **p**ressure, **B**MI higher than 35 kg/m^2, **a**ge older than 50 years, **n**eck circumference greater than 40 cm, and male **g**ender, with likelihood of the presence of OSA increasing as the number of affirmative responses increases. Sensitivity (proportion of patients with OSA correctly identified by the STOP-Bang to have OSA) and specificity (proportion of patients without OSA correctly identified by the STOP-Bang to not have OSA) for moderate to severe OSA defined by AHI higher than 15 in patients with BMI of 35 kg/m^2 or greater preparing for nonbariatric operations depends on the cutpoints selected: score of 3 or higher: 97% (sensitivity) and 7% (specificity); score of 4 or higher: 86% and 28%; 5 or higher: 65% and 65%; and 6 or higher: 42% and 86%.[39] These figures highlight the tradeoff inherent in creating and implementing OSA prediction tools: As the threshold for the number of required OSA features is increased, those incorrectly labeled as having OSA (false positives) decreases, but detection of patients with OSA (true positives) also decreases. In general, most OSA prediction tools are more sensitive than specific, favoring detection of true positives (presence of OSA in patients identified as having the disorder) and thereby minimizing false negatives, at the expense of false positives (designating many patients as having OSA when in fact they do not). Severe OSA is unlikely to missed by OSA prediction tools, but reported sensitivities and specificities for a given

screening tool have varied in the hands of different investigators, and the ideal preoperative OSA screening instrument has not been identified.[40]

Guidelines emerging from the anesthesiology literature[41-43] recommend incorporating an OSA prediction tool or checklist in the preoperative assessment of patients preparing for any surgery. The STOP-Bang is highlighted in several of these strategies[41,42]; the American Society of Anesthesiologists guideline for the perioperative management of patients with known or suspected OSA contains its own OSA prediction checklist.[43] Those patients judged to be at low risk for having OSA as determined by the screening instrument are cleared to proceed directly to surgery without further sleep testing, whereas those deemed to be at intermediate or high risk should proceed either to a formal sleep assessment or to surgery, with adjustment of their perioperative care for presumptive OSA, depending on the clinical status and the urgency of the surgical issue.

Attempts to create new OSA screening tools[37,44] or to validate existing[45,46] tools in bariatric surgery populations have not yielded powerfully discriminative sensitivity and specificity characteristics for the instruments tested. In the study by Gasa and colleagues,[44] the addition of sleep oximetry data did significantly enhance the sensitivity and specificity of their initial model based on anthropometric and clinical factors alone (age, waist circumference, systolic blood pressure, and witnessed apneas). The latest AACE-TOS-ASMBS bariatric surgery clinical practice guidelines[36] recommend "standardized screening" for OSA, with "confirmatory polysomnography to follow if screening tests are positive," but do not elaborate on the phrase, "screening tests." The guidelines label predictive modeling attempts "encouraging" but not definitive.[36] The current understanding of the role for OSA screening tools in the bariatric population will continue to evolve. At present, incorporation of an OSA screening tool into the preoperative evaluation of all bariatric surgery candidates by the surgical team should be considered as an initial minimum means to consistently ensure that OSA is deliberately contemplated in an organized manner and to enhance detection of the most severe OSA cases. Because the "best" OSA screening tool is not known, the decision about which tool to use must be made locally by the bariatric surgery team, ideally with guidance from their sleep medicine colleagues. The designation of OSA status by the screening tool output must be integrated with other pertinent information, such as collateral observations about the patient's breathing during sleep from the bed partner, history of airway difficulties with previous anesthetics, anticipated surgical approach (open versus laparoscopic), and comorbidity burden.

Laboratory-based, technologist-attended polysomnography remains the diagnostic "gold standard" modality for diagnosis of OSA.[47] A routine role for preoperative sleep testing (polysomnography or home sleep apnea testing) continues to be debated. Proponents for such a practice cite the high prevalence of OSA among severely obese patients, the potential for perioperative complications from unrecognized OSA, and the limited accuracy of clinical impression alone in OSA diagnosis. Opponents point to the lack of data demonstrating improved postoperative outcomes with preoperative initiation of CPAP in patients undergoing bariatric surgery, the uncertainty regarding the relative contribution of OSA to postoperative complications in this population, and potential clinical overuse of such testings in patients deemed to be at low risk by clinical impression or OSA screening tool. The AACE-TOS-ASMBS bariatric surgery clinical practice guidelines[36] are vague: "routine preoperative screening with polysomnography should be considered." Performing sleep studies in all bariatric surgery candidates without exception, however, seems overly rigid. The reality is that if an OSA screening tool is systematically implemented, most patients will face the prospect of sleep testing before bariatric surgery because they will be judged to be at high risk by the OSA screening tool. For instance, in the study examining the performance of STOP-Bang in obese patients,[39] just 5% of preoperative patients with a BMI of 35 kg/m^2 or higher scored less than 2. Additionally, sleep testing may be needed to establish OSA as a weight-related comorbid condition in building justification for bariatric surgery, or if therapy for OSA is desired regardless of whether bariatric surgery is ultimately performed. In those patients without collateral sleep history or who are suspected of downplaying OSA symptoms, overnight oximetry monitoring may be an intermediate step between OSA screening by history and physical examination and a formal sleep study—that is, polysomnography. In bariatric surgery programs in which preoperative polysomnography or home sleep apnea testing is not routinely performed in every patient, preoperative consultation with a sleep specialist about the need for further sleep testing often is appropriate and is specifically recommended in situations of ambiguity about OSA status or its perioperative importance. Private insurers are increasingly mandating home sleep apnea testing in cases of suspected OSA, and the adult bariatric surgery candidate with a high pretest probability of having moderate to severe OSA and without significant comorbid cardiopulmonary disease may be an appropriate candidate for home sleep apnea testing followed by initiation of autoadjusting CPAP.[48] However, laboratory-based polysomnography is indicated and typically is covered by insurance carriers for the very obese suspects with OSA (BMI greater than 45 to 50 kg/m^2) or those with suspected OHS, because of the possible need for attended titration of modalities other than CPAP, such as bilevel (i.e., biphasic) positive airway pressure (BiPAP) and supplemental oxygen.

A high index of clinical suspicion should be maintained for the presence of OHS, because affected patients require more careful presurgical consideration. For several reasons, patients with OHS would be expected to be at higher risk than eucapnic obese patients with OSA during bariatric surgery. The diminished ventilatory responsiveness to hypoxia and hypercapnia in this patient group leads to increased sensitivity to sedatives and opioids, potentially greater problems with weaning from mechanical ventilation, and development of life-threatening obstructive apnea events, as well as acute worsening of hypercapnia with supplemental oxygen therapy unaccompanied by any ventilatory support such as BiPAP.[49] Rates of comorbid conditions such as systemic hypertension, pulmonary hypertension, cor pulmonale, and angina are higher in patients with OHS than in eucapnic obese patients.[50] OHS is a risk factor for development of venous thromboembolic (VTE) disease, which is a leading cause of postoperative death in bariatric surgery.[51]

OHS can be challenging to diagnose before bariatric surgery because patients may not always appear dramatically different than eucapnic obese patients with OSA. Patients

with OHS more commonly have lower extremity edema, report moderate to severe dyspnea on exertion, exhibit higher AHIs and more profound minimum oxyhemoglobin saturations during sleep, spend greater time with an oxyhemoglobin saturation lower than 90% during sleep, have lower awake oxyhemoglobin saturations, are afflicted with higher BMIs, demonstrate greater restrictive changes on pulmonary function testing, and use more health care resources compared with eucapnic obese patients with OSA[4,50] Serum bicarbonate of 27 mEq/L or more (reflecting metabolic compensation for chronic respiratory acidosis) is a sensitive but not specific marker for OHS in the obese patient with OSA.[52] If OHS is suspected on the basis of any or all of these factors, the following tests are recommended: arterial blood gas (hypoventilation manifests as hypercapnia, the severity of which should be determined), pulmonary function tests and chest radiograph (to search for other causes of chronic hypoventilation), echocardiogram (to assess the right heart pressures and function), complete blood count (to detect erythrocytosis), thyroid function tests (to rule out hypothyroidism, if not already done as part of routine testing of the obese patient), and polysomnography[4] (see Chapters 113 and 114).

PAP therapy for moderate to severe OSA (AHI of 15 or higher) should be initiated preoperatively. Case-by-case decisions about CPAP are necessary in milder forms of OSA (e.g., position-dependent OSA); CPAP may be recommended preoperatively for a patient with AHI of 5 to 14 who also is hypersomnolent or in whom the degree of desaturation during obstructive apneas/hypopneas is of greater clinical concern, perhaps in the presence of comorbid conditions such as pulmonary hypertension. The elective nature of bariatric surgery should allow for follow-up assessment of titrated PAP therapy in the patient with newly diagnosed OSA or OHS before surgery. In many bariatric surgery centers, adherence to the preoperative PAP regimen for the candidate with OSA is a mandatory prerequisite; accordingly, failure to comply with this recommended therapy is a deal-breaker, because it points to the likelihood of poor adherence to other postoperative care requirements. The minimal preoperative PAP trial duration for achieving PAP acclimatization and garnering improvement in physical status is not known,[53] but because patterns of PAP use may be established within the first week of therapy,[54] close follow-up monitoring in the first few weeks to document adherence, address problems, and assess response is advised.[55] In the patient with OHS, a repeat measurement of arterial blood gases after 4 weeks of PAP therapy[4] may allow discontinuance of supplemental oxygen with confirmation of an adequate therapeutic response or may indicate the need for tracheostomy with or without ventilation when PAP fails to effect improvement.

Preoperative Assessment for Bariatric Surgery in the Patient with Established Obstructive Sleep Apnea

Patients with a known diagnosis of OSA and who are already on an established PAP regimen at the time they begin to explore the option of bariatric surgery should be asked preoperatively about suboptimal compliance, technical difficulties, persistent symptoms despite PAP, and increases in weight since their last sleep evaluation. Identification of any of these issues should prompt referral to a sleep specialist. Follow-up polysomnography is indicated for patients with substantial weight gain (i.e., 10% of baseline body weight or more) and recurrent OSA symptoms despite PAP adherence.[47] Asymptomatic, PAP-adherent patients generally can proceed to surgery. They should be advised that PAP use will be required postoperatively and that they should bring their equipment to the hospital. Settings for PAP with or without supplemental oxygen usually are maintained postoperatively at preoperative levels, although acute adjustments may be necessary depending on the cumulative effects of factors such as opioid requirements and postoperative pulmonary disorders (e.g., VTE, pneumonia). Those patients with OSA treated previously with upper airway surgery who remain symptomatic, or in whom objective evidence of sleep-disordered breathing resolution is lacking, should be assumed to remain at risk for residual OSA and may benefit from a preoperative evaluation by a sleep specialist before bariatric surgery.[42,56] Such patients should be advised that temporary application of PAP may be necessary in the immediate postoperative period if upper airway obstruction occurs. The possibility of temporary outpatient use of PAP after surgery also should be discussed with bariatric surgery candidates who use an oral appliance for OSA management, because it may not be feasible to use their dental device immediately postoperatively.

Common Bariatric Surgical Procedures

Bariatric surgery procedures have been historically grouped into three categories based on anatomic components: predominately malabsorptive procedures, predominately restrictive procedures, and procedures with both malabsorptive and restrictive components. A majority of these operations are now performed by a less invasive, small-incision, laparoscopic approach. Ongoing research in animal models and human trials is aimed at further elucidation of the underlying mechanisms of action of bariatric surgery, which may ultimately allow a more sophisticated grouping of the surgical procedures based on their impacts on endocrine, neuronal, and behavioral physiologic variables.[57]

Roux-en-Y gastric bypass (RYGB) (Figure 121-2) combines creation of a small gastric pouch with modest intestinal or small bowel bypass. A traditional RYGB consists of transection of a small (15-mL) proximal gastric pouch along the lesser curvature of the stomach from the larger gastric segment, combined with a modest (encompassing 60 to 150 cm) intestinal bypass. The Roux-en-Y configuration allows biliopancreatic secretions and digestive juices to pass through the bile duct into the duodenum and then merge with the alimentary stream passing down from the stomach at the Y-type connection. The lengths of both the Roux and biliopancreatic limbs can be varied to produce more malabsorption. Most weight is lost in the first year, with long-term weight loss stabilizing at 2 to 3 years. Approximately 80% of patients typically experience weight stabilization, usually slightly above weight nadir, approximately 3 years after surgery. The remaining 20% of patients slowly regain excess weight over longer-term follow-up, and they risk becoming surgical failures.

The *laparoscopic adjustable gastric band* is an inflatable silicone prosthetic device that is placed around the top portion of the stomach, just below the esophagus (Figure 121-3), and restricts the upper stomach size to a small volume. The band is attached to a reservoir, with a port placed under the skin on the abdominal wall, and the inner lining of the band is a balloon that is adjustable by the addition or removal of saline

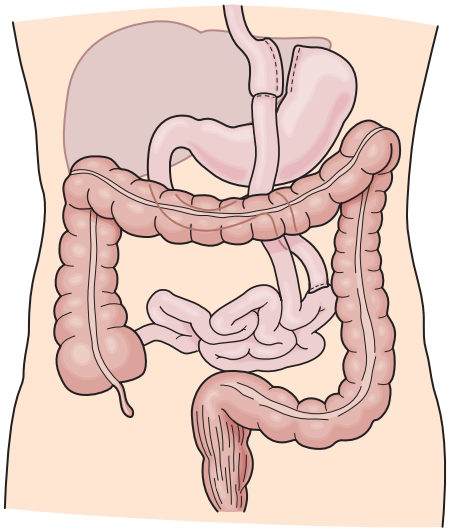


Figure 121-2 Roux-en-Y gastric bypass.

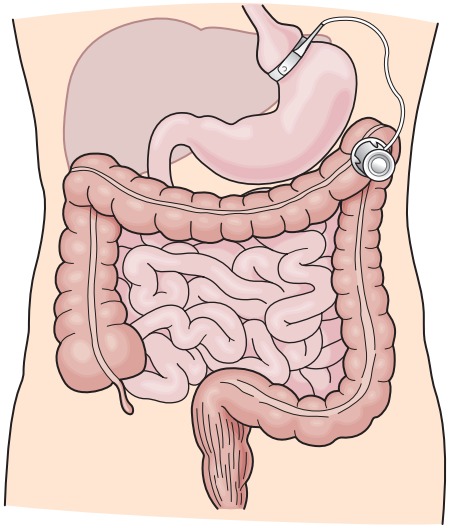


Figure 121-3 Gastric band procedure.

from the reservoir port. Inflation of the band increases the restriction of gastric outlet size and food flow. Postoperative management for patients who have undergone this procedure entails frequent follow-up visits for band adjustments/reservoir fills and strict adherence to dietary guidelines and lifestyle modification to achieve consistent weight loss. The weight loss trajectory after banding procedures is more gradual, with less weight lost than after RYGB. The favorable aspects of this procedure are that it is less invasive and requires less operating time, and that the band is both adjustable and removable. Use of the band procedure is declining with the increase in number of alternative surgical options being offered.[58]

The most recent major bariatric procedure to be introduced and growing in usage is the *vertical sleeve gastrectomy* (VSG).[59] This operation is a 70% vertical gastric resection, creating a long and narrow tubular gastric reservoir with no intestinal bypass. In many ways, VSG is intermediate between bypass and banding in terms of complexity, risk, and weight loss results. Data on long-term results (beyond 1 to 2 years) with VSG are lacking.[60]

The less commonly used *biliopancreatic diversion* (BPD) and *BPD with duodenal switch* (BPDDS) procedures, which result in an extreme degree of malabsorption, are reserved for the treatment of "super-obese" patients. BPD combines a partial, subtotal gastrectomy and a very long Roux-en-Y anastomosis with a short common channel for nutrient absorption. With this procedure, patients can eat much larger quantities of food and still achieve and maintain weight loss. Disadvantages to the procedure include loose and foul-smelling stools, intestinal ulcers, anemia, vitamin and mineral deficiencies, and possible protein-calorie malnutrition. Because of these potential problems, patients who undergo BPD require lifelong dietary supplementation and close follow-up monitoring. With similar weight loss and complications, the BPDDS is a hybrid operation that combines a gastric sleeve resection with a long intestinal bypass in the Roux-en-Y configuration. In this procedure, ulcer rate is reduced, and dumping syndrome (the constellation of nausea, vomiting, abdominal pain or cramping, diarrhea, bloating, fatigue, palpitations, lightheadedness, sweating, and anxiety beginning within 15 to 30 minutes after eating) is eliminated by leaving intact the first portion of the intestine in the alimentary stream. BPD and BPDDS procedures are the most major and technically difficult procedures performed for weight loss and consequently should be offered only by experienced surgeons, and to patients who are able to undertake lifelong follow-up.

Debate continues regarding the selection of a specific procedure type for any given patient, and predictive data to meaningfully guide these decisions are lacking. Patients are provided with general guidelines about the different potential mechanisms of action, percent weight loss over time, and morbidity profile among procedures when making a final decision regarding surgery. The optimal choice of procedure depends in part on the expertise of the surgeon and the clinical facility, patient preference, and risk stratification.[36]

CLINICAL COURSE

Management of Obstructive Sleep Apnea Immediately after Bariatric Surgery

Many details regarding optimal care of the patient with OSA immediately after bariatric surgery remain unclear. The following general recommendations are based on experience, consensus expert opinion for generic surgical care,[41-43,56] and a limited peer-reviewed literature.

Airway extubation after bariatric surgery should be performed only when the patient is fully awake and alert and has demonstrated evidence of return of neuromuscular function (as evidenced by sustained head-lift for more than 5 seconds) and adequate vital capacity and peak inspiratory pressure.[43] Removal of the endotracheal tube should take place in the operating room, postanesthesia care unit (PACU), or special care unit so that airway control can be monitored closely and expertly addressed if lost.[43]

In the PACU, the patient should be maintained in the semiupright or lateral, not supine, position, if possible. Supplemental oxygen typically is provided under continuous pulse oximetry monitoring and titrated to the lowest level to maintain adequate oxygenation, especially in patients with OHS.

Ventilation also must be specifically monitored as supplemental oxygen may maintain adequate oxyhemoglobin saturation despite medication-exacerbated hypoventilation. Ventilation monitoring may include capnography, arterial blood gas testing, and scheduled assessments for respiratory events by PACU staff. In a study of a non–bariatric surgery perioperative patient population, recurrent respiratory events in the PACU powerfully predicted postoperative respiratory complications.[61] Respiratory events were scored during three consecutive 30-minute periods immediately after extubation and were defined as bradypnea (three or more episodes of fewer than 8 breaths/minute), apnea (one or more episodes of 10 seconds or longer of breathing cessation), desaturations (three or more episodes of oxyhemoglobin saturation below 90%), and pain-sedation mismatch (one or more episodes of high pain score and simultaneously high sedation score). Recurrent respiratory events meant that one or more of any of the PACU respiratory events occurred in at least two separate 30-minute time blocks and were associated with an odds ratio of 21 for postoperative respiratory complications.[61] In the PACU, PAP is instituted at the level prescribed before surgery in those patients who were using it preoperatively. In patients whose preoperative CPAP settings are not known or in whom initiation of CPAP is desired to address recurrent respiratory events emerging in the PACU, CPAP in autoadjusting mode can be applied or started at an empirically chosen level of 8 to10 cm H_2O and adjusted as needed, although acute initiation of such therapy in the PACU can be challenging in the PAP-naive patient. BiPAP, usually with oxygen, may be initiated to address acute hypoventilation. PACU staff must be capable of monitoring and managing PAP therapy, including addressing interface leaks and observing diligently for signs of breakthrough upper airway obstruction despite PAP, such as snoring, choking, witnessed apneas, cardiac dysrhythmias, or repetitive oxygen desaturation.

In the first 24 postoperative hours after bariatric surgery, patients are likely to be the most vulnerable to potential OSA-related complications,[62] although OSA propensity may be increased for at least several days after bariatric surgery because of the aggregate effects of ongoing sleep deprivation, rapid eye movement sleep rebound, and medication synergies.[56] Fortunately, the length of hospital stay usually is short (3.5 days and 1.6 days for gastric bypass and restrictive procedures, respectively).[63] Clinicians must consistently keep the possibility of OSA in mind in all patients as they consider postoperative analgesia, monitoring, oxygenation, and patient positioning[43] for the duration of the hospitalization. Systemic opioids should be used cautiously because of their ability to depress the respiratory drive and cause subsequent oxygen desaturation. The use of patient-controlled analgesia is controversial, although it may be an option if used without a basal rate and with restricted dosing. Nonsteroidal antiinflammatory agents may help decrease opioid dosing as recovery progresses but should be used cautiously in the postsurgical patient because of the enhanced potential for bleeding complications. Benzodiazepines should be avoided because of their negative effects on the respiratory control and upper airway musculature. Access to PAP should be available at all times during recovery—a seemingly obvious recommendation but one that may be overlooked by busy house staff, nurses unfamiliar with this ventilatory technique, and patients distracted or impaired by postoperative pain or pharmacologic obtundation. Properly trained health care staff should be readily available to assist patients in PAP device placement, to troubleshoot interface problems, to observe for breakthrough upper airway obstruction, and to reassure patients struggling with a new PAP regimen.

Continuous pulse oximetry monitoring after discharge from the PACU is recommended for all post–bariatric surgery patients for as long as they are deemed to be at increased risk, which may be defined as the duration of intravenous opioid use or an oral opioid dose of greater than 60 mg of codeine every 4 hours.[42] Oximetry data should be continuously observed at the bedside in a critical care or stepdown unit, by telemetry on a hospital ward, or by a dedicated, trained observer in the patient's room. Choosing the optimal monitoring site will depend on the interplay of multiple factors. It is reasonable to consider intensive care unit (ICU) care for the first 24 to 48 hours after bariatric surgery in patients with one or more of the following features: age older than 50 years, BMI greater than 60 kg/m^2, significant comorbid cardiopulmonary disease, brittle diabetes mellitus, severe OSA/OHS with worrisome record of suboptimal PAP compliance, sluggish emergence from anesthesia, and intraoperative complications. In the University HealthSystem Consortium evaluation, a review of the bariatric programs at 29 academic medical centers in the United States, 7.7% of patients undergoing gastric bypass and 1.1% of patients undergoing a restrictive procedure required ICU support postoperatively.[5] Operative approach to bariatric surgery also must be considered. Case series from experienced surgery teams have reported that patients with established and treated OSA undergoing laparoscopic bariatric procedures do not require routine postoperative admission to the ICU.[64,65]

Incentive spirometry should be encouraged. If oxygen desaturations occur despite an appropriate PAP regimen, supplemental oxygen should be added while the provider searches for an explanation, such as transient worsening of upper airway obstruction requiring adjustment of PAP settings, VTE event, atelectasis, aspiration, pneumonia, or anastomotic leak. Caution with the use of supplemental oxygen without PAP during sleep is advised, because this strategy provides no protection against upper airway obstruction and will blunt detection of a disordered breathing event by oximetry monitoring. Postoperative supine positioning also should be avoided; instead, the head of the bed should be kept elevated in a semi-Fowler position (to at least 30 degrees) at all times. All post–bariatric surgery patients should be considered to be at moderate to high risk for VTE events; accordingly, thromboprophylaxis with low-molecular-weight heparin or low-dose unfractionated heparin, along with application of intermittent pneumatic compression stockings, is indicated in all patients.[36,66] The frequency of VTE after bariatric surgery with thromboprophylaxis is low at less than 1%.[66] Because most VTEs occur after hospital discharge, the AACE-TOS-ASMBS advises extended chemoprophylaxis (duration unspecified) for patients at higher risk for such events, such as those with a history of VTE or reduced activity level.[36] Prolonged respiratory failure after bariatric surgery is uncommon, occurring in less than 1% of cases, according to data from the American College of Surgeons' National Surgical Quality Improvement program encompassing approximately 32,000 patients who underwent bariatric surgery between 2006 and 2008; the impact of OSA on the rate of

respiratory failure is not known, because OSA was not a risk factor assessed in the analysis.[67]

Laparoscopic bariatric surgery is performed in some patients in an ambulatory setting. A consensus statement from the Society for Ambulatory Anesthesia[41] warns against use of outpatient surgical procedures in patients with OSA if it is accompanied by a a nonoptimized comorbid condition; if an inability to control pain predominantly with nonopioid analgesic techniques can be anticipated; or if the patient displays unwillingness or inability to use PAP. Patients on PAP should be advised to bring their device to the ambulatory care facility for use during recovery. If the patient experiences recurrent respiratory events while in the PACU, hospital discharge should be delayed until the patient is observed to maintain adequate oxygenation (on PAP if necessary, if used preoperatively) in an unstimulated environment, preferably while sleeping.[43] Patients not on PAP should be advised to sleep exclusively nonsupine, and PAP users should wear their device during all sleep periods, including naps, for "several days" after surgery; all are advised to minimize use of opioids.

Benefits of Bariatric Surgery

Postoperative weight loss typically is reported as the mean percentage of excess weight loss, defined by the following formula:

$$(\text{Weight loss} \div \text{excess weight}) \times 100$$

where excess weight equals total preoperative weight minus ideal weight. In a review of 136 studies involving 22,000 bariatric surgery patients, Buchwald and colleagues[68] reported that the mean percentage of excess weight loss with bariatric surgery was 61.2%: 47.5% for gastric banding, 68.2% for gastric bypass (principally RYGB), and 70.1% for BPD and BPDDS. The mean decrease in BMI was 14.2 kg/m^2, whereas the mean decrease in absolute weight was 39.7 kg, similar to the 20- to 30-kg weight loss reported in the meta-analysis by Maggard and colleagues.[69] Comorbid conditions correspondingly decreased in severity with weight loss. Overall, diabetes mellitus completely resolved in 76.8%, hyperlipidemia decreased in degree in 70%, and arterial hypertension lessened in severity or resolved in 78.5%.[68] Weight loss, improvement with respect to comorbid conditions, and quality of life at 1 year, as well as rates of severe postoperative complications, are similar for RYGB and for VSG, both performed laparoscopically.[70] A retrospective cohort study comparing long-term mortality rates among 7925 patients who underwent gastric bypass and 7925 age-, sex-, and BMI-matched control subjects randomly selected from a state driver's license applicant registry demonstrated a 40% reduction in adjusted long-term mortality with bariatric surgery during a mean follow-up period of 7.1 years.[71]

No large randomized trials have compared bariatric surgery with medical management of obesity. The Swedish Obesity Study[72] was a large, prospective, nonrandomized, controlled trial that compared outcomes for 2010 obese subjects treated with bariatric surgery and for 2037 contemporaneously matched obese control subjects treated conventionally. At 2 years, weight had decreased by 23.4% in the surgery group but had increased by 0.1% in the control group, and after 10 years, weight had increased by 16.1% over presurgical weight in the surgery group but had increased by 1.6% in control subjects ($P < .001$ at both time points). Improvements in clinical indices of diabetes, hypertriglyceridemia, and hypertension were more favorable in the surgery group, and the surgery group exhibited lower 2- and 10-year incident rates of diabetes than those for the control group. Maximal weight loss in the surgery group was assessed after 1 to 2 years, and at 10 years the maximal average losses were 32% for gastric bypass, 25% for vertical banded gastroplasty, and 20% for banding. Overall mortality was lower in the surgery group than in the control group.[73]

Long-term Impact of Bariatric Surgery on Obstructive Sleep Apnea

Weight loss induced by bariatric surgery is consistently associated with reductions in AHI.[74] Buchwald and coworkers' meta-analysis[68] of selected bariatric surgery outcomes reported that OSA resolved or decreased in severity in 83.6%. The weighted (i.e., weighting results by sample size) mean change in the AHI was 40 (events/hour), with a range of 16 to 52.8. Enthusiasm over these results must be tempered by several methodologic concerns. *Improvement* and *resolution* with respect to OSA were not explicitly defined. The studies included in the meta-analysis are not entirely specified, but a review of studies from the inclusion period (1990 to June 2003) revealed that reduction in OSA symptoms probably was sufficient in some studies to assess OSA response (i.e., postoperative polysomnography was not required in all subjects), the timing of polysomnography after surgery was nonuniform, and the results probably were variably reported (e.g., only preoperative and postoperative apnea indices were described, not AHIs).

Studies published since June 2003[75-78] corroborate earlier series reporting that surgically induced weight loss is associated with symptomatic improvement in OSA when reassessment occurs approximately 1 year or longer after surgery. However, many patients have residual OSA. Even though gastric banding resulted in a mean AHI reduction of 23.4 (events/hour), Lettieri and colleagues[79] found that 23 patients (96%) still met criteria for OSA (AHI higher than 5), 20 (83%) continued to experience transient nocturnal hypoxia (oxyhemoglobin saturation below 90%), and 13 (54%) had persistent sleepiness (Epworth Sleepiness Scale scores higher than 10) despite an average weight loss of 54 kg at a mean of 418 days postoperatively. In a metaanalysis of 12 studies involving 342 patients, Greenburg and colleagues[80] found that the pooled mean BMI decreased by 17.9 kg/m^2 and AHI decreased by 38.2, but the residual AHI averaged 15.8. Since that meta-analysis, a small randomized, controlled trial pitting bariatric surgery (laparoscopic adjustable gastric banding) ($n = 30$ patients) against a conventional weight loss program (individualized dietary, physical activity, and behavioral programs) ($n = 30$ patients) demonstrated that although bariatric surgery produced significantly greater mean weight loss at 2 years than that achieved in the nonsurgical program (27.8 kg versus 5.1 kg), the reductions in AHI were statistically similar (25.5 versus 14; $P = .18$).[81] The mean residual AHI in the bariatric group was 39.5, and 73% of the bariatric group continued to have an AHI of 15 or higher. Accordingly, health care providers must remain vigilant for persistent OSA with a systematic postoperative follow-up program, because even dramatic changes in weight and symptoms do not guarantee objective cure of OSA. The optimal timing for postoperative polysomnography is not clear but depends in part on the

patient's weight loss evolution. The CPAP requirement for residual OSA is likely to fall by at least 2 to 4 cm H_2O in the year after surgery.[82] Autotitrating CPAP after surgery may bridge the patient to polysomnography and obviate subjective pressure reductions or serial sleep studies.

The American Academy of Sleep Medicine concluded that bariatric surgery may be adjunctive in the treatment for OSA, but it rates this recommendation as an option, meaning that bariatric surgery is of uncertain clinical use in the management of OSA.[83] This designation is based on the lack of data at the Sackett level of evidence I to III and the potential for perioperative complications.

Risks and Complications of Bariatric Surgery

The Longitudinal Assessment of Bariatric Surgery (LABS) Consortium conducted a prospective, observation study of outcomes of bariatric surgical procedures at 10 clinical sites in the United States from 2005 to 2007. The rate of operative mortality, defined as death within the first 30 days, was 0.3% among 4610 consecutive patients who underwent RYGB or laparoscopic adjustable gastric banding: None of the 1198 patients who had a laparoscopic adjustable gastric band procedure died, 0.2% of the 2975 patients who underwent laparoscopic RYGB died, and 2.1% of the 437 patients who underwent open RYGB died.[84] A composite endpoint of death, deep vein thrombosis or venous thromboembolism, reintervention, and failure to be discharged by 30 days after surgery was reported in 4.1% of patients: 1% in the laparoscopic adjustable gastric banding group, 4.8% in the laparoscopic RYGB group, and 7.8% in the open RYGB group. Factors that were each independently associated with an increased risk of the composite endpoint were a history of venous thromboembolic disease, impaired functional status, and extreme values of BMI. Box 121-1 lists the postoperative adverse events from bariatric surgery, which can be grouped as early and late. A dreaded complication is anastomotic leak: In the University HealthSystem Consortium evaluation,[63] the anastomotic leak rate for gastric bypass procedures was 1.6%.

The extent to which OSA is linked to complications after bariatric surgery is not fully known. In a review of data for more than 3000 patients, OSA, older age, male sex, and revision gastric bypass were found to be independent predictors for anastomotic leak,[85] whereas OSA, hypertension, and less surgeon experience were identified by multivariate analysis as predictors of postoperative complications in a series of nearly 200 patients undergoing laparoscopic RYGB.[86] In the LABS analysis, OSA also was independently associated with increased risk for an adverse outcome, with a composite of such outcomes defined as the endpoint.[84] Accordingly, OSA has been linked to increased cost of postoperative care[87] and higher risk for prolonged postoperative hospital stay.[88] In other studies, however, the investigators have not identified OSA as an independent predictor of complications after bariatric surgery.[89-91] These reports are challenging to interpret and compare because of differences in procedures used and uncertainty over how aggressively OSA was pursued preoperatively and managed postoperatively, and many are single-center and possibly underpowered retrospective reviews, thus providing a lower grade of evidence. PAP initiated immediately after bariatric surgery does not appear to increase risk for anastomotic leaks[92,93]

Box 121-1 COMPLICATIONS OF BARIATRIC SURGERY

Complications Common to All Bariatric Procedures

Early (up to 30 days after surgery)

Venous thromboembolic disease
Bleeding
Anastomotic leaks
Wound infections
Persistent nausea/vomiting, dehydration
Regional abdominal organ trauma
Incisional and internal hernias
Bowel obstruction
Atelectasis
Pneumonia
Cardiac dysrhythmias
Urinary tract infection
Death

Late (beyond 30 days after surgery)

Incisional and internal hernias
Bowel obstruction from adhesions
Nutritional deficiencies
Anastomotic strictures and marginal ulcers or erosions
Cholelithiasis
Anemia
Persistence or recurrence of obstructive sleep apnea
Need for body contouring
Weight regain

Procedure-Unique Complications/Adverse Effects

Roux-en-Y Gastric Bypass

Dumping syndrome

Laparoscopic Adjustable Gastric Banding

Band slippage or erosion
Port or device malfunction

Vertical Sleeve Gastrectomy

Refractory reflux

Biliopancreatic Diversion

Loose, foul-smelling stools
Protein-calorie malnutrition

PITFALLS AND CONTROVERSY

Controversy remains about the impact of OSA on bariatric surgery complications and thus to what extent must OSA be sought and treated preoperatively. Definitive data are not available. Bariatric surgery clinical practice guidelines[36] stipulate that OSA should be considered in all bariatric surgery candidates—but does that mean polysomnography is mandatory and that if it yields a positive result, PAP is required? The answer to both of these questions is likely to be "no." Instead, the history and physical findings pertinent to OSA, perhaps initially organized by a screening tool sensitive to OSA so that the search is consistent and systematized, must be combined with consideration of a host of other factors, both patient-related (comorbidity burden; OSA symptom severity; OHS likelihood) and procedure-related (open versus laparoscopic; inpatient versus ambulatory; anticipated postoperative opioid requirements), in deciding how to proceed. Those patients

judged to be at low risk for having OSA as indicated by the collective clinical information (e.g., STOP-Bang score less than 3) can proceed directly to bariatric surgery without further sleep testing[42] *provided that* postoperative precautions are in place (e.g., careful monitoring in the PACU; minimization of opioid/sedative use; head of bed elevation; incentive spirometry) and the bariatric team is prepared to address OSA should it manifest during the immediate postoperative period. Patients deemed to be at intermediate or high risk for having OSA should proceed to a formal sleep assessment, with the sleep specialist guiding the ordering of sleep testing and the interpretation of findings.[42] PAP is started preoperatively for treatment of moderate to severe OSA (AHI of 15 or higher) and OHS. Preoperative PAP initiation allows the patient to begin accruing its neurobehavioral and cardiovascular benefits while working through other preparatory steps typically required for bariatric surgery. Furthermore, retrospective data suggest that it may decrease the risk of post–bariatric surgery complications[94] and is recommended by the AACE-TOS-ASMBS guidelines.[36]

CLINICAL PEARLS

- The sleep clinician should be mindful that bariatric surgery candidates are likely to have OSA, which requires careful consideration during the preoperative evaluation as well as in the postoperative period.
- Post–bariatric surgery patients should be expected to lose 20 to 50 kg by 1 to 2 years postoperatively, which current studies indicate should be accompanied by a 50% to 75% reduction in AHI and a drop in required PAP levels.
- Autotitrating CPAP may be a useful management modality as weight decreases after surgery.
- Despite dramatic weight loss, OSA may persist in many patients, so follow-up polysomnography is advised to reassess for this condition and to guide decisions about longer-term PAP use.

SUMMARY

The prevalence of obesity, a leading cause of preventable disease and death, is increasing. Nearly 70% percent of Americans currently are overweight or obese. Rates of overweight and obesity are higher among Mexican Americans and non-Hispanic black Americans than among non-Hispanic whites and Asian Americans. Excess weight is the strongest risk factor for OSA because of its adverse impact on upper airway neuromuscular function and anatomy. Bariatric surgery, comprising a variety of procedures that limit food absorption or restrict intake (or both), is indicated for severely obese patients in whom an adequate exercise and diet program has failed to achieve results and who have either a BMI of 40 kg/m^2 or greater or a BMI of 35 kg/m^2 or above in conjunction with one or more obesity-related severe comorbid conditions. The mean percentage of excess weight loss with bariatric surgery is approximately 60%, and in patients with major obesity-related conditions, such as diabetes mellitus and hypertension, consistent improvement in clinical indices is to be expected. Thirty-day mortality rate for bariatric surgery is less than 1%. OSA is almost universally present in bariatric surgery candidates, sometimes in the context of OHS. The immediate post–bariatric surgery setting may exacerbate OSA, whereas OSA and its associated conditions may exacerbate challenges to the patient's immediate postoperative well-being. Systematic screening for OSA should therefore be a required component of preparation for bariatric surgery, with addition of a formal sleep evaluation for patients deemed to be at higher risk for OSA. Symptomatic improvement follows, but the OSA does not usually resolve with bariatric surgery–induced weight loss. The sleep clinician plays an important role in the bariatric surgery process, in preoperatively collaborating with the surgical team to identify OSA and by helping to define its significance, determining which patients need OSA treatment, and initiating or optimizing OSA therapy to postoperatively determine the degree of residual OSA and the need for additional treatment.

Selected Readings

Abbas M, Cumella L, Zhang Y, et al. Outcomes of laparoscopic sleeve gastrectomy and roux-en-Y gastric bypass in patients older than 60. *Obes Surg* 2015;**25**(12):2251–6.

Arterbarn DE, Courcoulas AP. Bariatric surgery for obesity and metabolic conditions in adults. *BMJ* 2014;**349**:g3961.

Dudley KA, Tavakkoli A, Andrews RA, et al. Interest in bariatric surgery among obese patients with obstructive sleep apnea. *Surg Obes Rel Dis* 2015;[Epub ahead of print].

Kushner RF, Ryan DH. Assessment and lifestyle management of patients with obesity. *JAMA* 2014;**312**:943–52.

Lettieri CJ, Eliasson AH, Greenburg DL. Persistence of obstructive sleep apnea after surgical weight loss. *J Clin Sleep Med* 2008;**4**:333–8.

The Longitudinal Assessment of Bariatric Surgery (LABS) Consortium. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med* 2009;**361**:445–54.

Mechanick JI, Youdim A, Jones DB, et al. Clinical practice guidelines for the perioperative nutritional, metabolic, and nonsurgical support of the bariatric surgery patient—2013 update: cosponsored by American Association of Clinical Endocrinologists, The Obesity Society, and American Society for Metabolic and Bariatric Surgery. *Obesity* 2013;**21**:S1–27.

Nguyen NT, Nguyen B, Gebhart A, Hohmann S. Changes in the makeup of bariatric surgery: a national increase in use of laparoscopic sleeve gastrectomy. *J Am Coll Surg* 2013;**216**:252–7.

North American Association for the Study of Obesity and the National Heart, Lung, and Blood Institute. *The practical guide: identification, evaluation, and treatment of overweight and obesity in adults*. NIH publication 00-4084. Bethesda (Md.): National Institutes of Health; 2000.

Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-12. *JAMA* 2014;**732**:806–14.

Piche ME, Auclair A, Harvey J, et al. How to choose and use bariatric surgery in 2015. *Can J Cardiol* 2015;**31**:153–66.

Schwartz AR, Patil SP, Laffan AM, et al. Obesity and obstructive sleep apnea: pathogenic mechanisms and therapeutic approaches. *Proc Am Thorac Soc* 2008;**5**:185–92.

Sjostrom L, Lindroos A-K, Peltonen M, et al. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med* 2004;**351**:2683–93.

Sjostrom L, Narbro K, Sjostrom CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007;**357**:741–52.

Stefater MA, Wilson-Perez HE, Chambers AP, et al. All bariatric surgeries are not created equal: insights from mechanistic comparisons. *Endocr Rev* 2012;**33**:595–622.

A complete reference list can be found online at ExpertConsult.com.

Chapter 122

Sleep and Breathing at High Altitude

Vahid Mohsenin; Shahrokh Javaheri; Jerome A. Dempsey

Chapter Highlights

- Exposure to high altitude imposes significant strain on the cardiopulmonary system and the brain. As a consequence, sojourners to high altitude frequently experience sleep disturbances, often reporting restless and sleepless nights.
- At altitudes above 3000 meters, almost all healthy subjects develop periodic breathing, especially during non–rapid eye movement (NREM) sleep.
- Sleep architecture gradually improves after acclimatization to altitude, with increased NREM and rapid eye movement (REM) sleep despite persistence of periodic breathing.
- The primary reason for periodic breathing at altitude is a hypoxia-induced increase in chemoreceptor sensitivity to changes in Pa_{CO_2}—both above and below eupnea, leading to periods of apnea and hyperpnea.
- Acetazolamide improves sleep consolidation and periodic breathing through development of metabolic acidosis and induced hyperventilation, decreasing the plant gain and widening the P_{CO_2} reserve. This widening of the P_{CO_2} reserve impedes development of central apneas during sleep.
- Benzodiazepines and other gamma-aminobutyric acid receptor antagonists such as zolpidem improve sleep without affecting breathing pattern or cognitive function.
- Promising effects of other modalities to improve high-altitude periodic breathing, including noninvasive ventilation and increasing dead space breathing, remain to be further investigated in humans.

EXPOSURE TO HIGH ALTITUDE

Each year several million people worldwide travel from areas of low elevation to altitudes over 2500 meters (8200 ft). Empirically, and because of higher risk for altitude illnesses, 2500 meters has been used as the threshold for high-altitude illnesses. However, polysomnography at altitudes of both 1630 meters (5348 ft) and 2590 meters (8497 ft) shows decreased non–rapid eye movement (NREM) stage N3 sleep (slow wave sleep) and an increased apnea-hypopnea index (AHI) in the form of period breathing compared with the altitude of 490 meters (1608 ft). Interestingly, these healthy male subjects reported no changes in their total sleep time, excessive sleepiness, or acute mountain sickness (AMS) symptoms.[1]

With the increasing popularity of mountain sports such as skiing, climbing, and snowshoeing, and the latest trend in adventure travel to places like the Andes and Himalayas, it is expected that the incidence of high-altitude exposure will continue to grow. These often rapid ascents of nonacclimatized individuals place them at an increased risk for AMS, insomnia, and sleep-disordered breathing.

This chapter focuses on the effect of high altitude on sleep and respiratory control systems, the clinical disorders associated with such effects, and the optimal treatments. Relevant features of acute physiologic adjustment to high altitude are summarized; then the characteristics of the sleep disturbance at high altitude, its pathogenesis, and therapeutic interventions are reviewed. Although much of the focus is on sleep after acute ascent to high altitude, alterations in sleep during long-term altitude exposure are also mentioned. Information in this chapter may also be relevant to the pathophysiology of central sleep apnea at low altitude (see Chapters 15, 16, and 110).

PHYSIOLOGIC RESPONSES TO HIGH ALTITUDE

Primary among the changes in physical environment that attend the ascent to high altitude is a decrease in barometric pressure such that, although the fractional concentration of O_2 is similar to that at sea level, O_2 tension—the product of fractional concentration and barometric pressure—is reduced (Figure 122-1). This decreased O_2 tension of ambient air presents a threat to arterial and tissue oxygenation and elicits a series of responses that act to minimize tissue hypoxia. These responses consist of early increases in ventilation and cardiac output and, during more prolonged exposure, rises in circulating red cell concentration and adaptive changes in peripheral tissue, including increased spatial density of capillaries and mitochondria. Also see Chapter 18 for a detailed review of the physiology related to high altitude.

Increased Ventilation

The earliest, and one of the most important, of these responses is increased ventilation, which acts to minimize the extent of alveolar hypoxia and arterial hypoxemia in the face of a decrease in ambient O_2 tension (Figure 122-2). The acute

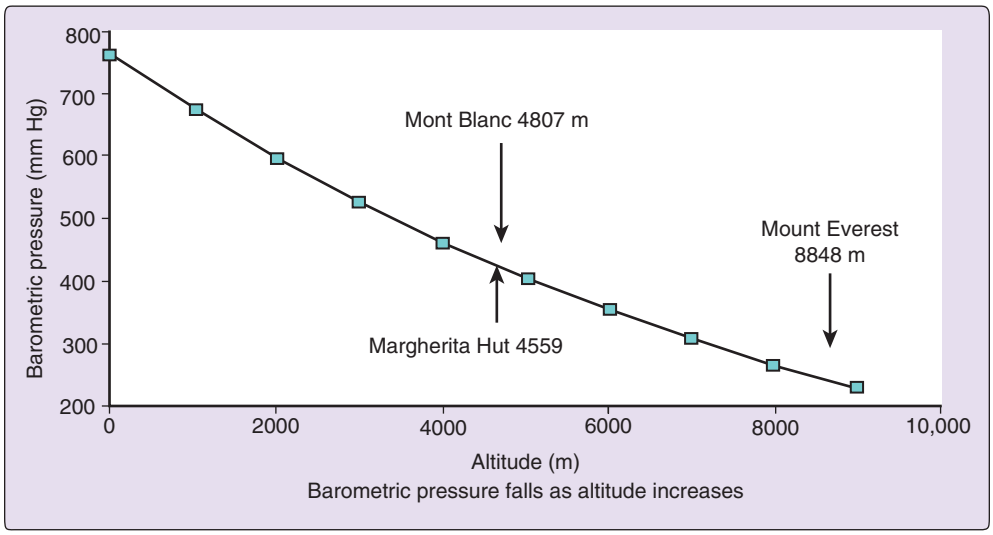


Figure 122-1 Relationship between altitude and barometric pressure. Inspired oxygen pressures at Margherita Hut, Italy, on the summit of Mont Blanc France, and Mount Everest are 84, 82, and 43 mm Hg, respectively.

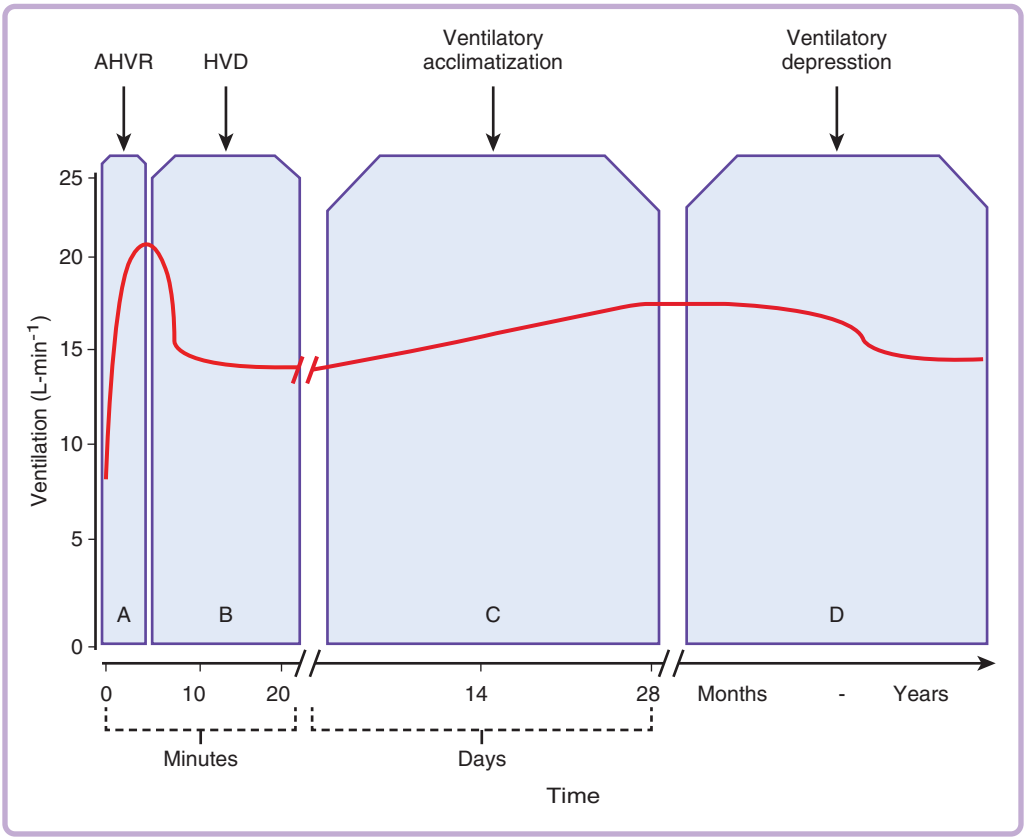


Figure 122-2 Temporal changes in ventilation on acute and prolonged exposure to high altitude. The acute hypoxic ventilatory response (AHVR) is followed within minutes by hypoxic ventilatory decline (HVD) before ventilatory acclimatization to hypoxia occurs. (From Ainslie PN, Lucas SJ, Burgess KR. Breathing and sleep at high altitude. *Respir Physiol Neurobiol* 2013;188:233–256.)

ventilatory response to hypoxia (acute hypoxic ventilator response; AHVR) is followed within minutes by hypoxic ventilatory decline (HVD) before ventilatory acclimatization to hypoxia (VAH) occurs. The exact mechanisms causing HVD are uncertain; however, its occurrence could result from elevations in cerebral blood flow (CBF) increasing CO_2 and hydrogen ion [H^+] washout, resulting in decreased central chemoreceptor sensitivity and/or neural stimulus response.[2,3] The development of HVD has also been attributed, at least in part, to an increased peripheral chemoreflex threshold to the isocapnic hypoxic stimulus.[4] The magnitude of the ventilatory response to hypoxia increases with increasing altitude, but it

also varies considerably among individuals at a fixed altitude. This variability, in part, reflects intrinsic, interindividual differences in the strength of the basal (preascent) ventilatory response to hypoxia.[5]

Ventilation progressively increases over several days after ascent to high altitude (see Figure 122-2). This gradual increase occurs despite the fact that the increasing ventilation is lessening hypoxia, the presumed stimulus to breathing, as well as increasing hypocapnic alkalosis, which is a ventilatory inhibitor. This is the phenomenon of ventilatory acclimatization to high altitude, which is manifested as a progressive decrease in arterial P_{CO_2} (P_{aCO_2}) with increasing ventilation over several days. On restoration of normoxia, hyperventilation continues but slowly dissipates over several days.[6] Although the mechanism of such acclimatization is debated, studies in humans and animals suggest that increased hypoxic sensitivity of the carotid body may be a major contributor.[7] In any case, it is during the early phase of altitude adjustment, shortly after ascent, that sleep disturbances appear to be most marked; they tend to improve during the period of acclimatization.

Periodic Breathing

In the early and mid-nineteenth century, Cheyne and Stokes, respectively, described the crescendo-decrescendo breathing pattern in cardiac patients that now bears their names (see Chapter 129). That periodic breathing is frequent during sleep in normal individuals at high altitude was observed shortly thereafter by Tyndall in 1857, by Egli-Sinclair and Mosso in 1893 and 1894, and by Douglas and colleagues[8] in 1913, and it continues to be a consistent finding in current studies of sleep after ascent to high altitude.[9-14]

The major effects of sleep on ventilatory control include removal of the "wakefulness stimulus"—which is specifically associated with inhibition of motor output to dilator musculature of the pharyngeal airway, resulting in significant increases in upper airway resistance; critical dependence of ventilatory control on P_{aCO_2} and the unmasking of an apneic threshold for P_{CO_2} residing within a few millimeters of mercury (mm Hg) below waking levels of eupneic P_{aCO_2}, and a propensity for periodic breathing in the form of ventilatory overshoot and undershoot.[15] The temporal pattern of periodic breathing and its linkage to sleep stages show night-to-night variation and considerable intersubject differences.[11,13,16-21] Periodicity is usually evident early in sleep, during light sleep stages (NREM sleep N1 to N2), and can persist despite improvement in sleep architecture with increases in slow wave sleep, rapid eye movement (REM) sleep, and a reduction in the arousal index (Figure 122-3).[22] Periodic breathing at high altitude may also occur in wakefulness, especially during periods of drowsiness.[23-25] Periodic breathing at high altitude is different from the typical crescendo-decrescendo pattern in tidal volume observed in heart failure[26] (see later), although has some resemblance to the irregular pattern of breathing associated with opiate use.[27]

Hyperventilation during NREM or REM sleep begins immediately on hypoxic exposure and intensifies with time.[9,28] Within the initial 10 minutes of hypoxia in the sleeping human, tidal volume begins to oscillate, ultimately resulting in cluster-type periodic breathing with a few augmented tidal volumes interspersed with apneas (Figure 122-4). Hypoxemia augments CO_2 chemosensitivity below and above eupnea. Consequently, P_{aCO_2} is closer to the apneic threshold P_{CO_2}, and therefore, a small transient fall in P_{aCO_2} (e.g., after an arousal with hyperpnea) results in an apnea. During these periodic cycles, arterial oxygen saturation (S_{aO_2}) also oscillates and often—depending on the altitude—falls to a point on the steep portion of the oxygen dissociation curve. The periodic breathing pattern characteristically has a short cycle length, ranging from 12 to 34 seconds, which progressively shortens with increasing altitude[29-31] and in this manner differs from the longer cycles of 40 to 90 seconds in patients with heart failure.[32] The relatively shorter cycle length of periodic breathing at high altitude compared with that observed in heart failure is thought to be attributable to the absence of a prolonged circulation time seen in heart failure patients. During the breathing clusters, the most obvious aspect of the periodicity is the oscillation of tidal volume, with a less obvious change in breathing frequency.[29] Like periodic breathing at sea level, periodicity at altitude is often initiated by movement, arousal, or a deeper breath with a resultant transient decrease in the prevailing P_{aCO_2}.

The most striking influence of sleep stage on periodic breathing at high altitude observed in most,[13,22,33] but not all,[20,34] studies is that the breathing periodicity promptly and consistently decreases in REM sleep. This is similar to the relative rarity of periodic breathing and central apnea in heart failure[26] and also in association with opioids.[27] This is attributable to a widening of the P_{CO_2} reserve in REM sleep, making it less likely for the prevailing P_{CO_2} to reach the apneic threshold P_{CO_2}.[26]

Effects of Hypoxia and Hypercapnia

During sleep, ventilation and oxygenation fall below waking values, and these relative changes are similar at high altitude and at sea level.[35] However, the important difference is that at high altitude, basal (awake) oxygenation is lower, chemosensitivity to CO_2 is increased, and P_{aCO_2} is closer to the apneic threshold P_{aCO_2}. Arterial oxygen tensions fall closer to the descending limb of the oxygen dissociation curve, where values are nearer the threshold for stimulation of ventilation. Similarly, because of increased CO_2 chemosensitivity below eupnea, CO_2 tensions fall to values nearer the apnea threshold, below which breathing ceases during sleep.[36] As a result, small variations in gas tensions have much greater effects on ventilation at high altitude, with greater stimulation by hypoxia and greater inhibition by transient hypocapnia.

Central apnea and periodic breathing during sleep occur within minutes of hypoxic exposure (Figure 122-5). As shown in the sleeping sojourner to 4300 meters (14,108 ft), hyperventilation in response to the reduced P_{aO_2} occurs first, followed by oscillations in tidal volume and minute ventilation ($\dot{V}E$) between hypopnea and hyperpnea and then—usually following a single augmented inspiration—full-blown periodic breathing occurs with clusters of large breaths interspersed with apneas at regularly occurring cycle lengths averaging about 20 to 25 seconds. This periodic pattern continues over several nights at high altitude. If normal oxygenation is suddenly restored (via increased inspired oxygen concentration, F_{IO_2}), periodicity continues for a brief period with prolonged apneic length; then, as hyperventilation abates and P_{aCO_2} rises, periodicity resolves, and a rhythmic, stable breathing pattern is restored. The causes of this hypoxia-induced cluster-type periodic breathing pattern involve responsiveness to CO_2 in both the ventilatory overshoot and

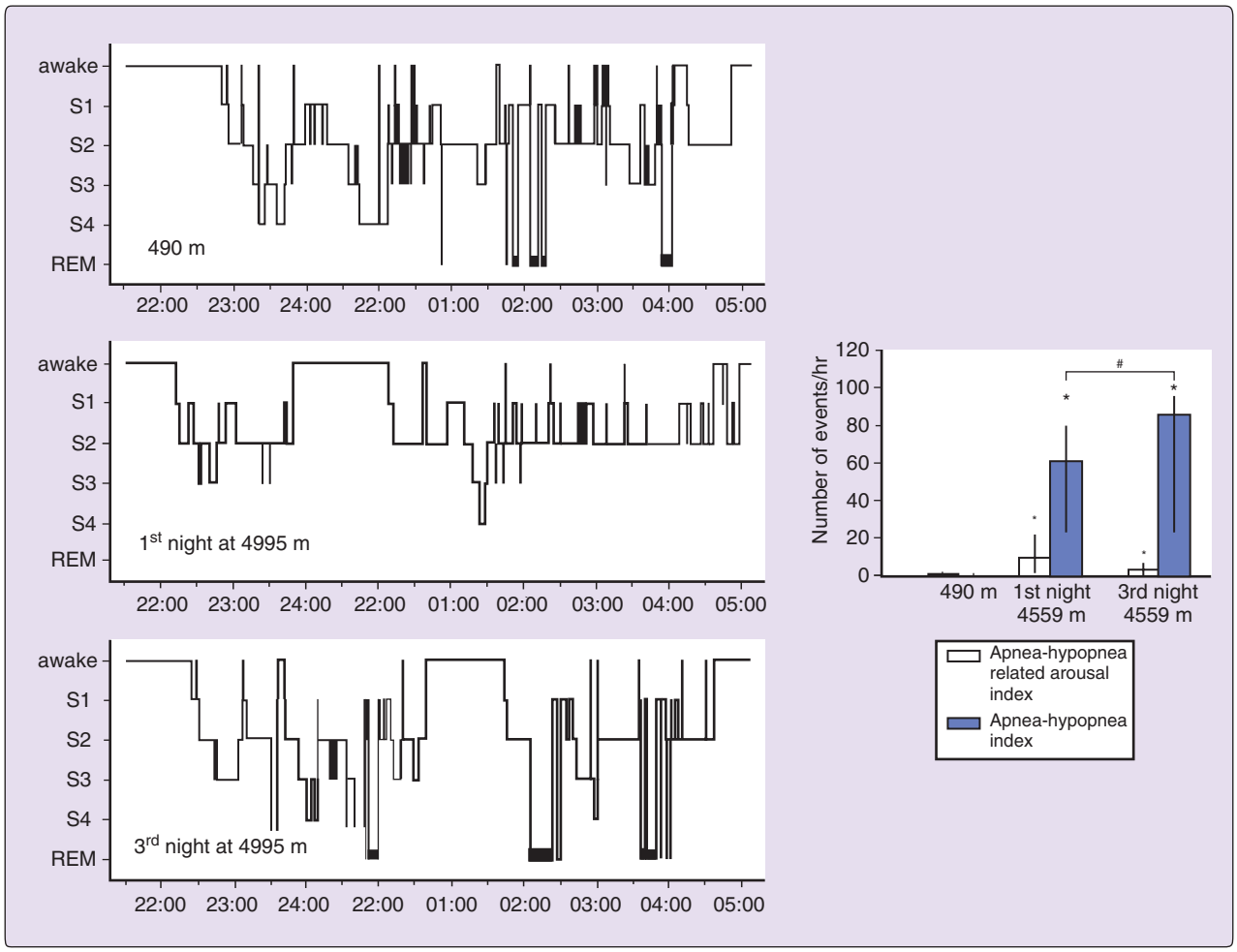


Figure 122-3 A normal subject who climbed to Regina Margherita Hut at 4995 meters (16,388 ft) within 24 hours, who had a reduction in total sleep time, slow wave sleep, and REM sleep and an increased number of arousals on polysomnography during the first night at 4995 meters (16,388 ft) compared with 490 meters (1607 ft). Three days of acclimatization resulted in improvement in sleep architecture, including increases in slow wave sleep and REM sleep and a reduction in the arousal index despite a further increase in apneas and hypopneas (*inset*), suggesting that periodic breathing was not the predominant cause of the sleep disturbances at altitude. (From Nussbaumer-Ochsner Y, Ursprung J, Siebenmann C, et al. Effect of short-term acclimatization to high altitude on sleep and nocturnal breathing. *Sleep* 2012;35:419–23.).

undershoot portions of the periodic pattern. First, as described earlier (see Figure 122-5), variations in $Paco_2$ correlate closely with the development of apnea and periodic breathing with hypoxia and its relief on restoration of normoxia. Second, if inspired fraction of CO_2 ($Fico_2$) is raised at hypoxia onset to prevent hypocapnia, periodic breathing is prevented, and if $Paco_2$ is raised during periodic breathing to elicit 1 to 2 mm Hg increases in $Paco_2$, rhythmic breathing is restored.[9]

The theoretical basis for the importance of $Paco_2$ changes depends on the effect of hypoxia on two types of "gains" that are the key determinants of the tendency toward ventilatory instability[26,37]: chemosensitivity (or controller) gain, defined by the slope of the ventilatory increase or decrease in response to hypercapnia or hypocapnia, respectively ($\Delta\dot{V}E/\Delta PaCO_2$); and plant gain, or the efficiency with which changes in ventilation eliminates CO_2 ($\Delta PaCO_2/\Delta\dot{V}E$) (Figure 122-6). Hypoxic exposure affects these gains in opposite directions; steady-state hypocapnic hyperventilation reduces plant gain, which by itself promotes ventilatory stability; whereas hypoxemia increases CO_2 chemosensitivity both above and below eupnea and narrows the difference between eupneic and apneic threshold $Paco_2$ (see Figure 122-6). The latter mechanism is quite pronounced, overwhelming the former and therefore increasing the likelihood of periodic breathing.

During ventilatory bursts of periodic breathing, the oxygen dissociation curve is shifted to the left, favoring O_2 uptake by the lungs; alternately, during apneas the curve is shifted to the right, favoring O_2 release to the tissues. These findings indicate that satisfactory gas exchange may persist during periodic breathing.

At altitudes above 3000 meters (9842 ft) and with Sao_2 less than 90%, almost all healthy subjects have sufficiently increased chemosensitivity to cause periodic breathing, especially during NREM sleep. With increased duration of hypoxic exposure and ventilatory acclimatization, plant gain is further reduced and chemoreceptor sensitivity further

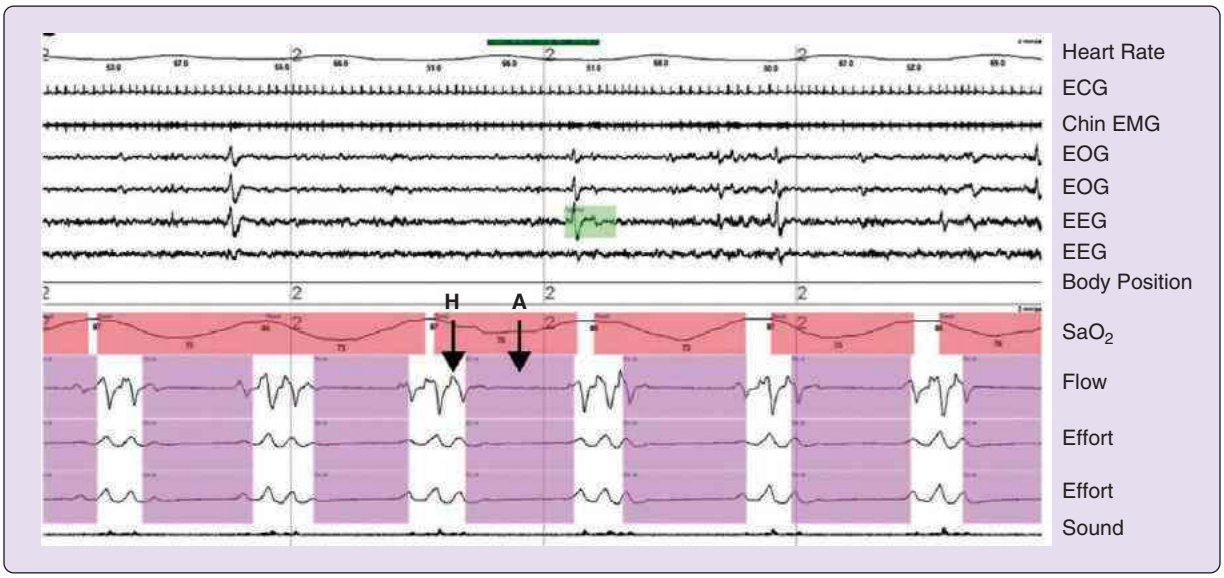


Figure 122-4 A 2-minute epoch from a polysomnogram from one subject during sleep at 5050 meters (16,568 ft) showing periodic breathing with central sleep apnea. *Arrow H* indicates the period of hyperpnea, and *arrow A* the period of apnea. Of note, not all apneas were followed by electroencephalogram (EEG) arousals. Arterial oxygen saturation (Sa_{O_2}) reading showing periods of desaturation. Nasal airflow was measured by pressure transducer. Respiratory effort readings by piezoelectric bands. ECG, Electrocardiogram; EMG, electromyogram; EOG, electro-oculogram. (From Ainslie PN, Lucas SJ, Burgess KR. Breathing and sleep at high altitude. *Respir Physiol Neurobiol* 2013;188:233–256.)

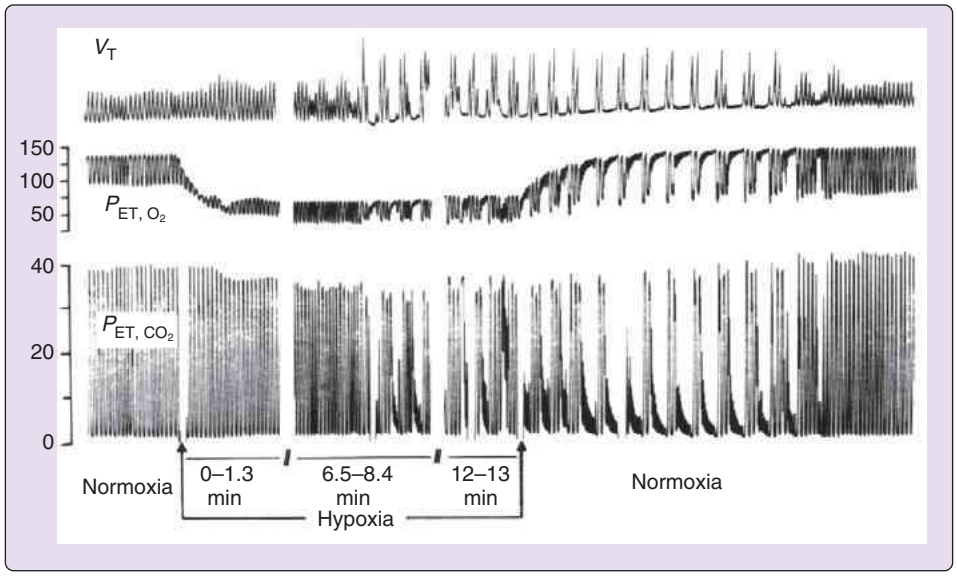


Figure 122-5 Time course of development of periodic breathing on abrupt exposure to simulation of the environmental hypoxia present at 4300 m altitude during NREM sleep and its relief on restoration of normoxia (see description in text). (From Berssenbrugge A, Dempsey J, Iber C, et al. Mechanisms of hypoxia-induced periodic breathing during sleep in humans. *J Physiol* 1983;343:507–24.)

increased. Thus a major—but not the only (see later)—determinant of whether longer durations of hypoxic exposure will result in periodic breathing is the balance struck between further changes in plant versus controller gains. Reports are mixed concerning the question of whether duration of hypoxic exposure increases, decreases, or has no effect on periodicity of breathing.[38] The stabilizing effect on respiratory rhythm of adding F_{ICO_2} in hypoxia—even if P_{ACO_2} is increased by only 1 to 2 mm Hg[9]—is likely attributed to reductions in plant gain,[37] as is the effect of adding such ventilatory stimuli as acetazolamide[39] or progesterone, so long as these added stimuli do not increase chemoreceptor gain.[40] Thus we propose that the net effect of changes in chemosensitivity versus plant gain is the major determinant of periodic breathing in hypoxia

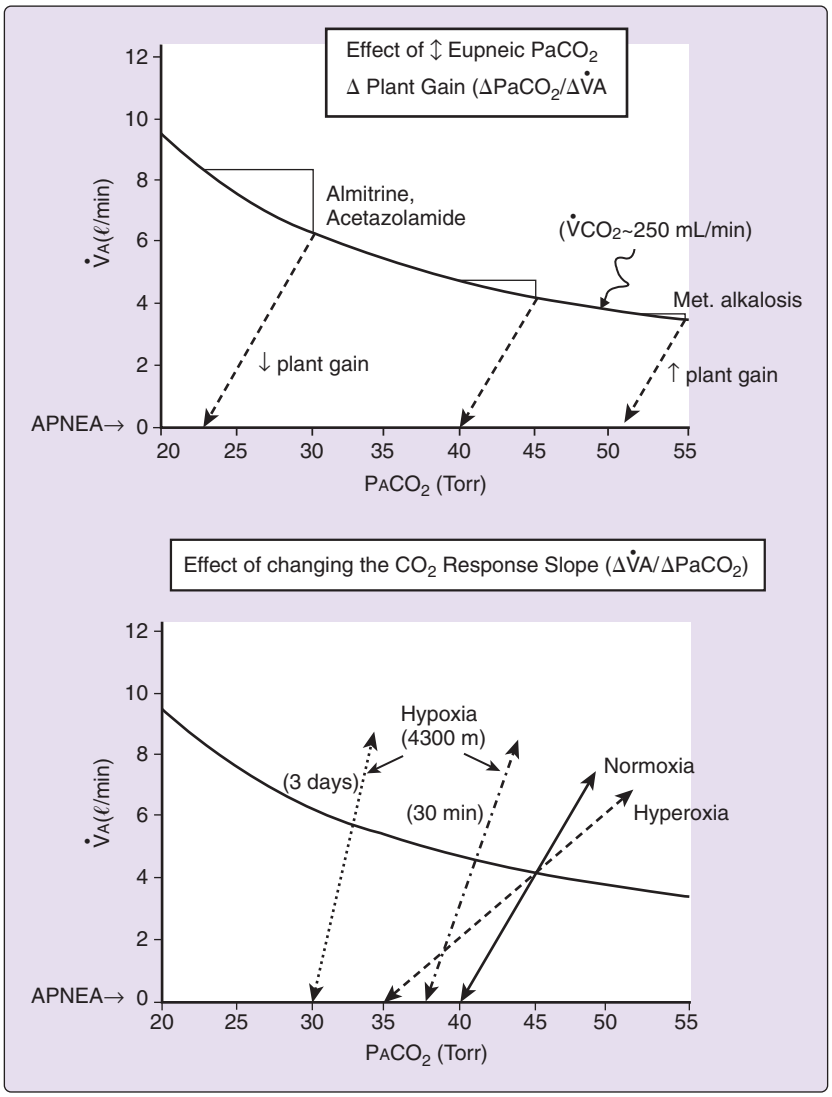


Figure 122-6 Graphic presentation of the relationship of $\dot{V}_A$ to Pa_{CO_2} at a fixed resting $\dot{V}_{CO_2}$ to illustrate how alterations in plant gain or controller gain effect the CO_2 reserve (eupneic Pa_{CO_2} – apneic threshold Pa_{CO_2}) and the propensity for apnea and instability. *Top panel*: changing plant gain by stimulating or reducing eupneic ventilation displaces Pa_{CO_2} along the isometabolic line relating $\dot{V}_A$ to Pa_{CO_2}, thereby altering the $\Delta Pa_{CO_2}/\Delta\dot{V}_A$ ratio and changing the CO_2 reserve and the susceptibility to apnea/periodicity. *Bottom panel*: altering controller gain ($\Delta\dot{V}_A/\Delta Pa_{CO_2}$ slopes) via acute hyperoxia or by acute and chronic hypoxia in normal subjects changes CO_2 reserve and susceptibility to apnea and breathing periodicity. Note that in hypoxia, controller and plant gains will change together. The hyperventilation and increased slope of the CO_2 response observed within the initial 30 minutes of hypoxia are each increased after 3 days in hypoxia. The stabilizing effect of reduced plant gain (decreased eupneic Pa_{CO_2}) is outweighed by the increased controller gain, and the CO_2 reserve is markedly reduced, leading to breathing instability and apnea. $\dot{V}_A$, Alveolar ventilation; $\dot{V}_{CO_2}$, CO_2 production. (From Dempsey JA, Smith CA, Blain GM, et al. Role of central/peripheral chemoreceptors and their interdependence in the pathophysiology of sleep apnea. *Adv Exp Med Biol* 2012;758:343–9.)

at high altitude.[40] However, other important secondary modulators of periodic breathing are also likely to contribute, as outlined later.

Secondary Modulators of Periodic Breathing

Changes occur in CBF in response to hypoxia, and especially in response to transient changes in $Paco_2$ during periodic breathing (averaging ~3% ΔCBF/mm Hg $\Delta Paco_2$) (Figure 122-7).[41] This highly sensitive cerebral vascular reactivity serves to regulate the Pco_2 difference between arterial blood and brain or cerebrospinal fluid, thereby minimizing changes in cerebral extracellular fluid Pco_2 and $[H^+]$ for any given change in $Paco_2$. On arrival to high altitude (5050 meters or 16,568 ft) the CBF velocity increases during NREM sleep compared with before sleep onset, with large oscillations secondary to varying cerebrovascular reactivity to $Paco_2$ and periodic breathing (Figure 122-8).[42] After 2 weeks of acclimatization the mean CBF returns to that seen at sea level but continues to have large oscillations because of the persistent periodic breathing. When the sensitivity of the cerebrovascular response to CO_2 is reduced experimentally (via cyclooxygenase inhibition), the slope of the $Paco_2$ versus

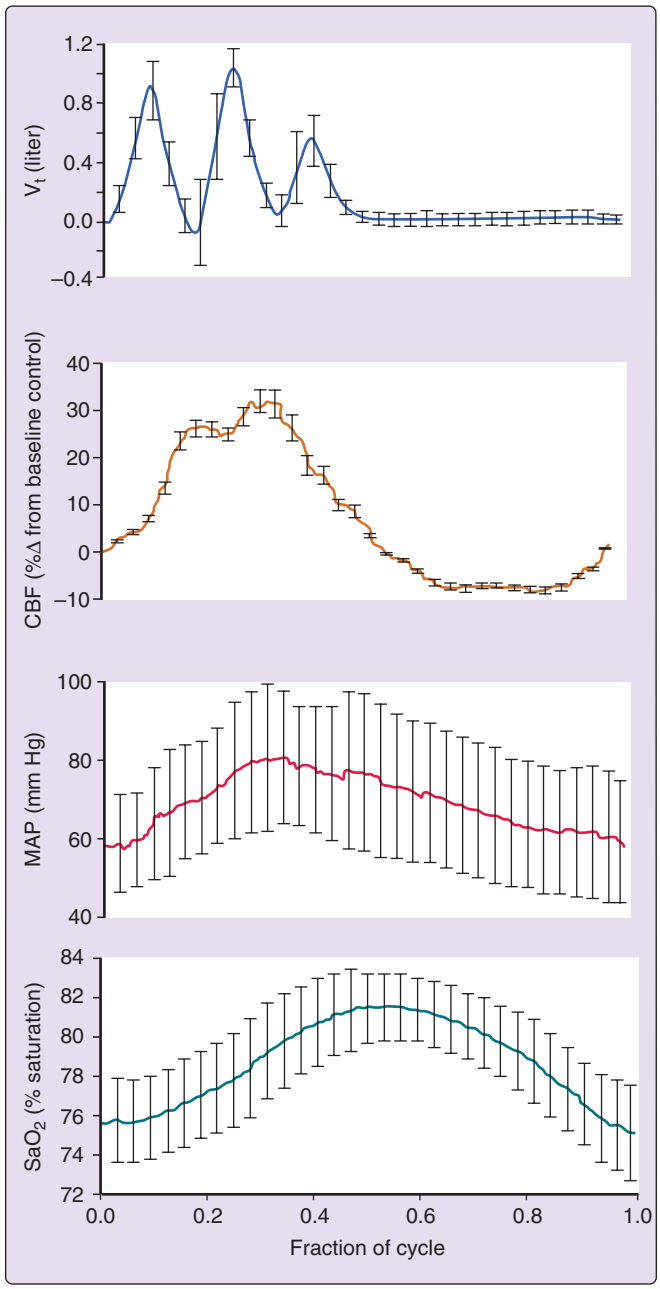


Figure 122-7 The effects of one night of hypoxic exposure (F_{IO_2} = 0.11) on periodic breathing and middle cerebral artery (MCA) blood flow during NREM sleep in the healthy human. Shown are signal-averaged data over 150 periodic breathing cycles in a single healthy subject. These data are representative of those found in six additional subjects. Each periodic breathing cycle averaged 22 seconds in duration and consisted of three hyperpnea tidal breaths followed by an 8- to 12-second apnea. Note the average 30% increase in MCA blood velocity (CBF) as determined by Doppler ultrasound measurements and the increase in mean arterial pressure (MAP), which began at the termination of the apnea and peaked during the ventilatory overshoot phase. Immediately following the ventilatory overshoot, CBF fell 10% below baseline control. (From Dempsey JA, Smith CA, Przybylowski T, et al. The ventilatory responsiveness to CO(2) below eupnoea as a determinant of ventilatory stability in sleep. *J Physiol* 2004;560:1–11.)

$\dot{V}E$ response is increased and the CO_2 reserve reduced[43]; at high altitudes periodic breathing is enhanced.[38] Conversely, acute (via intravenous acetazolamide) elevations in CBF velocity and reactivity to P_{aCO_2} are related to improvements in breathing stability at high altitude during wakefulness[44] and sleep.[42]

Experimental acute increases in left atrial and pulmonary vascular pressures in the naturally sleeping canine model in normoxia will stimulate breathing, enhance the CO_2 response gain and controller gain, and reduce CO_2 reserve.[45] It is not known whether these same influences contribute to ventilatory control instability as pulmonary vascular resistance increases in humans with altitude-related hypoxia.

A powerful stabilizing influence on breathing occurs through the phenomenon of short-term potentiation (STP) of phrenic nerve and respiratory motor output, which occurs following abrupt removal of a chemoreceptor (or other types) of ventilatory stimuli. This results in a gradual dissipation of ventilatory drive back to control levels.[46,47] When apnea occurs during NREM sleep following a transient ventilatory overshoot, this is likely a manifestation of hypocapnic inhibition, including that which is secondary to vagal influences from lung stretch, which has overridden the stimulatory STP effect.[47,48] Acute systemic hypoxemia also has the effect of reducing STP, thereby contributing to apnea and periodicity during sleep in hypoxia.[47]

The modulating roles of sex hormones on sleep-disordered breathing have been suggested by their protective roles against obstructive sleep apnea.[49,50] Sex hormones may directly contribute to ventilatory control through their effects on central respiratory centers, upper airways structure and function, lung dynamics, modulation of chemoreflex sensitivity, and plant gain.[51-53] The potential protective roles of sex hormones on periodic breathing at high altitude were tested in 23 men and 14 premenopausal women who had a sleep study at sea level and then at 3400 meters (11,155 ft) and 5400 meters (17,717 ft). At sea level, a normal breathing pattern was observed in all subjects throughout the night. At 3400 meters (11,155 ft) there was a significant and large difference in the mean AHI between men (AHI = 40/hour: central apneas 77.6%, central hypopneas 22.4%) and women (AHI = 2/hour: central apneas 58.2%, central hypopneas 41.8%). However, with further ascent to the altitude of 5400 meters, AHI increased in both men and women (AHI = 87/hour: central apneas 60.0%, central hypopneas 40.0%; AHI = 41/hour: central apneas 73.2%, central hypopneas 26.8%, respectively). Interestingly, there was no difference in nocturnal S_{aO_2} between the genders in these data despite the large differences in the central respiratory event indexes.[54] In another study, at an altitude of 4559 meters, men more frequently than women exhibited increased nocturnal periodic breathing. Increased periodic breathing directly correlated with heightened hypoxic chemosensitivity that had been assessed at sea level.[55] These data suggest gender differences in hypoxic chemosensitivity and periodic breathing at high altitude, with men being at higher risk for the latter.

In summary, the pathogenesis of periodic breathing in altitude-related hypoxia is clearly multifaceted, although the key element appears to be hypoxia-induced increased chemosensitivity to CO_2, likely involving both primary carotid body stimulation and secondary effects of enhancing the central chemoreceptor CO_2 response.[56,57]

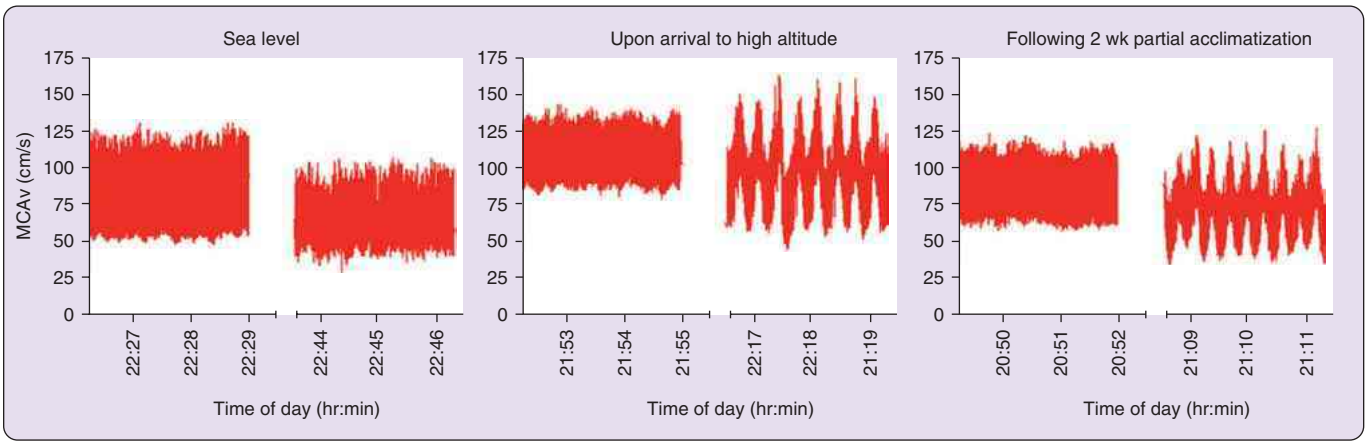


Figure 122-8 The changes in middle cerebral artery blood flow velocity (MCAv) on arrival to high altitude (*middle graph*) and following 2 weeks of partial acclimatization. The decrease in MCAv from wakefulness (left-hand trace) to stage N2 sleep (right-hand trace) was similar on arrival and following partial acclimatization to high altitude compared with sea level (≈10 cm/s). However, both awake and asleep MCAv levels were elevated on arrival to high altitude with marked oscillations compared with sea-level values. (From Burgess KR, Lucas SJ, Shepherd K, et al. Worsening of central sleep apnea at high altitude—a role for cerebrovascular function. *J Appl Physiol* 2013;114:1021–8.)

Sleep at High Altitude

In a comprehensive evaluation of normal subjects ascending to higher elevations (4995 meters, 16,388 ft), altitude-induced hypoxemia during the first night at this altitude was associated with a reduction in total sleep time, N3 sleep, REM sleep, and an increased number of arousals.[22] Breathing during sleep was characterized by increased minute ventilation, periodic breathing, and cyclical oxygen desaturation. Three days of acclimatization resulted in partial improvements of Sao_2 and of alterations in sleep architecture on the third night at 4559 meters, despite a further increase in central apneas and hypopneas, suggesting that periodic breathing was not the predominant cause of sleep disturbances at high altitude (see Figure 122-3). There seems to be no consistent change in the amount of REM sleep at high altitude, which is variably found to be either unchanged (field study),[13,33,58] increased (field study),[59,60] or decreased (hypobaric chamber or field studies).[16,20,22] The inconsistencies of the effect of altitude on REM sleep among these studies are likely related to differences in settings (hypobaric chamber vs. field studies), rate of ascent, and degree of physical exertion (being transported vs. climbing) and sleeping altitude. There is a disparity between subjective evaluation of sleep quality and the objective findings of normal sleep duration. After few nights at high altitude, there is improvement in total sleep time and sleep architecture; however, the subjects continue to complain about sleeplessness. This is most likely due to sleep fragmentation at altitude despite normal cumulative sleep duration that produces the impression of sleeplessness in that setting.[13,61-63]

In summary, on the first two nights following ascent to altitude, sleep is typically of near-normal duration or of decreased duration, with increases in NREM sleep stages N1 and N2 ("light" sleep) and decreases in stage N3 ("deep" sleep). Sleep quality improves after the third night at the same altitude, with increased slow wave sleep and REM sleep. Periodic breathing is present in a substantial proportion of sleep, with frequent arousals and disruption of sleep continuity. Of note, sleep architecture improves over two to three nights at the same altitude but periodic breathing persists, suggesting that periodic breathing is not the predominant cause of sleep disturbances at high altitude.

HIGH-ALTITUDE ILLNESSES

Sleep Apnea in High Altitude

The respiratory pauses and hypopneas in sleep at high altitude are mainly of central origin, unassociated with snoring or other evidence of sleep-related upper airway obstruction, and accompanied by decreased rib cage and abdominal activity.[9,11,33,34,64] One study of subjects with moderate obstructive sleep apnea with AHI of 26/hour at low altitude (60 meters, 197 ft) showed, in fact, that at a simulated altitude of 2750 meters (9022 ft) obstructive events were entirely replaced by central apneas. The authors believed that the obstructive sleep apneas resolved because of an increased respiratory rate and an increase in upper airway tone, whereas central sleep apneas developed because of hypocapnia.[65] However, in a study of patients with moderate to severe obstructive sleep apnea using a randomized crossover design, it was shown that a moderate altitude exposure (up to 2590 meters, 8497 ft) in these untreated patients aggravated hypoxemia and increased the frequency of central apneas and hypopneas, but obstructive apneas persisted.[66] The transformation of obstructive sleep apneas to central type at high altitude in some studies likely reflects an increase in hypoxic ventilatory drive, which may augment the activity of muscles of the upper airway.[67] Conversely, sleep studies of subjects living at moderate altitude (above 2400 meters, 7874 ft) that were conducted at a lower altitude (1370 meters, 4495 ft) in comparison to the altitude of residence, show that at the lower altitude the AHI was reduced. This was likely largely due to a decrease in central events; there was no change in frequency of obstructive apneas, although the latter were more prolonged at a low altitude.[68]

Acute Mountain Sickness and High-Altitude Pulmonary Edema

Rapid altitude ascent is often associated with two well-recognized clinical syndromes of acute altitude maladaptation: (1) AMS manifested by headache, loss of appetite, nausea, vomiting, decreased mental acuity, and insomnia; and (2) high-altitude pulmonary edema (HAPE). The pathophysiology of HAPE is related to acute exposure to high-altitude hypoxia with increased pulmonary artery pressure and increased pulmonary capillary permeability and resultant leakage of plasma proteins and red blood cells into the alveolar air space. Although these are most common in the early post-ascent period when sleep disturbance and respiratory periodicity are also most pronounced, most studies find that the characteristic periodic breathing in sleep is not correlated with the development or severity of these syndromes. Indeed, in subjects with pronounced AMS and HAPE, periodic breathing tends to be replaced by an irregular, nonperiodic pattern.[17,34] However, in one study, periodic breathing was found to be more frequent in those with HAPE than in those with AMS or in control subjects.[69] This suggests that HAPE may be associated with periodic breathing through stimulation of intrapulmonary afferents.[45] In naturally sleeping dogs, increasing pulmonary capillary pressure by inflating a balloon placed in the left atrium resulted in periodic breathing. AMS, conversely, is more closely associated with hypoxemia rather than periodic breathing during sleep, as suggested by a study in trekkers that found a trend toward an association between sleep hypoxemia and severity of AMS.[70] However, the role of sleep-disordered breathing in the symptoms of AMS is questioned by a study of high-altitude expedition members, which found that headache, a common feature of AMS, occurred only 26% of the time in association with sleep or awakening from sleep.[71]

There is evidence for cognitive decline at high altitude in nonacclimatized climbers, as a facet of AMS, which most likely reflects in part the central nervous system effects of hypoxemia compounded by the cerebral vasoconstrictor effects of hypocapnia and sleep fragmentation.[72,73] However, in acclimatized climbers there seems to be no evidence for significant cognitive impairment even at extreme altitudes of 7500 meters (24,606 ft).[74]

SLEEP AT HIGH ALTITUDE AFTER LONG-TERM ADAPTATION

In the Andes, South America, at an altitude of 4330 meters (14,206 ft) healthy natives have sleep duration and distribution of stages comparable to those of subjects at lower altitude.[75] However, these native highlanders have periodic breathing with cyclical oxygen desaturation and elevated hematocrit.[76] Similar observations have been made in Sherpas native to high altitude but not in Sherpas native to low altitude.[64] It is possible that Sao_2 induced by high altitude shifts Sao_2 during sleep to the steeper portion of the dissociation curve and thereby amplifies the influence of ventilatory dysrhythmia on Sao_2. The potential contribution of ethnic or genetic differences is suggested by a study comparing Tibetan and Chinese Han residents of 4000 meters (13,123 ft). Sleep was studied in a hypobaric chamber at simulated altitudes of 2261 (7418 ft) and 5000 meters (16,404 ft). At the higher altitude, Tibetans had more periodic breathing, higher Sao_2, and better sleep structure than did the Han subjects.[77]

Natives and long-term residents of high altitude exhibit chronic mountain sickness or Monge disease, a syndrome of severe polycythemia with headache, dizziness, breathlessness, and sleep disturbance. The pathophysiology of the syndrome is debated, but it likely is induced by increased hypoxemia reflecting the combined effects of altitude, decreased ventilatory drive, and lung dysfunction. Compared with normal subjects, individuals with chronic mountain sickness exhibit exaggerated hypoxemia during sleep without an increase in respiratory disturbance index.[78-80] These subjects also exhibit greater daytime hypoxemia, and thus the role of sleep-associated desaturation remains uncertain.

The adverse effects of high-altitude exposure on physiologic systems are complex but primarily involve the central nervous system and cardiopulmonary regulation. The rate of ascent and hence the degree of hypoxia exposure is the main determinant for the development of AMS, sleep disturbances, periodic breathing, and the risk for HAPE. Acclimatization to high altitude does not completely return the physiologic functions to sea level or low altitudes, even in long-term high-altitude dwellers. With acclimatization AMS and sleep architecture improve but periodic breathing persists with minimal effects on sleep continuity.

PREVENTION AND TREATMENT OF HIGH-ALTITUDE ILLNESSES

The prophylaxis and treatment of sleep abnormalities at high altitude and AMS are similar. Staged, gradual ascent to high altitude is an effective way to blunt sleep-related symptoms and to prevent AMS, but this may be inconvenient or not possible, as in flying to high-altitude research facilities. Pharmacologic approaches include carbonic anhydrase inhibitors and hypnotic agents. Noninvasive positive pressure ventilation or dead space breathing masks have also been evaluated as a possible treatment.[81]

Sleep Disturbances

Acetazolamide, a carbonic anhydrase inhibitor, is the best-studied agent used for amelioration of sleep disturbance at high altitude; it has the advantage of also reducing symptoms of AMS.[82-85] Acetazolamide improves both the mean level and the stability of arterial oxygenation during sleep at high altitude and markedly reduces the proportion of sleep time during which periodic breathing occurs.[14,86,87] Acetazolamide is a respiratory stimulant that causes metabolic acidosis and hyperventilation by increasing renal excretion of bicarbonate.[39] These changes mimic the natural process of acclimatization. An effective prophylactic dose of acetazolamide is 125 mg twice daily to be taken a day before ascent and continued for 2 days after the highest sleeping altitude.

Several studies suggest the safety and potential utility of benzodiazepines for the treatment of sleep disturbance of high altitude.[88-90] Temazepam shortened sleep latency, decreased arousals, increased sleep efficiency, and increased REM sleep, with subjectively better quality sleep in climbers sleeping above 4000 meters (13,123 ft).[89,90] The nonbenzodiazepine sedative agents zolpidem and zaleplon have each been found to be effective in improving sleep architecture and

consolidation at high altitude. A study of these agents at a simulated altitude of 4000 meters (13,123 ft) and another in trekkers at 3613 meters (11,854 ft) showed that, compared with placebo, both agents increased sleep efficiency and decreased wakefulness and that zolpidem increased slow wave sleep. Neither drug, however, had an effect on nocturnal respiratory pattern or Sao_2 nor significantly affected daytime cognitive or physical performance.[91,92]

Periodic Breathing

Acetazolamide improves both the mean level and the stability of arterial oxygenation during sleep at high altitude and markedly reduces the proportion of sleep time during which periodic breathing occurs.[14,86,87] The beneficial effect of acetazolamide on periodic breathing is seen in both men and women but is more pronounced in men.[55] Acetazolamide is a respiratory stimulant that causes metabolic acidosis and hyperventilation by increasing renal excretion of bicarbonate.[39] These changes mimic the natural process of acclimatization. An effective prophylactic dose of acetazolamide is 125 mg twice daily to be taken a day before ascent and continued throughout the ascent and for 2 additional days after the highest sleeping altitude. As mentioned previously benzodiazepines at low doses (temazepam 7.5 mg or 10 mg) improve sleep but have neither beneficial nor adverse effects on breathing at high altitude.

Oxygen supplementation improves periodic breathing by widening the Pco_2 reserve as mentioned earlier. However, the issues with logistics of using oxygen at high altitude preclude its routine use except for the treatment of HAPE and high-altitude cerebral edema. Promising effects of other modalities to improve high-altitude periodic breathing, including noninvasive ventilation and increasing dead space breathing, remain to be further investigated in humans.

CONCLUSIONS

Ascent to high altitude is characterized by frequent awakenings from sleep, which in part reflects sleep fragmentation by respiratory dysrhythmia typically consisting of periodic breathing. Such periodic breathing, an abnormal ventilatory pattern in which central apneas and hypopneas alternate with periods of hyperventilation, is induced by hypoxic stimulation of the peripheral chemoreceptors at high altitude, coupled with enhanced CO_2 chemosensitivity, which facilitates the development of central apneas when $Paco_2$ falls below the eupneic threshold.

Sleep disruption, although not its associated periodic breathing, decreases with time (acclimatization) at moderate altitude but not high altitude. Therapies that have been shown to improve the periodic breathing of high altitude and its associated sleep disruption and clinical syndrome of AMS include pretreatment with acetazolamide, an inhibitor of carbonic anhydrase, which induces metabolic acidosis and widens the Pco_2 reserve, opposing the effects of hypoxia on the peripheral chemoreceptors. Benzodiazepines and other hypnotic agents may improve sleep quality without apparent beneficial or adverse effects on breathing in sleep.[93] Positive airway pressure and increased dead space breathing each also have been used to improve altitude-associated periodic breathing.

CLINICAL PEARLS

- Sleep at high altitude is initially disturbed by the hypoxic stimulation of the peripheral chemoreceptors, resulting in widening of the Pco_2 reserve and proximity of the prevailing Pco_2 to the apneic threshold. This leads to periodic breathing, central apneas, hypopneas, and frequent arousals.
- During acclimatization, sleep architecture improves but periodic breathing persists.
- Effective treatments include acetazolamide, increasing dead space, or blunting of hypoxic stimulation with certain benzodiazepines.

SUMMARY

Exposure to high altitude imposes significant strain on the cardiopulmonary system, including central ventilatory control. As a consequence, sojourners to high altitude frequently experience sleep disturbances, often reporting restless and sleepless nights. Others describe a feeling of suffocation or shortness of breath on awakening from sleep. Objective observations show that sleep stages are generally shifted from deeper toward lighter sleep, along with a characteristic waxing and waning breathing pattern known as periodic breathing that accompanies sleep at high altitude. Such periodicity typically consists of two to four breaths, separated by a central apnea from the next burst of two to four breaths.

The principal reason for apnea and periodic breathing during sleep in hypoxic environments is believed to be elevations in controller or feedback gain, as evidenced by the steep increase in the ventilatory-CO_2 response slope above and below eupnea. Periodic breathing at altitude seems to reflect the respiratory dilemma of acute altitude ascent in which the stimulatory effects of hypoxia are opposed by the inhibitory action of hypocapnic alkalosis. The outcome is respiratory oscillation. With an apnea, Pco_2 rises; this in turn lessens alkalotic inhibition and augments hypoxic stimulation. This triggers hyperpnea, which lessens respiratory stimulation by decreasing hypoxia and increasing alkalosis, leading to recurrent apnea. The occurrence of altitude related apnea with lessening of ventilatory stimuli is enhanced during sleep.

On balance, the poor subjective quality of sleep seems to reflect the fragmentation of sleep by frequent arousals linked to the marked changes in respiratory pattern of periodic breathing. Arousals commonly occur at the transition from the end of apnea to the onset of hyperpnea. Subsequent acclimatization to altitude is associated with lessening of periodic breathing and better-quality sleep.

The most common treatment for periodic breathing of high altitude and its associated syndrome of AMS is prophylactic administration of acetazolamide, an inhibitor of carbonic anhydrase, which likely works by increasing Pco_2 reserve below eupnea. The results of recent studies suggest that benzodiazepines and other sleep-promoting agents may improve sleep quality without apparent beneficial or adverse effects on breathing.

Selected Readings

Ainslie PN, Lucas SJ, Burgess KR. Breathing and sleep at high altitude. *Respir Physiol Neurobiol* 2013;**188**:233–56.

Beall CM. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integr Comp Biol* 2006;**46**:18–24.

Bloch KE, Latshang TD, Ulrich S. Patients with obstructive sleep apnea at altitude. *High Alt Med Biol* 2015;**16**(2):110–16.

Dempsey JA, Smith CA, Blain GM, et al. Role of central/peripheral chemoreceptors and their interdependence in the pathophysiology of sleep apnea. *Adv Exp Med Biol* 2012;**758**:343–9.

Mohsenin V. Common high altitudes illnesses. A primer for healthcare provider. *Brit J Med Medical Res* 2015;**7**:1017–25.

Shah NM, Hussain S, Cooke M, et al. Wilderness medicine at high altitude: recent developments in the field. *J Sports Med* 2015;**6**:319–28.

A complete reference list can be found online at ExpertConsult.com.

Section 15

Cardiovascular Disorders

Chapter 123

Sleep and Cardiovascular Disease: Present and Future

Shahrokh Javaheri; Luciano F. Drager; Geraldo Lorenzi-Filho

Chapter Highlights

- Cardiocerebrovascular disorders, including systemic hypertension, coronary artery disease, congestive heart failure, stroke, and transient ischemic attacks, are prevalent and associated with excess morbidity and mortality as well as huge economic costs.
- One of the most significant developments in the field has been the recognition that sleep disorders such as sleep apnea are extremely common among patients with established cardiovascular disease and when present could contribute to a worsening outcome.
- Importantly, sleep disorders are also a potential cause of various cardiovascular diseases. This bidirectional relationship is well established for congestive heart failure and stroke, which can cause sleep disorders such as insomnia and sleep apnea (Figure 123-1). This chapter provides an overview of this section consisting of several chapters of sleep and cardiovascular diseases.

CARDIOVASCULAR DISEASE

Cardiovascular disorders have a high prevalence and are associated with excessive morbidity and mortality and huge economic costs (Table 123-1).[1] Each year, the American Heart Association, in conjunction with the Centers for Disease Control and Prevention, the National Institutes of Health, and other government agencies, updates statistics on the morbidity and mortality of cardiovascular disease. According to the 2014 update,[1] approximately 85 million people—35% of the U.S. population—have some form of cardiovascular disease.

Hypertension alone, a disorder proved to be caused by obstructive sleep apnea (OSA; see Chapter 127), affects 78 million Americans. Many of these patients are erroneously diagnosed as having essential hypertension because of the underdiagnosis of OSA. Congestive heart failure (see Chapter 129) and stroke (see Chapter 93), disorders frequently associated with both central and OSA, are also highly prevalent, each affecting approximately 5.1 to 6.8 million Americans (see Table 123-1).

According to the 2014 update, the actual number of cardiovascular disease deaths per year declined by 16.7%, but cardiovascular diseases still accounted for about 32% or 1 of every 3 deaths in the United States. On the basis of 2010 death rate data, more than 2150 Americans die of cardiovascular disease each day, an average of 1 death every 40 seconds. In fact, since 1900, cardiovascular disease has been the number

Table 123-1 Prevalence, Mortality, and Economic Burden of Cardiovascular and Cerebrovascular Disorders in the United States*

Population Group	Prevalence 2010	Mortality 2010	Hospital Discharges 2010	Cost 2012 (Billion Dollars)
Total	84 million (35% of all adults)	787,650 (32% of all deaths)	5.8 million	315
Women	43 million (34%)	502,200 (54%)	4 million	—
Men	41 million (37%)	390,600 (46%)	3 million	—
Age ≥60 yr	42 million	—	—	—
Hypertension	78 million(33% of Americans)	63,119	488,000	46
Coronary heart disease	15.4 million	380,000	1.8 million	165
Myocardial infarction	7.6 million	—	—	—
Angina	7.8 million	—	—	—
Congestive heart failure	5.1 million (825,000 new heart failure cases annually)	57,757 (any-mention mortality; 280,000)	1,023,000	30.7
Stroke	6.8 million	130,000	1,000,000	36.5

*Numbers are rounded.
Data from American Heart Association. Heart disease and stroke statistics—2014 update. *Circulation* 2014;129:e28–e292.

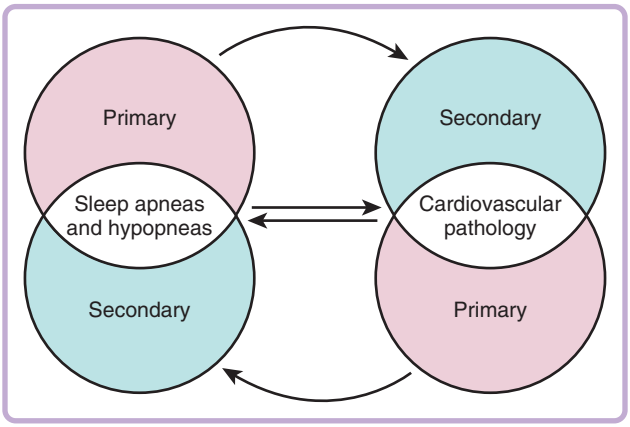


Figure 123-1 The relationship between obstructive sleep apnea (a primary sleep disorder), which secondarily could result in cardiocerebrovascular diseases, and a primary cardiovascular disease, specifically congestive heart failure or stroke, which secondarily could result in sleep-related breathing disorders.

Box 123-1 POTENTIAL CARDIOVASCULAR AND CEREBROVASCULAR COMPLICATIONS OF OBSTRUCTIVE SLEEP APNEA

Endothelial dysfunction
Hypertension
Systemic
Pulmonary (cor pulmonale)
Heart failure (reduced or preserved ejection fraction)
Arrhythmias
Coronary artery disease
Carotid artery atherosclerosis
Stroke; transient ischemic attack
Neuropsychological dysfunction
Dementia
Death (including sudden death)

one killer every year except 1918. In 2010, the total direct and indirect cost of cardiovascular disease and stroke in the United States was estimated to be $315.4 billion.[1]

Since sleep apnea was recognized as a common medical disorder after the widespread use of full polysomnography, the association of sleep apnea, both obstructive and central, with cardiocerebrovascular disorders is well recognized (see Figure 123-1, Box 123-1, and Video 123-1).[2-7]

SLEEP APNEA

The explosion of basic science and physiologic studies in both experimental animals and humans, as well as epidemiologic and clinical studies, supports the bidirectional linking of sleep apnea to a variety of cardiovascular disorders (see Figure 123-1 and Box 123-1; also see Chapters 127 and 129).[2-7] Although much more research needs to be done in this area, it is interesting to note that cor pulmonale was recognized as a feature of pickwickian syndrome before it became known that the underlying pathologic process of cor pulmonale is a sleep disorder.

There is now evidence that OSA is extremely common among patients with established cardiocerebrovascular and metabolic disease. For instance, among patients with hypertension, coronary artery disease, atrial fibrillation, type 2 diabetes, and metabolic syndrome, the prevalence of OSA ranges from 30% to 87%.[8-12] Moreover, the more severe the underlying cardiovascular disease, the higher the prevalence of OSA. On the other hand, there is also evidence that OSA is largely underdiagnosed and undertreated.[13,14] In addition to the low awareness of the medical community, there is also evidence that the typical symptoms associated with OSA observed in patients referred to the sleep specialist, such as excessive daytime sleepiness, are frequently not present among patients with established cardiocerebrovascular disease.[15]

OSA is associated with a number of biochemical and cellular abnormalities. Obstructive apneas result in neurohormonal activation, release of inflammatory mediators such as

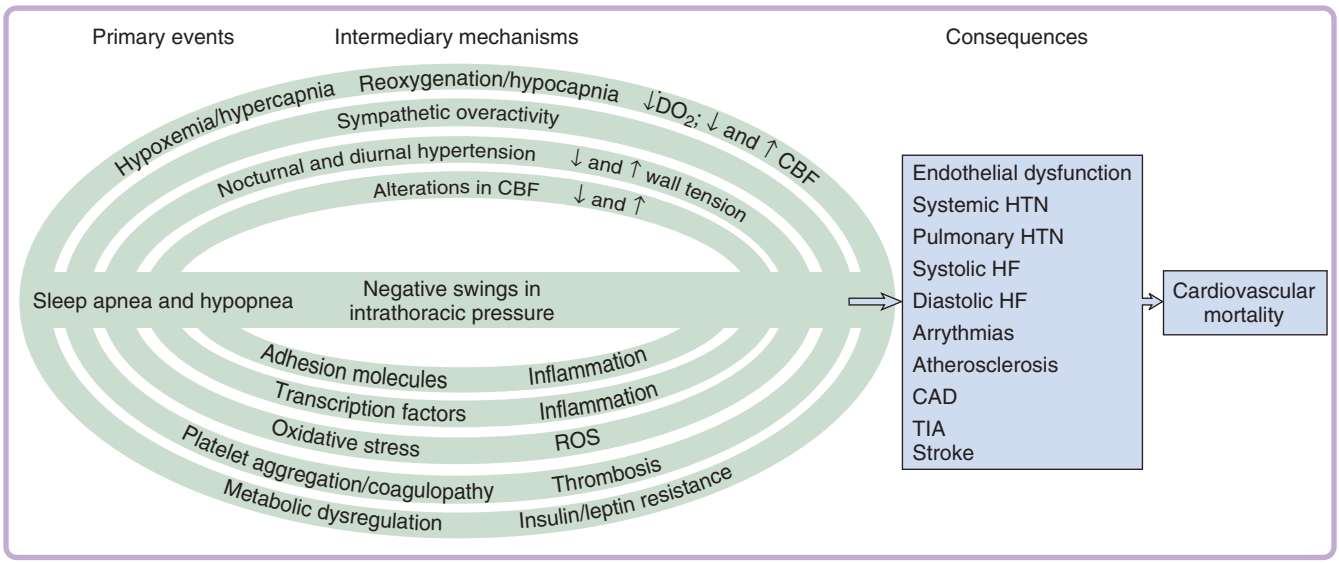


Figure 123-2 The mechanisms by which sleep apnea may result in endothelial dysfunction and cerebrovascular and cardiovascular disorders. CAD, Coronary artery disease; CBF, coronary/cerebral blood flow; $\dot{D}o_2$, oxygen delivery; HF, heart failure; HTN, hypertension; ROS, reactive oxygen species; TIA, transient ischemic attacks; ↑, increase; ↓, decrease. (From McNicholas WT, Javaheri S. Pathophysiological mechanisms of cardiovascular disease in obstructive sleep apnea. *Clin Sleep Med* 2007;2[4]:539–47.)

cytokines, and increased expression of adhesion molecules, resulting in attachment of white blood cells to endothelial cells and their transmigration, as well as oxidative stress. Through increased production of reactive oxygen species,[16] a number of transcription factors are activated, increasing the expression of redox-sensitive genes and resulting in the production of vasoactive and inflammatory proteins. These reactions underlie the pathologic processes involved in endothelial dysfunction syndrome, the underlying pathophysiologic mechanism for atherosclerosis, hypertension, stroke, heart failure, and coronary artery disease (Figure 123-2).[15]

Because of the aforementioned abnormalities, and along with cyclic changes in blood pressure resulting in wall stress, changes in coronary and cerebral blood flow, and diminished oxygen delivery, OSA could play a causative role or contribute to the development of atherosclerosis (see Figure 123-2).[17] In this context, treatment of OSA with nasal continuous positive airway pressure (CPAP) results in reversal of a number of biochemical abnormalities and attenuates surrogate markers of atherosclerosis.[18,19] Furthermore, a number of studies also demonstrate that the treatment of OSA with CPAP results in a reduction in systemic and pulmonary hypertension (see Chapter 127).[4,20,21] Systemic hypertension is the best studied relationship between OSA and cardiovascular disease. Several studies showed that the treatment of OSA with CPAP, even for short periods, is associated with a significant drop in blood pressure. The most beneficial therapeutic effects are observed in patients with severe OSA who are compliant with CPAP[20,21] as well as in patients with resistant hypertension[22-25] (see Chapter 127). Although the fall in blood pressure is frequently small, it has been shown that even small reductions in blood pressure over the long term significantly decrease the incidence of cerebrovascular and cardiovascular diseases.[26] For instance, in prospective studies of 420,000 patients with a mean follow-up of 10 years, drops in diastolic blood pressure of 5, 7.5, and 10 mm Hg were associated with, respectively, at least 34%, 46%, and 56% fewer strokes and at least 21%, 29%, and 37% less coronary heart disease.[26] Therefore, in patients with OSA, even a small drop in blood pressure, which could be maintained with long-term use of CPAP, is clinically, and also from a public health point of view, quite meaningful. Furthermore, treatment of OSA with CPAP may result in additional protection against vascular disorders because OSA may contribute to cardiovascular and cerebrovascular disease by a variety of mechanisms other than hypertension (see Figure 123-2). Regardless of the precise mechanisms, several studies consistently suggest not only that OSA is independently associated with increased cardiovascular mortality but also that the treatment of OSA with CPAP is associated with decreased cardiovascular mortality.[27-30]

METABOLIC SYNDROME

Metabolic syndrome is the clustering of few cardiometabolic risk factors related to abdominal obesity and insulin resistance and not surprisingly a multicomponent risk factor for cardiovascular disease, type 2 diabetes, and also OSA, although the latter is not as well recognized.

With the emergence of metabolic syndrome, a risk factor for incident cerebrovascular and cardiovascular diseases, a new epidemic is evolving.[31-33] Metabolic syndrome goes hand in hand with obesity and is characterized by elevated blood pressure, hyperglycemia, insulin resistance, and hypertriglyceridemia.[34] Metabolic syndrome is a proinflammatory and prothrombotic condition and is associated with increased serum concentrations of high-sensitivity C-reactive protein, fibrinogen, and von Willebrand factor and with increased platelet aggregation. Metabolic syndrome has a high prevalence, affecting one third of all adults.[1,33] Its prevalence increases with age, peaking among those aged 60 to 69 years.[33] Because metabolic syndrome is a precursor of incident cerebrovascular and cardiovascular disorders (relative risk, 1.78; 95% confidence interval, 1.58 to 2.0),[1] its early recognition and targeted therapy have been emphasized by different

medical societies.[34] OSA is also tightly linked to obesity, and OSA itself is associated with a large number of biochemical abnormalities that are markers of metabolic syndrome. As an example, through sympathetic stimulation, OSA contributes to insulin resistance and hypertension. Therefore metabolic syndrome, obesity, and OSA are entangled and are components of a vicious cycle.[15] In concert with the emphasis on early recognition of metabolic syndrome,[33,34] early recognition of OSA as a comorbid condition must be emphasized, particularly because the treatment of OSA with CPAP may reverse some of the abnormalities associated with metabolic syndrome.[35,36] Longitudinal studies should be conducted to determine whether early recognition and treatment of OSA as a companion of metabolic syndrome will prevent incident cerebrovascular and cardiovascular diseases.

SLEEP IN PATIENTS WITH HEART FAILURE

Another major development in the field is the recognition that central sleep apnea (CSA) associated with Hunter-Cheyne-Stokes breathing in patients with congestive heart failure, both with reduced and preserved ejection fraction, is extremely common (see Chapter 129).[37] Interestingly, the discovery of this respiratory pattern of breathing dates back to John Hunter,[38,39] 37 years before John Cheyne's description in 1818. There has been an explosion of physiologic and clinical research studies in this field.[40]

We must emphasize, however, that in addition to CSA, OSA is commonly observed in patients with congestive heart failure,[40,41]

Although some studies have shown that treatment of sleep apnea with CPAP improves mortality in patients with heart failure[42] (see Chapter 129), a major breakthrough in the field has been the advancement in positive airway pressure device algorithms and the development of adaptive servoventilation devices. The most recent generation of devices are equipped with automatic end-expiratory positive airway pressure algorithms that operate to eliminate obstructive disordered breathing events.[43,44] This feature, along with automatic variable inspiratory pressure support and the backup rate, makes these devices very effective for treatment of complex sleep-related breathing disorders.[43,44] As will be reviewed in Chapter 129, a number of studies using servoventilation devices show a reduction in hospital readmission and improved survival in patients with heart failure and sleep apnea. However a recent RCT with ASV showed a negative outcome with treatment.[45] Additionally, a transvenous phrenic stimulator has been developed to treat CSA in heart failure.[45,46] This is quite exciting because the pacemaker is placed intravenously by a cardiologist and primes the phrenic nerve according to a set algorithm to eliminate central apnea during sleep. Adherence should be virtually complete, in contrast to mask therapy with positive airway pressure devices.

IMPACT

The recognition of the association of sleep-related breathing disorders and cardiocerebrovascular diseases is important for several reasons. First, as noted previously (see Table 123-1), these disorders pose a great burden to patients and society. Second, sleep disorders are extremely common among patients with established cardiocerebrovascular disease. Third, recent studies show that treatment of OSA and CSA is associated with improvement in the cardiovascular morbidity, readmission, and probably mortality as well. However, long-term randomized trials that are adequately powered to prove that sleep apnea is a cause of cardiovascular mortality are lacking. Meanwhile, the design of such studies is complicated by a number of factors, such as inclusion and exclusion criteria; for example, should the most severe cases of OSA and excessive daytime sleepiness be excluded because of ethical issues of not treating patients with recognized OSA? This subgroup of patients may be the most compliant with CPAP, and their response to treatment may have the most favorable impact on mortality, the primary end point of the study. Other factors that complicate a long-term randomized clinical trial of OSA include cost, compliance with CPAP (particularly when patients do not feel short-term symptomatic benefit from it), and lack of a perfect placebo for CPAP.

Because of the importance of the relationship between sleep-related breathing disorders (both OSA and CSA) and cardiovascular disorders, this sixth edition devotes a series of chapters to cardiocerebrovascular diseases and sleep. In Chapters 124 and 125, a number of cardiovascular disorders related to sleep but unrelated to sleep apnea are reviewed. The emphasis in Chapter 124 is on nocturnal myocardial ischemia and infarction, and Chapter 125 emphasizes arrhythmias as they relate to changes in autonomic nervous system and sleep stages. The remaining chapters in the section are devoted to OSA and CSA and their relationship to cardiocerebrovascular diseases.

WHERE THE FIELD IS GOING AND WHERE WE ARE

The ability of sleep apnea treatment to reduce cardiovascular morbidity and mortality has not been rigorously demonstrated. Such considerations underscore the importance of well-designed, multicenter, randomized controlled trials (RCTs) to evaluate the impact of sleep apnea treatment on cardiovascular morbidity and mortality. The Spanish Sleep and Breathing Network, started in the mid-1990s, has been very active in conducting clinical trials based on the approach of a central committee that identifies relevant clinical questions, establishes funding streams, and selects participating centers for each project, providing technical support and guidance to researchers and presubmission review of manuscripts. This network has recently concluded an RCT to assess a composite incident cardiovascular disease and hypertension end point in 724 nonsleepy patients with sleep apnea. The Oxford Group reported on the Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular Trial (MOSAIC), which investigated the effects of CPAP on cardiovascular risk score, sleepiness, and quality of life in 391 minimally symptomatic patients.

The largest ongoing RCT, the Sleep Apnea Cardiovascular Endpoints (SAVE) study, has recruited more than 2000 patients with minimally symptomatic OSA and ischemic heart disease or cerebrovascular disease to study the impact of CPAP therapy on combined vascular events and mortality. Reflecting a cautious interest in the area of sleep apnea cardiovascular trials, the U.S. National Institutes of Health have recently funded two planning grants, the Heart Biomarker Evaluation in Apnea Treatment (HeartBEAT) and Best Apnea Interventions in Research (BestAIR), to evaluate

design approaches for a large scale clinical trial of CPAP for cardiovascular risk reduction, including feasibility of recruitment, methods for optimizing adherence, use of control treatments, and use of oxygen as an alternative to CPAP.

There are trials in heart failure and sleep apnea (see also Chapter 129); these include the SERVE-HF study using an adaptive servoventilation device for treatment of sleep-disordered breathing with predominant CSA and the ADVENT-HF trial using a different adaptive servoventilation device for treatment of OSA and CSA in heart failure. The former study was terminated because of negative outcomes with treatment.[47] A third study is using a transvenously placed phrenic nerve pacemaker to treat CSA.[47]

SUMMARY

Cardiocerebrovascular disorders and sleep apnea are common conditions, affecting each other in a bidirectional way. There are many observational studies correlating the impact of both OSA and CSA on cardiocerebrovascular outcomes. Similarly, many studies show that treating sleep apnea with positive airway pressure devices improves the outcome. Most of these studies, except those for systemic hypertension and OSA, have been observational. However, several RCTs are in progress (Table 123-2). Among these are trials using new-generation adaptive servoventilation devices to treat sleep apnea in heart failure.

CLINICAL PEARLS

- Sleep-related breathing disorders are common in patients with cardiovascular disorders and may either play a causative role or contribute to the progression of the cardiovascular pathologic processes. OSA is associated with increased incidence of coronary heart disease, stroke, heart failure, and mortality. Treatment with CPAP improves survival.
- Similarly, CSA and OSA are common in heart failure, with both reduced and preserved ejection fraction. Sleep apnea contributes to excess hospital readmission and premature mortality in patients with heart failure, and effective treatment with CPAP or adaptive servoventilation may increase left ventricular ejection fraction, decrease readmission, and improve survival.

Table 123-2 Ongoing Randomized Controlled Trials Assessing the Effect of Treatment of Obstructive Sleep Apnea on Cardiovascular Outcomes or Intermediate Mechanisms

Acronym	Country	Patient Condition	Intervention	Primary Outcome	Number Enrolled
Cardiovascular Events					
SAVE	Multinational	Stable (>3 mo) cardiovascular or cerebrovascular disease	CPAP vs. conservative treatment	Serious cardiovascular end points or cardiovascular mortality	2500
ISAACC	Spain	Acute coronary artery disease	CPAP vs. conservative treatment	Serious cardiovascular end points or cardiovascular mortality	1864
SERVE-HF*	Multinational	Sleep apnea and congestive heart failure	Adaptive servoventilation vs. conservative treatment	Mortality or hospitalization	1116
ADVENT-HF*	Multinational	Sleep apnea and heart failure	Adaptive servoventilation vs. conservative treatment	Death or hospitalization for a cardiovascular cause	860
RICCADSA	Sweden	Coronary artery disease (with recent percutaneous coronary intervention or coronary artery bypass graft)	CPAP vs. conservative treatment	Serious cardiovascular end points or cardiovascular mortality	510

Table 123-2 Ongoing Randomized Controlled Trials Assessing the Effect of Treatment of Obstructive Sleep Apnea on Cardiovascular Outcomes or Intermediate Mechanisms—cont'd

Acronym	Country	Patient Condition	Intervention	Primary Outcome	Number Enrolled
Intermediate Mechanisms					
BestAIR	USA	Moderate to severe OSA patients with cardiovascular risk factors presenting to a sleep disorders clinic	Active CPAP with RT support vs. active PAP with behavioral modification CPAP vs. conservative medical therapy vs. sham PAP	24-hr systolic blood pressure	150
COMET	USA	OSA and arterial hypertension	CPAP vs. oral appliance	24-hr ambulatory blood pressure monitoring	238
The remedē System Pivotal Trial	Multinational	Central sleep apnea	The remedē System therapy vs. optimal medical management	AHI reduction of 50% in patients with treatment vs. optimal medical management	173
MORPHEOS	Brazil	OSA and uncontrolled hypertension	CPAP vs. nasal strips	Central blood pressure	200
OPTIMAL-HF*	USA	Sleep apnea in chronic heart failure	CPAP vs. ASV vs. nocturnal supplemental oxygen vs. healthy lifestyle and sleep education	Left ventricular ejection fraction	161
OSA and and vasculopathy (no acronym)	China	Men with newly diagnosed severe OSA	CPAP vs. sham CPAP	Vascular reactivity of brachial artery and pulse wave velocity	150
PAC-IC-SAOS	France	OSA in heart failure patients undergoing coronary artery bypass surgery	CPAP vs. sham CPAP	Percentage of ventricular function recovery	69
RAP	France	OSA and resistant hypertension	CPAP vs. CPAP plus physical activity program	Systolic arterial blood pressure assessed by 24-hr home blood pressure monitoring	100
SASS	USA	OSA	CPAP vs. sham CPAP	Oxidative stress	250
Sleep Tight	USA	OSA patients with transient ischemic attack or stroke	Standard CPAP intervention vs. enhanced CPAP intervention vs. usual care	Reduction in five domains of cardiovascular risk markers (inflammatory, autonomic, insulin resistance, endothelial, and atherosclerotic)	255

*Included healthy lifestyle, sleep education, and supplemental oxygen.
AHI, Apnea-hypopnea index; ASV, adaptive servo ventilation; CPAP, continuous positive airway pressure; OSA, obstructive sleep apnea; PAP, positive airway pressure; RT, respiratory therapist.
Modified from Sánchez-de-la-Torre M, Campos-Rodriguez F, Barbé F. Obstructive sleep apnoea and cardiovascular disease. *Lancet Respir Med* 2013;1(1):61–72.

Selected Readings

Abraham WT, Jagielski D, Oldenburg O, et al; Remedē Pilot Study Investigators. Phrenic nerve stimulation for the treatment of central sleep apnea. *JACC Heart Fail* 2015;**3**(5):360–9.

Aggarwal S, Nadeem R, Loomba RS, et al. The effects of continuous positive airways pressure therapy on cardiovascular end points in patients with sleep-disordered breathing and heart failure: a meta-analysis of randomized controlled trials. *Clin Cardiol* 2014;**37**(1):57–65.

Campos-Rodriguez F, Martinez-Garcia MA, de la Cruz-Moron I, et al. Cardiovascular mortality in women with obstructive sleep apnea with or without continuous positive airway pressure treatment: a cohort study. *Ann Intern Med* 2012;**156**(2):115–22.

Chirinos JA, Gurubhagavatula I, Teff K, et al. CPAP, weight loss, or both for obstructive sleep apnea. *N Engl J Med* 2014;**370**(24):2265–75.

Cowie MR, Woehrle H, Wegscheider K, et al. Adaptive servo-ventilation for central sleep apnea in systolic heart failure. *N Engl J Med* 2015;**373**(12):1095–105.

de Oliveira AC, Martinez D, Massierer D, et al. The antihypertensive effect of positive airway pressure on resistant hypertension of patients with obstructive sleep apnea: a randomized, double-blind, clinical trial. *Am J Respir Crit Care Med* 2014;**190**(3):345–7.

Drager LF, Bortolotto LA, Figueiredo AC, et al. Effects of continuous positive airway pressure on early signs of atherosclerosis in obstructive sleep apnea. *Am J Respir Crit Care Med* 2007;**176**(7):706–12.

Drager LF, Lopes HF, Maki-Nunes C, et al. The impact of obstructive sleep apnea on metabolic and inflammatory markers in consecutive patients with metabolic syndrome. *PLoS ONE* 2010;**5**(8):e12065.

Javaheri S, Blackwell T, Ancoli-Israel S, et al. Sleep disordered breathing and incident heart failure in older men. *Am J Respir Crit Care Med* 2015 Oct. [Epub ahead of print].

Javaheri S, Brown L, Randerath W. Positive airway pressure therapy with adaptive servo-ventilation. Part 1: operational algorithms. *Chest* 2014;**146**:514–53.

Javaheri S, Brown L, Randerath W. Positive airway pressure therapy with adaptive servo-ventilation. Part 2: clinical applications. *Chest* 2014;**146**:855–68.

Javaheri S, Javaheri S, Javaheri A. Sleep apnea, heart failure, and pulmonary hypertension. *Curr Heart Fail Rep* 2013;**10**(4):315–20.

Lavie L. Obstructive sleep apnoea syndrome: an oxidative stress disorder. *Sleep Med Rev* 2003;**7**:35–51.

Martínez-García MA, Capote F, Campos-Rodríguez F, et al. for the Spanish Sleep Network. Effect of CPAP on blood pressure in patients with obstructive sleep apnea and resistant hypertension: the HIPARCO randomized clinical trial. *JAMA* 2013;**310**(22):2407–15.

Pedrosa RP, Drager LF, de Paula LK, et al. Effects of OSA treatment on BP in patients with resistant hypertension: a randomized trial. *Chest* 2013;**144**(5):1487–94.

Zhang J, Hobkirk J, Carroll S, et al. Exploring quality of life in patients with and without heart failure. *Int J Cardiol* 2016;**202**:676–84.

A complete reference list can be found online at ExpertConsult.com.

Chapter 124

Sleep-Related Cardiac Risk

Richard L. Verrier; Murray A. Mittleman

Chapter Highlights

- The brain, in subserving its need for periodic reexcitation during rapid eye movement sleep and dreaming, imposes significant demands on the heart by inducing bursts of sympathetic nerve activity, which reaches levels higher than during wakefulness. In patients with cardiac disease, such neural activity may compromise coronary artery blood flow, as metabolic demand outstrips supply, and may trigger sympathetically mediated life-threatening arrhythmias in response to functional myocardial ischemia.
- An additional challenge is presented by non–rapid eye movement sleep, when hypotension may lead to malperfusion of the heart and brain as a result of a lowered blood pressure gradient through stenosed vessels.
- Impairment of ventilation by sleep-related breathing disorders, including obstructive sleep apnea and central sleep apnea, which afflict millions of Americans, can generate reductions in arterial oxygen saturation and other pathophysiologic sequelae. Obstructive sleep apnea has been strongly implicated in the etiology of hypertension, myocardial ischemia, arrhythmias, myocardial infarction, heart failure, and sudden death in individuals with coexisting ischemic heart disease. Similarly, central sleep apnea has been associated with a variety of atrial and ventricular arrhythmias.
- Atrial fibrillation may be triggered by autonomic or respiratory disturbances during sleep in certain patient populations. Medications that cross the blood-brain barrier may alter sleep structure and provoke nightmares with severe cardiac autonomic discharge.

In healthy individuals, sleep is usually salutary and restorative. Ironically, during sleep in patients with respiratory or heart disease, the brain can precipitate breathing disorders, myocardial ischemia, arrhythmias, and even death. Our observation that 20% of myocardial infarctions (MIs) and 15% of sudden deaths occur during the period from midnight to 6:00 AM projects to an estimated 300,000 nocturnal MIs and 45,000 nocturnal sudden deaths annually in the U.S. population.[1] Thus, sleep is not a fully protected state. Furthermore, the nonuniform distribution of deaths and MIs during the night is consonant with provocation by pathophysiologic triggers. The two main factors implicated in nocturnal cardiac events are sleep state–dependent surges in autonomic activity[2] and depression of respiratory control mechanisms,[3] which affect a vulnerable cardiac substrate. Precise characterization of their interaction in precipitating nocturnal cardiac events is, however, incomplete. Although sudden death during sleep can be presumed to be painless, in many cases it is premature because it occurs in infants and adolescents and in adults with ischemic heart disease, for whom the median age is 69 years. Populations at risk for nocturnal cardiorespiratory events include several large patient groups (Table 124-1). For example, arrhythmias, including paroxysmal atrial fibrillation and nonsustained ventricular tachycardia, during sleep are markedly increased shortly after a sleep-disordered breathing event.[4]

It is an insidious component of the problem of nocturnal risk that many people are unaware of their respiratory or cardiac distress at night and therefore take no corrective action. Thus, sleep presents unique autonomic, hemodynamic, and respiratory challenges to the diseased myocardium that cannot be monitored by daytime diagnostic tests. The importance of nocturnal monitoring of patients with cardiac disease extends beyond identifying sleep state–dependent triggers of cardiac events because nighttime myocardial ischemia, arrhythmias, autonomic activity, and respiratory disturbances carry predictive value for daytime events (Box 124-1).

This chapter discusses the pathophysiologic mechanisms responsible for sleep-related cardiac morbidity and mortality. For a review of mechanisms and treatment of nocturnal arrhythmias, see Chapter 125. Effects of sleep-disordered breathing and apnea on the cardiovascular system are discussed in Chapters 126–129.

AUTONOMIC ACTIVITY AND CIRCULATORY FUNCTION DURING SLEEP

The generalized decrease in mean heart rate and arterial blood pressure at the onset of sleep and throughout non–rapid eye movement (NREM) sleep, which occupies 80% of sleep time, has prompted the assumption that sleep is a period of relative autonomic inactivity. NREM sleep, the initial stage,

Table 124-1 Patient Groups at Potentially Increased Risk for Nocturnal Cardiac Events

Indication (U.S. Patients per Year)	Possible Mechanism
Angina, myocardial infarction (MI), arrhythmias, ischemia, or cardiac arrest at night; 20% of myocardial infarctions (~300,000 cases/yr) and 15% of sudden deaths (~48,750 cases/yr) occur between midnight and 6:00 AM.	The nocturnal pattern suggests a sleep state–dependent autonomic trigger or respiratory distress.
Unstable angina, Prinzmetal angina	Nondemand ischemia and angina peak between midnight and 6:00 AM.
Acute MI (1.5 million)	Disturbances in sleep, respiration, and autonomic balance may be factors in nocturnal arrhythmogenesis. Nocturnal onset of MI is more frequent in older and sicker patients and carries a higher risk for congestive heart failure.
Heart failure (5.3 million)	Sleep-related breathing disorders are pronounced in the setting of heart failure and may contribute to its progression and to mortality risk.
Atrial fibrillation (2.5 million)	Twenty-nine percent of episodes occur between midnight and 6:00 AM. Respiratory and autonomic mechanisms are suspected.
Sleep apnea in patients with coronary disease (5 to 10 million patients with sleep apnea)	Patients with hypertension or atrial or ventricular arrhythmias should be screened for the presence of sleep apnea.
Long QT syndrome	The profound cycle-length changes associated with sleep may trigger pause-dependent torsades de pointes in these patients.
Sudden infant death syndrome (SIDS) (2000-2500 cases or 8% of infant deaths)	SIDS commonly occurs during sleep with characteristic cardiorespiratory symptoms.
Brugada syndrome in Western populations; Asians with warning signs of sudden unexplained nocturnal death syndrome (SUNDS)	SUNDS is a sleep-related phenomenon in which night terrors may play a role. The Brugada syndrome is genetically related to the long QT syndrome.
Patients with epilepsy (2.3 million)	Sudden unexpected death in epilepsy (SUDEP) occurs primarily at night and in patients with a history of nocturnal seizures.
Patients on cardiac medications (13.5 million patients with cardiovascular disease)	Beta blockers and calcium channel blockers that cross the blood-brain barrier may increase nighttime risk because poor sleep and violent dreams may be triggered. Medications that increase the QT interval may conduce to pause-dependent torsades de pointes during the profound cycle-length changes of sleep. Because arterial blood pressure is decreased during NREM sleep, additional lowering by antihypertensive agents may introduce a risk for ischemia and infarction due to lowered coronary perfusion.

is characterized by marked stability of autonomic regulation with a high degree of parasympathetic neural tone and prominent respiratory sinus arrhythmia.[2,5] Baroreceptor gain is high and contributes to the stability of arterial blood pressure and to overall cardiovascular homeostasis.[6] Muscle sympathetic nerve activity is stable, falls with the transition from awake to NREM sleep, and decreases progressively with depth of sleep,[2,7] reaching half the awake value during N3 sleep.[2] Short-lasting increases in muscle sympathetic nerve activity, heart rate, and arterial blood pressure accompany the appearance of high-amplitude K-complexes during N2 sleep.[2,7] Heart rate accelerations may even precede the electroencephalographic arousals of N2 and rapid eye movement (REM) sleep.[8] During transitions from NREM to REM sleep, bursts of vagus nerve activity may result in pauses in heart rhythm and frank asystole. Transitions between REM and NREM sleep elicit posture shifts that are associated with varying degrees of autonomic activation and attendant changes in heart rate and arterial blood pressure.[9] These shifts in body position increase in frequency as individuals age and sleep becomes fragmented.

Autonomic nervous system activity is dramatically altered when REM sleep is initiated (Figure 124-1). REM sleep is marked by profound muscle sympathetic nerve activation, in terms of both frequency and amplitude,[2,6] which attains levels significantly higher than in wakefulness.[2] Sympathetic nerve activity is concentrated in short, irregular periods that are most striking when accompanied by intense eye movements.[2] These bursts trigger intermittent increases in heart rate and arterial blood pressure to levels similar to those in wakefulness, with increased variability.[2,7,8] Significant surges and pauses in heart rate during REM sleep have been described in several species, including humans.[7,8] Cardiac efferent vagal tone and baroreceptor regulation[6] are generally suppressed during REM sleep, and breathing patterns may become highly irregular and may lead, in susceptible individuals, to oxygen desaturation. Thus, while subserving the neurochemical functions of the brain, REM sleep can disrupt cardiorespiratory

Box 124-1 PREDICTIVE VALUE OF NOCTURNAL CARDIORESPIRATORY STATUS

- Because parasympathetic nerve activity is elevated during sleep in healthy individuals, lack of circadian pattern of heart rate variability and baroreflex sensitivity may be readily monitored for increased risk for cardiac events.
- Nondemand nocturnal ischemic episodes may disclose a critical underlying coronary lesion, coronary vasospasm, or transient coronary artery stenosis.
- In elderly subjects, nighttime multifocal ventricular ectopic activity predicts increased mortality from cardiac causes independent of clinically evident cardiac disease.
- Sleep apnea, which may be screened by heart rate variability analysis, conduces to hypertension, left ventricular remodeling, myocardial ischemia, and atrial and ventricular arrhythmias and is a risk factor for lethal daytime cardiac events, including myocardial infarction.
- Hypertensive patients with less than a 10% nocturnal decline in blood pressure (remaining higher than 101/65 mm Hg) are at increased risk for total and cardiovascular mortality and all cardiovascular end points, myocardial ischemia, frequent or complex ventricular arrhythmias, cerebrovascular insult, and increased organ damage, including left ventricular hypertrophy.
- Elevated nocturnal heart rates are associated with increased cardiac mortality.

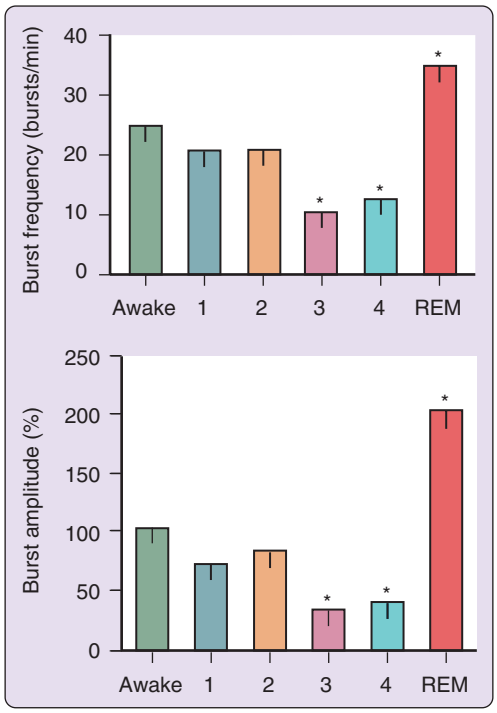


Figure 124-1 Sympathetic nerve burst frequency and amplitude during wakefulness, NREM sleep (eight subjects), and REM sleep (six subjects). Sympathetic nerve activity was significantly lower during stages 3 and 4 ($P < .001$). During REM sleep, sympathetic nerve activity increased significantly ($P < .001$). Values are means ± standard error of the mean. (From Somers VK, Dyken ME, Mark AL, et al. Sympathetic nerve activity during sleep in normal subjects. *N Engl J Med* 1993;328:303–7, with permission from the Massachusetts Medical Society. All rights reserved.)

homeostasis. The brain's increased excitability during REM sleep can also trigger major surges in sympathetic nerve activity to the skeletal muscular beds, accompanied by muscular twitching,[2] which interrupts the generalized skeletal atonia of REM.[9] The peripheral autonomic status characterized by muscle sympathetic nerve recording is compatible with reduced neuronal activity in the brainstem and other regions of the brain and reduced cerebral blood flow during NREM sleep and, during REM sleep, with increased brain activity in several discrete regions to levels higher than waking values.[10]

The decline in autonomic activity during sleep is also evident in peroneal muscle sympathetic nerve activity[2,7] and peripheral levels of epinephrine and norepinephrine and mirrors the generalized sleep-induced decline in heart rate and arterial blood pressure.[11] A nocturnal nadir in plasma catecholamines is evident 1 hour after sleep onset. Plasma cortisol is also depressed during sleep; increased levels are initiated at 5:00 AM.

In the absence of readily achieved, direct measures of cardiac-bound nerve activity, analysis of heart rate variability (HRV) has emerged as a widely accepted method for measuring cardiac sympathetic versus parasympathetic neural dominance. High-frequency (HF) HRV is a general indicator of cardiac parasympathetic tone and includes the effects of respiration. The low- to high-frequency (LF/HF) ratio is widely accepted as an approximation of cardiac-bound sympathetic nerve activity, as validated by studies involving β-adrenergic receptor blocking agents. Decreased HRV, indicating a decline in parasympathetic nerve activity, is an established indicator of risk for sudden cardiac death after MI. HRV analysis reveals a generalized increase in vagus nerve activity and a decrease in cardiac sympathetic nerve activity across the sleep period,[12,13] probably reflecting the dominance of total sleep time by NREM sleep. HRV studies using 5-minute intervals provide results consistent with muscle nerve recording, indicating increased HF and decreased LF (or parasympathetic nerve dominance) in NREM sleep but decreased HF and increased LF (or predominant sympathetic nerve activity) in REM sleep and during wakefulness.[8] In healthy individuals, the increase in HRV measures of cardiac sympathetic nerve activity at onset of REM sleep is initiated before[8,13] the transition from NREM sleep as classically defined from the polysomnographic record.

The typical circadian pattern of decreased nocturnal cardiac sympathetic nerve activity as described by heart rate[14] and HRV studies is altered in patients with coronary artery disease,[15,16] MI,[12,17] and diabetes,[18] suggesting either increased nocturnal cardiac sympathetic nerve activity or decreased parasympathetic nerve activity compared with healthy subjects. The HF component has been observed to decrease approximately 10 minutes before onset of nocturnal myocardial ischemia.[16] In unmedicated patients with a recent MI, the LF/HF ratio was significantly increased during both REM and NREM sleep, in contrast to healthy subjects, in whom this ratio during REM sleep is similar to awake levels and higher than during NREM sleep (Figure 124-2).[12] The conclusions were reached that MI decreases the capacity of the vagus nerve to be activated during sleep, resulting in unbridled cardiac sympathetic nerve activity,[12] and that loss of rise in the HF component is characteristic of patients after an MI and with residual myocardial ischemia.[17]

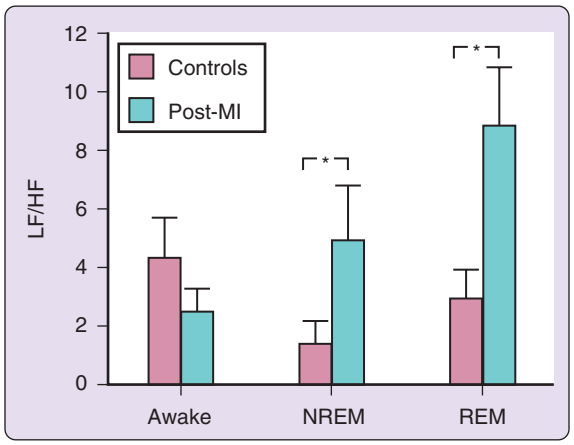


Figure 124-2 Bar graphs indicating low- to high-frequency (LF/HF) ratio of heart rate variability during the awake state (*left*), during NREM sleep (*middle*), and during REM sleep (*right*) in healthy subjects and in post-myocardial infarction (MI) patients ($P < .01$ when comparing control subjects and post-MI patients). Values are means ± standard error of the mean. (From Vanoli E, Adamson PB, Ba-Lin, et al. Heart rate variability during specific sleep stages: a comparison of healthy subjects with patients after myocardial infarction. *Circulation* 1995;91:1918–22, with permission from the American Heart Association.)

These sleep state–dependent profiles of autonomic activity have significant potential to affect coronary function and cardiac electrical stability in patients with ischemic heart disease.

NOCTURNAL CARDIOVASCULAR EVENTS

Nocturnal Myocardial Ischemia and Angina

Accurate assessment and treatment of nocturnal angina has been a subject of concern for more than two centuries. Heberden in 1768 described angina that "will often oblige [the patients] to rise up out of their bed every night for many months altogether." John Hunter, the well-known eighteenth-century surgeon, reported chest pains that "seized him in his sleep so as to awaken him."[19] As early as 1923, MacWilliam[20] postulated that the mechanisms of nocturnal ventricular fibrillation and angina were stimulation of sympathetic nerves and increased arterial blood pressure. He described "reflex excitations, dreams, nightmares, etc., sometimes accompanied by extensive rises of arterial blood pressure (hitherto not recognized), increased heart action, changes in respiration, and various reflex effects" and noted "the suddenness of development of the functional disturbances in arterial blood pressure, heart action, etc., in the dreaming state." He documented greater stress on the circulatory system during dreaming than during wakefulness, with arterial blood pressures reaching 200 mm Hg. In the middle of the twentieth century, the renowned cardiologists Paul Dudley White and Samuel Levine remarked on the frequency of MI and angina in sleep and suggested an association with dreams.

Ischemic activity is an important prognostic marker in patients with cardiac disease, and characteristics of both REM and NREM sleep may conduce to nocturnal myocardial ischemia and angina. The few studies in patients with cardiac disease that have used sleep staging have concluded that in the absence of significant depression of left ventricular function, nocturnal ischemic events occur primarily during REM sleep,[21,22] which is characterized by increased sympathetic nerve activity, metabolic demands, and heart rate surges. In patients with stable coronary artery disease, myocardial ischemia is largely attributable to bouts of sympathetically mediated surges in heart rate and resultant metabolic demands in flow-limited, stenotic coronary arteries.[4,16,22-26] Nowlin and coworkers[22] attributed nocturnal angina to heightened blood pressure after performing detailed, multisession polysomnographic analysis of four patients with advanced coronary artery disease and nocturnal angina pectoris. They established that attacks of nocturnal angina occurred predominantly during REM sleep (32 of 39 recordings) and were associated with heart rate acceleration. Dream content, in patients who could describe it, included awareness of chest pain and involved strenuous physical activity or emotions of fear, anger, or frustration.

Nocturnal myocardial ischemia may be generated by mechanisms in addition to sympathetic nerve activity and unsatisfied metabolic demands. This possibility is suggested by the finding that nighttime ischemic events remain, although they are less frequent, in patients receiving β-adrenergic receptor blockade therapy, the primary therapy that effectively reduces the overall incidence of and suppresses the morning peak in cardiac events by containing sympathetic nerve activity and demand-related myocardial ischemia.[26,27] The main factors that may contribute to nondemand-related myocardial ischemia during NREM sleep are decreased coronary perfusion pressure as the result of hypotension[4,22-29] and increased coronary vasomotor tone.[28] These influences decrease the metabolic threshold for induction of nocturnal myocardial ischemia, which has a nadir between 1:00 and 3:00 AM.[23,28,30] During these hours in patients with stable coronary disease, Benhorin and colleagues[28] observed that myocardial ischemia can be provoked at heart rates of 83 beats/minute, in contrast to 96 beats/minute during midday, and that its incidence was not affected by β-adrenergic receptor blockade. Patel and colleagues[27] noted that nocturnal myocardial ischemia is attended by heart rate elevations of 6 beats/minute or less in patients with unstable angina receiving β-adrenergic receptor blocking agents. Mancia[5] hypothesized that the hypotension of NREM sleep is a major contributor to nocturnal myocardial ischemia and MI because it "reduces the volume and velocity of blood flow, favoring the development of thrombi and embolic and ischemic phenomena before and after arousal." It has also been postulated that myocardial ischemia provoked by transient thrombus formation[31] is attributable to the nocturnal nadir in endogenous fibrinolytic activity[31] as well as to peaks in serum levels of plasminogen activator inhibitor[31] and tissue plasminogen activator antigen, increasing blood viscosity or hypercoagulability at night, and free-radical generation.

Nondemand nocturnal myocardial ischemia is prevalent among patients with more severe coronary disease,[26,30,32] acute coronary syndromes,[17] or diabetes, populations with significant endothelial dysfunction. Indeed, it has been concluded that nondemand nocturnal ischemic episodes disclose a critical underlying coronary lesion, coronary vasospasm, or transient coronary artery stenosis.[27] Patel and colleagues[27] documented a nocturnal peak in ischemic events in their study of 256 hospitalized patients with the acute coronary syndromes of unstable angina and non–Q-wave MI (Figure 124-3). Electrocardiograms were recorded within hours after patients' admission for chest pain to the coronary care unit for

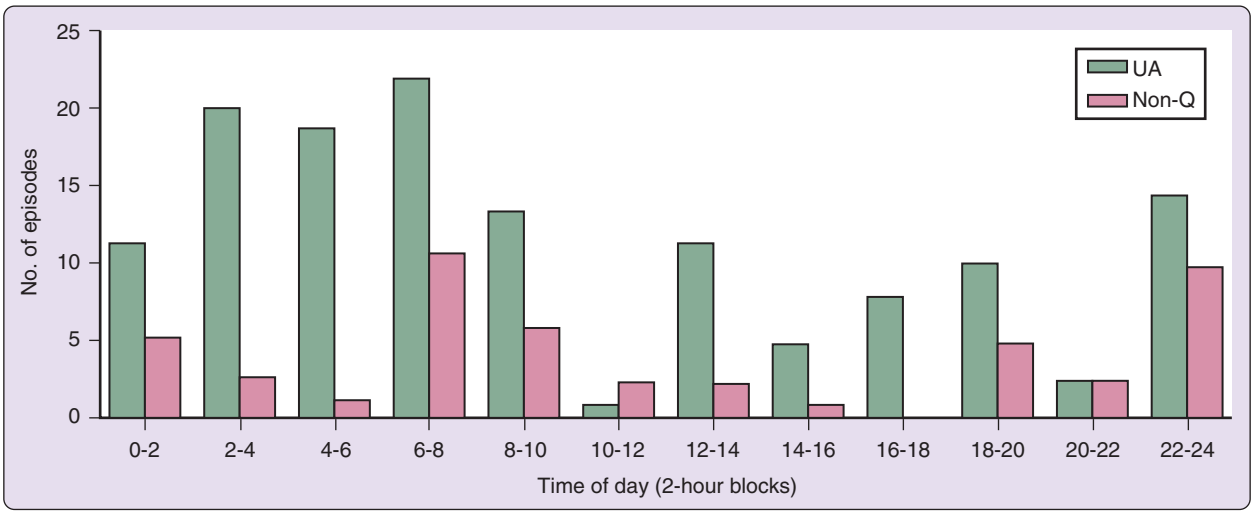


Figure 124-3 The circadian variation of ischemic activity based on 2-hour time blocks for the in-hospital study population. There is a single peak of ischemic activity at night between 10:00 PM and 8:00 AM, and no morning peak in ischemic activity is apparent. More than 64% of episodes occurred during this period ($P < .001$ compared with daytime). The circadian distribution of ischemic episodes in unstable angina (UA) and non–Q-wave myocardial infarction (Non-Q) is similar to the overall pattern of ischemic activity. (From Patel DJ, Knight CJ, Holdright DR, et al. Pathophysiology of transient myocardial ischemia in acute coronary syndromes: characterization by continuous ST-segment monitoring. *Circulation* 1997;95:1185–92, with permission from the American Heart Association.)

new-onset angina, sudden acceleration of previously stable angina, or angina within 1 month of MI. In the hospital, they received optimal medical therapy aimed at containing demand-related myocardial ischemia. It is important to note, however, that the peak in out-of-hospital onset of the syndromes followed the usual circadian pattern, as reported by Cannon and coworkers[33] in the Thrombosis in Myocardial Infarction (TIMI) III Study of 3318 patients. By contrast, in patients with longstanding diabetes or with documented autonomic nervous system dysfunction, there is no nocturnal decrease in myocardial ischemia or onset of acute MI.

Demand-related ischemic episodes can be effectively contained by β-adrenergic receptor blockade,[27] but antihypertensive treatment does not reduce the nocturnal incidence of nondemand-related myocardial ischemia.[34] The use of vasodilators to treat nondemand episodes resulting from endothelial dysfunction is the subject of debate. The lack of sleep staging and arterial blood pressure monitoring in patients with nocturnal myocardial ischemia leaves unidentified any contribution by autonomic and hemodynamic activity dictated by sleep states. Such monitoring would also disclose the prevalence of the established proischemic influences of nocturnal arousal and rising from bed.[15,35]

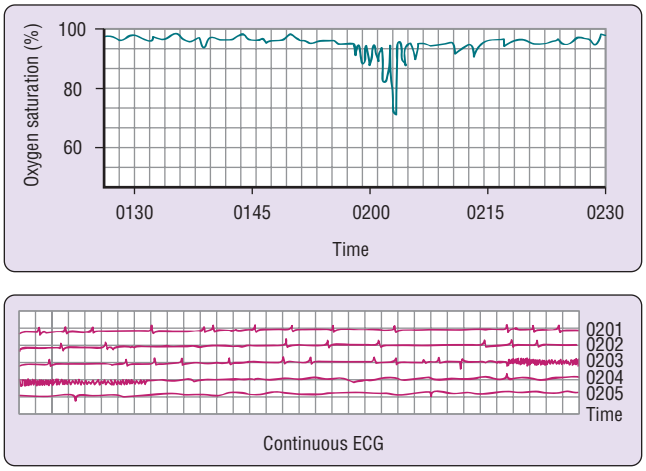


Figure 124-4 Importance of monitoring nocturnal oxygen saturation in patients who have sustained a myocardial infarction. Nonsustained ventricular tachycardia (*bottom*) and hypoxemia measured by pulse oximetry (*top*) occurred simultaneously in a patient on the third night after infarction. The patient died on the following day of cardiogenic shock. ECG, electrocardiogram. (From Galatius-Jensen S, Hansen J, Rasmussen V, et al. Nocturnal hypoxemia after myocardial infarction: association with nocturnal myocardial ischaemia and arrhythmias. *Br Heart J* 1994;72:23–30, with permission from the British Cardiac Society.)

Post–Myocardial Infarction Patients

During the first weeks after MI, sleep is significantly disturbed,[29,36] and nocturnal oxygen desaturation, especially in patients with impaired left ventricular function, may be generalized or episodic and may directly provoke tachycardia, ventricular premature beats, and ST-segment changes (Figure 124-4).[36-39] Both the duration and number of nighttime ischemic events are increased, consonant with increased cardiac sympathetic nerve activity[27,40] or decreased parasympathetic nerve activity (see Figure 124-2),[12] particularly in patients with residual myocardial ischemia.[17] Nocturnal levels of norepinephrine are increased, and nocturnal secretion of melatonin, an endogenous hormone that suppresses sympathetic nerve activity, is impaired.[41] These symptoms become normal over time so that within the first 6 months, ventricular tachycardia during sleep is relatively rare. Improved cardiac function early after MI is associated with decreased sleep apnea.[42]

The most detailed study to date of sleep in post-MI patients was performed in 1978 by Broughton and Baron,[29] who

reported on the sleep and cardiovascular condition of 12 patients, aged 33 to 70 years, after severe MI, first during their stay in the intensive care unit and then in the hospital ward. They noted a "marked disturbance of nocturnal sleep patterns … characterized by high amounts of wakefulness, stage 1, and number of awakenings, and REM density and low amounts of REM sleep, shorter REM periods with prolonged REM latencies. Sleep efficiency was substantially reduced."[29] All of these sleep-quality parameters improved in parallel with time after MI until on day 9 the only remaining abnormal feature was a high content of NREM N3 sleep. REM density peaked on post-MI nights 3 and 4 and NREM sleep on night 4. On subsequent hospital visits after discharge, the patients described terrifying dreams, suggesting that REM suppression was followed by REM rebound more than 2 weeks after the crisis. Importantly, Broughton and Baron observed that NREM sleep provoked nocturnal angina and awakening. They postulated that the hypotension associated with NREM sleep resulted in a diminution in perfusion pressure of the major coronary and collateral vessels supplying the mechanically compromised myocardium. The decreased heart rates typical of NREM sleep, however, were not observed, and heart rates were higher in NREM sleep than in wakefulness on half the nights recorded, indicating enhanced cardiac sympathetic nerve activity even in NREM sleep. In half of the cases, the electrocardiogram amplitude decreased during anginal attacks. In the context of nighttime monitoring, it is of interest to note that T-wave alternans, an electrocardiographic (ECG) phenomenon indicating vulnerability to lethal arrhythmias,[43] was documented in the nighttime electrocardiogram of a patient with left dysfunction enrolled in the ambulatory ECG arm of the Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (EPHESUS) (Figure 124-5).[44]

Nocturnal Myocardial Infarction

Although only 20% of MIs occur between midnight and 6:00 AM, their nonuniform distribution implicates pathophysiologic triggers.[1] The dynamic perturbations in autonomic nervous system activity both independent of and in conjunction with apnea[45] are likely to constitute important triggers of MI at night. REM-induced surges in sympathetic nerve activity have the potential to provoke tachycardia and hypertension, alterations that carry the potential for inducing MI secondary to coronary artery plaque rupture as well as to inappropriate decreases in the myocardial oxygen supply-demand relationship or α-adrenergically mediated coronary vasoconstriction.

Alternatively, in a starkly opposite manner, the hypotension of slow wave sleep may lead to malperfusion of the myocardium because of reduced coronary perfusion pressure through stenotic vessel segments (Figure 124-6).[46] Several investigators[27,47,48] have attributed nocturnal MI and myocardial ischemia to the relative hypotension of NREM sleep, which "reduces the volume and velocity of blood flow, favoring the development of thrombi and embolic and ischemic phenomena before and after arousal."[5] Mancia therefore advocated avoiding drugs that enhance the hypotension of NREM sleep and prescribing antihypertensive medications only for daytime therapy.[5] He echoed the argument of Floras,[34] who observed that antihypertensive treatment did not reduce the incidence of nocturnal MI and myocardial ischemia. Further evidence of the risk for hypotension-induced infarction has been provided by Kleiman and colleagues,[47] who reported that the incidence of subendocardial MI clustered at 2:00 to 4:00 AM, simultaneously with the nadir in arterial blood pressure. Other factors known to contribute to MI are operative during sleep, including increased ventricular diastolic pressures and volumes caused by the fluid shifts resulting from assuming a supine posture, unfavorable alterations in the balance of fibrinolytic and thrombotic factors,[31] and chronic or episodic oxygen desaturation.[29,36-39]

Specific patient groups experience an increased incidence of nighttime MIs, particularly those with poor ventricular function, advanced age, or diabetes.[49,50] The risk for development of congestive heart failure is higher for nighttime than daytime MIs,[51] potentially because of either the pathologic process or a delay in obtaining high-quality care.

Hypertension

Patients whose nighttime arterial blood pressure declines less than 10% from day to night (called "nondippers") are at increased risk for total and cardiovascular mortality[52] as well as all cardiovascular end points,[53] frequent or complex ventricular arrhythmias,[54] myocardial ischemia,[55] cerebrovascular insult,[56] and increased organ damage, including cardiac

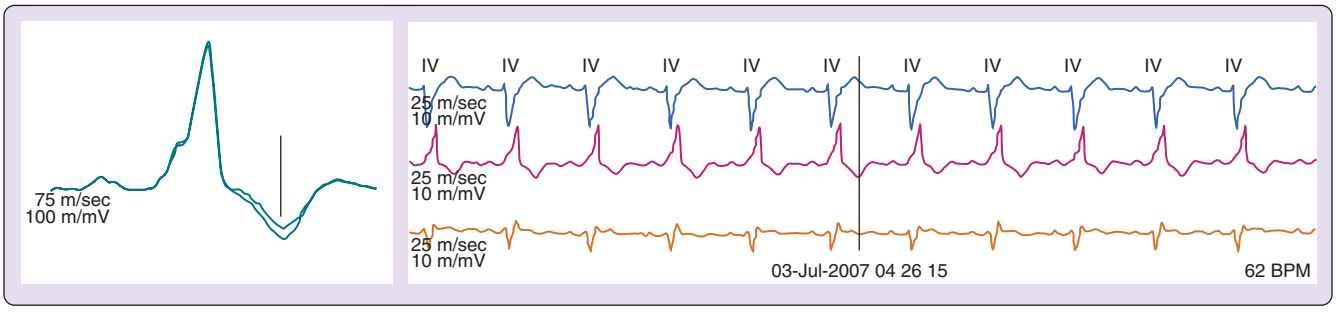


Figure 124-5 High-resolution template showing T-wave alternans (65 μV) in precordial lead V_3 in superimposed electrocardiographic (ECG) waveforms from a nighttime ambulatory ECG recording of a patient with heart failure and left ventricular dysfunction enrolled in the ambulatory electrocardiographic (AECG) substudy of the Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (EPHESUS). The associated ECG strip is also provided. (From Stein PK, Sanghavi D, Domitrovich PP, et al. Ambulatory ECG-based T-wave alternans predicts sudden cardiac death in high-risk post-MI patients with left ventricular dysfunction in the EPHESUS Study. *J Cardiovasc Electrophysiol* 2008;19:1037–42, with permission from Wiley.)

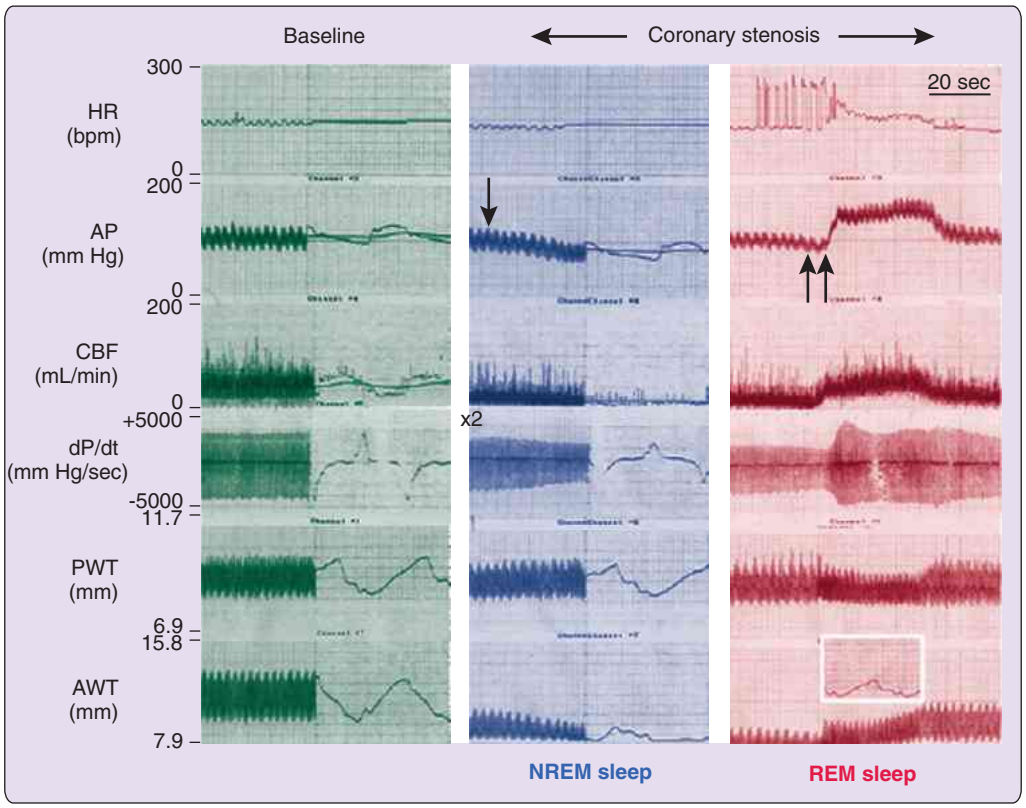


Figure 124-6 Representative tracings for hemodynamic variables at baseline (before stenosis) and during NREM (*single arrow*) and REM (*double arrow*) sleep in the presence of coronary stenosis. NREM sleep initiated a decrease in arterial pressure (AP), resulting in akinesis in the anterior wall (stenotic region). REM sleep induced a rapid increase in heart rate (HR), AP, and dP/dt_{max} (rate of change of left ventricular pressure). The onset of REM sleep increased coronary blood flow, which returned anterior wall function to the poststenotic condition before the onset of NREM sleep. (See expanded tracing in box.) AWT, Anterior wall thickness; CBF, coronary blood flow; PWT, posterior wall thickness. (From Kim SJ, Kuklov A, Kehoe RF, et al. Sleep-induced hypotension precipitates severe myocardial ischemia. *Sleep* 2008;31:1215–20, with permission from the Sleep Research Society and the American Academy of Sleep Medicine.)

hypertrophy.[57] The absence of a nocturnal decline in blood pressure may be an important marker of complications among patients with type 1 diabetes,[58] and it may be reflected in the significant incidence of death at 2:00 to 4:00 AM among hypertensive women reported by Mitler and colleagues (Figure 124-7).[59] Faulty baroreceptor activation may account for the fact that arterial blood pressure during sleep remains significantly elevated in these hypertensive patients, who typically show evidence of central hypersympathetic nerve activity with an increased number of microarousals, a reduced length and depth of NREM sleep, and a shortened REM latency. Blunted endothelium-dependent vasodilation is also implicated.

Heart Failure

Excess mortality risk attends chronic congestive heart failure, particularly among the 40% to 80% of heart failure patients with either obstructive or central sleep apnea. These sleep-related breathing disorders may contribute to the severity of disease, specifically to remodeling of cardiac chambers and left ventricular diastolic dysfunction as well as to T-wave alternans.[43,60] In patients with systolic heart failure, central sleep apnea, smoking,[61] severe right ventricular systolic dysfunction, and low diastolic blood pressure are associated with increased nocturnal ventricular arrhythmias and mortality.[62] Treatment of apnea with continuous positive airway pressure, medications, or devices frequently lessens heart failure symptoms and mortality risk. Diagnostic and treatment strategies are discussed in Chapter 129.

Patients with Epilepsy

Whereas most patients with epilepsy exhibit a low incidence of seizure at night,[63] sudden unexpected death in epilepsy (SUDEP) occurs primarily in patients with a history of nocturnal seizures.[64-66] More than 99% of nocturnal seizures arise in NREM sleep.[67] Sleep disorders are common among patients with epilepsy; sleep apnea worsens seizures, and its treatment improves seizure control.[67,68]

Elderly Patients

Elderly individuals' (particularly women's) reports of daytime sleepiness, suggesting poor quality of sleep, are associated with mortality, cardiovascular morbidity and mortality, MI, and congestive heart failure.[69] Depression, poor health, daytime angina, a limited activity level, and cardiac arrhythmias may accompany disturbed sleep in elderly individuals. Initiating a moderately intense exercise program significantly improves sleep quality[70] and autonomic status[71] in formerly sedentary older people. Nocturnal myocardial ischemia is not uncommon in older patients with vascular disease who experience regular episodes of oxygen desaturation and increased heart

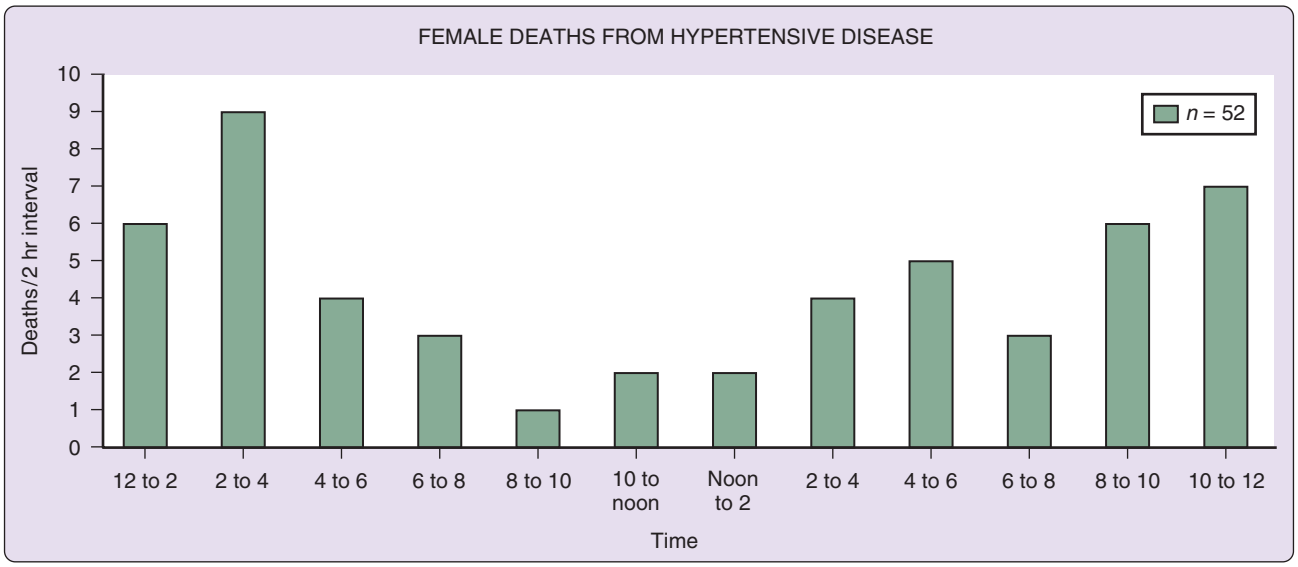


Figure 124-7 The temporal distribution of female deaths attributed to hypertensive disease peaked at 2:00 to 4:00 AM. The temporal concentration was statistically significant ($P < .01$). Data were derived from a 4600-person (>8%) sample of deaths due to disease in New York City in 1979. (Reprinted from Mitler MM, Hajdukovic RM, Shafor R, et al. When people die: cause of death versus time of death. *Am J Med* 1987;82:266–74. Copyright 1987, with permission from Excerpta Medica, Inc.)

rate. Conflicting evidence has been presented of increased risk for nighttime compared with daytime MI and sudden cardiac death in elderly people.[48,49,72] Impaired baroreceptor sensitivity,[73] a measure of the capacity for reflex vagus nerve activation,[6] and increased low-frequency power of HRV[74] are evident at night in susceptible elderly patients. Given this autonomic background, it is not surprising that nighttime multifocal activity in elderly patients is a predictor of cardiac mortality.

CLINICAL PEARLS

- Sleep exerts a major impact on the health of the patient with cardiac disease through both direct cardiovascular influences and sleep-disordered breathing.[75] In a sense, the diseased heart and lungs are unwitting victims of the needs of the sleeping brain, which commands dramatic alterations in autonomic and respiratory activity.
- A sizeable population experiences cardiac events during sleep, with identifiable high-risk groups (see Table 124-1).
- Sleep presents unusual opportunities to monitor the patient with cardiac disease because there is growing appreciation of the fact that nighttime heart rate, blood pressure, myocardial ischemia, arrhythmias, and respiratory disturbances carry predictive value for daytime events (Box 124-1).
- Daytime tests cannot substitute for nighttime monitoring of the patient with cardiac disease because exercise treadmill testing and daytime ambulatory monitoring cannot replicate the autonomic, hemodynamic, or respiratory challenges that uniquely accompany sleep.
- Improved identification of the precise triggers of nocturnal cardiac events may be anticipated when technologies are integrated for monitoring sleep state, respiration, oxygen desaturation, and cardiovascular variables.

SUMMARY

Specific patient groups exhibit elevated sleep-related cardiac risk, including those with a history of cardiac arrhythmias, myocardial infarction, angina, heart failure, or sleep apnea; family members of a person who died from SIDS; and patients with Brugada syndrome, long-QT syndrome, or epilepsy. Improved risk assessment is possible through noninvasive monitoring of autonomic variables and repolarization abnormalities.

ACKNOWLEDGMENTS

The authors thank Sandra Verrier for her editorial contributions.

Selected Readings

Broughton R, Baron R. Sleep patterns in the intensive care unit and on the ward after acute myocardial infarction. *Electroencephalogr Clin Neurophysiol* 1978;**45**:348–60.

Hung J, Whitford EG, Parsons RW, et al. Association of sleep apnoea with myocardial infarction in men. *Lancet* 1990;**336**:261–4.

Mancia G. Autonomic modulation of the cardiovascular system during sleep. *N Engl J Med* 1993;**328**:347–9.

Shamsuzzaman AS, Somers VK, Knilans TK, et al. Obstructive sleep apnea in patients with congenital long QT syndrome: implications for increased risk of sudden cardiac death. *Sleep* 2015;**38**:1113–19.

Somers VK, Dyken ME, Mark AL, et al. Sympathetic nerve activity during sleep in normal subjects. *N Engl J Med* 1993;**328**:303–7.

Trimer R, Cabiddu R, Mendes RG, et al. Heart rate variability and cardiorespiratory coupling during sleep in patients prior to bariatric surgery. *Obes Surg* 2014;**24**(3):471–7.

Verrier RL, Klingenheben T, Malik M, et al. Microvolt T-wave alternans: physiologic basis, methods of measurement, and clinical utility. Consensus guideline by the International Society for Holter and Noninvasive Electrocardiology. *J Am Coll Cardiol* 2011;**44**:1309–24.

A complete reference list can be found online at ExpertConsult.com.

Chapter 125

Cardiac Arrhythmogenesis During Sleep: Mechanisms, Diagnosis, and Therapy

Richard L. Verrier; Mark E. Josephson

Chapter Highlights

- The pronounced sleep state–dependent changes in autonomic nervous system activity and respiration can provoke both atrial and ventricular arrhythmias in patients with cardiovascular disease.
- Mortality due to ventricular arrhythmias is most frequent during sleep in the distinct syndromes of sudden infant death, sudden unexplained nocturnal death, and the Brugada syndrome, each of which has been linked to genetic abnormalities.
- The proarrhythmic potential of class III antiarrhythmic agents (potassium channel blockers) for patients with significant heart rate pauses and the sleep-disrupting effect of medications must be considered.

Cardiac arrhythmias are prevalent in the 13.5 million Americans with heart disease, with potentially severe consequences. Approximately 15% of sudden cardiac deaths, which result from lethal ventricular arrhythmias, occur during sleep, and most atrial arrhythmias in patients younger than 61 years have their onset at nighttime. Sleep apnea profoundly alters autonomic nervous system activity and increases risk of arrhythmia, hypertension, and myocardial infarction (see Chapters 126–128). Lack of streamlined technology for concurrent monitoring of sleep state, electrocardiogram (ECG), oxygen saturation, and respiration continues to hamper diagnosis and evaluation of therapy of these arrhythmias. We will review the current state of knowledge regarding epidemiology, risk factors, pathogenesis, and treatment options for each nocturnal arrhythmia type.

VENTRICULAR ARRHYTHMIAS

Malignant ventricular arrhythmias are usually suppressed during sleep, as is evidenced by the nocturnal trough in incidence of myocardial infarction, sudden cardiac death, implantable cardioverter–defibrillator discharge, myocardial ischemic events, and arrhythmias in patients with ischemic heart disease.[1,2] This decrement coincides with lessened metabolic demands during non–rapid eye movement (NREM) sleep, which occupies approximately 80% of sleep time. However, sleep is not entirely free of risk because the nocturnal incidence of sudden cardiac death, which is attributable to ventricular fibrillation, has been calculated at approximately 15%[3] or 45,000 cases annually in the United States alone. Moreover, the nonuniformity of the nighttime distribution of these events (Figure 125-1)[3] suggests physiologic triggering that may be amenable to monitoring for improved diagnosis and therapy.

Surges in cardiac sympathetic nerve activity during rapid eye movement (REM) sleep have been implicated in nocturnal ventricular arrhythmias and myocardial ischemia[4-7] (see Chapter 124). The specific mechanisms of REM-induced cardiac events include direct effects on electrophysiologic status or indirect consequences of heart rate and arterial blood pressure accelerations, which may disrupt plaques and lead to intra-arterial platelet aggregates, releasing proarrhythmic constituents such as thromboxane A_2.[8] Myocardial ischemia or other changes in cardiac substrate and mechanical function resulting from disease,[9] infarction,[10] or aging[11] can amplify nocturnal electrical instability. Oxygen desaturation may trigger nighttime ventricular tachycardia in patients with cardiac disease in the subacute phase after myocardial infarction[10] or in those with heart failure.[9] Hypoxemia and tachycardia frequently occur together during sleep after major surgery and may promote myocardial ischemia.[12] Frequent or complex arrhythmias are also characteristic of hypertensive patients in whom the typical nocturnal trough in blood pressure is not observed.[13] The nocturnal increase in QT-interval dispersion among survivors of sudden cardiac death,[14] acute myocardial infarction,[15] and heart failure[15] provides evidence of their increased vulnerability to cardiac arrhythmias at night.

REM-related nocturnal arrhythmogenesis may have a significant affective component. REM sleep dreams, which may be vivid, bizarre, and emotionally intense, commonly generate the emotions of anger and fear. Because these emotions have been linked in wakefulness to the onset of myocardial infarction and sudden death,[16] it is reasonable to hypothesize that when these affective states are evoked during dreaming, they may trigger lethal events. This possibility is illustrated by a case report of recurrence of ventricular fibrillation in a 39-year-old man with normal coronary arteries and cardiac function while sleeping. A subsequent sleep study determined that ventricular premature beats were substantially increased during REM and that dreams at the same hour that fibrillation had occurred were emotionally charged.[17]

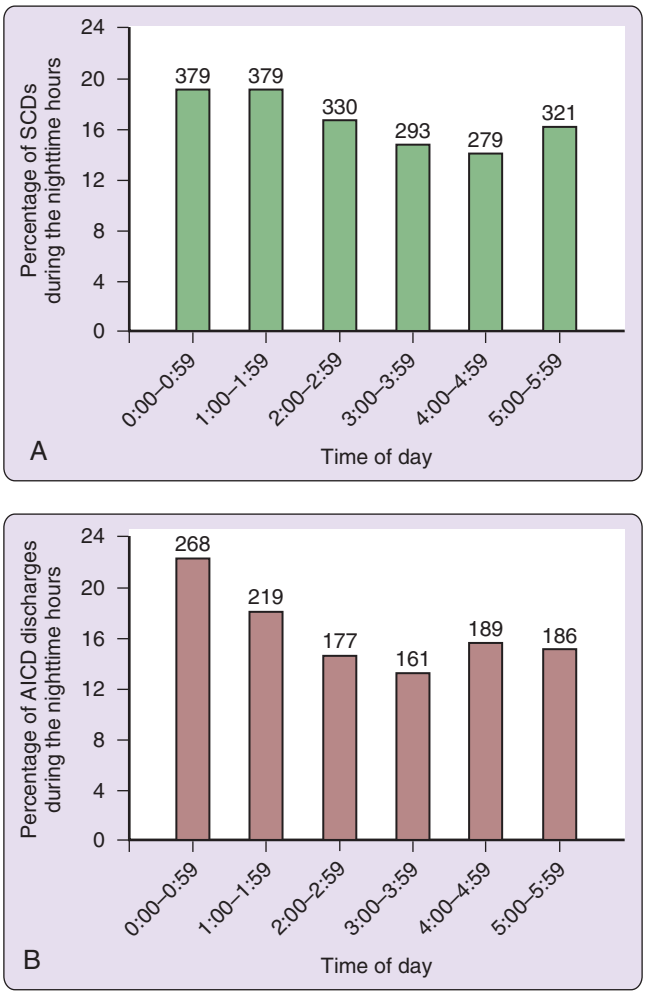


Figure 125-1 A, Hourly incidence of sudden cardiac death (SCD) onset between midnight and 5:59 AM from 12 studies enrolling 1981 patients. The number of sudden cardiac deaths observed each hour is indicated above each bar. **B,** Hourly incidence of automatic implantable cardioverter-defibrillator (AICD) discharge between midnight and 5:59 AM from seven studies enrolling 1197 patients, who experienced 1200 discharges during the nocturnal period. The number of discharges observed each hour is indicated above each bar. (From Lavery CE, Mittleman MA, Cohen MC, et al. Nonuniform nighttime distribution of acute cardiac events: a possible effect of sleep states. *Circulation* 1997;5:3321–7.)

In some cases, arrhythmia frequency may be enhanced during NREM sleep, when latent automatic foci are exposed by the generalized reduction in heart rate after withdrawal of overdrive suppression, or when hypotension exacerbates impaired coronary perfusion.

Therapy

In most cases, an electrically unstable substrate underlies the propensity to develop nocturnal ventricular arrhythmias, and treatment is similar to that for daytime arrhythmias. If surges in sympathetic nerve activity, which typically occur during REM sleep and dreaming, are suspected, beta-adrenergic receptor blockade therapy may prove helpful, with careful attention to avoiding medications that disrupt sleep.[18]

In treating hypertensive patients, it is important to appreciate Mancia's[19] suggestion that pharmacologic therapy that exacerbates the hypotensive effect of NREM sleep may introduce the potential risk of thrombosis and embolism in patients with stenotic lesions in the heart or brain. Floras[20] determined that the nocturnal incidence of myocardial infarction was not diminished in patients treated with antihypertensive agents and suggested that the agents induced nocturnal hypotension. Thus, special attention should be given to the hemodynamic effects of antihypertensive drugs and vasodilators to avoid precipitating cardiac events by inducing profound hypotension. The importance of ruling out "white coat" hypertension is underscored because more than 30% of individuals with elevated blood pressure readings in the physician's office or hospital prove to be normotensive during daily life, as documented by ambulatory blood pressure monitoring.[21]

Nighttime onset of ventricular arrhythmias may also indicate provocation by disturbed breathing, which can be treated by continuous positive airway pressure.[22] For example, sleep-disordered breathing provokes ventricular arrhythmias and appropriate firing of implantable cardioverter–defibrillators, suggesting apnea-screening and therapy to reduce these events.[23] Risk of apnea-induced cardiac events is not limited to the nocturnal period and can be monitored from continuous ECGs by T-wave alternans,[24] a marker of risk for lethal ventricular arrhythmias.[25]

NOCTURNAL ASYSTOLE AND QT-INTERVAL PROLONGATION

Sinus pauses of less than 2 seconds, prolonged atrioventricular (AV) conduction, Wenckebach AV block, and bradycardia are well documented in normal populations during sleep and are attributed to effects of increased parasympathetic activity on AV node conduction.[26,27] These asystoles are more frequent in individuals who are young[28,29] or physically fit, such as athletes[30,31] and heavy laborers.[32] More extreme cases were observed by Guilleminault and colleagues,[32a] who reported periods of sinus arrest of up to 9 seconds during REM sleep in young adults with apparently normal cardiac function. It was concluded that the nocturnal asystoles were the result of exaggerated, if not abnormally elevated, vagal tone, because muscarinic receptor blockers significantly reduced the duration of the nocturnal asystoles but did not prevent them. No further therapeutic intervention was warranted.

However, in patients with cardiac disease, especially those taking class III antiarrhythmic drugs (potassium channel–blocking agents), nocturnal asystolic events can set the stage for ventricular arrhythmias. Such prolongation of cycle length can facilitate the development of early afterdepolarizations and the lethal arrhythmia torsades de pointes. In patients with damaged endothelium resulting from coronary atherosclerosis, the acetylcholine released by surges in vagus nerve activity could result in vasoconstriction rather than vasodilation because of impaired release of endothelium-derived relaxing factor.[33]

Nocturnal heart rate pauses may be particularly arrhythmogenic in subsets of patients with the long QT syndrome, specifically LQT2 and LQT3, who have mutations on the sodium channel, voltage-gated, type V, alpha gene (*SCN5A*).[34] The lethal arrhythmias occur almost exclusively at rest or during sleep, when the QT interval is typically prolonged.[35] (See later section on sudden infant death syndrome.)

Therapy

Ascertaining whether patients exhibit nocturnal heart rate pauses is important when treating individuals for whom class III antiarrhythmic drugs (potassium channel blockers) are the primary option.

ATRIAL FIBRILLATION

In the 2.5 million U.S. patients with atrial fibrillation, which has serious consequences in terms of increased morbidity and mortality,[36] it is likely that 10% to 25% of the arrhythmias are facilitated by vagal influences. This has been termed *vagally mediated atrial fibrillation*. Several investigators have reported nocturnal peaks in onset of paroxysmal atrial fibrillation.[37-39] A significant midnight to 2:00 AM peak in atrial fibrillation onset and a higher average nocturnal incidence were documented by Rostagno and colleagues[37] in their review of records from 10 years of responses by mobile coronary care units staffed by cardiologists in Florence, Italy. This arrhythmia was also found to exhibit a peak in frequency of onset at midnight in a Japanese population of 60 years of age or younger. The maximal duration of the arrhythmia (77±27 minutes per episode) was also greatest between midnight and 6:00 AM (Figure 125-2).[38] Other investigators characterized a 4:00 AM to 5:00 AM peak in onset of paroxysmal atrial fibrillation that was refractory to antiarrhythmic drugs in a 3-month study of 67 patients with implantable cardioverters.[39] The 514 recorded episodes with an atrial rate of greater than 220 beats per minute lasted more than 1 minute before termination by pacing or spontaneous reversion. A potential contribution of sympathetic nerve activity is implicated by the timing of these bouts of atrial fibrillation, which occurred during a period of sleep when REM typically emerges. However, the potential of REM sleep to trigger the arrhythmia was not discussed.

Patients with more frequent atrial tachycardia are more likely to develop atrial tachycardia/fibrillation at night.[40] Records of concurrent monitoring of sleep and nocturnal onset of atrial fibrillation are rare (Figure 125-3).[41] In the case illustrated, disruption of the nocturnal trough in heart rate disclosed sleep-related atrial fibrillation. Available evidence indicates that nocturnal atrial fibrillation is provoked during periods of intense vagus nerve activity, as indicated by heart rate variability studies,[42,43] and the presence of bradycardia,[44] in individuals with structurally normal hearts. Enhanced adrenergic activity may interact in a complex manner with changes in vagal tone to affect atrial refractoriness and dispersion of repolarization and alter intra-atrial conduction, thus increasing the propensity to develop this arrhythmia.[36,44] The high level of vagus nerve tone maintained during slow wave sleep has the capacity to exacerbate atrial fibrillation in patients whose atria are particularly prone to the arrhythmogenic influence of acetylcholine.[36]

Risk of atrial fibrillation is doubled if breathing during sleep is disordered,[45] because apnea can provoke nocturnal hypoxemia, sympathetic nerve activity, and hemodynamic stress.[22,45] Obstructive apnea–induced surges in blood pressure distend atrial chambers and can activate stretch receptors. Incidence of atrial fibrillation is strongly predicted by nocturnal oxygen desaturation in subjects younger than 65 years old

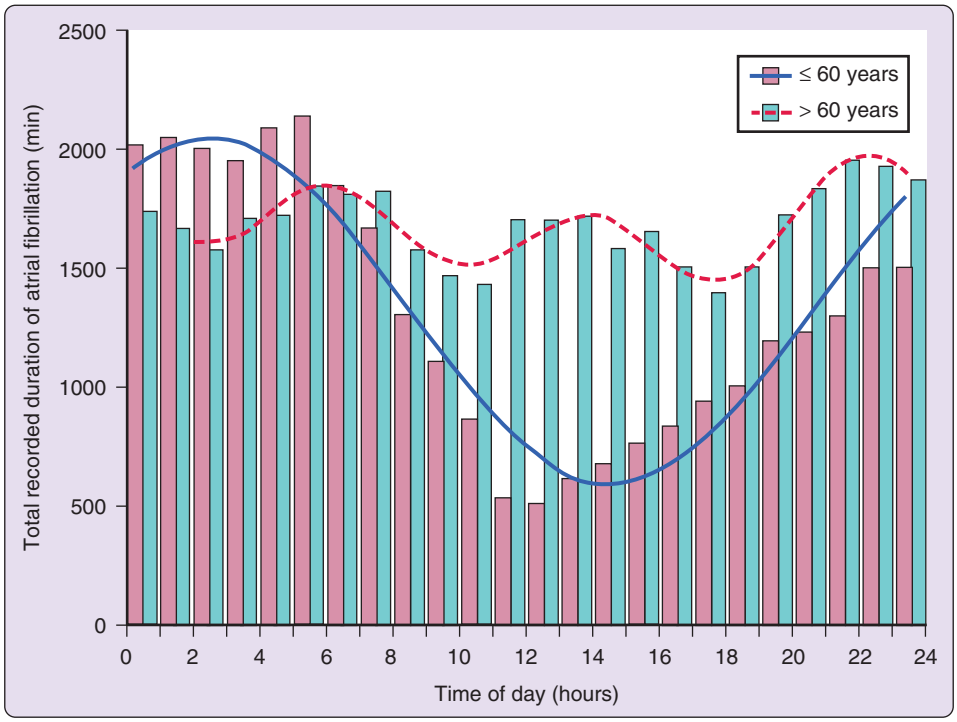


Figure 125-2 Hourly total duration of paroxysmal atrial fibrillation in younger (<60 yr; *red bars*) and older patients (*green bars*). The single harmonic fit of the data from the younger patients is shown by the *unbroken line*. The triple harmonic fit in the older patients is shown by the *broken line*. A prominent monophasic circadian rhythm is present in younger patients, in contrast to a toneless triphasic rhythm in older patients. (From Yamashita T, Murakawa Y, Hayami N, et al. Relation between aging and circadian variation of paroxysmal atrial fibrillation. *Am J Cardiol* 1998;82:1364–7.)

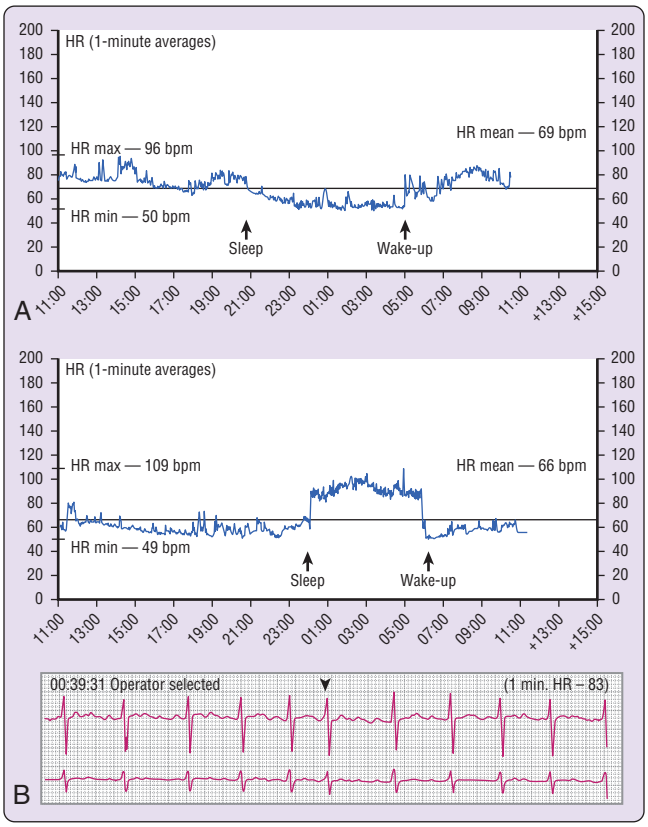


Figure 125-3 **A,** Heart rate (HR) trend from an ambulatory electrocardiogram (AECG) showing a normal circadian rhythm with a sleep-induced decrease in heart rate. **B,** Heart rate trend from an AECG in our patient shows a nocturnal increase in heart rate caused by paroxysmal atrial fibrillation at the onset of sleep and a drop in heart rate after awakening, resulting from spontaneous conversion to sinus rhythm. The ECG (*below*) documents atrial fibrillation during the sleep period. (From Singh J, Mela T, Ruskin J. Images in cardiovascular medicine: sleep [vagal]-induced atrial fibrillation. *Circulation* 2004;110: e32–3.)

and by heart failure in older subjects.[45] Moreover, ablation for atrial fibrillation is effective in patients appropriately treated for sleep-disordered breathing but of limited value in patients whose apnea is not treated.[46] Furthermore, patients with atrial fibrillation and severe sleep apnea are less likely to respond to antiarrhythmic medical therapy.[47]

Therapy

Medical therapy is similar to that for patients whose arrhythmias occur during the day, including therapy to control rate or terminate the arrhythmia pharmacologically or with an atrial cardioverter–defibrillator. Because nighttime atrial fibrillation is classified as *vagally mediated*, anticholinergic agents, such as disopyramide and flecainide, are sometimes helpful to prevent recurrences; adrenergic blocking drugs or digitalis sometimes worsen symptoms.[48] In addition, individuals with nocturnal onset of atrial fibrillation should be monitored for the presence of sleep-disordered breathing, which can be effectively treated by continuous positive airway pressure.

SUDDEN INFANT DEATH SYNDROME

Sudden infant death syndrome (SIDS), the leading cause of mortality in infants between 1 week and 1 year of age, occurs during sleep.[49] The syndrome is a diagnosis of exclusion; that is, it includes all causes that remain unexplained after a thorough case investigation, including an autopsy, examination of the death scene, and review of the clinical history. Thus, SIDS, which took a toll of 2234 infants in 2001 in the United States[50] or 8.1% of infant deaths, may be attributable to a variety of etiologies that challenge the developing cardiorespiratory system. The fatal event in SIDS victims is characterized by hypotension and bradycardia[51] and appears to be attributable to a deficit in the normal reflex coordination of heart rate, arterial blood pressure, and respiration during sleep.[52] This failure to respond to cardiorespiratory challenges during sleep may result from a binding deficit in the arcuate nucleus of SIDS infants,[52] because muscarinic cholinergic activity in this structure at the ventricular medullary surface is postulated to be involved in cardiorespiratory control. Heart rates in infants who later died of SIDS are generally higher and exhibit a reduced range, suggesting altered autonomic control.[53] Autonomic instability has also been documented in NREM sleep in infants with aborted SIDS events.[54]

Repolarization abnormalities have also been observed. Recent evidence from a 19-year, prospective, multicenter observational study of 34,442 infants determined that significant prolongation (35 milliseconds or more) of the QT interval characterized the 24 (0.07%) infants who died of SIDS within the first year of life.[55] These results suggest that some SIDS cases may be attributed to a genetic defect that produces a developmental abnormality in cardiac sympathetic innervation and alters repolarization to increase the risk of ventricular arrhythmia. These repolarization abnormalities typify infants and children with the long QT syndrome genotype linked to chromosome 3 (LQT3). Mutations in the sodium channel gene *SCN5A* are the most common causes of long QT syndrome and are responsible for the arrhythmias and reduced heart rates. The genetic locus of the defect and the length of the QT interval are independent predictors of risk.[56] T-wave alternans, an electrocardiographic indicator of heightened vulnerability to sudden cardiac death,[25] has been reported in infants who became SIDS victims[57,58] or who were successfully treated with pacing[59] or beta-blockade therapy.[60,61] The latter therapy diminished T-wave alternans, indicating antiarrhythmic efficacy.

Among environmental influences, the increased risk of SIDS during the winter season is well documented[62,63] and is not related to bronchiolitis.[64] A genetic susceptibility that may interact with environmental factors has been implicated by a 5.8-fold increase in recurrence of SIDS within families.[65] Tishler and colleagues[66] reported a significant incidence of deficits in ventilatory responses to hypoxia in families with apnea. Conflicting evidence has been provided regarding the relative increase in risk attributable to prone (face-down) sleeping,[67-71] and decreased incidence of SIDS has been attributed to the Back-to-Sleep campaign, which advocates placing infants in a supine position for sleeping.

Passive cigarette smoking is a highly significant modifiable risk factor in SIDS. A reduction of 61% in the number of SIDS deaths has been projected if smoking were eliminated from infants' environments.[66-69,72] A dose-dependent effect has been demonstrated.[67] Maternal smoking during gestation is also implicated.[72-74] Established SIDS risk factors of preterm birth and low birth weight increased risk more than 15-fold

among smokers[74] but not at all among nonsmokers. Illegal drug use increases risk of SIDS by more than fourfold.[75] The mechanisms may include impairment in chemoreceptor responsiveness, resulting from decreased sensitivity to carbon dioxide in infants of substance-abusing mothers.[75] The increase in SIDS due to passive smoking may be attributable to nicotine's adverse effect on chemoreceptor activation of respiration,[76] dulling the arousal response to hypoxia.[77] Nicotine and its metabolites have been found at autopsy in the pericardial fluid of SIDS infants.[78,79] Epicardial nicotine is associated with hypopnea[80] and affects the sinoatrial node and epicardial neural fibers to induce hypotension and bradycardia,[81,82] the documented symptomatology of the final event in SIDS infants.[51]

Therapy

This profile suggests some straightforward opportunities for intervention, including placing infants in a supine (face-up) position for sleeping and avoidance of maternal smoking during gestation and passive smoking during infancy. Theoretically, sodium channel blockade[34] or cardiac pacing[34,59] might be useful in treating infants diagnosed with the long QT3 syndrome, but prospective studies are required. Beta blockade is the current treatment of choice.[34,60,61] Assessment of vulnerability to arrhythmias by QT-interval prolongation has been suggested in a prospective study[55] and by T-wave alternans in multiple clinical reports based on ambulatory electrocardiographic (AECG) recordings.[60,61]

THE BRUGADA SYNDROME AND SUDDEN UNEXPLAINED NOCTURNAL DEATH SYNDROME

The striking phenomenon of sudden death during sleep has been reported in Western adults diagnosed with the Brugada syndrome, which strikes men almost exclusively, and in young, apparently healthy Southeast Asian men with the sudden unexplained nocturnal death syndrome (SUNDS). The latter syndrome is named *lai-tai* ("sleep death") in Laos, *pokkuri* ("sudden and unexpected death") in Japan, and *bangungut* ("to rise and moan in sleep") in the Philippines. These syndromes probably represent the same disorder, which is characterized by right precordial ST segment elevation.[56,83] Deaths are due to lethal ventricular arrhythmias.

The Brugada syndrome is considered responsible for 4% to 12% of all sudden cardiac deaths and for approximately 20% of deaths in patients with structurally normal hearts.[56] The electrocardiographic abnormality is estimated to be present in approximately 5 per 10,000 inhabitants, and, apart from accidents, in geographic regions where it is widespread, this inherited syndrome is the leading cause of death of men younger than 50 years of age. A single sodium channel mutation in the *SCN5A* gene identified in an eight-generation kindred with a high incidence of nocturnal sudden cardiac death, QT-interval prolongation, and Brugada-like ECG characterizes 20% of Brugada patients; other mutations are suspected. Genetic defects in the sodium channel are also associated with progressive conduction system disease attended by bradycardia. A mechanistic link with enhanced presynaptic norepinephrine recycling has been described.[84] Presynaptic sympathetic cardiac dysfunction has been hypothesized, based on abnormal iodine-123 metaiodobenzylguanidine uptake, with bradycardia-dependent QT prolongation, intrinsic sinus node dysfunction, conduction abnormalities, and absence of ventricular ectopy.[85]

In the United States, 117 SUNDS deaths were registered among male Southeast Asian immigrants or their descendants from 1981 to 1988.[86] Autopsies of those who died of SUNDS have established that cardiovascular disease is absent, but, in some instances, conduction pathways are developmentally abnormal.[83] Companions have reported that the immediate symptoms are onset of agonal respirations during sleep along with vocalization; violent motor activity; non-arousability; rapid, irregular deep breathing; perspiration; heart rate surges; and severe autonomic discharge. Several victims revived by vigorous massage reported sensations of airway obstruction, chest discomfort or pressure, and numb and weak limbs. When these symptoms recurred within weeks to months, they culminated in death.[87] Three victims who had been resuscitated from ventricular fibrillation then experienced recurring fibrillation in the hospital during sleep accompanied by similar moaning vocalizations. In these three patients, there was no evidence of atherosclerosis or structural abnormalities and no sleep apnea, but creatine kinase levels were markedly elevated, and potassium was depressed. Vagal tone is lower in SUNDS survivors than in healthy individuals, particularly at night.[88]

The Brugada syndrome is genetically related to the long QT syndrome, which shares risk for lethal nocturnal ventricular arrhythmias,[89] a period when these patients exhibit abnormal levels of T-wave alternans.[25,90]

Therapy

Development of effective therapy for these syndromes has been particularly challenging. Currently, implantation of cardioverter–defibrillators appears to be the most effective approach in patients with Brugada syndrome[56] or risk of SUNDS.[83]

SLEEP-DISRUPTING EFFECTS OF CARDIAC MEDICATIONS

Several important medications that are widely prescribed for patients with cardiac disease, including antihypertensive agents and beta blockers that cross the blood–brain barrier, have the potential to disrupt sleep.[18] In particular, the lipophilic beta blockers (pindolol, propranolol, and metoprolol) increase the total number of awakenings and total wakefulness compared with placebo and with the nonlipophilic atenolol. Penetration of the blood–brain barrier occurs with prolonged therapy, when these distinctions may become less apparent. In addition, pindolol, which has intrinsic sympathomimetic activity, increases REM latency and, as a result, decreases REM sleep time. Sleep disruption may provoke daytime fatigue and lethargy, symptoms widely reported by patients taking beta blockers, which may prompt discontinuation of the medication or noncompliance. It has been postulated that the mechanism of sleep disruption by beta-blocking agents is their well-known tendency to deplete endogenous melatonin,[91] a key sleep-regulating hormone that modulates sympathetic nerve activity. An additional important side effect of these beta blockers[18] is their potential to provoke nightmares. Despite these effects, it is the lipophilic beta blockers (propranolol, metoprolol, carvedilol) that have been shown to reduce the risk of sudden cardiac death. Sleep disturbance has also been documented in conjunction with the widely used

antiarrhythmic agent amiodarone.[92,93] Neurologic side effects were attributed to amiodarone in 20% to 40% of patients. Optimal antiarrhythmic management with this agent to minimize side effects dictates prescription of lower dosages and close patient monitoring and follow-up.

CLINICAL PEARL

Diagnosing and treating patients with nocturnal arrhythmias have been hampered by a paucity of information about concurrent autonomic nervous system activity, cardiac electrical instability, oxygen desaturation, and breathing disturbances. It is now possible to assess autonomic nervous system activity by AECG monitoring of noninvasive markers, such as heart rate variability, a measure of autonomic nervous system tone, and heart rate turbulence, an indicator of baroreceptor function based on the pattern of heart rhythm recovery after a ventricular premature beat.[94] Simultaneous measurement of these indicators, along with clinical history and analysis of cardiac electrical instability with QT-interval dispersion[14,15] or T-wave alternans[25] on continuous ECG monitoring, promises to provide valuable information regarding vulnerability to nocturnal arrhythmias and potential provocation by the autonomic nervous system. Concurrent monitoring of oxygen saturation and respiratory patterns will provide essential information. Increased survival from in-hospital nighttime cardiac arrest can be anticipated with improved monitoring.[95]

SUMMARY

Because the etiology of nocturnal arrhythmias is multifactorial, their management necessitates a comprehensive approach and consideration of a host of cardiovascular and respiratory factors. Treatment must be tailored to contain neurally induced arrhythmias while avoiding exacerbation of hypotension and myocardial ischemia during NREM sleep.

Selected Readings

Bitter T, Fox H, Gaddam S, et al. Sleep-disordered breathing and cardiac arrhythmias. *Can J Cardiol* 2015;**31**(7):928–34.

Guilleminault CP, Pool P, Motta J, et al. Sinus arrest during REM sleep in young adults. *N Engl J Med* 1984;**311**:1006–10.

Kinney HC, Filiano JJ, Sleeper LA, et al. Decreased muscarinic receptor binding in the arcuate nucleus in sudden infant death syndrome. *Science* 1995;**269**:1446–50.

Lin GM, Colangelo LA, Lloyd-Jones DM, et al. Association of sleep apnea and snoring with incident atrial fibrillation in the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 2015;**182**:49–57.

Linz D, Linz B, Hohl M, Böhm M. Atrial arrhythmogenesis in obstructive sleep apnea: therapeutic implications. *Sleep Med Rev* 2015;**26**:87–94.

Maan A, Mansour M, Anter E, et al. Obstructive sleep apnea and atrial fibrillation: pathophysiology and implications for treatment. *Crit Pathw Cardiol* 2015;**14**(2):81–5.

Mancia G. Autonomic modulation of the cardiovascular system during sleep. *N Engl J Med* 1993;**328**:347–9.

Priori SG, Wilde AA, Horie M, et al. HRS/EHRA/APHRS Expert Consensus Statement on the Diagnosis and Management of Patients with Inherited Primary Arrhythmia Syndromes. *Heart Rhythm* 2013;**10**:1932–63.

Schwartz PJ, Stramba-Badiale M, Segantini A, et al. Prolongation of the QT interval and the sudden infant death syndrome. *N Engl J Med* 1998;**338**:1709–14.

Verrier RL, Klingenheben T, Malik M, et al. Microvolt T-wave alternans: physiologic basis, methods of measurement, and clinical utility. Consensus guideline by the International Society for Holter and Noninvasive Electrocardiology. *J Am Coll Cardiol* 2011;**44**:1309–24.

A complete reference list can be found online at ExpertConsult.com.

Cardiovascular Effects of Sleep-Related Breathing Disorders

Chapter 126

Virend K. Somers; Shahrokh Javaheri

Chapter Highlights

- The cycle of apnea and recovery causes hypoxemia and reoxygenation, hypercapnia and hypocapnia, changes in intrathoracic pressure, and arousals. These consequences of sleep apnea, both obstructive and central apnea, adversely affect cardiovascular function. The cardiovascular effects of sleep apnea may be mediated by redox-sensitive gene activation, altered autonomic nervous system activity, oxidative stress, and release of inflammatory mediators. Pathophysiologic consequences of sleep apnea elicit acute and chronic cardiovascular changes.
- Hypoxemia has direct (decreased myocardial oxygen delivery) and indirect (activation of sympathetic nervous system, promotion of endothelial cell dysfunction, and pulmonary arteriolar vasoconstriction) cardiac and vascular effects. Reoxygenation may cause additional damage through further production of free radical species. Hypoxemia-reoxygenation, with intermittent and profound alterations in the partial pressure of oxygen (P_{O_2}), may occur hundreds of times during sleep.
- Because of potentiated chemoreflex responses to hypoxemia-hypercapnia, the sympathetic and consequent presser responses to hypoxemia-hypercapnia, particularly in the absence of inhibitory effects of breathing, are marked. Nighttime sympathetic activation carries over into daytime wakefulness.
- Large negative intrathoracic pressures are generated during episodes of obstructive apnea. Negative intrathoracic pressure increases the transmural pressure (pressure inside minus pressure outside) of the intrathoracic vascular structures, including aorta, pulmonary vascular bed, and ventricles.
- Bradycardias may be especially severe and are elicited because of activation of the diving reflex by the combination of hypoxemia and apnea. Episodes of up to 10 seconds or more of sinus arrest may occur because of chemoreflex-mediated vagal activation.
- Sleep apnea has been implicated in systolic and diastolic heart failure, ventricular arrhythmias, and atrial fibrillation. However, whether treating sleep apnea prevents heart failure and arrhythmias or improves survival remains to be determined from randomized controlled trials.

Hemodynamic changes have been most studied in patients with obstructive sleep apnea (OSA), and in these studies, acute apnea-induced hemodynamic changes have been documented. Chronic exposure may also result in left ventricular systolic and diastolic dysfunction and in increased atrial volume. A limited number of studies have shown that treatment of OSA with nasal continuous positive airway pressure (CPAP) devices or tracheostomy can result in reversal of left ventricular dysfunction and arrhythmias. Whether treatment of sleep apnea reduces cardiovascular events or cardiovascular mortality remains to be demonstrated in randomized control trials. However, several observational studies have reported that treatment of sleep apnea improves survival primarily because cardiovascular events are reduced.

Periodic breathing is characterized by cyclic changes in tidal breathing with intervening episodes of obstructive or central apnea or hypopnea. These disordered breathing events result in three basic pathophysiologic consequences: (1) intermittent arterial blood gas abnormalities characterized by hypoxemia-reoxygenation and hypercapnia-hypocapnia, (2) arousals and a shift to light sleep stages, and (3) large negative swings in intrathoracic pressure (Figure 126-1).[1-3] These pathophysiologic consequences of apnea and hypopnea, both obstructive and central, adversely affect cardiovascular function, acutely and chronically.

ARTERIAL BLOOD GAS ABNORMALITIES AND THEIR CONSEQUENCES

Periodic breathing consists of cyclic changes in breathing pattern that include episodes of apnea and hypopnea, resulting in hypoxemia and hypercapnia. After apnea and hypopnea,

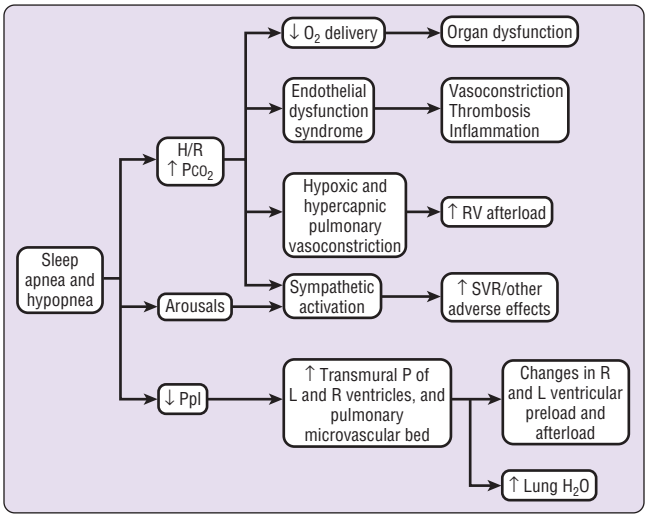


Figure 126-1 Pathophysiologic consequences of sleep apnea and hypopnea. Pleural pressure (Ppl) is a surrogate of the pressure surrounding the heart and other vascular structures. H/R, Hypoxia-reoxygenation; L, left; P, pressure; R, right; RV, right ventricular; SVR, systemic vascular resistance; ↑, increased; ↓, decreased. (Modified from Javaheri S. Sleep-related breathing disorders in heart failure. In: Mann DL, editor. *Heart failure: a companion to Braunwald's heart disease.* Philadelphia: Saunders; 2003. p. 478.)

hyperpnea ensues, resulting in reoxygenation and hypocapnia. These alterations in blood gases affect the cardiovascular system in different ways.

Hypoxemia and Reoxygenation

Hypoxemia has direct (decreased myocardial oxygen delivery) and indirect (activation of sympathetic nervous system, promotion of endothelial cell dysfunction, and pulmonary arteriolar vasoconstriction) cardiac and vascular effects. Hypoxemia with reoxygenation may be analogous to ischemia with reperfusion, and reoxygenation may cause additional damage through further production of free radical species. Biochemical injury due to hypoxemia-reoxygenation has considerable relevance to sleep apnea-hypopnea, where intermittent and profound alterations in the partial pressure of oxygen (Po_2) may occur hundreds of times during sleep.

Direct Effects of Hypoxia on Myocardium

Decreased myocardial oxygen delivery may result in an imbalance between myocardial oxygen consumption and demand, resulting in myocardial hypoxia, particularly if there is already coronary artery disease. At the same time, myocardial oxygen demand may be elevated because of concomitant tachycardia. Potential clinical consequences of myocardial hypoxia include nocturnal angina, nocturnal myocardial infarction,[4] arrhythmias, and even nocturnal sudden death.[5] Hypoxia may also impair myocardial contractility and cause diastolic dysfunction.[6]

Hypoxemia-Reoxygenation and Coronary Endothelial Dysfunction

Coronary vessel endothelial cells play a central role in vasoregulation, coagulation, and inflammation.[7] Blood flow and coagulation are modulated by production and release of vasoactive substances that include vasodilators and platelet deaggregators (e.g., nitric oxide, prostacyclin) and vasoconstrictors and platelet aggregators (e.g., endothelin and thromboxane). The balance between vasoregulatory agents is important in modulating coronary blood flow and coagulation status in both health and disease.

Through activation of certain transcription factors such as hypoxia-inducible factor-1 and nuclear factor-κB,[8,9] hypoxia increases the expression of a number of genes such as those encoding endothelin-1, a potent vasoconstrictor with proinflammatory properties, vascular endothelial growth factor, and platelet-derived growth factor. In contrast, it suppresses the transcriptional rate of endothelial nitric oxide synthase,[10] resulting in decreased production of nitric oxide, which is vasodilatory and has antimitogenic properties. Hypoxia also enhances expression of adhesion molecules and promotes leukocyte rolling and endothelial adherence,[11] and it is involved in induction of endothelial and myocyte apoptosis.[12]

Some of the aforementioned adverse effects of sustained hypoxia have also been observed with intermittent hypoxia (i.e., hypoxia-reoxygenation).[13-23] In this context, intermittent hypoxia has been proposed to be more deleterious than sustained hypoxia.[18,19] Reoxygenation through delivery of oxygen molecules provides a substrate for additional production of oxygen radicals and may contribute to oxidative stress.

The pathophysiologic consequences of hypoxemia-reoxygenation could lead to vascular inflammation and remodeling, similar to atherosclerosis.[7,23] Endothelial dysfunction has been demonstrated in a number of cardiovascular disorders, including hypertension, myocardial infarction, and stroke. Interestingly, these disorders have been also associated with OSA. It is therefore conceivable that endothelial dysfunction caused by sleep-related breathing disorders may contribute to worsening of atherosclerosis, atherothrombosis, and left ventricular dysfunction.[1,24]

The inflammatory and neurohormonal (see Obstructive Sleep Apnea and Systolic Heart Failure, later) consequences of altered blood gas chemistry have been best studied in patients with OSA, which is associated with increased sympathetic activity, high concentrations of endothelin, adhesion molecules, inflammatory cytokines, activation of white blood cells, oxidative stress, endothelial dysfunction, and hypercoagulopathy.[1,22,24-39] These autonomic, biochemical, and functional alterations may be reversed with use of nasal CPAP to treat OSA. However, such systematic studies are lacking for central sleep apnea, with the exception of studies showing increased overnight and morning sympathetic activity and increased concentration of endothelin and brain natriuretic peptide in patients with heart failure with central sleep apnea compared with those without central sleep apnea (for details, see Chapter 129).[40]

Hypoxemia-Hypercapnia and the Autonomic Nervous System

Sleep apneas and hypopneas, both obstructive and central (Figures 126-2 and 126-3), increase sympathetic activity through complex mechanisms. Hypoxemia stimulates the peripheral arterial chemoreceptors in the carotid bodies, triggering reflex increases in sympathetic activity.[41,42] Hypercapnia stimulates the peripheral and the central chemoreceptors located in the region of the brainstem, also increasing sympathetic activity.

Both hypoxemia and hypercapnia increase ventilation, which, acting through thoracic afferents, buffers the increases

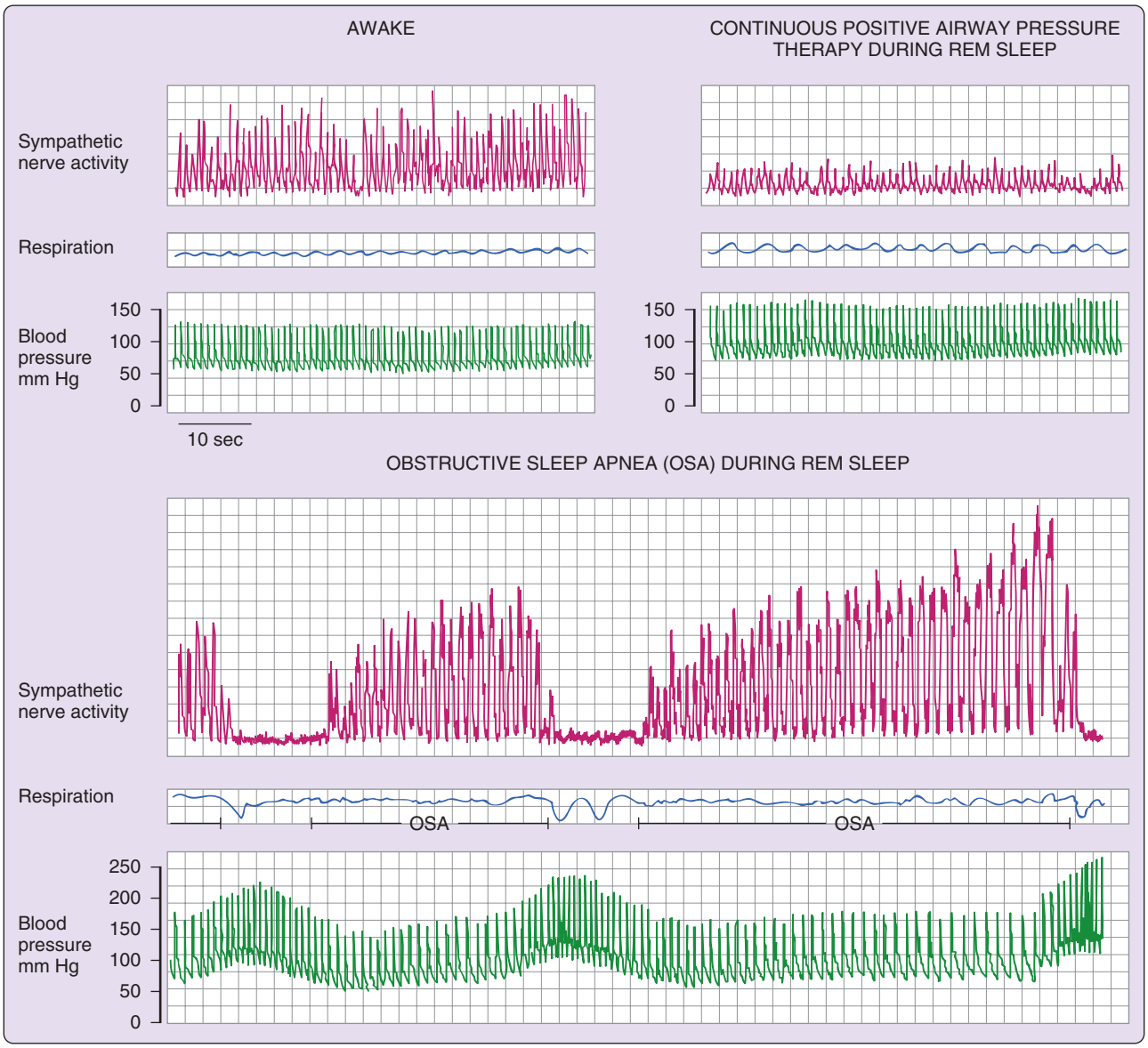


Figure 126-2 Recordings of sympathetic nerve activity, intraarterial blood pressure, and breathing in a normotensive patient with obstructive sleep apnea (OSA) during resting normoxic wakefulness (*top left*). The patient was free of any other overt cardiovascular disease and on no medications. Note the high levels of sympathetic nerve traffic even in the absence of apneic events. During REM sleep (*bottom*), the repetitive hypoxemia and hypercapnia elicit chemoreflex-mediated sympathetic activation and vasoconstriction. At the end of apneas, with increases in cardiac output and severe vasoconstriction, intraarterial blood pressure can reach levels from 130/60 mm Hg during wakefulness to a peak of 220/130 mm Hg during apneas. At the end of apneas, there also is abrupt inhibition of sympathetic traffic because of the increase in blood pressure acting through the baroreflexes and the sympathetic inhibitory effects of the thoracic afferents. After treatment of OSA with continuous positive airway pressure (*top right*), there is a marked reduction in sympathetic traffic and in blood pressure. (From Somers VK, Dyken ME, Clary MP, et al. Sympathetic neural mechanisms in obstructive sleep apnea. *J Clin Invest* 1995;96:1897–904.)

in sympathetic drive during hypoxemia and to a lesser extent during hypercapnia.[41,42] Thus, when hypoxemia or hypercapnia occurs during apnea, the absence of ventilatory inhibition results in a potentiation of sympathetic activation and consequent vasoconstriction and blood pressure surges. In this context, and especially when there are potentiated chemoreflex responses to hypoxemia-hypercapnia,[43,44] the sympathetic and consequent pressor responses to hypoxemia-hypercapnia, particularly in the absence of inhibitory effects of breathing, are marked.

Nighttime sympathetic activation carries over into daytime wakefulness. Repetitive hypoxemia may be implicated because after 2 weeks of chronic intermittent hypoxemia, healthy normal subjects manifested an increase in sympathetic outflow, together with increased chemoreflex gain and blunted baroreflex function.[45]

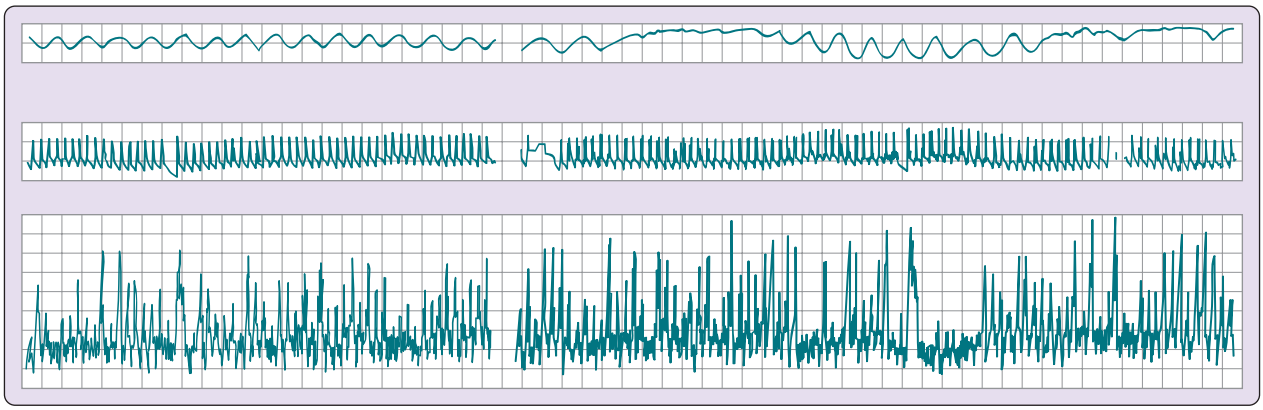


Figure 126-3 Recordings of breathing (*top*), beat-by-beat blood pressure (*middle*), and muscle sympathetic nerve activity (MSNA) (*bottom*) in a patient with severe congestive heart failure, during normal breathing on the *left* and during Cheyne-Stokes breathing on the *right*. Oxygen saturation was 94% during normal breathing and oscillated between 97% and 90% during Cheyne-Stokes breathing. MSNA total burst amplitude increased from 1533 arbitrary units per minute during normal breathing to 1759 arbitrary units per minute during Cheyne-Stokes breathing. Mean blood pressure was 70 mm Hg during normal breathing and peaked at 82 mm Hg during the hyperventilation that followed central apnea. Patients with heart failure have high levels of sympathetic drive even during normal breathing. During central apneas, there is a modest but significant further increase in sympathetic activity. (From Van de Borne P, Oren R, Abouassaly C, et al. Effect of Cheyne-Stokes respiration on muscle sympathetic nerve activity in severe congestive heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *Am J Cardiol* 1998;81:432–6.)

Alveolar Hypoxia-Hypercapnia and Pulmonary Arteriolar Vasoconstriction

Alveolar hypoxia, in part through release of endothelin, and hypercapnia cause pulmonary arteriolar vasoconstriction and hypertension, which could adversely affect right ventricular function (see Chapter 127).

Hypocapnia

Episodes of hyperpnea after apneas and hypopneas result in hypocapnia. Hypocapnia may impair myocardial oxygen delivery and uptake by coronary artery vasoconstriction[45] and shifting of the oxygen-hemoglobin dissociation curve to the left. Hypocapnia may also contribute to arrhythmogenesis.

Arousals, Shift to Light Sleep Stages, and the Autonomic Nervous System

Compared with wakefulness, the balance of activity of sympathetic and parasympathetic nervous system reverses in normal sleep.[46,47] Normally, there is a progressive reduction in sympathetic nerve traffic, heart rate, and blood pressure during the deepening stages of non–rapid eye movement (NREM) sleep, such that sympathetic activity, heart rate, and blood pressure in stage 4 sleep are substantially lower than during supine resting wakefulness.[46,47] During phasic rapid eye movement (REM) sleep, there is an abrupt increase in sympathetic activity, resulting in intermittent and brief surges in blood pressure and heart rate. On average, blood pressure and heart rate during REM sleep are similar to levels recorded during wakefulness. Thus, during normal sleep, there is a well-regulated pattern of alteration in autonomic and hemodynamic measures, modulated by changes in sleep stage. These organized responses to normal sleep are disrupted in patients with sleep-related breathing disorders, both obstructive and central sleep apnea. Sleep architecture is dramatically altered in patients with OSA-hypopnea and also in patients with heart failure and central sleep apnea. There is a shift to light sleep stages. Most important, however, apneas and hypopneas commonly result in arousals that are also associated with an increase in sympathetic activity and a decrease in parasympathetic activity,[47,48] and increasing blood pressure and heart rate. In OSA, arousals occur at the end of the apnea and with resumption of breathing. In patients with central sleep apnea and Hunter-Cheyne-Stokes breathing pattern, arousals occur at the peak of hyperventilation.

In addition to arousals, sleep-related breathing disorders may increase sympathetic activity by hypoxemia, hypercapnia, and changes in ventilation, as noted previously.

There are multiple adverse cardiac consequences of sympathetic activation. These include increased systemic vascular resistance and left ventricular afterload, venoconstriction with increased right ventricular preload, increased myocardial contractility, hypertrophy, tachycardia, and arrhythmias. Furthermore, increased myocardial norepinephrine may cause myocyte toxicity and apoptosis.[49,50]

Central sleep apnea and OSA increase sympathetic activity as measured by either microneurography or blood and urinary norepinephrine levels.[51-56] Treatment of obstructive[54-56] and central sleep apnea[52,57] decreases sympathetic activity, with important implications. First, with regard to central sleep apnea in heart failure, increased sympathetic activity is associated with poor survival; therefore a reduction in sympathetic activity should have favorable prognostic implications. OSA causes nocturnal increases in sympathetic activity and blood pressure, which carry over into the daytime. OSA is a known cause of hypertension, and in some patients blood pressure decreases relatively quickly with effective treatment of OSA with CPAP (see Chapter 127).

In summary, pathophysiologic consequences of sleep-related breathing disorders, such as increased periods of wakefulness (interruption insomnia), arousals, hypoxemia, and hypercapnia, collectively contribute to increased sympathetic activity.

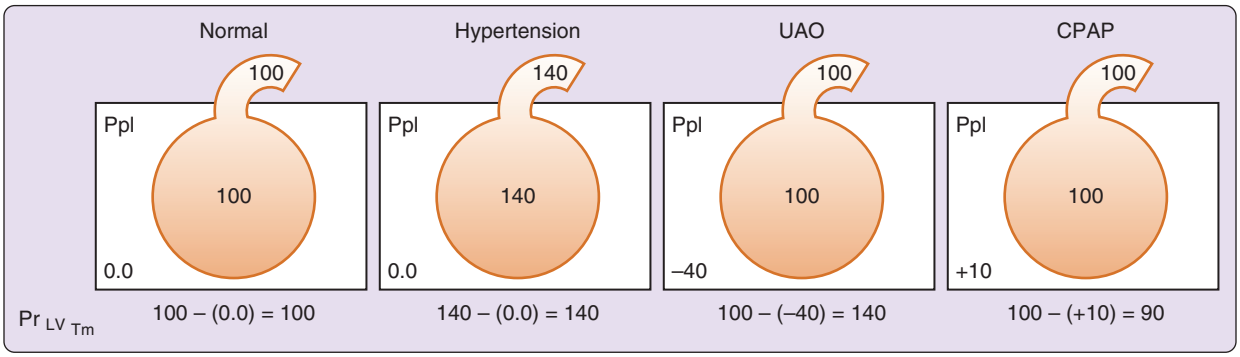


Figure 126-4 Transmural (Tm) pressure (Pr) of the left ventricle (LV) during systole. Because of an obstructive apnea (upper airway occlusion [UAO]), a negative pleural pressure (Ppl) of –40 mm Hg is generated. This increases left ventricular transmural pressure from 100 to 140 mm Hg, which is equivalent to an increase in systolic aortic blood pressure from 100 to 140 mm Hg. Note the reduction in left ventricular transmural pressure with application of nasal continuous positive airway pressure (CPAP). (Modified from Javaheri S. Sleep-related breathing disorders in heart failure. In: Mann DL, editor. *Heart failure: a companion to Braunwald's heart disease.* Philadelphia: Saunders; 2003. p. 480.)

Exaggerated Negative Intrathoracic Pressure and Its Consequences

Large negative intrathoracic pressures are generated during episodes of obstructive apnea. In central sleep apnea, relatively large negative pressure deflections occur during hyperpnea, particularly in the face of less compliant (stiff) lungs (due to heart failure). However, pleural pressure changes are usually more pronounced in obstructive than in central sleep apnea.

A number of studies have addressed the cardiovascular consequences of both negative and positive pressure deflections affecting right and left ventricular function.[58,59] Negative intrathoracic pressure increases the transmural pressure (pressure inside minus pressure outside) (Figure 126-4) of the intrathoracic vascular structures, including aorta, pulmonary vascular bed, atria, and ventricles.

According to Laplace's law, increased transmural myocardial pressure increases wall tension and myocardial oxygen consumption. Furthermore, negative intrathoracic perivascular pressure could increase extravascular lung water by favoring fluid transudation across the pulmonary microvascular bed and by diminishing lymph outflow from the lung.[60] This may account in part for cases of flash pulmonary edema reported in OSA, and sleep apnea may contribute to excess lung water and pulmonary edema in congestive heart failure. In addition, decreased intrathoracic pressure increases venous inflow, resulting in increased right ventricular diastolic filling, which in turn may decrease left ventricular compliance and volume, a phenomenon called ventricular interdependence. Application of nasal CPAP to treat sleep apnea, both obstructive and central, reduces transmural pressure by two mechanisms. First, and most important, it decreases or eliminates apneas, desaturation, and arousals, which as noted previously collectively increase sympathetic activity and result in cyclic surges in arterial blood pressure. Second, nasal CPAP not only attenuates steep surges in intrathoracic pressure but also actually increases the pleural pressure, thus decreasing transmural pressures across intrathoracic structures (see Figure 126-4).

ACUTE HEMODYNAMIC EFFECTS OF SLEEP APNEA

The circulatory responses to individual apneas and hypopneas are governed by the interaction of stresses and physiologic consequences described previously.[61,62] Hemodynamic changes are related to development of hypoxemia, hypercapnia, presence or absence of breathing, changes in intrathoracic pressure, and the consequent mechanical effects.

Hemodynamic changes have been best studied in human OSA.[61,63,64] The evolution of a cycle of apnea and recovery is complex and represents an unsteady hemodynamic state. For these reasons, hemodynamic changes occur during the course of an apnea, and these changes are different from those occurring during the immediate or late postapneic periods. During recovery, arousals and ventilation further affect hemodynamics. Cyclic changes in heart rate and systemic and pulmonary arterial blood pressure paralleling periodic breathing occur commonly.[61,63-66] In some patients, there is a very clear and progressive bradycardia toward the end of apnea, with abrupt development of tachycardia with resumption of breathing, because of the vagolytic effects of lung inflation and arousals. This manifests as a pattern of repetitive bradycardias or tachycardias during sleep, which may be evident on Holter monitoring and may signify the presence of OSA. In experimental sleep apnea, decreases in heart rate are more severe during central than obstructive apnea, reflecting lack of activation of thoracic afferents.[62]

The bradycardias may be especially severe,[65,66] and they are elicited because of activation of the diving reflex by the combination of hypoxemia and apnea. Episodes of up to 10 seconds or more of sinus arrest may occur because of the chemoreflex-mediated vagal activation. The consequent absence of perfusion, because of asystole, may have implications for patients with preexisting severe cerebral or cardiac ischemia.

At the termination of obstructive apneas, there are surges in blood pressure. This cyclic change in blood pressure is one of the most consistent hemodynamic findings in patients with OSA. Multiple mechanisms are involved. During apnea, the

increased hypoxemia and hypercapnia, acting through the chemoreflexes, progressively elicit sympathetic activation and vasoconstriction.[53] With resumption of breathing, because of the inspiratory increase in right ventricular filling, stroke volume may increase. Vagolytic effects of inspiration result in tachycardia. The increased stroke volume and heart rate result in an increased cardiac output entering a vasoconstricted peripheral circulation, with consequent acute increases in blood pressure.[53] However, just after termination of an obstructive apnea, there is abrupt inhibition of sympathetic activity to the peripheral blood vessels, in part because the deep breathing inhibits sympathetic activity through thoracic afferents and in part because of baroreflex inhibition of sympathetic activity secondary to the postapneic blood pressure surge. Nevertheless, despite the interruption in sympathetic nerve traffic, vasoconstriction persists for several seconds after termination of the sympathetic nerve discharge because of the kinetics of norepinephrine uptake, release, and washout at the neurovascular junction.

Another consistent finding is a mild reduction in stroke volume during obstructive apnea, which has been documented using noninvasive techniques for measuring beat-to-beat cardiac output.[61] This probably results from a decrease in left ventricular preload and an increase in afterload. Changes in stroke volume after termination of the apnea depend on where in the recovery cycle it is being measured.[61]

Obstructive Sleep Apnea, Left Ventricular Dysfunction, and Heart Failure

The relationship between central sleep apnea and heart failure is discussed in Chapter 129. In this section, we review OSA as a cause of heart failure.

Obstructive Sleep Apnea and Systolic Heart Failure

In a canine model mimicking severe OSA,[67] within a 1- to 3-month period of exposure to apneas during sleep, left ventricular systolic dysfunction developed. Left ventricular ejection fraction, measured during the daytime, decreased significantly because of an increase in left ventricular systolic volume.

In humans, there are two kinds of studies relating left ventricular systolic dysfunction and OSA—first, studies in which patients with OSA have been assessed for the presence of left ventricular dysfunction,[68-71] and second, studies in patients with established left ventricular systolic dysfunction who have been assessed to determine the prevalence of OSA.[72,73] In some studies,[74-75] changes in left ventricular ejection fraction in response to treatment for OSA have also been described.

Results of studies assessing left ventricular systolic function in OSA patients are conflicting.[68-70] However, in the two studies[69,70] in which technetium-99m was used to assess left ventricular systolic function, OSA was associated with left ventricular systolic dysfunction. Use of radionuclide ventriculography to assess left ventricular function is important because in obese subjects, echocardiography, which has been used in some studies, may be associated with technical difficulties.

Alchanatis and colleagues[69] studied 29 patients with severe OSA (apnea-hypopnea index [AHI] greater than 15/hour; mean AHI, 54/hour; lowest arterial oxygen saturation, 62%) and 12 control subjects (AHI, 9/hour; lowest saturation, 92%). The subjects were without known cardiovascular disease. The mean left ventricular ejection fraction was significantly lower in patients with OSA compared with the control group (53% vs. 61%; $P < .003$). Six months after treatment with CPAP, left ventricular ejection fraction increased significantly to 56% ($P < .001$). Left ventricular diastolic dysfunction also improved significantly (see later).

In a large study[70] of 169 patients with OSA (AHI greater than 10/hour; mean AHI, 47/hour), 13 subjects (8%) had left ventricular systolic dysfunction (range, 32% to 50%). Left ventricular systolic dysfunction was not the result of ischemic disease as evidenced by echocardiography and dipyridamole stress testing. In seven patients who were treated for OSA (six with CPAP and one with upper airway surgery), 1 year after therapy, mean left ventricular ejection increased significantly from 44% to 63%.[70]

In the cross-sectional analysis of more than 6000 patients enrolled in the Sleep Health Heart Study,[71] the presence of OSA increased the likelihood of having a history of heart failure by an odds ratio of 2.5. Furthermore, there was a significant dose-dependent correlation between AHI and the prevalence of heart failure.

In studies of patients with established left ventricular systolic dysfunction undergoing polysomnography (reviewed in Chapter 129), the prevalence of OSA, defined as an AHI of at least 15/hour, ranged from 12% to 32%.[76] This wide range is not particularly surprising. The prevalence depends on a number of factors, including the number of obese patients with heart failure enrolled in each study and the different polysomnographic criteria used by various investigators for diagnosis of OSA. Another important issue is the difficulty in accurately classifying hypopneas into central versus obstructive, which is a determinant of prevalence of the phenotype of sleep-disordered breathing.

In a prospective study[72] of 81 patients with known systolic dysfunction and in whom no question was asked regarding snoring or other symptoms associated with OSA, 11% had OSA, with a mean AHI of 36/hour and a lowest arterial oxygen saturation of 72%. In a retrospective study[73] of 450 patients with systolic dysfunction who were referred for a sleep study because of snoring and other symptoms of sleep apnea, 32% had OSA. From the aforementioned studies, however, it cannot be determined whether OSA preceded heart failure. Yet, as is discussed later, treatment of OSA with nasal CPAP increases left ventricular ejection fraction,[74,75] indicating that OSA contributes to worsening of left ventricular systolic dysfunction.

The mechanisms by which OSA may impair left ventricular systolic function are multiple. Hypoxemia plays a critical role, both by impairing myocardial contractility and through a host of neurohormonal mechanisms. In addition, increases in left ventricular wall stress and transmural pressure occur because of additive effects of the excess negative juxtacardiac pressure (during obstructive apneas) and development of hypertension.

The effects of positive airway pressure therapy on left ventricular ejection fraction in patients with OSA and systolic heart failure have been reported in five randomized clinical trials, two of which were double blind (Table 126-1). In three of the studies in which CPAP was used, including the only two double-blind randomized clinical trials, the rise in left ventricular fraction was minimal or not at all. It should be

Table 126-1 Effects of Positive Airway Pressure Therapy on Left Ventricular Ejection Fraction in Patients with Obstructive Sleep Apnea and Systolic Heart Failure

Variable	Kaneko Open	Mansfield Open	Egea DB	Smith DB	Khayat Open	Khayat Open
n	12	19	20	23	11	13
AHI (*n*/hr)	40	25	44	36	30	34
LVEF (%)	25	35	29	30	29	26
Increase in LVEF (%)	9*	5*	2.2*	0.0	0.5	8.5*
Duration	4 wk	3 mo	3 mo	6 wk	3 mo	3 mo
PAP titration	CPAP yes	CPAP yes	CPAP yes	Auto CPAP	CPAP yes	Bilevel yes
Compliance (hr)	6.2	5.6	NR	3.5	3.6	4.5

*Indicates a statistically significant change.
AHI, Apnea-hypopnea index; CPAP, continuous positive airway pressure; DB, double blind; NR, not reported; PAP, positive airway pressure.
Data from Kaneko Y, Flores JS, Usui K, et al. Cardiovascular effects of continuous positive airway pressure in patients with heart failure and obstructive sleep apnea. *N Engl J Med* 2003;348:1233–41; Mansfield DR, Gollogly, NC, Kaye DM, et al. Controlled trial of continuous positive airway pressure in obstructive sleep apnea in heart failure. *Am J Respir Crit Care Med* 2004;169:361–6; Egea CJ, Aizpuru F, Pinto JA, et al. Cardiac function after CPAP therapy in patients with chronic heart failure and sleep apnea: a multicenter study. *Sleep Med* 2008;9:660–6; Schmidt LA, Vennelle M, Gardner RS, et al. Autotitrating continuous positive airway pressure therapy in patients with chronic heart failure and obstructive sleep apnea: a randomized placebo controlled trial. *Eur Heart J* 2007;28:1221–7; Khayat RN, Abraham WT, Patt B, et al. Cardiac effects of continuous and bilevel crowded airway pressure for patients with heart failure and obstructive sleep apnea: a pilot study. *Chest* 2008;134:1162–8.

noted, however, that in at least two of these studies compliance with CPAP was also limited. In the two open studies in which compliance hours with CPAP were more than those in the double-blind studies, ejection fraction increased between 5% and 9%. In one open randomized clinical trial of CPAP versus a bilevel device, the ejection fraction increased significantly only with bilevel therapy.

Obstructive Sleep Apnea and Diastolic Heart Failure

Isolated left ventricular diastolic heart failure with relative preservation of left ventricular systolic function is the most common form of heart failure in elderly subjects. The pathophysiologic consequences of this form of heart failure relate to a hypertrophied, noncompliant left ventricle, shifting the pressure-volume curve upward and to the left. Therefore, for a given left ventricular volume, left ventricular end-diastolic pressure increases, resulting in elevated left atrial and pulmonary capillary pressure and in pulmonary congestion and edema.

As noted previously, hemodynamic studies[63,64] of patients with OSA have documented that pulmonary capillary pressure increases during the course of an obstructive apnea, indicating development of diastolic dysfunction. During obstructive apnea, left ventricular transmural wall tension increases because of an increase in aortic blood pressure and a simultaneous decrease in juxtacardiac pressure. Furthermore, hypoxemia may impair left ventricular relaxation, further impairing diastolic function.[77] Repeated exposure to nocturnal hypertension and hypoxemia and consequent development of OSA-induced systemic hypertension and increased left ventricular mass may also contribute to left ventricular diastolic dysfunction.

Most studies show that OSA is associated with an increase in left ventricular mass,[78-81] and suggest that the OSA-related cardiac structural changes may resolve with CPAP treatment.[80] An early study[78] reported that OSA may cause left ventricular hypertrophy even in the absence of daytime systemic hypertension. This finding was later supported by another study[79] comparing patients with OSA (AHI >20/hour) and those without OSA (AHI <20/hour).

In the largest study,[81] consisting of 2058 Sleep Heart Health Study participants, left ventricular mass was associated with both apnea-hypopnea and hypoxemia indexes after adjustment for age, sex, ethnicity, study site, body mass index, smoking, systolic blood pressure, antihypertensive medication use, diabetes mellitus, myocardial infarction, and alcohol consumption. Although there are considerable data[72,73,82] regarding the prevalence of sleep apnea in patients with systolic heart failure (reviewed by Javaheri[76]), the prevalence of OSA in diastolic heart failure has been studied only in one large systematic study.[83] Bitter and colleagues evaluated 244 consecutive patients (87 women) with heart failure with a preserved ejection fraction (HFpEF). All underwent polygraphy, right heart catheterization, and echocardiography. The two major causes of HFpEF were systemic hypertension (44%) and coronary artery disease (33%). Forty-eight percent had an AHI of 15 or more per hour, a prevalence similar to that seen in patients with heart failure with a reduced ejection fraction (HFrEF). Among patients with an AHI of 15/hour or more, 23% had central sleep apnea. Consistent with the observation in HFrEF, patients with HFpEF and central sleep apnea had lower Pco_2 and higher left ventricular end-diastolic and pulmonary capillary wedge pressure than OSA patients.

As noted earlier, isolated diastolic heart failure is highly prevalent in elderly subjects. Furthermore, elderly subjects have a high prevalence of OSA. It is speculated that OSA could be the cause of diastolic heart failure, or the presence of OSA could contribute to the worsening of left ventricular diastolic dysfunction. In this regard, a preliminary study reported that treatment of OSA improves left ventricular diastolic dysfunction,[69] an observation confirmed by the only randomized, placebo (sham CPAP)-controlled trial[80] showing that after 12 weeks on effective CPAP therapy, there was a significant increase in E/A ratio (the ratio of early to late diastolic filling) and a significant decrease in isovolumic relaxation and mitral deceleration. These observations are similar

to the improvement seen in systolic function when patients with heart failure and OSA are treated with CPAP (see Table 126-1).[74-76]

Arrhythmias in Obstructive Sleep Apnea

Obstructive Sleep Apnea Predisposing to an Arrhythmogenic Substrate

Repetitive nocturnal apneas elicit severe derangements in cardiovascular homeostasis. Hypoxemia, hypercapnia, acidosis, adrenergic activation, increased afterload, and rapid fluctuations in cardiac wall stress would reasonably be expected to be conducive to tachycardia-brachycardia oscillations and atrial and ventricular arrhythmias (Figures 126-5 and 126-6). A variety of atrioventricular arrhythmias, including complete heart block and ventricular asystole during sleep, have been observed in patients with OSA[84-86] and have been eliminated by either tracheostomy or use of nasal CPAP.[84,85] Profound OSA-induced arrhythmias can occur in the absence of any major structural abnormalities in the conduction system.[87]

Although the normal heart would be less likely to manifest malignant arrhythmias in the setting of severe obstructive apnea, the ischemic, hypertrophied, or failing heart may be more susceptible.[88] Nevertheless, activation of the diving reflex[66,89] during apneas can often elicit severe bradyarrhythmias, even in the setting of a normal myocardium and normal cardiac electrophysiologic function.

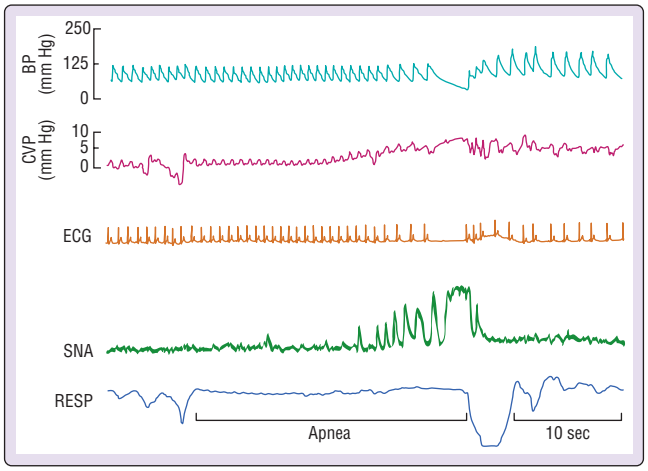


Figure 126-5 Recordings of intraarterial blood pressure (BP), central venous pressure (CVP), electrocardiogram (ECG), sympathetic nerve activity (SNA), and respiratory patterns (RESP) in a healthy subject during voluntary end-expiratory apnea. During apnea, there is a progressive increase in the RR interval on the ECG with eventual sinus pause and atrioventricular block. Accompanying this is increased sympathetic activity. The simultaneous sympathetic activation to peripheral blood vessels and vagal activation of the heart is characteristic of the diving reflex. Note the rapid increase in heart rate and sympathetic inhibition during resumption of breathing. This occurs in part because thoracic afferents activated by inspiration inhibit both sympathetic traffic and vagal cardiac drive. (From Somers VK, Dyken ME, Mark AL, Abboud FM. Parasympathetic hyperresponsiveness and bradyarrhythmias during apnea in hypertension. *Clin Auton Res* 1992;2:171–6.)

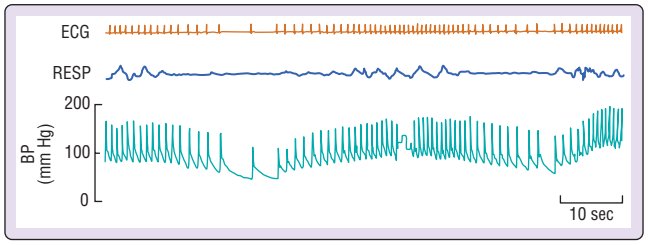


Figure 126-6 A patient with sleep apnea manifesting prolonged and profound bradyarrhythmias with absence of either atrial or ventricular contraction. The beat-by-beat blood pressure (BP) recording confirms the absence of any perfusion during the bradycardia. ECG, electrocardiogram; RESP, respiratory pattern. (From Somers VK, Dyken ME, Mark AL, Abboud FM. Parasympathetic hyperresponsiveness and bradyarrhythmias during apnea in hypertension. *Clin Auton Res* 1992;2:171–6.)

Tachycardia-Bradycardia Oscillations

Patients undergoing Holter monitoring may be noted to have repetitive cyclic episodes of tachycardias and bradycardias during the night.[90,91] These cyclic fluctuations may be attributable to obstructive apneas, although this cannot be confirmed because standard Holter monitoring does not incorporate simultaneous measurements of either breathing pattern or oxygen saturation.

These oscillations in cardiac rate are for the most part explained by changes in cardiac autonomic drive related to breathing pattern. During the course of apnea, incremental hypoxemia elicits the diving reflex so that bradycardia becomes progressively more marked. With termination of apnea, hyperpnea occurs with consequent activation of thoracic afferents, which is vagolytic.[92] Thus, with resumption of breathing, abrupt lung inflation interrupts vagal drive to the heart, resulting in rapid-onset tachycardia. Furthermore, increased cardiac-bound sympathetic drive and withdrawal of parasympathetic activity because of arousals should also contribute to the tachycardia seen with termination of obstructive apnea. It is interesting that tachycardia persists even though blood pressure increases strikingly with termination of apnea. The vagolytic effects of inspiration and the arousal-associated changes in the autonomic nervous system not only interrupt the chemoreflex-mediated cardiac vagal drive but also blunt the expected cardiac vagal drive that would occur secondary to baroreflex activation by the postapneic surge in blood pressure.

Because of the repetitive nature of nocturnal apneas, Holter or other electrocardiographic monitoring at night manifests as a tachycardia-bradycardia pattern. This cardiac rate oscillation is less apparent in patients with autonomic dysfunction, such as patients with long-standing diabetes or cardiac transplant recipients with denervated hearts. Although the changes in cardiac rate are predominantly reflex mediated, breathing-related changes in cardiac filling, as well as rapid changes in cardiac transmural pressures resulting from the Müller maneuver, also modulate heart rate by variations in stretch of cardiac conduction tissue.

Bradyarrhythmias

The primary response to hypoxia is bradycardia.[89] When hypoxia is accompanied by the action of breathing, the bradycardic response is masked because of inhibition of cardiac vagal drive by ventilation.[66] The sympathetic response to hypoxemia, although evident to some extent during breathing, is also attenuated by ventilation and is therefore potentiated during apnea.[93,94] Patients with OSA may be particularly susceptible to hypoxia-induced bradyarrhythmias because their peripheral chemoreflex is heightened, so that even during

voluntary apneas, hypoxemia elicits greater bradycardia than is seen in closely matched control subjects.[95] The arterial baroreflexes serve as an important buffer to diminish chemoreflex gain.[96] Impaired baroreflex sensitivity, such as is seen in hypertension[97] and heart failure,[98] may be associated with further increased chemoreflex drive. Thus patients with hypertension or heart failure who have OSA may manifest even greater sympathetic, and perhaps bradycardic, responses to obstructive apneas.

Profound bradyarrhythmias may have important consequences, particularly in patients with underlying cardiovascular disease. As an example, in the absence of recognition of OSA as a potential cause of the bradyarrhythmia, patients may receive pacemaker implantation, even though their cardiac conduction system may be completely normal and the bradyarrhythmias could be abolished by effective treatment with CPAP.[83,84,99] Second, prolonged episodes of asystole result in absence of perfusion (see Figure 126-6). Absence of perfusion in the setting of apnea-induced hypoxemia, occurring repetitively through the night, may have important implications for ischemic damage to end organs in which there may already be preexisting circulatory compromise.

Ventricular Arrhythmias

There is an extensive literature on sleep apnea inducing nocturnal angina and cardiac ischemia evidenced by ST-segment depression.[100,101] Thus there is a potential contribution of OSA to ventricular arrhythmias through ventricular ectopy during profound bradycardia as well as polymorphic ventricular tachycardia due to cardiac hypoxia-ischemia. These episodes occur primarily with severe desaturation,[83,84] are more common in patients with coronary heart disease,[86] and are virtually eliminated with treatment.[83,84,99] The prevalence of these arrhythmias is low in patients without premorbid cardiorespiratory disease or severe desaturation.[102]

Atrial Fibrillation

In patients cardioverted for atrial fibrillation, those with polysomnographically proven OSA who were not receiving effective CPAP treatment had a 12-month recurrence rate of 82% compared with a 42% recurrence rate in patients with OSA receiving effective CPAP.[103] In patients cardioverted for atrial fibrillation in whom no sleep study had been done, the recurrence rate was 53%. This risk for recurrence in the patients with atrial fibrillation without a previous sleep study suggests that undiagnosed OSA may be present in a large proportion of patients with atrial fibrillation. In addition, among the untreated patients with OSA, those experiencing a recurrence of atrial fibrillation had more severe nocturnal hypoxemia than those without a recurrence. Furthermore, the increased recurrence in patients with untreated OSA could not be explained by factors such as antiarrhythmic medication, body mass index, hypertension, cardiac function, or atrial size.

Mooe and colleagues[104] observed that after coronary artery bypass surgery, patients with OSA were more likely to experience postoperative atrial fibrillation. However, it is not clear whether this was explained by other variables in the patients with OSA.

In a recent longitudinal study of several thousand patients, those with OSA had an increased risk for developing new-onset atrial fibrillation compared with those who did not have OSA. This risk was evident in patients aged 65 or younger, and it was especially marked in those with more severe nocturnal hypoxemia.[105]

There are many reasons that OSA may be conducive to atrial fibrillation. Hypoxemia, presser surges, and sympathetic activation are all potential mechanisms leading to atrial fibrillation. High levels of C-reactive protein may also independently predict the development of atrial fibrillation.[106] Patients with OSA may have increased levels of C-reactive protein.[107-110] Furthermore, abrupt and dramatic changes in intrathoracic negative pressures may especially affect the atria because of their relatively thin walls compared with the ventricles. Increased pressure gradients with consequent increased atrial wall stretch, occurring repetitively through the night, may be expected to induce mechanical and electrical changes that are also conducive to atrial fibrillation.[111-113] Autonomic mechanisms may be pivotal. Animal models suggest that ganglionated plexus ablation may profoundly inhibit the development of atrial fibrillation in response to hypoxemia and apnea.[114]

About 50% of patients presenting for cardioversion have a high risk for sleep apnea compared with 30% of patients from a general cardiology clinic.[115] Even in patients with comorbid OSA undergoing pulmonary vein isolation, recurrence of atrial fibrillation is more than twofold greater in those not treated with CPAP compared with those receiving CPAP therapy.[116]

CLINICAL PEARLS

- Apnea and recovery cycles result in three basic abnormalities: alterations in blood gases, arousals, and changes in intrathoracic pressure.
- Hypoxemia-reoxygenation has deleterious effects on the cardiovascular system. This activates redox-sensitive genes, resulting in synthesis of vasoconstrictor and inflammatory mediators; increases sympathetic activity; and causes oxidative stress. These alterations have been best studied in patients with OSA.
- Untreated OSA may increase the risk for recurrence of atrial fibrillation after cardioversion.
- Sleep apnea can induce severe bradyarrhythmias, including prolonged periods of asystole and heart block, even in the setting of a normal myocardium and cardiac electrophysiologic function.
- OSA should be considered in patients who have ST-segment depression or angina occurring primarily at night.
- Heart failure may be significantly linked to the presence of either central sleep apnea or OSA.

SUMMARY

Sleep-related breathing disorders affect cardiovascular function in a variety of ways. OSA and central sleep apnea act through multiple mechanisms to elicit acute circulatory responses, which have implications for the development of chronic vascular and cardiac dysfunction. The acute responses to apnea are mediated in large part by the effects of apnea on blood gas chemistry, which exerts important cardiovascular effects directly on the myocardium and blood vessels and also acts through reflex mechanisms. Acute neural, circulatory, endothelial, inflammatory, and other responses to repetitive nocturnal hypoxemia and hypercapnia may act to induce

long-term damage to the myocardium and to the coronary and other vascular beds. With the development of functional and structural cardiovascular disease, the consequences of acute apneas are magnified. For example, severe hypoxemia in the setting of sleep apnea is more easily tolerated by an overtly healthy cardiovascular system compared with one in which myocardial ischemia or left ventricular dysfunction is present, with consequent diminished cardiovascular reserve. Small, short-term studies have suggested that effective prevention of recurrent apneas may favorably affect surrogates of cardiovascular disease outcome, such as sympathetic activity, blood pressure, and left ventricular ejection fraction. The importance of large randomized controlled trials in establishing the benefits, if any, of treating sleep apnea in patients with heart failure are highlighted by the results of the recently completed SERVE-HF Study.[117] In patients with stable systolic heart failure (LVEF ≤ 45%) and predominantly central sleep apnea, treatment with adaptive servo-ventilation (ASV) versus usual care was not accompanied by any reduction in the primary endpoint of hospitalization for worsening heart failure or mortality. This was despite a significant improvement in central sleep apnea with ASV. In fact, there was an increase in all cause mortality and in cardiovascular mortality in the treated group. It is important to note that these findings cannot be extended to similar patients with unstable heart failure, to patients with heart failure with preserved ejection fraction, or to heart failure patients with OSA.

Selected Readings

Cowie MR, et al. Adaptive sero-ventilation for central sleep apnea in systolic heart failure. *N Eng J Med* 2015 Sep 1.

de Burgh Daly M, Angell-James J, Elsner R. Role of carotid-body chemoreceptors and their reflex interactions in bradycardia and cardiac arrest. *Lancet* 1997;**1**:764–7.

Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden cardiac death in obstructive sleep apnea. *N Engl J Med* 2005;**352**:1206–14.

Ghias M, Scherlag BG, Lu Z, et al. The role of ganglionated plexi in apnea-related atrial fibrillation. *J Am Coll Cardiol* 2009;**54**:2075–83.

Hoyos CM, Melehan KL, Liu PY, et al. Does obstructive sleep apnea cause endothelial dysfunction? A critical review of the literature. *Sleep Med Rev* 2015;**20**:15–26.

Javaheri S. A mechanism of central sleep apnea in patients with heart failure. *N Engl J Med* 1999;**341**:949–54.

Javaheri S. Heart failure. In: Kushida CA, editor. *The encyclopedia of sleep*, vol. 3. Waltham, MA: Academic Press; 2013. p. 374–86.

Javaheri S, Parker TJ, Liming JD, et al. Sleep apnea in 81 ambulatory male patients with stable heart failure: types and their prevalences, consequences and presentations. *Circulation* 1998;**97**:2154–9.

Jelic S, Padeletti M, Canfield S, et al. Inflammation, oxidative stress and repair capacity of the vascular endothelium in obstructive sleep apnea. *Circulation* 2008;**117**:2270–8.

Kaneko Y, Floras JS, Usui K, et al. Cardiovascular effects of continuous positive airway pressure in patients with heart failure and obstructive sleep apnea. *N Engl J Med* 2003;**348**:1233–41.

Mansukhani MP, Wang S, Somers VK. Sleep, death, and the heart. *Am J Physiol Heart Circ Physiol* 2015;**309**(5):H739–49.

Punjabi NM, Beamer BA. C-reactive protein is associated with sleep disordered breathing independent of adiposity. *Sleep* 2007;**30**:29–34.

Somers VK, Dyken ME, Clary MP, et al. Sympathetic neural mechanisms in obstructive sleep apnea. *J Clin Invest* 1995;**96**:1897–904.

Somers VK, Dyken ME, Mark AL, et al. Sympathetic-nerve activity during sleep in normal subjects. *N Engl J Med* 1993;**328**:303–7.

Somers VK, White DP, Amin R, et al. Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation scientific statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council On Cardiovascular Nursing. In collaboration with the National Heart, Lung, and Blood Institute National Center on Sleep Disorders Research (National Institutes of Health). *Circulation* 2008;**118**:1080–111. and *J Am Coll Cardiol* 2008;**19**;(52):686–717.

A complete reference list can be found online at ExpertConsult.com.

Systemic and Pulmonary Hypertension in Obstructive Sleep Apnea

Chapter 127

F. Javier Nieto; Terry Young; Paul E. Peppard; Shahrokh Javaheri

Chapter Highlights

- Cross-sectional and prospective cohort studies in both population and clinical settings show an association between obstructive sleep apnea (OSA) and risk for systemic hypertension that appears to be independent of obesity, age, and other potential confounding factors.
- The strength, consistency, and dose-response relationship shown across studies suggest that the association is causal.
- In support of a causal relationship, recent meta-analysis of randomized trials shows that treatment with positive airway pressure results in a reduction of blood pressure among hypertensive patients with OSA that is likely to be of clinical and therapeutic significance. This effect is most pronounced in patient with baseline elevated blood pressure, those who have severe OSA, and those who are adherent to therapy.
- Using the current definition of pulmonary hypertension, about 10% of patients with OSA have mean pulmonary artery pressure of 25 mm Hg or greater. Mild pulmonary arterial hypertension may occur in patients with OSA without daytime hypoxemia or chronic obstructive pulmonary disease, although pulmonary hypertension could be more severe in the presence of chronic lung disease, heart failure, and obesity hypoventilation.
- Studies, mostly observational, suggest that treatment of OSA improves pulmonary hypertension.

Although the clinical association between obstructive sleep apnea (OSA) and hypertension has long been reported[1-3] in sleep medicine, the potential importance of OSA in patients with elevated blood pressure and cardiovascular disease is gaining recognition beyond the field of sleep research. As early as 1998, a committee of experts gathered by the World Health Organization (WHO) recognized OSA as a likely cause of secondary pulmonary arterial hypertension.[4] A few years later, the increasing evidence that OSA has a causal role in the development of hypertension was discussed in two influential taskforce statements.[5,6] Furthermore, the 2013 guidelines for the management of hypertension by the European Society of Hypertension and the European Society of Cardiology include OSA as one of the "special conditions" that need to be evaluated and treated in hypertensive patients.[7] In the United States, even though OSA was recognized as an identifiable cause of hypertension in the seventh report of the Joint National Committee for the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure,[8] surprisingly, the recently updated eighth report does not include reference to OSA—or any sleep-related disorder—in relation to hypertension management.[9]

An accurate estimate of the fraction of systemic hypertension that can be causally attributed to OSA is lacking, and data on OSA and pulmonary hypertension (PH) are sparse; however, clinical recognition of both the high prevalence of hypertension in people with OSA[10,11] and the high occurrence of OSA in hypertensive patients[12-14] is imperative. The aim of this chapter is to present the epidemiologic and clinical evidence in support of a role of OSA in systemic and PH and to describe the clinical issues in identification and treatment of patients with OSA and hypertension.

SYSTEMIC HYPERTENSION

Epidemiologic Evidence for a Role of Obstructive Sleep Apnea in Systemic Hypertension

The early observations of hypertension in patients with sleep apnea stimulated several cross-sectional clinic- and community-based studies that attempted to determine whether there was an association between OSA and hypertension that was not explained by excess body weight or other factors common to both OSA and hypertension (see Chapter 61).[11] Results were mixed, but many of the studies had methodologic shortcomings, such as inadequate sample size, flawed comparison groups, substantial measurement error, or limited statistical analysis.[15,16] Since then, findings from both population and clinical studies have shed important new light on this association.

Reports from several well-designed epidemiology studies, summarized in Table 127-1, generally show associations of polysomnography (PSG)-determined OSA and hypertension that remain significant after adjustment for potential confounding factors.[17-19] The strongest epidemiologic evidence for a causal association comes from longitudinal analyses of data from the Wisconsin Sleep Cohort Study of middle-aged state employees.[20,21] The incidence of new hypertension, defined as systolic blood pressure of at least 140 mm Hg, diastolic blood pressure of at least 90 mm Hg, or use of antihypertensive medication at follow-up, was significantly dependent on

Table 127-1 Associations of Polysomnographically Determined Sleep-Disordered Breathing and Hypertension in Four Population Studies

Study Design	Participants (*n*)	Odds Ratio* for Hypertension† (95% CI) AHI Category <1.0‡	1 to 4.9	5 to 14.9	15 to 30	≥30
Wisconsin Sleep Cohort Study,[20] state employees, ages 30 to 65 years, prospective, 4–8 years' follow-up	709	1.0	1.2 (1.1–1.8)	2.0 (1.3–3.2)	2.9 (1.5–5.6)	
Sleep Heart Health Study,[19,25] multicenter, ages 40–97 years						
a. Cross-sectional[25]	6132	1.0	1.1 (0.9–1.3)	1.2 (1.0–1.4)	1.3 (1.9–1.6)	1.4 (1.0–1.8)
b. Prospective, 2- and 5-year follow-up[19]	2470	1.0		0.9 (0.7–1.2)	1.1 (0.8–1.5)	1.5 (0.9–2.5)
Southern Pennsylvania,[23] population sample through random-digit dialing, ages 20–100 years, cross-sectional	1741	1.0		2.3§ (1.4–3.6)	6.9§ (2.0–26.4)	
Vitoria-Gasteiz, Spain,[24] random census sample, ages 30–70 years, cross-sectional						
a. Cross-sectional[24]	552	1.0 (RDI <3)	2.5 (1.1–5.8) (3 ≤ RDI <7)	1.3 (0.5–4.1) (7 ≤ RDI <14)	2.3 (0.9–5.7) (RDI ≥14)	
b. Prospective, 7.5-year follow-up[22]	1180	1.0	1.1 (0.8–1.5)	0.9 (0.6–1.3)	1.0 (0.6–1.6)	

*Odds ratios are all adjusted for age, sex, body mass index (BMI), neck circumference, alcohol intake, and cigarette smoking. Additional adjustments are made for baseline hypertension and waist circumference in the Wisconsin study; for ethnicity and waist-to-hip ratio in the Sleep Heart Health Study; for ethnicity, menopause, and hormone replacement therapy in the Southern Pennsylvania study; and for coffee consumption and fitness level in the prospective Spanish study.
†Defined by systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or use of antihypertensive medication.
‡Reference category for odds ratio.
§Estimated at the mean age and BMI of the sample.
AHI, Apnea-hypopnea index; CI, confidence interval; RDI, respiratory disturbance index.

baseline level of OSA. After considering confounding factors, the odds of developing new hypertension over 4 years was twofold greater for those with an apnea-hypopnea index (AHI) of 5 to 15 events/hour and threefold greater for those with an AHI of greater than 15 at baseline, compared with participants without OSA at baseline (i.e., AHI <1). Longitudinal analyses of OSA as a predictor of 5-year incidence of hypertension in the Sleep Heart Health Study (SHHS)[19] and 7.5-year incidence of hypertension in the Vitoria-Gasteiz (Spain) Cohort[22] did not find the same strong association as was found in the Wisconsin cohort; however, the SHHS findings were consistent with about a 50% increased risk for incident hypertension in persons with severe OSA.

Cross-sectional analyses of baseline data from three population cohorts (Southern Pennsylvania,[23] Spain,[24] and the SHHS[25]) using measurements, definitions, and statistical adjustment models similar to those used in the Wisconsin Sleep Cohort have also shown OSA to be a statistically significant risk factor for hypertension (Figure 127-1; see Table 127-1). Collectively, the relationship between OSA and hypertension has been assessed with state-of-the-art measurements in more than 10,000 men and women from the general population, and results have been generally consistent. Confidence in the validity of the findings is increased as a result of the care taken by investigators to evaluate the effects of methodologic limitations.[20,26] If the epidemiology findings do reflect a causal relationship, as is suggested by most of the population-based data, a harder look must be taken at the importance of preventing OSA or at treating even mild OSA, discussed more fully below.

The largest population study, the cross-sectional examination of OSA and prevalent hypertension of Nieto and colleagues,[22] had sufficient power to describe the strength of the association across a wide spectrum of moderate to severe OSA. The findings suggest a dose-response association between OSA and hypertension up to a moderate level of OSA severity, with the association "flattening" across the severe range (AHI >30). It is possible that a flattening threshold exists, but methodologic limitations may account for it as well. This feature may reflect a survival bias against persons with high-severity OSA and cardiovascular disease, thus potentially missing subjects with the strongest associations from the analysis, as well as greater measurement error of OSA severity at higher AHI levels. Alternatively, a flattening threshold may be due to the presence of pathophysiologic consequences of severe OSA, such as heart failure, that could reduce blood pressure. In regard to heart failure, severe OSA has been associated with left ventricular systolic dysfunction[27]; in the SHHS, there was a dose-dependent relation between AHI and prevalence of heart failure.[28] If left ventricular systolic dysfunction occurs as OSA becomes more severe, adequate left ventricular stroke volume may not be maintained to

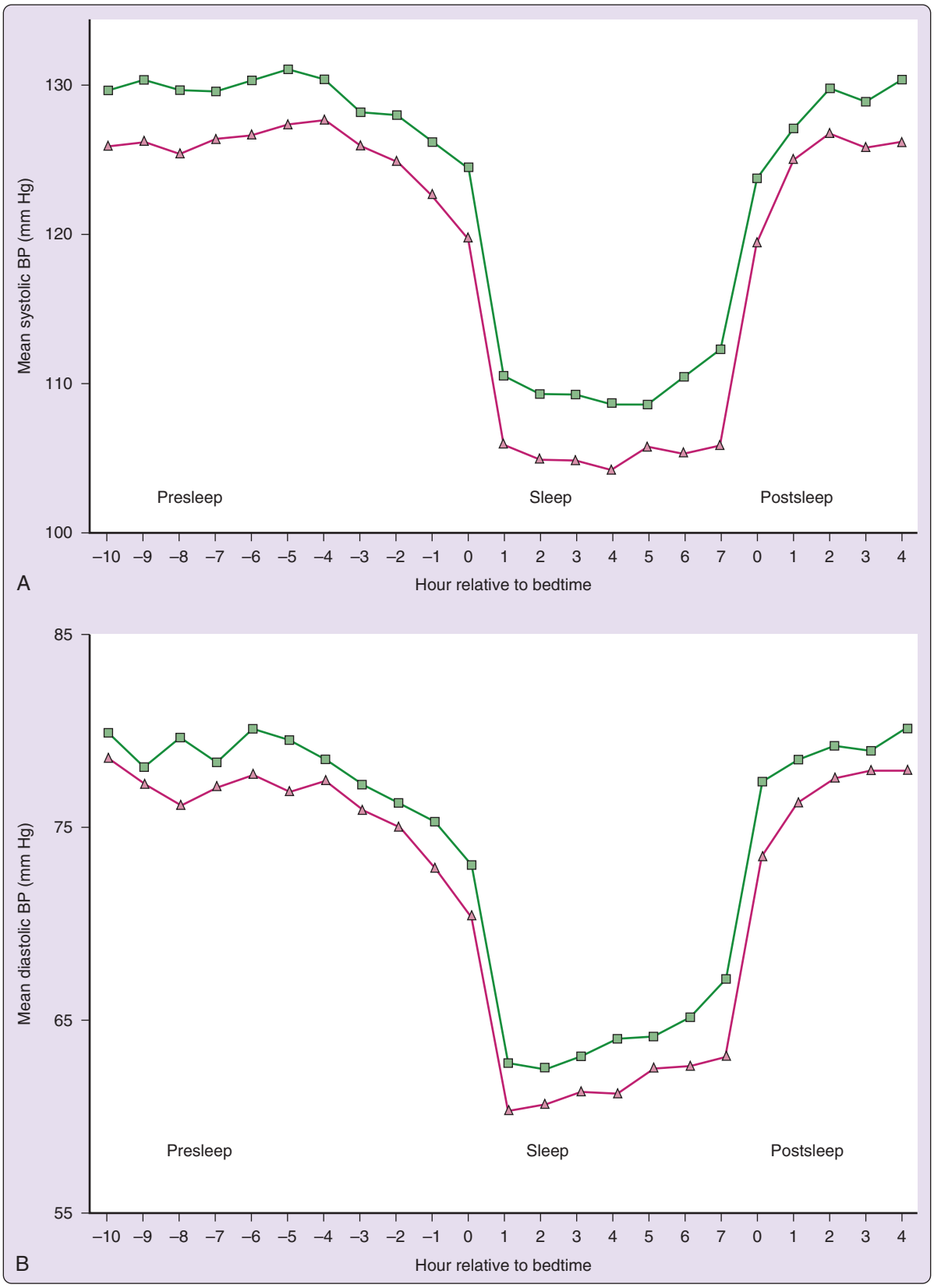


Figure 127-1 Ambulatory blood pressure (BP) during presleep, sleep, and postsleep by apnea-hypopnea index (AHI) category (Wisconsin Sleep Cohort Study). *Triangles:* AHI ≤5 events/hour; number of participants, 537. *Squares:* AHI ≥5, number of participants, 231. **A,** Systolic blood pressure. **B,** Diastolic blood pressure. Mean blood pressure values are adjusted for age, sex, and body mass index.

sustain a high blood pressure. Additionally, mechanisms mediating hypertension in OSA may be become saturated at greater OSA severity levels. In a two-arm study[29] of therapeutic and subtherapeutic positive airway pressure (PAP) in OSA, there were no changes in systemic arterial blood pressure from baseline in the subtherapeutic PAP trial arm despite a decrease in average AHI from 65 to 33. These results are consistent with a flattening of the OSA-hypertension association at greater OSA severity levels.

Obstructive Sleep Apnea and Hypertension in Population Subgroups

Determining the association of OSA with hypertension by subgroups (e.g., sex, age, ethnicity, body habitus) will increase understanding of physiologic mechanisms and may help target health care to appropriate subgroups. Most population cohort studies have reported a lack of sex differences in the association of OSA and hypertension. This message is particularly important because of past underdiagnosis and undertreatment of OSA in women; less-aggressive evaluation and treatment of OSA in women may lead to a relative survival disadvantage for them.[30]

Few studies have included sufficient population diversity to determine whether the association of OSA and hypertension varies by ethnicity. Recent population studies of OSA and hypertension in older adults have expanded the focus beyond OSA in middle age. Cross-sectional regression analyses of two large cohort studies with age ranges that include sufficient older people have suggested a negative interaction between age and AHI with respect to hypertension. Bixler and coworkers found that odds ratios (ORs) for hypertension decreased as age increased when patients with an AHI of 30 or greater were compared with those with an AHI of zero, and there was essentially no risk for hypertension associated with OSA for those 70 years and older.[23] Similar findings were reported by the SHHS: after stratifying their sample with a cutoff age of 65 years, the OR for hypertension with an AHI of greater than 30 versus an AHI of less than 1.5 was lower and not statistically significant for ages 65 years and older (OR = 1.23), compared with ages 40 to 65 years (OR = 1.64).[25] Further analysis of the SHHS data revealed that OSA was not associated with isolated systolic hypertension, the most common form of hypertension in older people.[31]

In contrast, cross-sectional analysis of the Bay Area Sleep Cohort of 129 older adults found a significant association between frequent (10 or more) apnea and hypopnea events during rapid eye movement sleep and diastolic blood pressure greater than 90 mm Hg.[32] This study of adults with a mean age of 72 years also linked occult sleep-disordered breathing to markers of cardiovascular disease. The authors concluded that a role of OSA in cardiovascular disease in older adults should not be ruled out.

Although it is possible that OSA outcomes differ in older versus younger people, methodologic difficulties arising from high comorbidity and survival bias are inherent in studying health effects of OSA in older populations and could cause a spurious age effect. Clinical practice should not be influenced by the notion that health risks of OSA in older people are lower until there are conclusive data—especially considering that even if relative risks relating OSA and hypertension are attenuated in older adults, corresponding absolute risk differences may well not be.

The prevalence of being overweight, a strong risk factor for both OSA and hypertension, is increasing worldwide. Most studies, using multiple regression analysis, have demonstrated that even after consideration of the strong associations of OSA and obesity, they both persist as independent factors in the development of hypertension.[5] However, understanding of specific mechanisms as well as precise effect sizes are limited by many analytic issues, including identification of relevant measures of body habitus, how the variables should be modeled in analyses, and the time course for OSA to have an effect on blood pressure. Adjusting for a marker of obesity (e.g., body mass index [BMI]) in multiple regression analysis of the relationship between OSA and hypertension is justified under the assumption that obesity is a confounder (i.e., obesity is a risk factor for both OSA and hypertension); however, if the relationship between OSA and obesity is bidirectional (e.g., if metabolic changes associated with OSA lead to an increase in body weight), adjustment for obesity in regression models of OSA and hypertension may lead to bias.

Furthermore, the interaction of obesity and OSA with respect to the development of hypertension is not clear. That is, does the strength of the OSA-hypertension association vary by body habitus? Findings from two of the large cohort studies have indicated that OSA is more strongly predictive of hypertension in leaner than in obese individuals: BMI was a significant modifier of the OSA-hypertension association in studies by Young and coworkers[33] and Bixler and coworkers.[23] The OR for hypertension and OSA increased in magnitude with decreasing BMI, indicating that in leaner people, those with OSA compared with those without may be at particularly high risk for hypertension. This finding has clinical implications, particularly in primary care, where sleep apnea is not likely to be suspected in nonobese patients, even in the presence of symptoms.

Obstructive Sleep Apnea and Diurnal Blood Pressure

Hypertension defined by chronic elevated daytime blood pressure is generally considered to be the outcome of interest; however, OSA-related nocturnal perturbations also include repeated spiking of pressures that exceed hypertension cutoff points (blood pressure load) and elevated average nighttime blood pressure as well as a carryover effect resulting in elevated daytime blood pressure.[5] Early studies of hemodynamics in patients with OSA involved invasive blood pressure monitoring; more recent studies of circadian patterns of blood pressure rely on ambulatory monitors that sample blood pressure at 15- to 30-minute intervals with arm cuff inflation–based methods.

Most studies with ambulatory blood pressure monitoring have been conducted on patients with sleep apnea. A Marburg study of 93 OSA patients[34] showed that the number of oxygen desaturation events per hour of presumed sleep was linearly related to both daytime and nighttime systolic and diastolic blood pressures, and it was related more strongly to nighttime pressures. In Great Britain, Davies and colleagues[35] performed ambulatory blood pressure studies on 45 pairs of sleep apnea patients and community controls matched to the patients on age, BMI, and treated hypertension. Sleep apnea patients had significantly higher diastolic pressures during the day and

night, higher systolic pressures during the night, and a notably smaller nocturnal dip.

Data from population studies with PSG and nocturnal blood pressure measures are sparse. In a preliminary study of 147 Wisconsin Sleep Cohort Study subjects,[36] OSA was consistently associated with elevated blood pressure measured by 24-hour ambulatory monitoring. Average systolic and diastolic blood pressures during wake and sleep, and systolic blood pressure load during wake and sleep, were all statistically significantly higher in the participants with mild to more severe OSA. Findings from an update with a much larger cohort sample (n = 768) are shown in Figure 127-1. Blood pressures, adjusted for confounding factors, were higher at every hour before, during, and after sleep for those with an AHI of greater than 5 versus an AHI of less than 5.

Nocturnal blood pressure relative to daytime pressure has been of special interest because studies have shown that the lack of the normal nighttime decline (nondipping) in blood pressure, usually considered to be a nighttime drop of at least 10% of the daytime pressure, is related to adverse cardiovascular outcomes independent of hypertension.[37] Findings from relatively small clinic-based samples do suggest that patients with OSA have a smaller nocturnal decline in blood pressure.[34,35,37-42] Prospective data from the Wisconsin Sleep Cohort study strengthen the evidence that acute effects of apnea and hypopnea episodes lead to a less favorable nighttime pattern of blood pressures.[43] Baseline and 7-year follow-up data from PSG and ambulatory blood pressure on a subsample of 328 participants in the Wisconsin cohort, free of nondipping at baseline, were analyzed. There was a dose-response increase in the odds of developing a pattern of nondipping in systolic blood pressure: Compared with subjects with an AHI of less than 5 at baseline, ORs, adjusted for confounding factors, were 3.1 (95% confidence interval [CI], 1.3 to 7.7) and 4.4 (95% CI, 1.2 to 16) for an AHI of 5 to 14 and for an AHI of 15 or greater, respectively.

Collectively, the ambulatory blood pressure studies do support an association between OSA and elevated blood pressure during both the nighttime and the daytime, and some studies suggest that the effect is greater during nighttime. These findings are particularly important because elevated ambulatory blood pressure has been shown to predict cardiovascular events independently of office-measured blood pressure and other cardiovascular risk factors.[44]

Blood Pressure Changes in Patients with Obstructive Sleep Apnea after Positive Airway Pressure Treatment

Studies of the effect of PAP on blood pressure have evolved from clinical observations of treated patients to sophisticated, randomized, double-blind trials, with sham PAP as the control condition and objectively measured compliance. Results from these studies are important, even though they do not necessarily directly address the basic question of whether OSA caused the initiation (incidence) of hypertension. This is because the effect of OSA on vasculature hemodynamics might be different at different stages in the natural history of OSA and hypertension, both chronic conditions evolving over a lifetime. However, even with this important caveat, clinical trial results generally show that successful treatment of OSA in patients who are also hypertensive tends to result in lower blood pressure levels.

The results of these studies are not uniformly consistent, likely because of differences in study design, sample size, patient populations, treatment regimens, and compliance from study to study. However, as reviewed in the following paragraphs, the pattern emerging from the overall body of literature on the subject, and especially when the most recent and highly powered studies are considered, strongly suggests that treatment of OSA does have a measurable effect in lowering blood pressure.

With respect to observational evidence, a prospective study that followed 55 patients (35 of whom were men) with both OSA and hypertension for 2 years,[45] a significant decrease was shown only for diastolic blood pressure (–2.2 mm Hg [95% CI, –4.2 to –0.1]), but not in systolic or 24-hour mean arterial blood pressure. Subgroup analyses, however, showed that 24-hour mean blood pressure did decrease significantly in patients with incompletely controlled hypertension at entry (–4.4 mm Hg, P = .01) as well as in those with high PAP compliance (–5.3 mm Hg, P = .01).

The predictors of change in blood pressure associated with PAP treatment in patients with OSA were examined in another observational prospective cohort study that recruited 86 patients treated for daytime sleepiness with PAP.[46] After 6 months, the average fall in 24-hour mean blood pressure was 4.9 mm Hg (95% CI, –7.1 to –2.1); the main predictors of blood pressure fall were the change in level of self-reported sleepiness and baseline BMI. These results are consistent with another observational study addressing a slightly different question: whether PAP treatment prevents new-onset hypertension. In this study, the hypertension incidence was assessed among 1889 patients who were referred for a PSG study and were free of hypertension at baseline and followed for up to 17 years (median, 12.2 years).[47] Compared with controls in this study, the adjusted hazard ratio of hypertension was 1.33 in patients with OSA who were not eligible for PAP treatment, 1.96 among OSA patients who declined PAP treatment, and 1.78 among patients not adherent to PAP (all P < .05). In contrast, patients with OSA who received PAP treatment had a statistically significant lower risk for hypertension incidence than controls (hazard ratio, 0.71; 95% CI, 0.53 to 0.94).

Evidence from randomized controlled trials has been growing in recent years, and even though not all the studies are consistent, the evidence supports the notion that PAP treatment results in a moderate but clinically significant lowering of blood pressure. Some of the early studies showed a significant lowering of blood pressure in patients with OSA, even though most of these studies did not select according to hypertension status.[29,48-50] In some of these studies, the effects were stronger or exclusively observed among the PAP-compliant patients.[29,49]

Other randomized trials, however, have shown minimal or nonsignificant effects of PAP in blood pressure.[51-56] Aside from the small sample size and limited statistical power, a potentially important limitation of some of these studies is that many of the patients included in the trial were not hypertensive, and thus there might be limited room for a lowering of blood pressuring effect as well as limited compliance with the treatment.

To address the latter, Hla and coworkers[57] investigated the effect of 3 weeks of PAP on blood pressure in hypertensive subjects with and without OSA to control for effects that PAP

might have on blood pressure independent of an effect from elimination of apnea and hypopnea events. Newly diagnosed, unmedicated hypertensive men from primary care settings were assessed for OSA by laboratory PSG to identify 14 men with an AHI of greater than 5 (mean AHI, 25) and 10 men with an AHI of less than 5 (mean AHI, 1). The OSA group received therapeutic PAP, and the non-OSA group received PAP at a pressure of 5 cm H_2O for 3 weeks. Ambulatory blood pressure monitoring indicated that nocturnal blood pressure in the OSA group dropped significantly with PAP (–10.3 mm Hg systolic, –4.5 mm Hg diastolic) but was essentially unchanged with PAP in the non-OSA group. There was a greater but statistically nonsignificant difference in blood pressure drop in the OSA versus non-OSA group in daytime blood pressure (–2.4 mm Hg systolic, –0.6 mm Hg diastolic).

In a meta-analysis of 16 trials conducted between 1996 and 2006 and including a total of 818 patients,[58] a small but statistically significant mean net change in systolic (–2.5 mm Hg; 95% CI, –4.3 to –0.6) and diastolic (–1.8 mm Hg; 95% CI, –3.0 to –0.6) blood pressures was observed. Net reductions in blood pressure were not statistically different between daytime and nighttime. This meta-analysis had some limitations that might reduce its generalizability (e.g., most studies included predominantly or exclusively obese middle-aged men, some studies were not blinded, and compliance was often limited). Furthermore, the duration of PAP treatment in all these studies was relatively short, ranging from 2 to 24 weeks[58]; longer treatment may be associated with different effects on systemic blood pressure.

These limitations notwithstanding, it is remarkable that a more recent meta-analysis including 28 studies published between 1980 and 2012 (representing 1948 patients) obtained results that were highly consistent with those of the previous study[59]: The weighted mean difference in diurnal systolic blood pressure (–2.58 mm Hg; 95% CI, –3.57 to –1.59 mm Hg) and diastolic blood pressure (–2.01 mm Hg; 95% CI, –2.84 to –1.18 mm Hg) both significantly favored PAP treatment over control arms. The effects were stronger in studies that included patients who were younger, were sleepier, had more severe OSA, and exhibited a higher degree of adherence to PAP.[59] The estimated effects were even stronger (3- to 5-mm Hg decrease in 24-hour blood pressure values) in another meta-analysis focusing on studies of resistant hypertension[60] and in another recent multicenter study in Spain.[61]

In one of the largest and more carefully conducted randomized controlled trials to date, Gottlieb and colleagues recruited 318 patients with cardiovascular disease or multiple cardiovascular risk factors from cardiology practices, that is, patients who were presumably subject to proper blood pressure control.[62] Among the 281 patients who had 24-hour blood pressure data at baseline and 12-week follow-up in this study, mean 24-hour arterial blood pressure was significantly lower in the group receiving PAP than in the control group (–2.4 mm Hg; 95% CI, –4.7 to –0.1; $P = .04$). Remarkably, the estimated reduction in blood pressure achieved by PAP is almost identical to that estimated in the two meta-analyses described previously.[58,59]

The focus of most of the previous studies and meta-analyses was primarily on changes in blood pressure associated with different types and length of PAP treatment in a variety of patient populations (e.g., some with OSA, some with hypertension, combinations of both). The results of one of the few randomized studies looking at the effect of PAP treatment on the *incidence* of hypertension were inconclusive.[63] In this Spanish multicenter study, 725 patients with OSA but without baseline hypertension and symptoms of daytime sleepiness were randomized to PAP treatment or no active treatment and followed up for a median of 4 years; even though a slight reduction (17%) in the combined incidence of hypertension and cardiovascular events was observed in the study, the result was not statistically significant, probably owing to the small number of events and limited statistical power.

A possible mechanism that might explain the putative improvement in blood pressure associated with PAP treatment is an improvement in vascular function. OSA may increase sympathetic activity, which is reversed by therapy with PAP. Other mechanisms are likely to be involved. In a randomized controlled trial including 29 patients with OSA associated with desaturation, compared with placebo, 6 weeks of PAP was associated with an improved forearm blood flow in response to both endothelial and non–endothelium-dependent stimuli.[64] Previous non–placebo-controlled studies have also demonstrated an association between PAP and either biochemical markers of vascular function[65] or forearm flow-mediated dilation measures.[55] Overall, the results of both observational and experimental studies on the effect of PAP therapy are consistent with evidence from both animal and human epidemiologic (particularly longitudinal) studies showing OSA as a possible cause of hypertension (see above). All this evidence, coupled with biologic plausibility studies defining mechanisms linking OSA to hypertension and the reversal of such mechanisms with PAP, support the hypothesis of a contribution of OSA to elevated blood pressure.

Clinical Relevance of the Role of Obstructive Sleep Apnea in Hypertension

In contrast to the state of evidence two decades ago, findings in support of a role for OSA in the development of hypertension are now difficult to dismiss as spurious.[66-68] Translating these findings to clinical settings, however, poses challenges. These findings, as well as new guidelines on hypertension detection and treatment, support case-finding for OSA in patients with hypertension in primary care settings, but what is the next step? Although diagnosis and treatment of a patient with hypertension and symptomatic OSA are priorities, a critical and much-debated question is the course of action that should be taken if a hypertensive patient has mild OSA without daytime symptoms of sleepiness. Is treatment of mild, asymptomatic OSA without complaints of sleepiness warranted on the basis of potential cardiovascular consequences? Arguments against treatment point to the lack of conclusive data. Based on the average effect of PAP on change in blood pressure, the predicted effect on an individual is likely to be small and of uncertain clinical significance and thus would not warrant replacing antihypertensive drugs for controlling blood pressure.[69]

On the other hand, there are arguments for considering PAP treatment in nonsleepy patients with mild, asymptomatic OSA and coexisting hypertension.[70,71] Worsening of mild sleep apnea over time is likely, and cardiovascular consequences that might be attributable to severe OSA could be prevented.[71] Some data suggest that PAP treatment for

patients with OSA and drug-resistant hypertension may be of benefit in lowering blood pressure.[60,72] Also, as noted by Gottlieb and colleagues,[62] even an average reduction of 2 mm Hg in systolic blood pressure—the approximate mean effect of PAP treatment from meta-analyses of OSA treatment trials described previously—might be expected to substantially lower stroke rates (by ~10%) and heart disease mortality rates (by ~7%) in treated populations.[73]

Primary health care providers increasingly recognize markers for OSA, such as snoring, and are becoming more aware of the OSA-hypertension association. Consequently, referrals of patients with mild, asymptomatic OSA for sleep evaluations will very likely increase, heightening the dilemma of whom to treat in the face of limited medical resources. Further understanding of the costs and benefits of OSA treatment in preventing the incidence and progression of hypertension and other cardiovascular diseases is needed before this problem can be satisfactorily addressed.

Meanwhile, clinicians providing care for OSA patients using PAP should underscore the importance of full compliance as well as adequate control of disordered breathing events with PAP, as evidenced by studies showing the importance of treatment adherence in lowering blood pressure.[29,48] As indicated earlier, long-term small changes in blood pressure have major preventive effects. In prospective studies of 420,000 persons, a decrease of 5 mm Hg in diastolic blood pressure lessened the incidence of stroke and coronary heart disease by approximately 34% and 21%, respectively, and there was a dose-dependent reduction in blood pressure and incident cardiovascular diseases.[74] Therefore, in patients with OSA, long-term compliance with PAP may be effective in preventing adverse cerebrovascular and cardiovascular diseases.

PULMONARY HYPERTENSION

Obstructive Sleep Apnea as a Cause of Pulmonary Hypertension

In 1998, the second WHO conference on pulmonary arterial hypertension[4] recognized sleep-disordered breathing as a secondary cause of PH. The classification of PH has been revised by a more recent WHO conference.[75,76] There are five groups, each with a number of subgroups consisting of various causes of PH. The first group, pulmonary arterial hypertension, includes idiopathic pulmonary arterial hypertension. The second class is pulmonary venous hypertension, which is most commonly due to elevated left heart filling pressures such as left ventricular diastolic dysfunction. PH secondary to OSA falls into the third category, which also includes chronic obstructive pulmonary disease (COPD), interstitial lung diseases, and PH chronic exposure to high altitude. The basic pathophysiologic mechanism underlying PH in this group of disorders is hypoxemia. However, as will be emphasized later, OSA can cause PH through left ventricular diastolic dysfunction, as it occurs in group 2. Group 4 is PH due to thromboembolic pathologic disorders, and group 5 consists of miscellaneous disorders that cannot be easily classified in the other four groups.

The gold standard for diagnosis of PH is right heart catheterization. As noted earlier, the WHO defines the presence of PH as resting mean pulmonary artery pressure of 25 mm Hg or greater. Investigators in the field of sleep apnea have mostly used a mean pulmonary artery pressure of 20 mm Hg or greater, a threshold that is lower than that defined by WHO. However, a resting mean pulmonary artery pressure of 20 mm Hg or greater is generally considered abnormal. In this chapter, we defined PH according to the WHO criteria, which have become the rule independent of the cause of the PH. We also emphasized that right heart catheterization is essential for phenotyping PH, assessing its severity and the targeted therapy.

In patients with OSA, the prevalence of abnormal mean pulmonary artery pressure varies considerably, from 15% to 70%.[77-84] This variation in part depends on inclusion in some studies of patients with COPD, hypercapnic OSA (obesity hypoventilation syndrome), and obesity, which contribute to increased frequency, prevalence, and severity of PH in OSA. Meanwhile, in patients with OSA without comorbid disorders, PH is usually mild, although it could also be severe in advanced OSA, resulting in cor pulmonale, a feature of pickwickian syndrome (hypercapnic OSA syndrome).

An early study[1] of 12 patients with OSA who had undergone right heart catheterization showed cyclic changes in pulmonary artery pressure coinciding with episodes of OSA. A marked degree of hypoxemia and hypercapnia was associated with these hemodynamic abnormalities. In some of these patients, systolic pulmonary artery pressure exceeded 60 mm Hg. During wakefulness, four patients had abnormal mean pulmonary artery pressure ranging from 20 to 22 mm Hg. One of these four patients had an elevated pulmonary capillary wedge pressure of 16 mm Hg. With exercise, most of the patients had mean pulmonary artery pressure of about 30 mm Hg. In some of these patients, the wedge pressure increased with exercise, unmasking left ventricular diastolic dysfunction. This study indicated that OSA impaired the physiologic processes that normally operate to enable pulmonary circulation and left ventricular function to maintain pulmonary artery pressure close to normal in the face of increases in cardiac output.

Since the study of Tilkian and colleagues,[1] many studies have demonstrated presence of PH in patients with OSA. Box 127-1 summarizes the four largest studies in which full-night PSG and right heart catheterization, the gold standard for diagnosis of OSA and PH, were performed. In a French study[78] involving 220 consecutive patients with an AHI of greater than 20, 37 patients (17%) had a resting mean pulmonary artery pressure of at least 20 mm Hg (range, 20 to 44 mm Hg), and in 17 patients (8%) the mean pressure was 25 mm Hg or greater. Patients with a resting mean pulmonary artery pressure of at least 20 mm Hg had more severe OSA, a higher $Paco_2$, a higher BMI, and a lower Pao_2 than patients without PH. Furthermore, these patients had higher prevalence of both obstructive and restrictive pulmonary defects. $Paco_2$ and forced expiratory volume in 1 second (FEV_1) were the two major predictors of high resting mean pulmonary artery pressure. When PH was defined as a mean pressure of 30 mm Hg with exercise, virtually all patients met this criterion; in 23 patients (62%) the mean pressure exceeded 40 mm Hg.

In an Australian study[79] of 100 consecutive patients with an AHI of 20 or more, 42% had elevated pulmonary artery pressure, with the mean pressure ranging from about 20 to 52 mm Hg. Some patients had overlap syndrome. In 24% of the patients, the mean pressure was more than 25 mm Hg. In this study, $Paco_2$, Pao_2, and FEV_1 accounted for about 33% of

Box 127-1 STUDIES ON PULMONARY HYPERTENSION IN PATIENTS WITH OBSTRUCTIVE SLEEP APNEA

Chaouat A, Weitzenblum E, Krieger J, et al.[78]

- 220 consecutive French patients with AHI >20 events/hr
- 17% had mean PAP >20 mm Hg
- 8% had PH with mean PAP ≥25 mm Hg; I patient with mean PAP ≥40 mm Hg
- Patients with PAP >20 mm Hg had more severe OSA, higher $Paco_2$ and BMI, lower Pao_2, and a more obstructive and restrictive defect
- $Paco_2$ and FEV_1 were independent predictors of mean PAP

Laks L, Lehrhaft B, Grunstein RR, et al.[79]

100 consecutive Australian patients with AHI >20
42% had mean PAP >20 mm Hg; range, 20 to 52 mm Hg; 24% had PH with mean PAP ≥25 mm Hg
5% had PH with mean PAP >40 mm Hg
$Paco_2$, Pao_2, and FEV_1 accounted for 33% of the variability in PH
Six patients (6%) with mean PAP ranging from 20 to 52 mm Hg had normal Pao_2

Sanner BM, Doberauer C, Konermann M, et al.[80]

92 consecutive German patients with OSA and AHI >10; range, 10 to 100/hr
COPD was an exclusion criterion
20% had mean PAP ranging from 20 to 26 mm Hg; 2 with PH had mean PAP ≥25 mm Hg
Eight patients had increased PCWP; all had systemic hypertension
PCWP and time spent at <90% saturation were independent predictors of mean PAP

Bady E, Achkar A, Pascal S, et al.[81]

44 patients with OSA and AHI >5
COPD (FEV_1/FVC ratio <60%) was an exclusion criterion
27% had mean PAP >20 mm Hg with mean pressure ≥28.5 mm Hg
All with PCWP ≤15 mm Hg
18% had PH with mean PAP ≥25 mm Hg

AHI, Apnea-hypopnea index; BMI, body mass index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; PAP, pulmonary artery pressure; PCWP, pulmonary capillary wedge pressure; PH, pulmonary hypertension.

Box 127-2 MECHANISMS OF PULMONARY HYPERTENSION IN OBSTRUCTIVE SLEEP APNEA

Precapillary Pulmonary Hypertension

Hypoxemia
Hypercapnia
Endothelial dysfunction or remodeling*
Changes in intrathoracic pressure

Postcapillary Pulmonary Hypertension

Left ventricular hypertrophy and diastolic dysfunction

*See Tilkian AG, Guilleminault C, Schroeder JS, et al. Hemodynamics in sleep-induced apnea: studies during wakefulness and sleep. *Ann Intern Med* 1976;85:714–9.

variability in pulmonary artery pressure. Six patients with abnormal pulmonary artery pressure had normal Pao_2.

In a German study[80] of 92 consecutive patients with an AHI of greater than 10 and with COPD as an exclusion criterion, 20% had a mean pulmonary artery pressure of 20 to 25 mm Hg. Only one patient met the current criterion of PH with a mean of 25 mm Hg. Eight patients had increased pulmonary capillary wedge pressure, and all of these patients had systemic hypertension that was presumably causing left ventricular diastolic dysfunction. Pulmonary capillary wedge pressure and time spent with a saturation of below 90% were the independent variables predicting PH.

The presence of PH in patients with OSA but without COPD was also confirmed in another French study (see Box 127-1).[81] In this study, however, COPD was defined by an FEV_1 of less than 70% predicted and a ratio of FEV_1 to forced vital capacity (FVC) of less than 60% predicted. The study involved 44 patients, 12 of whom (27%) had mean pulmonary artery pressure greater than 20 mm Hg, all with pulmonary capillary wedge pressure of less than 15 mm Hg. Importantly, 8 patients (18% of all patients) had PH with mean pulmonary artery pressure of 25 mm Hg or greater. The authors reported that mean pulmonary artery pressure was positively correlated with BMI and negatively correlated with Pao_2. Patients with elevated mean pulmonary artery pressure had significantly lower values for FVC and FEV_1. The mechanisms by which BMI positively correlated with PH could have been multifactorial and related to restrictive lung defect and hypoxemia.

Combining the results of the aforementioned four studies using PSG to determine the presence of OSA, and right heart catheterization to define PH, 51 of the 456 patients (11%) satisfy the current WHO criteria for PH.

In conclusion, mild PH is common in patients with OSA and may occur in the absence of COPD and daytime hypoxemia. However, severe OSA, severe hypoxemia, hypercapnia (obesity hypoventilation syndrome), obstructive or restrictive lung defects, and left heart disease are more commonly associated with PH and contribute to its severity. In addition, as noted, increased pulmonary artery pressure either becomes manifest or is augmented by exercise and can cause dyspnea and exercise intolerance.[82]

Mechanisms of Pulmonary Hypertension in Patients with Obstructive Sleep Apnea

Intermittent nocturnal rises in the pulmonary artery pressure in association with upper airway collapse have been well documented. Multiple mechanisms mediate nocturnal rises in pulmonary artery pressure.[85] These include alterations in blood gases (i.e. intermittent hypoxemia and hypercapnia), cardiac output, lung volume, intrathoracic pressure, compliance of pulmonary circulation, and left ventricular diastolic dysfunction. With time and in the long run, nocturnal PH spills over to diurnal hypertension.

Diurnal PH in patients with OSA could be precapillary, capillary, or postcapillary, depending in part on comorbid disorders that may contribute to the development of PH (Box 127-2). Postcapillary PH (pulmonary venous hypertension) is common and results primarily from elevated left heart filling pressures, specifically owing to their left ventricular hypertrophy and diastolic dysfunction caused by diurnal systemic hypertension and nocturnal consequences of OSA noted

previously. In regard to the latter, left ventricular hypertrophy could be present in patients with OSA even in the absence of daytime systemic hypertension,[86] presumably because of cyclic changes in systemic artery blood pressure and hypoxemia[84] during sleep. In the presence of a hypertrophied or noncompliant left ventricle, end-diastolic pressure increases, resulting in backward passive increase in pulmonary venous, capillary, and pulmonary artery systolic and diastolic pressures. This acute postcapillary PH is reversible, if the etiologic factor (e.g., OSA) is effectively treated. Otherwise, with persistent PH, remodeling of pulmonary vascular bed occurs and vascular resistance increases, which in time may become reversible even if the etiology of left heart disease is effectively treated.

As noted earlier, as a cause of PH, OSA is categorized in group 3 along with COPD and other lung diseases. Here, the underlying pathophysiology is hypoxemia. However, because of intermittent partial or complete pharyngeal collapse during sleep, repeated episodes of hypoxemia and hypercapnia occur, both of which have been shown to acutely induce pulmonary arteriolar vasoconstriction, increasing pulmonary vascular resistance. With time, however, distinct pathophysiologic sequelae ensue that may be irreversible. In any case this is the precapillary PH, which is another potential mechanism of OSA-induced PH. Therefore the combination of hypoxic-hypercapnic pulmonary arteriolar vasoconstriction and pulmonary venous hypertension could result in severe PH in patients with OSA. Similarly, when OSA is comorbid with COPD, which is in the same group with OSA as a potential cause for PH, the combination could result in severe PH.

Detailed molecular mechanisms underlying PH in OSA are beyond the scope of this chapter. However, production of mediators eventually results in endothelial cell damage, vascular cell proliferation, and aberrant vascular remodeling. In addition, with endothelial cell dysfunction, there is reduced nitric oxide production and increased endothelin,[86,87] both of which contribute to further PH. As noted earlier, the initial cascade of events is potentially reversible, emphasizing the importance of early recognition and treatment of OSA.

Loss of vascular surface area, as may occur in patients with COPD, is an important cause of capillary PH, and it may significantly contribute to PH in patients with OSA. Several studies[77-79] have shown that COPD and a low FEV_1 are predictors of PH in patients with OSA. COPD could also contribute to PH by way of arteriolar vasoconstriction due to hypoxemia and hypercapnia as noted previously.

An important mechanism mediating PH in patients with OSA is the presence of factors that cause constriction of pulmonary arterioles, leading to precapillary PH. The best-known stimulus is alveolar hypoxia, and it is not surprising that hypoxemia is an independent predictor of PH in OSA (see Box 127-1). However, hypercapnia could also increase pulmonary arterial blood pressure. The molecular mechanisms of PH in general are complex and multifactorial. Both acquired and genetic factors are involved. Disordered endothelial cell function, in part caused by hypoxia (and reoxygenation) and manifested biochemically by an imbalance between concentrations of local vasodilators (e.g., nitric oxide and prostacyclins) and vasoconstrictors (e.g., endothelin-1, thromboxane, serotonin), as occurs in endothelial dysfunction syndrome, appears to mediate the development of PH.[87,88] It is also conceivable that if OSA is long-standing, pulmonary vascular remodeling similar to that in COPD could occur because a number of mediators such as vascular endothelial growth factor are proliferative and angiogenic.

In summary, the consequence of OSA on pulmonary circulation may vary from those of cyclic nocturnal PH, which occurs in virtually all patients, to daytime PH, right ventricular dysfunction, and eventually cor pulmonale, a feature of pickwickian syndrome. However, even in the absence of cor pulmonale, which is the manifestation of long-standing severe PH, presence of PH increases right ventricular afterload and myocardial oxygen consumption. If PH develops as a result of increases in cardiac output (e.g., with exercise), it may cause dyspnea and exercise intolerance.

Changes in Pulmonary Artery Pressure after Positive Airway Pressure Treatment of Obstructive Sleep Apnea

Because mechanisms of PH in OSA are multifactorial (see Box 127-2), the behavioral response of pulmonary circulation to therapy for OSA probably depends on several factors. For example, if loss of vascular surface area due to the presence of COPD or other comorbid pulmonary disorders is contributing to PH in OSA, this component is irreversible.[89] Similarly, if remodeling of the pulmonary vascular bed has occurred, long-standing effective therapy is necessary to effect any reversal component (reverse remodeling). Therefore, if PAP is used to treat OSA, long-term compliance with therapy is critical and needs to be confirmed by covert monitoring. Large, long-term systematic studies considering these important factors are necessary to determine the effects of treatment of OSA on pulmonary circulation. Lack of such considerations may lead to serious underestimation of effects.

Effective treatment of OSA could improve PH. Here we review the studies that have implemented right heart catheterization both at baseline as well as long term. In an early study when tracheotomy was the best therapeutic option, Fletcher and colleagues[89] studied three groups of OSA patients. Nine patients with hypercapnic OSA and mostly with COPD underwent the operation, and repeat right heart catheterization was performed about 6 months later. Six of nine patients had baseline right heart catheterization, and five of them met the current criteria for PH. The resting mean pulmonary artery pressure was 39 mm Hg in these six patients. About 12 months later, the mean resting pulmonary artery pressure was 25 mm Hg. Mean pulmonary artery pressure decreased in five of the six patients. It did not change significantly in the single patient who had mean pulmonary artery pressure of about 24 mm Hg. Meanwhile there was a significant rise in right ventricular ejection fraction in association with a reduction in the pulmonary artery pressure and vascular resistance. Motta and coworkers[90] also performed tracheostomy on six patients with OSA. However, the mean pulmonary artery pressure did not change significantly in these patients. Importantly, only one patient met the current criterion of PH with a mean pressure of 28 mm Hg.

Another negative study was reported by Sforza and colleagues.[91] The authors treated 54 patients with OSA with CPAP. The mean pulmonary artery pressure did not change significantly. However, similar to the study of Motta and colleagues, none of the patients met the current criterion for PH.

Alchanatis and colleagues[92] studied 29 patients with OSA and without COPD. Three of the 29 patients met the current criterion for PH with baseline values of 30, 30, and 28 mm Hg.

Six months after therapy with CPAP, the mean pulmonary artery pressure had dropped below 25 mm Hg in these 3 patients. Respective values were about 23, 22, and 20 mm Hg.

In another French study of OSA patients who were treated with long-term CPAP, Chouat and associates[92] reported no significant changes in mean pulmonary artery pressure in 44 patients who had undergone right heart catheterization. The mean pulmonary artery pressure at baseline, however, was normal at 16 mm Hg. After an average of 64 months of CPAP use, the mean pressure was 17 mm Hg. The authors reported that in 11 patients whose average value of the mean pulmonary artery pressure was 24 mm Hg at baseline (authors considered a value of 20 mm Hg or greater as PH), the pressure decreased to 20 mm Hg, although this value was not statistically significant. It is not clear how many of the 11 patients met the current criterion for PH. However, if the mean pressure is within normal range it may not be expected to decrease significantly with intervention.

The last a study using right heart catheterization was reported by Sajkov's group[93] who studied 20 patients with OSA (average AHI, 49 or greater) before and 4 months after treatment with PAP. In this study, PAP compliance was objectively monitored, and the average was 5 hours per night. Patients had normal lung function. Five patients who had abnormal pulmonary artery pressure (range, 20 to 32 mm Hg) showed the most dramatic decrease to less than 20 mm Hg after 4 months of effective treatment with PAP. Two of the 5 patients met the current criterion for PH with mean pulmonary artery pressures of 31 and 27 mm Hg. Four months after therapy with CPAP, respective pressures were 18 and 13 mm Hg. Interestingly, the authors showed a time-dependent progressive decrease in mean pulmonary artery pressure with right heart catheterization, which was performed at 1 and 4 months of intervention. In a subject who was not compliant with CPAP, there was no change in pulmonary artery pressure. Although this was a single observation, this finding and those reported for systemic hypertension strongly indicate that effective use of PAP is necessary to lower systemic and pulmonary artery pressures.

We now review relevant randomized clinical trials. Arias and colleagues[94] randomized 23 middle-aged patients with severe OSA (AHI, 44 or greater) to either sham PAP or PAP therapy. In this crossover trial, after 12 weeks of PAP therapy, pulmonary artery systolic pressure decreased significantly from a mean of about 30 to 24 mm Hg. The reduction was greatest (8.5 mm Hg) in patients with PH defined as pulmonary artery systolic pressure of 30 or more determined by echocardiography. In the second randomized study[95] sham CPAP was used as placebo. Arias and colleagues[95] performed a crossover study of 12 weeks' duration in 27 consecutive newly diagnosed men with OSA and abnormal echocardiographic left ventricular filling pattern. Twelve weeks of effective CPAP therapy resulted in significant increases in E/A ratio and reduction in mitral deceleration isovolume relaxation times. We must emphasize that based on the design of this important study, the main and perhaps the only pathologic reason causing left ventricular diastolic dysfunction was OSA; among the exclusion criteria were presence of known hypertension, ischemic or valvular heart disease, diabetes, morbid obesity, and daytime hypoxemia. Therefore the results of this study demonstrate that OSA could be a cause of the reversible diastolic dysfunction. As noted earlier, impaired left ventricular filling could be an important cause of PH in patients with OSA.

The American College of Cardiology and American Heart Association expert consensus document recommends PSG to rule out OSA for all patients with PH. The recommendation is based on the idea that targeted therapy of OSA could either improve or prevent further deterioration in central hemodynamics.

CLINICAL PEARLS

- Epidemiologic studies support a causal role of OSA in systemic hypertension independent of BMI, measures of fat distribution, age, sex, and other possible confounding factors.
- Randomized double-blind placebo (sham PAP)–controlled trials of patients with hypertension demonstrate that effective treatment of OSA with PAP lowers blood pressure. A decrease in blood pressure is most pronounced in those with the most severe OSA and those who are the most compliant.
- Even small decrements in blood pressure, maintained for the long term, have been shown to significantly lessen the incidence of cerebrovascular and cardiovascular diseases. Thus the potential lowering of blood pressure from PAP treatment holds promise for decreasing cerebrovascular or cardiovascular disease. However, adequate control of OSA and compliance with PAP, particularly in patients with severe OSA, are critical.
- Several observational studies show that OSA, particularly when severe, is a cause of mortality. Treatment with PAP decreases mortality risk.
- OSA is a cause of secondary PH, and this has been recognized by international health organizations. PH as defined by a mean pulmonary artery pressure of 25 mm Hg or greater is usually mild, although it could be severe, particularly in the presence of severe OSA or OSA with hypercapnia, obesity, and comorbid disorders such as COPD. Treatment of OSA with PAP may improve or prevent further deterioration in pulmonary pressures. All patients with PH, independent of the cause, should undergo PSG and targeted therapy when indicated. This may halt deterioration in central hemodynamics.

SUMMARY

Findings from investigations based on diverse populations and different study designs support a role for OSA in systemic hypertension and PH. Population-based epidemiology studies have shown that persons with moderate to severe OSA (15 or more apnea or hypopnea events per hour) have greater probability of having or developing hypertension than persons who do not have OSA. These associations are only partly explained by confounding factors such as age or increased BMI. In epidemiologic studies that use 24-hour ambulatory blood pressure monitoring, OSA–blood pressure associations are seen with both sleep and wake blood pressures. PH, too, is prevalent in patients with OSA. Mild pulmonary arterial hypertension may occur in patients with OSA without daytime hypoxemia or COPD, but these comorbidities are more common in patients with severe OSA.

OSA treatment trials also support a causal association between OSA and hypertension, with most studies of

systemic or pulmonary blood pressures before and after PAP therapy demonstrating blood pressure reductions. Intervention trials generally show modest reductions in systemic blood pressure (2 to 10 mm Hg reductions), with the largest effects seen in effectively treated patients with severe OSA. Importantly, small changes in blood pressure, if maintained, have the potential to significantly decrease the population incidence of cerebrovascular and cardiovascular disease.

Selected Readings

Barbé F, Durán-Cantolla J, Sanchez de la Torre M, et al. Effect of continuous positive airway pressure on the incidence of hypertension and cardiovascular events in nonsleepy patients with obstructive sleep apnea: a randomized controlled trial. *JAMA* 2012;**307**:2161–8.

Campos-Rodriguez F, Perez-Ronchel J, Grilo-Reina A, et al. Long-term effect of continuous positive airway pressure on BP in patients with hypertension and sleep apnea. *Chest* 2007;**132**:1847–52.

Cano-Pumarega I, Durán-Cantolla J, Aizpuru F, et al. Obstructive sleep apnea and systemic hypertension: longitudinal study in the general population: the Vitoria Sleep Cohort. *Am J Respir Crit Care Med* 2011;**184**:1299–304.

Floras JS. Hypertension and sleep apnea. *Can J Cardiol* 2015;**31**:889–97.

Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med* 2014;**370**:2276–85.

Ismail K, Roberts K, Manning P, et al. OSA and pulmonary hypertension: time for a new look. *Chest* 2015;**147**(3):847–61.

Javaheri S, Javaheri S, Javaheri A. Sleep apnea, heart failure and pulmonary hypertension. *Curr Heart Fail Rep* 2013;**10**:315–20.

Marin JM, Agusti A, Villar I, et al. Association between treated and untreated obstructive sleep apnea and risk of hypertension. *JAMA* 2012;**307**:2169–76.

Martinez-Garcia MA, Capote F, Campos-Rodriguez F, et al. Effect of CPAP on blood pressure in patients with obstructive sleep apnea and resistant hypertension: the HIPARCO randomized clinical trial. *JAMA* 2013;**310**:2407–15.

Montesi SB, Edwards BA, Malhotra A, Bakker JP. The effect of continuous positive airway pressure treatment on blood pressure: a systematic review and meta-analysis of randomized controlled trials. *J Clin Sleep Med* 2012;**8**:587–96.

Motta J, Guilleminault C, Schroeder JS, et al. Tracheostomy and hemodynamic changes in sleep-induced apnea. *Ann Intern Med* 1978;**89**:454–8.

O'Conner GT, Caffo B, Newman AB, et al. Prospective study of sleep-disordered breathing and hypertension; the Sleep Heart Health Study. *Am J Respir Crit Care Med* 2009;**179**:1159–64.

Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000;**342**:1378–84.

A complete reference list can be found online at ExpertConsult.com.

Chapter
128

Coronary Artery Disease and Obstructive Sleep Apnea

Yüksel Peker; Karl A. Franklin; Jan Hedner

Chapter Highlights

- Obstructive sleep apnea is overrepresented in patients with coronary artery disease and occurs in about 50% (36% to 66%) of such patients, most of them without complaints of excessive daytime sleepiness.
- Sleep apnea is even more prevalent during the presentation of myocardial infarction and may explain a peak incidence of sudden cardiac death during midnight and early morning hours.
- Increased oxygen demand and reduced oxygen supply after obstructive apneas may trigger nocturnal angina in patients with low oxygen reserve because of lack of ventilation.
- Patients with sleep apnea and coronary artery disease have an increased risk for developing stroke, but it is still unclear whether they have an increased risk for early death independent of other comorbidities.
- Obstructive sleep apnea is suggested as an independent risk factor for atherosclerosis because of repeated apnea-induced hypoxemia and reoxygenation-induced oxidative stress with immediate and sustained sympathetic activation, endothelial dysfunction, and inflammation.
- Prospective studies report a reduction of nocturnal ischemia during elimination of obstructive events with continuous positive airway pressure treatment and lowering of the risk for recurrent myocardial infarction, without any reports of adverse events.
- Randomized controlled trials examining the effect of continuous positive airway pressure treatment on long-term cardiovascular outcomes are underway.

Epidemiologic data suggest that obstructive sleep apnea (OSA) is overrepresented in patients with coronary artery disease (CAD). Other studies suggest that the clinical course of CAD is initiated or accelerated by the presence of sleep-related breathing disorders. A rapidly evolving field of experimental data demonstrates that OSA, by phenomena such as hypoxemia and reoxygenation, may trigger a sequence of events involved in the development of atherosclerotic disease.

Development of vascular disease and CAD is influenced by several risk factors that also have been associated with OSA. Sleep apneic events induce a state of increased cardiac oxygen demand but are also often associated with low oxygen reserve because of lack of ventilation. Nocturnal angina can therefore be triggered by sleep apneas in patients with CAD. There is growing evidence that elimination of the sleep disorder can benefit patients with OSA at risk for CAD. Other data suggest that treatment of OSA improves prognosis in patients undergoing coronary revascularization. This chapter reviews the evidence of an association between these two conditions.

EPIDEMIOLOGY

The risk for experiencing angina pectoris or an acute coronary syndrome such as unstable angina, acute myocardial infarction (MI), or sudden cardiac death (SCD) has long been known to be increased during the late hours of sleep or in the hours soon after awakening.[1] This association may be explained in part by occurrence of OSA. A retrospective analysis showed an overrepresentation of peak time in sudden death from the cardiac causes during the sleeping hours in patients with OSA, which contrasted with a nadir in sudden death from cardiac causes in subjects without OSA and in the general population.[2] The same group addressed whether OSA independently increases the risk for SCD in a longitudinal follow-up of 10,701 consecutive adults.[3] During an average follow-up of 5.3 years, 142 patients had resuscitated or suffered from SCD (annual rate, 0.27%). In a multivariate analysis, SCD was associated with an apnea-hypopnea index (AHI) of at least 20 events/hour (hazard ratio [HR] = 1.60), mean nocturnal oxygen saturation of less than 93% (HR = 2.93), and oxygen nadir saturation of less than 78% (HR = 2.60; all $P < .001$). The authors concluded that OSA predicted incident SCD. Another small prospective study addressing the time of onset of MI showed a higher likelihood of having OSA in those with an onset of MI during midnight hours.[4] Moreover, a recent study also demonstrated that the incidence of MI onset between 6:00 AM and 12:00 PM was higher in OSA patients (AHI ≥5) than in control patients (38% vs. 25%; $P = .039$). Moderate to severe OSA (AHI ≥15) significantly enhanced this circadian variation

(odds ratio [OR] = 2.0) after adjustment for age, body mass index (BMI), and comorbidities.[5]

In general, there is a stronger relationship between OSA and CAD in clinical cohorts than in the general population because clinical cohort studies are particularly influenced by comorbidity and confounding factors, including obesity, diabetes mellitus, hypertension, smoking, and hyperlipidemia. This circumstance also suggests that OSA constitutes an additive or synergistic risk factor for development of CAD.

Prevalence of Obstructive Sleep Apnea and Coronary Artery Disease in the General Population

The largest study to date addressing OSA and CAD in the general population is the Sleep Heart Health Study.[6] The investigators performed a cross-sectional analysis of 6132 subjects undergoing unattended full-night home polysomnography. There was a modest risk increase (peaking at an OR of 1.27) for self-reported CAD when the highest and lowest AHI quartiles were compared. The weak association in this general population study may be explained by a proportionally high age and a low median AHI.

Prevalence of Coronary Artery Disease in Patients with Obstructive Sleep Apnea

Clinical studies of CAD in sleep clinic cohorts generally involve patients with OSA and with daytime symptoms. Consequently, compared with studies in the general population, these studies deal with symptomatic patients, those likely to suffer from more severe sleep apnea, and potentially patients with excess comorbidities such as diabetes, obesity, and cardiovascular disorders. Available data are to a large extent based on uncontrolled studies. For example, in a sleep clinic cohort of 386 subjects,[7] CAD was present in almost one fourth of subjects with OSA, and the percentage of patients with CAD was high among those with moderate to severe OSA.

In another study,[8] simultaneous polysomnography and electrocardiographic recordings demonstrated that episodes of nocturnal ischemia were more common in patients with OSA who also had CAD, and mainly so during rapid eye movement (REM) sleep, during periods of high apnea activity, and during sustained hypoxemia. Moreover, ST-segment depression on electrocardiography was not uncommon during sleep in patients with OSA but without a history of CAD, and these changes were eliminated by continuous positive airway pressure (CPAP).[9] Studies using invasive measures, including angiography, verified CAD in more than 20% of investigated subjects with OSA,[10] and an even higher prevalence (68%) was reported in a slightly larger study of unselected patients with OSA.[11] Collectively these data suggest a proportionally high prevalence of CAD in sleep clinic cohorts.

Prevalence of Obstructive Sleep Apnea in Patients with Coronary Artery Disease

Sleep-disordered breathing appears to be common in patients with CAD. An early small study demonstrated OSA or central sleep apnea with Cheyne-Stokes breathing in 13 of 17 male patients with angiographically verified CAD.[12] A subsequent Australian case-control study that investigated middle-aged male survivors of acute MI and age-matched controls provided the first clinic-based epidemiologic evidence of an increased prevalence of OSA in patients with CAD.[13] OSA (apnea index [AI] ≥5 events/hour) was found in approximately one third of the patients, compared with only 4% of age-matched healthy controls, and constituted an independent predictor of MI after adjustment for traditional risk factors.

A larger case-control study provided a similar OSA prevalence (31%), whereas the prevalence in the control group was 20%.[14] In this population, an AHI of 20 was associated with a history of MI (OR = 2.0; 95% confidence interval [CI], 1.0 to 3.8). In a tightly age-, sex-, and BMI-matched Swedish case-control study of 62 patients, OSA, based on an AHI of greater than 10, provided an independent OR of 3.1 (95% CI, 1.2 to 8.3) for CAD adjusted for several cardiovascular risk factors.[15] A recent matched Spanish case-control study found OSA (based on AHI ≥15) in 35% of patients with acute MI compared with 15% in the control group ($P < .001$).[16] The adjusted OR for acute MI was 12.2 (95% CI, 2.0 to 72.6), applying the AHI cutoff value of 15 for OSA diagnosis.

There are also data suggesting that the OSA and CAD association may be influenced by sex and age. In patients with angiographically verified CAD, an AI of greater than 10 was almost twice as common in men,[17] but three times more common in women[18] younger than 70 years, compared with age-matched controls. An uncontrolled study of 50 randomly selected CAD patients demonstrated OSA in 50% based on an AI of greater than 10.[19] Another uncontrolled German study reported an OSA prevalence of 35% applying AHI of 10 or more events/hour as the diagnostic criterion in 74 men with significant stenosis of one or more coronary arteries but failed to establish a significant relationship between AHI and number of coronary vessels involved.[20] A subsequent uncontrolled follow-up study found OSA (AHI ≥10) in 57% of 89 subjects with acute coronary syndrome undergoing percutaneous coronary intervention (PCI).[21] A similar high prevalence of OSA (66%) was reported in another investigation based on AHI of 10 or greater.[22] Two other uncontrolled studies performed in CAD patients undergoing PCI demonstrated OSA (AHI ≥15) in 43% and 66%, respectively.[23,24] One study based on a retrospective chart review of 798 consecutive patients with acute MI demonstrated that OSA was initially suspected only in 12% of the patient records, whereas after overnight polysomnography, 41% of patients presented an AHI of 15 or greater, suggesting that OSA was common but unrecognized in patients with CAD.[25] Moreover, baseline data of a randomized controlled trial (RCT) among 662 revascularized CAD patients in Sweden (the Randomized Intervention with CPAP in Coronary Artery Disease and Sleep Apnoea [RICCADSA] trial) revealed that 64% had an AHI of 15 or greater (Figure 128-1), but most did not report daytime hypersomnolence.[26] Of note, the occurrence of OSA in this cohort was more common than the prevalence of hypertension, diabetes, obesity, and current smoking. Finally, a similar prevalence (64%) of sleep-disordered breathing, defined as an oxygen desaturation index (ODI) of greater than 5 based on a WatchPAT-100 sleep study, was reported recently in prospective evaluation of 180 patients with acute MI.[27] However, no distinction was made between OSA and central sleep apnea or CSR in that cohort.

The possibility that OSA may trigger episodes of nocturnal angina in patients with disabling CAD was addressed in an interventional study.[28] OSA was found in 9 of 10 investigated patients with CAD who had nocturnal angina, and episodes of ischemia were reversed after elimination of the apneic

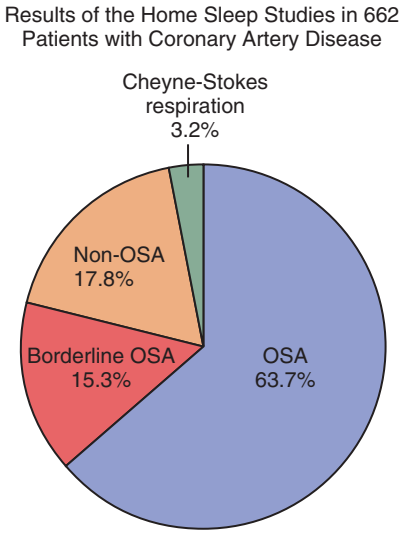


Figure 128-1 Classification of the groups based on the results of the unattended cardiorespiratory sleep recordings in patients with revascularized coronary artery disease. Obstructive sleep apnea (OSA) refers to apnea-hypopnea-index (AHI) of 15 or more events/hr; borderline OSA, AHI 5 to 14.9 events/hr; Non-OSA, AHI <5 event/hr. (Modified from Glantz H, Thunström E, Herlitz J, et al. Occurrence and predictors of obstructive sleep apnea in a revascularized coronary artery disease cohort. *Ann Am Thorac Soc* 2013;4:350–6.)

events with CPAP treatment. A subsequent larger cross-sectional study found signs of silent nocturnal myocardial ischemia in 31% of 226 patients with CAD but failed to demonstrate a general and immediate temporal relationship between OSA and episodes of myocardial ischemia.[29] However, a direct association could be documented in a small subgroup of patients, and, in general, episodes of silent ischemia appeared to be more frequent in those with more severe OSA. A retrospective evaluation of more than 200 patients undergoing electron-beam computed tomography within 3 years of an overnight sleep recording addressed the occurrence of subclinical coronary artery calcification.[30] With multivariate adjustment, the OR for coronary artery calcification was 3.3 in the most severe AHI quartile (mean, 63.4 events/hour).

The effect of OSA on the prognosis of CAD has been addressed in several studies. In a study of patients with CAD who were undergoing elective PCI, concomitant OSA was significantly related to increased late lumen loss and restenosis after an average follow-up time of 7 months.[31] In another study, the incidence of adverse cardiac events (cardiac death, reinfarction, and target vessel revascularization) was reported to be almost 24% among patients with OSA, compared with 5% among those without OSA, during a 6-month follow-up.[21] Moreover, patients with CAD who had concomitant OSA were found to have an increased risk for cardiovascular mortality over a 5-year period.[32,33] Conversely, another study demonstrated no effect of OSA on readmission rate of PCI-treated patients with CAD who had concomitant OSA during a 6-month follow-up period[22] and no significant difference regarding the 10-year survival rate for patients with CAD and with OSA compared with those without OSA at baseline.[34] Moreover, a recent clinical report (mentioned previously)[27] found no association between sleep apnea in the setting of acute MI and adverse clinical outcomes, including death, heart failure, and new MI, during a median follow-up of 68 months. On the other hand, the incidence of stroke was reported to be increased in patients with CAD who had concomitant OSA.[35] Stroke occurred in 18% of patients with CAD and sleep apnea, compared with 5% of those without sleep apnea, during 10 years of follow-up after a coronary angiography was performed. After adjustments for confounders, including hypertension and atrial fibrillation, the patients with sleep apnea had an adjusted HR of 2.9 (95% CI, 1.4 to 6.1) for a stroke.[35]

Hence OSA is common in patients with MI, but their mean AHI is relatively low in most published reports (Table 128-1).[13-15,17-22,27,36] Moreover, the prevalence of OSA is higher in patients with MI than in those with angina pectoris. This finding may be explained by the occurrence of Cheyne-Stokes breathing as a result of reduced ejection fraction.[19] Indeed the definition of the ideal timing for OSA screening after an acute MI remains unresolved. One study reported that 50% of CAD patients had an AHI of 15 or greater at the time of acute presentation in the coronary care unit, whereas 28% had remaining OSA based on the same AHI cutoff at least 6 weeks after hospital discharge.[37] A later study demonstrated occurrence of OSA (AHI ≥10) within the first 2 days after hospital admission in 54% of patients with acute coronary syndromes and preserved left ventricular ejection fraction, whereas 22 of 28 patients (79%) had residual OSA 1 month after the acute event, and only 6 of the 28 patients (21%) were diagnosed as having OSA at 6-month follow-up.[36] Thus OSA may be transient and, to some degree, related to the acute phase of the CAD, to more supine position of patients in the coronary care units, or to medications (sedatives and analgesics). Available studies accumulating some 2324 patients with CAD demonstrated an OSA prevalence of about 47% (see Table 128-1). Despite the differences in the diagnostic procedures (full polysomnography or cardiorespiratory sleep recordings) as well as varying cutoff values of AI, AHI, or ODI for definition of the sleep-disordered breathing, there seems to be enough evidence to advocate for overnight sleep recordings in individuals with MI given that concomitant OSA may worsen long-term outcomes in patients with CAD.

Incidence of Coronary Artery Disease in Longitudinal Studies

The first report on incident CAD data in a sleep clinic cohort was a smaller observational study demonstrating CAD development in almost one fourth of untreated patients with OSA during a 7-year follow-up period.[38] The corresponding numbers in treated sleep apnea patients and nonapneic snorers were 4% and 6%, respectively. Moreover, more than 50% of a normotensive cohort not treated for OSA developed at least one cardiovascular disease during the 7-year follow-up (Figure 128-2).[39] In that cohort, new CAD cases were also found among normotensive patients, suggesting that development of CAD in part may be independent of diurnal systemic hypertension induced by OSA. A larger observational study of a sleep clinic cohort, containing close to 1300 subjects with OSA, with a mean follow-up of 10 years, found a three to four times higher incidence of fatal and nonfatal cardiovascular events in patients with severe OSA compared with simple snorers.[40] Multivariate analysis showed that the risk for fatal cardiovascular events was significantly increased in severe untreated patients with OSA (OR = 3.2; 95% CI, 1.1 to 7.5) compared with healthy controls. Another prospective observational study of a sleep clinic cohort including 1436

Table 128-1 Prevalence of Obstructive Sleep Apnea in Patients with Coronary Artery Disease

Study	Patients (no.)	Sex	Prevalence (%)	Diagnostic Criteria (events/hr)	Controlled?
Hung et al., 1990[13]	101	Male	36	AI >5	Yes
Andreas et al., 1996[19]	50	Male, female	50	AI >10	No
Mooe et al., 1996[17]	142	Male	37	AHI ≥10	Yes
Mooe et al., 1996[18]	102	Female	30	AHI ≥10	Yes
Koehler & Schafer, 1996[20]	74	Male	35	AHI ≥10	No
Peker et al., 1999[15]	62	Male, female	31	AHI ≥10	Yes
Schafer et al., 1999[14]	223	Male	31	AHI ≥10	Yes
Skinner et al., 2005[37]	26	Male, female	50	AHI ≥15	No
Mehra et al., 2006[22]	104	Male, female	66	AHI ≥10	No
Nakashima et al., 2006[23]	86	Male, female	43	AHI ≥15	No
Yumino et al., 2007[21]	89	Male, female	57	AHI ≥10	No
Lee et al., 2009[24]	105	Male, female	66	AHI ≥15	No
Konecny et al., 2010[25]	74	Male, female	41	AHI ≥15	No
Schiza et al., 2012[36]	52	Male, female	54	AHI ≥10	No
Garcia-Rio et al., 2013[16]	192	Male, female	35	AHI ≥15	Yes
Glantz et al., 2013[26]	662	Male, female	64	AHI ≥15	No
Aronson et al., 2014[27]	180	Male, female	64	ODI >5	No
Total or mean	2324	—	47	—	—

AHI, Apnea-hypopnea index; AI, apnea index; ODI, oxygen desaturation index.

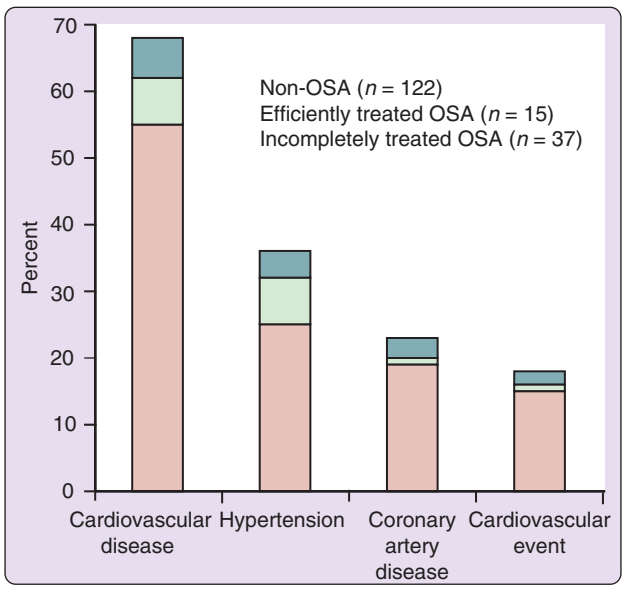


Figure 128-2 Incidence of cardiovascular disease during a 7-year follow-up in middle-aged men otherwise healthy at baseline. The fraction of individuals with incidence of cardiovascular disease, hypertension, coronary artery disease, and cardiovascular event (stroke, myocardial infarction, or cardiovascular death) is shown. Depicted are data from patients without OSA (Non-OSA) as well as from those incompletely or efficiently treated for their sleep and breathing disorder. (Modified from Peker Y, Hedner J, Norum J, et al. Increased incidence of cardiovascular disease in middle-aged men with obstructive sleep apnea: a seven-year follow-up. *Am J Respir Crit Care* 2002; 166:159–65).

consecutive subjects demonstrated that OSA (AHI ≥5) was associated with an increased risk for CAD events or death from cardiovascular causes (adjusted HR = 2.1; 95% CI, 1.1 to 3.9) during a follow-up period of almost 3 years.[41] Moreover, there was a dose-response relationship between AHI and composite outcome of CAD events or cardiovascular death (adjusted HR = 2.8; 95% CI, 1.5 to 5.5) in patients with severe OSA (AHI ≥30) compared with those without OSA (AHI <5).[41] Similarly, the 18-year follow-up study of the population-based Wisconsin Sleep Cohort sample reported an adjusted HR of 5.2 (95% CI, 1.4 to 19.2) for cardiovascular mortality in patients with severe OSA (AHI ≥30) and not using CPAP compared with those without OSA.[42] The longitudinal analysis of the Sleep Heart Health Study, including 1927 men and 2495 women free of CAD and heart failure at baseline, demonstrated a significant but weak association between severe OSA and incident CAD (adjusted HR of 1.7 for those with AHI ≥30 compared with those with AHI <5) in middle-aged men, but not in women.[43] In contrast to this population-based report, a recent observational follow-up study of 1116 women from two Spanish sleep clinic cohorts reported that untreated severe OSA was linked to increased cardiovascular mortality with an adjusted HR of 3.5 for those with AHI of 30 or greater compared to those with AHI of less than 10.[44] Another recent report from the same clinical cohorts demonstrated an adjusted HR of 2.8 (95% CI, 1.4 to 5.6) for the incidence of CAD or stroke in women with untreated OSA (AHI ≥10) compared with the control group without OSA.[45]

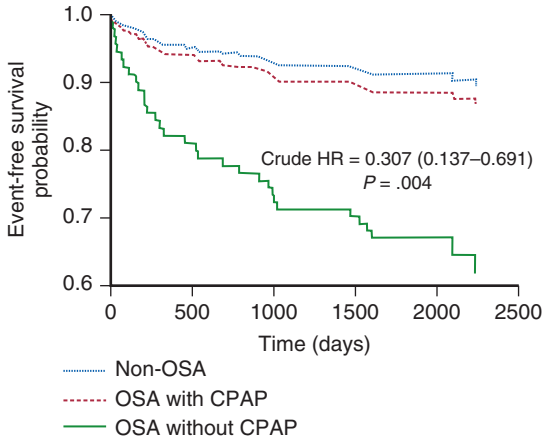
Effect of Obstructive Sleep Apnea Treatment on Coronary Artery Disease

The first-line treatment of OSA is CPAP, which is known to reduce daytime sleepiness and to improve quality of life.[46] A retrospective analysis of 55 subjects and comorbid OSA over an average follow-up time of 7.3 years showed a significantly lower occurrence of the composite end point of cardiovascular death, acute coronary syndrome, hospitalization for cardiac failure, or need for revascularization in those compliant with prescribed CPAP therapy.[47] In another follow-up study over 7.5 years, deaths from cardiovascular disease were less frequent in patients with OSA treated with CPAP compared with those without treatment.[48] Moreover, a review of 371 revascularized patients with OSA with concomitant CAD suggested a significantly lower cardiac death rate (3%) among 175 patients treated with CPAP compared with 10% among 196 untreated patients during a follow-up period of 5 years.[49] In the investigation of the sleep clinic cohort (mentioned earlier),[40] treatment with CPAP significantly reduced cardiovascular risk in men with severe OSA. Similar to that report, adequate CPAP treatment seemed to reduce the risk for composite end point of incident CAD or stroke in women with OSA.[45]

Plenty of evidence suggests that CPAP treatment reduces the number of ischemic events in patients with nocturnal angina and concomitant OSA in the short term.[28] A recent observational study addressed the effect of CPAP on recurrent episodes in patients with acute MI and concomitant OSA.[16] After adjustment for confounding factors, treated OSA patients who were compliant with CPAP had a lower risk for recurrent MI and revascularization (adjusted HRs = 0.16 and 0.15, respectively) than untreated patients and a similar risk to non-OSA patients (Figure 128-3). Of note, some patients with CAD and OSA may not experience daytime sleepiness (i.e., they are asymptomatic), and less is known regarding the adherence to CPAP therapy in these patients. However, a study of a sleep clinic cohort with concomitant CAD suggested a comparable compliance between sleepy and nonsleepy patients.[50] Observational studies might be criticized for potential bias because patients adhering to CPAP may have specific baseline characteristics, may be adhering to other medical therapy, or may otherwise have a healthier lifestyle. Observed benefits of CPAP in an observational study therefore provides a lower scientific evidence value compared with an RCT. Several RCTs examining the effect of CPAP on cardiovascular outcomes are underway.[51-54]

PATHOGENESIS

Obstructive sleep apnea is associated with considerable immediate hemodynamic change (see Chapter 119). During the cycle of the apneic event, there is increased work of breathing, considerable negative intrathoracic pressure swings, recurrent hypoxia and reoxygenation, and fluctuating autonomic activity (see Chapter 119). Heart rate and blood pressure also fluctuate through the cycle, but the relative contribution of each of these changes to development of cardiovascular disease is unknown. Increased oxygen demand and reduced oxygen supply (i.e., hypoxemia) after sleep-disordered breathing may trigger an attack of angina pectoris in patients with CAD, who already have reduced coronary flow reserve.[31] Nocturnal oxygen desaturations have been related to the severity of coronary atherosclerosis in patients with CAD[55] and may be an important contributor to coronary restenosis in patients with CAD who are treated with PCI.[56] Another study reported signs of apnea-induced ischemia predominantly during REM sleep,[8] a finding that may be explained by the often more prolonged and severe apneic events that commonly occur in this sleep stage. OSA is also associated with long-term alteration of cardiac structure, hemodynamic reflex function, and vascular structure or function. A recent report from the baseline echocardiographic investigations of the CAD patients



Patients at risk						
Non-OSA	63	60	59	57	31	12
OSA with CPAP	70	70	67	65	36	19
OSA without CPAP	52	38	35	35	17	9

Figure 128-3 Time until first recurrent myocardial infarction in the three groups of patients with coronary artery disease. Crude hazard ratio (HR) of treated versus untreated OSA is presented. CPAP, Continuous positive airway pressure; OSA, obstructive sleep apnea. (From Garcia-Rio F, Alonso-Fernandez A, Armada E, et al. CPAP effect on recurrent episodes in patients with sleep apnea and myocardial infarction. *Int J Cardiol* 2013;168:1328–35).

with preserved left ventricular ejection fraction in the RICCADSA cohort demonstrated a poorer diastolic function (OR = 1.9; 95% CI, 1.1 to 3.2) in patients with OSA (AHI ≥15) after adjustment for traditionally recognized risk factors.[57] The OSA is associated with immediate and sustained sympathetic activation.[58] Baroreceptor and chemoreceptor responsiveness is altered,[59] and vascular reactivity in terms of responsiveness to hypoxemia or vasoconstrictors appears to be elevated.[60] A series of studies demonstrated that vascular endothelial function, expressed in terms of nitric oxide vascular dilating capacity, appears to be reduced in OSA.[61] Changes are specific to OSA in the sense that they are reversed by CPAP (see Chapter 119).[62-66]

The mechanisms responsible for endothelial cell damage and dysfunction are not entirely understood. However, investigations have shown that oxidative stress, potentially as a result of periodic hypoxia and reperfusion, is enhanced.[67] Oxidative stress results in compromised nitric oxide bioavailability and leads to an activation of redox-sensitive gene expression. Ensuing steps in this chain of events include increased expression of adhesion molecules by an activated endothelium and leukocytes, which finally leads to acceleration of a vascular inflammatory cascade that promotes atherosclerosis and vascular dysfunction.

This hypothesis (see Chapter 117)[67] is supported by data from patients with OSA demonstrating increased free radical production,[63] increased plasma-lipid peroxidation, increased adenosine and uric acid levels,[67] and increased levels of redox-sensitive gene expression products, including vascular endothelial growth factor[68] and inflammatory cytokines.[69] Interestingly, there was an improvement of endothelial function in patients with OSA following inhibition of xanthine oxidase by allopurinol[70] or supplemental vitamin C.[71] Circulating levels of adhesion molecules[72] as well as adhesion molecule–dependent monocyte to endothelial cell avidity appear to be increased in OSA.[73] Finally, sleep apnea appears to provide an additive stimulus for adhesion molecule expression in patients with CAD.[74] Increased levels of circulating markers of inflammation, including tumor necrosis factor-α,[75] C-reactive protein (CRP),[69,76] and interleukin-6 (IL-6),[69] have been inconsistently found to be increased in OSA. A recent report from the Icelandic Sleep Apnea Cohort containing 454 untreated OSA patients (AHI ≥15) demonstrated that OSA severity was an independent predictor of levels of CRP and IL-6, but this association was found only in obese patients.[77] Another recent report from the RICCADSA cohort demonstrated an association between OSA and elevated CRP and IL-6 also in nonobese patients.[78] This suggests that established CAD may be associated with vascular inflammation in OSA patients irrespective of comorbid obesity and that determinants of these markers are, besides OSA, also influenced by multiple comorbid risk factors for cardiovascular disease.

However, given the knowledge about ischemic preconditioning as a cardioprotective maneuver for reducing experimental myocardial infarct size,[79] it has been proposed that intermittent nocturnal hypoxia in OSA may provide future protection against myocardial ischemic insults by the regulation of critical mechanisms in the coronary endothelium.[80] Indeed, a recent study in patients with acute MI demonstrated a greater mobilization of endothelial progenitor cells and increased endothelial growth factor expression in patients with mild to moderate sleep apnea.[81] Moreover, an observational study of 136 patients with an acute nonfatal MI found lower high-sensitivity troponin-T (hs-TnT) levels in patients with more severe OSA, suggesting a possible cardioprotective role by ischemic preconditioning in OSA.[82] On the other hand, there are data suggesting an independent association between increasing AHI and increasing hs-TnT levels in 1665 individuals without cardiovascular disease.[83] Over a median of 12.4 years' follow-up, hs-TnT was related to risk for death or incident heart failure in all OSA categories, suggesting that subclinical myocardial injury caused by OSA may play a role in the subsequent risk for cardiac disease.[83] Thus the short-term benefits of a possible ischemic preconditioning due to intermittent hypoxemia may be counteracted by other adverse outcomes in the long run.

The tentative association between OSA and CAD is supported by experimental data suggesting oxidative stress, endothelial dysfunction, and acceleration of vascular inflammation as a result of OSA. All these mechanisms facilitate the onset and progression of atherosclerosis. One study found signs of atherosclerosis in large arteries of OSA patients without other risk factors for CAD, and the severity of the signs of atherosclerosis was correlated to the severity of OSA.[84] In another study by the same researcher group, these signs of early atherosclerosis in OSA patients were responsive to CPAP.[85] Finally, a higher atherosclerotic plaque volume in the coronary vessels of the stable CAD patients with OSA was demonstrated using a three-dimensional intravascular ultrasound technique.[86] A more recent study found that the frequency of noncalcified or mixed plaques was much higher in patients with OSA than in non-OSA patients who were investigated by noninvasive coronary computed tomography angiography.[87] Thus atherosclerotic plaque formation may jeopardize coronary flow reserve and generate symptoms of nocturnal angina during periods of increased flow demand. Such episodes occur repeatedly in sleep apnea, and they are associated with hypoxemia that further enhances the vulnerability for ischemia. On the other hand, most heart attacks (i.e., acute MI, sudden cardiac death) stem from sudden rupture of less-obtrusive plaques, which triggers thrombus formation in coronary vessels.[88] As mentioned earlier, it is suggested that sleep-disordered breathing influences the circadian acute coronary event distribution. OSA may lead to a disproportionate number of events that occur during or soon after the sleeping period.

CLINICAL COURSE AND PREVENTION

Early recognition and treatment of OSA may be beneficial in terms of CAD prevention. A retrospective analysis of a sleep laboratory cohort followed over 7 years found a reduction (relative risk, 0.29; 95% CI, 0.10 to 0.82) of incident CAD in patients with effectively treated OSA compared with ineffectively treated or untreated patients.[38] On the other hand, in a group of patients with CAD followed for 5 years, mortality was higher in those with comorbid OSA (38%) than in those with no OSA (9%).[32] Although the higher mortality was in part explained by the presence of other traditional risk factors, there was an independent influence of the breathing disorder. Another study followed 408 patients with stable angina and angiographically verified CAD for 5 years after sleep apnea recordings. The risk for a cerebrovascular event,

including stroke and transient ischemic attack, was tripled in patients with CAD and concomitant OSA.[33,35]

Patients with CAD and nocturnal angina should be considered for sleep recording because nasal CPAP reduces angina attacks and nocturnal myocardial ischemia.[28] In the study of 10 severely disabled patients with a history of frequent nocturnal angina, 9 had sleep apnea.[28] Treatment with CPAP reduced episodes of nocturnal ischemia. There is no evidence to suggest that medication used for treatment of CAD affects the severity of the breathing disorder. A double-blind crossover study of nitrates in patients with OSA with or without CAD found lower oxygen saturation during apnea-associated ischemic episodes than during ischemia not associated with apnea (77.3% vs. 93.1%), and nitrate administration did not reduce the number of ischemic episodes associated with apnea.[89]

Although there is scientific support for a considerable effect of OSA on vascular structure and function, it is likely that development of CAD and other forms of vascular disease is determined by multiple genotypic and phenotypic factors. The absolute role of OSA in this concerted influence should evidently be better clarified. However, with the increasing recognition of OSA as an independent, additive, or even synergistic risk factor for CAD, we are facing a need for early identification of high-risk persons and a consensus on well-defined treatment strategies in such patients. Although more evidence is needed to address CPAP treatment as a useful therapy in patients with CAD and nonsymptomatic OSA, research in this field is growing rapidly, and the results from these ongoing RCTs[51-53] may soon add further insights.

CLINICAL PEARLS

- Recurrent apneas during sleep lead to a sequence of events that independently or in concert with other recognized risk factors are likely to have adverse effects on vascular structure and function.
- Not only may phenomena such as hypoxemia, reoxygenation, and recurrent vascular wall stress induce CAD, but also the events themselves may aggravate already-existing compromised coronary artery flow reserve.
- The adverse health effects of OSA in terms of CAD development, progression, and proneness to complications are likely to depend on genotypic and phenotypic factors. Markers or predictors for identification of high-risk persons in this context are still lacking.
- Almost 50% of patients with CAD have OSA defined according to conventional criteria. A large fraction of these patients do not exhibit daytime sleepiness. Additional data on different phenotypes of OSA as well as compliance with CPAP treatment, especially for nonsleepy patients with OSA and CAD, are needed.
- OSA identifies patients at risk for CAD and may represent a highly prevalent and modifiable risk factor.
- Recognition of the adverse effect of OSA on vascular disease will open a perspective of new primary and secondary prevention models for CAD that involve identifying and eliminating the sleep-disordered breathing.

SUMMARY

Recurrent apneas during sleep lead to a sequence of events that, independently or in concert with other recognized risk factors, are likely to have harmful effects on vascular structure and function. OSA-related phenomena, including hypoxemia, reoxygenation, and recurrent vascular wall stress, may induce CAD, and the events may aggravate already existing compromised coronary artery flow reserve. The epidemiologic support for a causal relationship between OSA and CAD is increasing but is not fully confirmed. This relationship is stronger in clinical cohorts than in the general population, which suggests that comorbid OSA in obese, hypertensive, smoking, and hyperlipidemic patients may provide an additive or synergistic risk factor for development of CAD.

Patients with CAD, including nocturnal angina, should therefore be considered for diagnostic sleep recording because elimination of apneas by nasal CPAP during sleep has been shown to reduce angina attacks and nocturnal myocardial ischemia. Moreover, prospective cohort data point to a reduction of recurrent myocardial infarction and revascularization in CAD patients treated with CPAP. The long-term tentative causal association between OSA and CAD is supported by experimental data suggesting endothelial dysfunction, acceleration of vascular inflammation, and development of atherosclerotic disease as a result of the breathing disorder. Increased recognition of the adverse effect of OSA on vascular disease may open a perspective of new primary and secondary prevention models for CAD that involve identification and elimination of the OSA.

Selected Readings

Arzt M, Hetzenecker A, Steiner S, Buchner S. Sleep-disordered breathing and coronary artery disease. *Can J Cardiol* 2015;**31**(7):909–17.

Franklin KA, Nilsson JB, Sahlin C, et al. Sleep apnea and nocturnal angina. *Lancet* 1995;**345**:1085–7.

Gami AS, Olson EJ, Shen WK, et al. Obstructive sleep apnea and the risk of sudden cardiac death. *J Am Coll Cardiol* 2013;**62**:610–16.

Garcia-Rio F, Alonso-Fernandez A, Armada E, et al. CPAP effect on recurrent episodes in patients with sleep apnea and myocardial infarction. *Int J Cardiol* 2013;**168**:1328–35.

Lutsey PL, McClelland RL, Duprez D, et al. Objectively measured sleep characteristics and prevalence of coronary artery calcification: the Multi-Ethnic Study of therosclerosis Sleep study. *Thorax* 2015;**70**(9):880–7.

Peker Y, Carlson J, Hedner J. Increased incidence of coronary artery disease in sleep apnoea: a long-term follow-up. *Eur Respir J* 2006;**28**:596–602.

Querejeta Roca G, Redline S, Punjabi N, et al. Sleep apnea is associated with subclinical myocardial injury in the community: the ARIC-SHHS study. *Am J Respir Crit Care Med* 2013;**188**:1460–5.

Thunström E, Glantz H, Fu M, et al. Increased inflammatory activity in nonobese patients with coronary artery disease and obstructive sleep apnea. *Sleep* 2015;**38**(3):463–71.

Valo M, Wons A, Moeller A, Teupe C. Markers of myocardial ischemia in patients with coronary artery disease and obstructive sleep apnea: effect of continuous positive airway pressure therapy. *Clin Cardiol* 2015;**38**(8):462–8.

A complete reference list can be found online at ExpertConsult.com.

Chapter 129

Heart Failure

Shahrokh Javaheri

Chapter Highlights

- Multiple studies from across the globe showing that about 50% of patients with heart failure, both heart failure with reduced ejection fraction and preserved ejection fraction, have moderate to severe sleep apnea with an apnea hypopnea index ≥15/hour.
- Both obstructive and central sleep apneas may occur concomitantly in the same patient. Therapy will depend, in part, on predominant apnea type, which is determined by polysomnography.
- Multiple studies also indicate that both obstructive and central sleep apneas are independently associated with readmission to the hospital and excess mortality. Furthermore, effective treatment has been shown to decrease the number of readmissions and premature mortality.
- For treatment of obstructive sleep apnea, continuous positive airway pressure therapy is the treatment of choice. Importantly, survival benefits are encountered in only those who are compliant with the device.
- For patients whose sleep apnea is not suppressed by continuous positive airway pressure, including almost 50% of patients with central sleep apnea, we recommend the use of adaptive servo-ventilation if LVEF is greater than 45%. Most of the studies using adaptive servo-ventilation have shown improvement in cardiac biomarkers, and readmission, but the impact on survival may be negative if LVEF is low.

Heart failure has been known for more than 2 centuries to be associated with abnormal breathing patterns, and John Cheyne and William Stokes have been credited for its description—hence the eponym *Cheyne-Stokes breathing*.[1,2] However, 37 years earlier, John Hunter,[3,4] a British physician, was the first to describe this breathing pattern, which is characterized by gradual crescendo–decrescendo changes in tidal volume, commonly with an intervening central apnea (Figure 129-1).[5-9] We therefore refer to this pattern as *Hunter-Cheyne-Stokes breathing* (HCSB). Periodic breathing is a pattern of breathing characterized by cyclic fluctuations in the amplitude of airflow and tidal volume.[10] It consists of recurring cycles of apnea or hypopnea, or both, followed by hyperpnea. The apneas and hypopneas may be obstructive (i.e., the result of upper airway occlusion) or central.[10] Obstructive sleep apnea (OSA)–hypopnea is the most common form of periodic breathing in persons without heart failure. However, in patients *with* heart failure, both obstructive and central periodic breathing are common and frequently occur together, although one phenotype is predominant.

HCSB is a form of periodic breathing with central sleep apnea (CSA) and hypopnea that occurs in patients with heart failure and has a long cycle time.[11] The latter is an important feature of HCSB breathing and reflects the prolonged circulation time that is a pathologic feature of heart failure. HCSB is a subjective description and is not readily quantifiable. For these reasons, the term *central sleep apnea* is preferable, and it also avoids misrepresentation, because credit for the discovery of breathing pattern has not been given to the original discoverer.

CSA observed in awake patients with heart failure has been considered an indicator of a terminal prognosis. However, like obstructive apnea, central apnea occurs primarily during sleep, and polysomnographic studies have reported a high prevalence of this disorder in ambulatory patients with stable heart failure.[8,9,12-14]

EPIDEMIOLOGY OF HEART FAILURE AND SLEEP-RELATED BREATHING DISORDERS

Heart failure has become a major public health problem.[15] In the United States, approximately 5.1 million adults 20 years and older have heart failure, and 0.8 million individuals are newly diagnosed each year.[15] Furthermore, it is projected that the prevalence of heart failure will increase 46% from 2012 to 2030, resulting in more than 8 to 9.5 million people ≥18 years of age with heart failure. It is estimated that heart failure may contribute directly or indirectly to about 280,000 deaths each year.[15] The death rate increases progressively with advanced symptomatology, with a 5-year survival rate approaching 50%.

Heart failure is the largest single Medicare expenditure because it is the leading cause of hospitalization for patients older than age 65 years. Annually, more than 1 million patients with heart failure need hospitalization. In 2012, total cost for heart failure was $30.7 billion dollars (Chapter 123).[15]

Left ventricular (LV) myocardial failure is the most common cause of heart failure in adults, and it could be predominantly diastolic, referred to as *heart failure with preserved ejection fraction* (HFpEF), or manifested by combined systolic and diastolic dysfunction, referred to as *heart failure with*

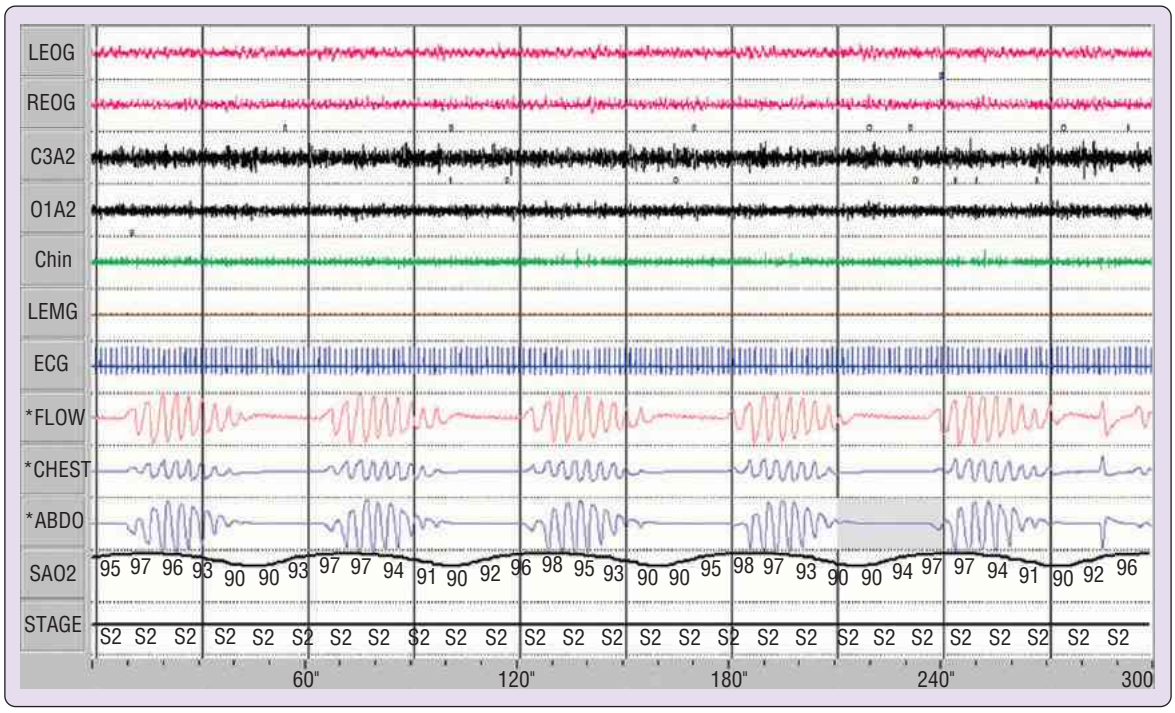


Figure 129-1 A 10-minute epoch of a patient with systolic heart failure and Hunter-Cheyne-Stokes breathing. Note recurrent hypoxia/reoxygenation as a result of central sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014.)

reduced ejection fraction (HFrEF). The principal hallmark of HFrEF is a depressed left ventricular ejection fraction (LVEF), which is commonly associated with an increase in end-diastolic and systolic volumes. The symptoms, which result from both diminished cardiac output and the concomitant diastolic dysfunction, include shortness of breath, orthopnea, nocturnal dyspnea, nocturia, fatigue, and exercise intolerance. Many of these symptoms overlap with those of sleep apnea, making it difficult to suspect sleep apnea when comorbid with heart failure.

Meanwhile, it is estimated that 20 million people may have asymptomatic LV systolic dysfunction, and, with time, these persons are likely to develop HFrEF. As will be discussed later, overall, about 50% of individuals with LV systolic dysfunction both asymptomatic or symptomatic suffer from moderate to severe sleep apnea.

The other phenotype, HFpEF, is the most common form of heart failure in elderly patients. The pathophysiologic consequence of LV diastolic dysfunction relates to a hypertrophied or a noncompliant left ventricle shifting the pressure volume curve upward and to the left. Therefore, for a given LV volume, LV end-diastolic pressure increases, resulting in elevated left atrial and pulmonary capillary pressures, pulmonary congestion, and edema. Similar to asymptomatic LV systolic dysfunction, which, with time, leads to HFrEF, asymptomatic LV diastolic dysfunction is also independently associated with incident overt heart failure and is predictive of all-cause death.[16]

Sleep Apnea in Heart Failure with Reduced Ejection Fraction

The prevalence of sleep-related breathing disorders has been systematically studied in patients with HFrEF.[5] Polysomnographic studies[6-14,17-25] and studies using respiratory channels[26-29] show a high prevalence of sleep apnea in this population. Most recent large studies of consecutive patients with HFrEF are depicted in Table 129-1.

High prevalence rates have been reported in patients who have HFrEF, HFrEF patients awaiting transplantation,[18] patients with valve heart disease,[19] and those with an implanted cardiac defibrillator.[20]

The most systematic prospective study of HFrEF[17] involved 100 ambulatory male patients with stable, treated heart failure. Using an apnea-hypopnea index (AHI) of 15 events per hour or greater as the threshold, 49 patients (49% of all patients) had moderate to severe sleep apnea–hypopnea, with an average AHI of 44. In comparison, a population study of subjects without heart failure[30] showed that 13% of working men and 6% women ages 30 to 70 years had an AHI of greater than 15. An AHI of 5 or greater has been used to define the presence of a significant number of disordered-breathing events in OSA–hypopnea syndrome.[23] Therefore, with a much higher prevalence of sleep apnea observed in patients with heart failure than in the general population, HFrEF should be the leading risk factor for sleep apnea in the general population. In our studies,[9,17] about 10% of the patients were on beta blockers; however, recent studies continue to show a high prevalence of sleep apnea, both central and obstructive, despite the widespread use of beta blockers. Table 129-1 shows the prevalence of central and OSA in the largest recent studies.[17,23-29] Combining the results of these recent series, which reported an AHI of 15 per hour of sleep as the threshold, 53% of 1607 patients with HFrEF had moderate to severe sleep apnea, 34% had CSA, and 19% had OSA (Figure 129-2). However, there has been considerable variation in the reported prevalence of these two forms of sleep

Table 129-1 Prevalence of Sleep Apnea in Recent Studies of Systolic Heart Failure

COUNTRY (year)	N	%AHI ≥15/h	%CSA	%OSA	%β Blockers
USA (06)* Javaheri	100	49	37	12	10
USA (08) McDonald	108	61	31	30	82
Canada (07)* Wang	218	46	21	26	80
UK (07)* Vazir	55	53	38	15	78
Germany (07) Oldenberg	700	651	32	19	85
Germany (09)* Hagenda	50	64	44	20	100
Germany (10)* Jilek	273	64	50	14	88
Portugal (10)* Ferreira	103	45	nr	nr	90
Total	1607	53	34	20	81

*In these studies, brain waves were not recorded.
AHI, Apnea–hypopnea index (the threshold used to define the presence of the disorder in each study); CSA, central sleep apnea; nr, not reported; OSA, obstructive sleep apnea.
Modified from Kryger MH. *Atlas of clinical sleep medicine*. 2nd ed. Philadelphia: Saunders; 2014.

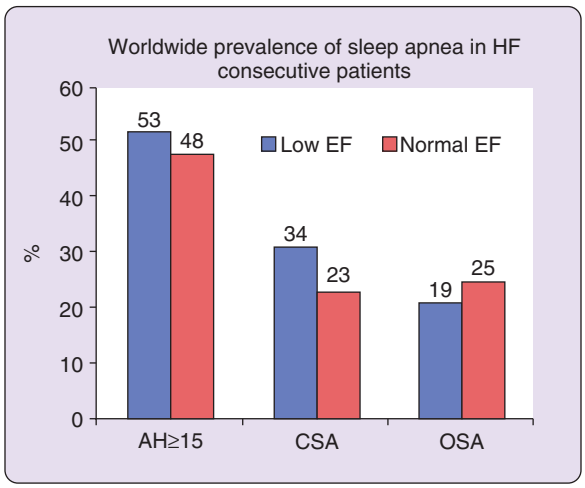


Figure 129-2 Prevalence of sleep apnea in heart failure with reduced ejection fraction (HFrEF) or preserved ejection HFpEF). The data presented combine a series of world sleep studies (see Table 129-1) in consecutive patients with HFrEF, and 244 consecutive patients with HFpEF.[31] AHI, Apnea–hypopnea index; CSA, central sleep apnea; OSA, obstructive sleep apnea. (From Kryger MH. *Atlas of clinical sleep medicine*. Philadelphia: Elsevier; 2014.)

apnea in patients with HFrEF (see Table 129-1), which depends on a number of issues. The major reasons are the criteria used to define hypopnea, the accuracy of classification of disordered-breathing events (obstructive versus central, particularly in regard to hypopneas), the criteria used to define predominant obstructive versus CSA, the number of obese patients with heart failure enrolled, the level of arterial Pco_2, and the severity of LV systolic dysfunction.

Sleep Apnea in Heart Failure with Preserved Ejection Fraction

There is also a high prevalence of sleep apnea in HFpEF.[31,32] The largest prospective study[31] to date evaluated 244 consecutive patients (87 women). All underwent polygraphy, right heart catheterization, and echocardiography. The two major causes of HFpEF were systemic hypertension (44%) and coronary artery disease (33%). Forty-eight percent had an AHI of 15 or more per hour, of whom 23% had CSA. Patients with CSA had lower Pco_2 but higher LV end-diastolic and pulmonary capillary wedge pressure. The latter finding is critical to the development of periodic breathing and CSA because increased wedge pressure and pulmonary congestion decrease the Pco_2 reserve, the major mechanism underlying CSA.[33] We must emphasize that there is a vicious cycle between HFpEF and sleep apnea. Hemodynamic studies show that pulmonary capillary pressure increases during the course of obstructive apnea, indicating the development of LV diastolic dysfunction (see Chapter 120). Chronic repetitive exposure to negative swings in intrathoracic pressure, cyclic nocturnal hypertension and hypoxemia, and diurnal systemic hypertension could eventually result in worsening the LV hypertrophy, dysfunction, and heart failure. In this regard, studies[34,35] suggest that OSA is associated with an increase in LV mass and dysfunction. In the largest study,[36] consisting of 2058 Sleep Heart Health Study participants, LV mass was associated with both apnea–hypopnea and hypoxemia indices after adjustment for age, sex, ethnicity, body mass index, smoking, systolic blood pressure, antihypertensive medication use, diabetes mellitus, prevalent myocardial infarction, and alcohol consumption. Furthermore, in an observational study,[34] treatment of OSA patients with CPAP resulted in a reversal of diastolic dysfunction. In another observational study,[35] an adaptive servo-ventilation (ASV) device was used to treat patients with HFpEF and HCSB and severe CSA. ASV treatment led to a significant decrease in left atrial diameter and early-to-atrial (E/A) filling velocity ratio, whereas the early filling to early diastolic mitral annular velocity ratio increased significantly. It therefore appears that treatment of both OSA and CSA results in remodeling of the left heart structures. These observational findings have been confirmed by a randomized placebo (sham CPAP)–controlled trial,[37] showing that, after 12 weeks on effective CPAP therapy, there was a significant increase in the E/A ratio, a significant decrease in isovolemic relaxation and mitral deceleration time. The results of these studies have important therapeutic implications for HFpEF because, to date, there are no approved therapies to reduce hospitalization or mortality for this disorder, which is on the rise, and the growing elderly population in whom HFpEF is the predominant phenotype of heart failure guarantees additional burden.[38] In addition, similar to asymptomatic LV

systolic dysfunction, which, as noted earlier, will eventually lead to HFrEF, it has been shown that LV diastolic dysfunction is a precursor to HFpEF.[16]

In summary, the prevalence of moderate to severe sleep apnea is about 50% in both forms of heart failure, HFrEF and HFpEF (Figure 129-2), and introduction of beta blockers in the therapeutic armamentarium of heart failure has had no impact on the prevalence of sleep apnea. In contrast to OSA (which is the predominant form of the disorder in the general population with a rare episode of central apnea in the pattern, in heart failure), central and OSA commonly occur together. Sleep physicians reviewing the polysomnogram, therefore, have to determine the predominant form of the disorder for therapeutic options. The predominant phenotype, obstructive versus central, is quite variable and, in large part, depends on the categorization of hypopneas into central or obstructive.

Sex and Sleep-Related Breathing Disorders in Heart Failure

In the general population, the prevalence of OSA is much higher in men than in women. This also holds true for CSA in HFrEF. Combining the results of several studies of patients with HFrEF,[12-14,18,39] 40% of the male patients and 18% of the female patients have CSA (Figure 129-3). A similar trend was found for OSA.

The results of population studies of subjects without heart failure (reviewed by Young and colleagues[40]) suggest that menopause may be a risk factor for OSA, and that the risk is probably reduced by hormone replacement therapy. In women with congestive heart failure and systolic dysfunction, the risk of CSA was six times higher in those ages 60 years and older than in those younger than 60 years.[12] A similar difference was also reported for OSA–hypopnea before and after age 60 years.[12] Thus, female hormonal status plays a role in the development of sleep-disordered breathing in women with and without heart failure.

Progesterone is a known respiratory stimulant, and its effects on the respiratory system may, in part, explain the lower prevalence of central and OSA in menstruating women. Progesterone increases ventilation[41] and the tone of the dilator muscles of the upper airway.[42] Furthermore, premenopausal women have a significantly lower apneic threshold than men.[43] This should decrease the probability of developing central apnea during sleep in female subjects (see the following section on mechanisms of CSA). In contrast to progesterone, administration of testosterone to premenopausal women results in diminution of the P_{CO_2} reserve,[44] which should increase the likelihood of developing apnea during sleep.[32] The results of these two studies suggest that the balance of progesterone/testosterone is critical in determining the P_{CO_2} reserve.

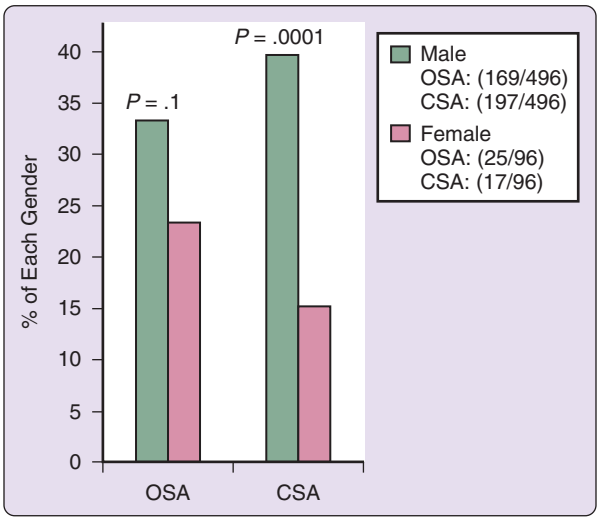


Figure 129-3 Prevalence of obstructive sleep apnea (OSA) and central sleep apnea (CSA) in men and women with systolic heart failure. The prevalence of CSA is much lower in women than in men. A similar trend is found in OSA, although it is not statistically significant. (From Javaheri S. Sleep related breathing disorders in heart failure. In: Mann DL, editor. *Heart failure: a companion to Braunwald's heart disease*. Philadelphia: Saunders; 2004. p. 471–87.)

MECHANISMS OF SLEEP-RELATED BREATHING DISORDERS IN HEART FAILURE

Mechanisms of Central Sleep Apnea in Heart Failure

The mechanisms of periodic breathing and CSA in heart failure are complex and multifactorial (see Chapter 100).[5,33,45,46] In heart failure, alterations occur in various components of the negative feedback system controlling breathing that increase the likelihood of developing periodic breathing, during both sleep and wakefulness. In addition, there are specific sleep-related mechanisms that explain the genesis of CSA and the reason periodic breathing becomes so prevalent during sleep.

Mathematical models of the negative feedback system predict that increased arterial circulation time (which delays the transfer of information regarding changes in P_{O_2} and P_{CO_2} from pulmonary capillary blood to the chemoreceptors, referred to as *mixing gain*), enhanced gain of the chemoreceptors, and enhanced plant gain (e.g., decreased functional residual capacity), which are the three components of the loop gain, collectively increase the likelihood of periodic breathing.[10,11,33,45-48]

Loop gain is the engineering term that defines the tendency of the negative feedback loop toward instability in response to a ventilatory disturbance. As an example, normally, a short pause in breathing, an apnea, or hypopnea causes a compensatory increase in ventilation. If the magnitude of the increase in ventilation is greater than or equal to the magnitude of the preceding respiratory disturbance, that is, loop gain is ≥1, the system becomes unstable and will fluctuate between underventilation and overventilation.

Delay in transfer of information due to prolonged circulation time, that is, increased mixing gain, plays a fundamental role in destabilization of a negative feedback system.[11,33,47] It has the potential to convert a negative feedback system to a positive feedback system. In heart failure, arterial circulation time may be increased for a variety of reasons, including dilation of cardiac chambers, increased pulmonary blood volume, and decreased cardiac output. However, patients with heart failure invariably have increased circulation time. Therefore, although increased circulation time is necessary to develop periodic breathing, it does not explain why only some heart failure patients have periodic breathing. The second component of the loop gain that increases the likelihood of occurrence of periodic breathing (and also central apnea during sleep) is the gain of the chemoreceptors.[8] In persons with increased sensitivity to CO_2 (or hypoxia), the chemoreceptors elicit a large ventilatory response whenever the P_{CO_2} rises (or the P_{O_2} decreases). The consequent intense hyperventilation,

by driving the P_{CO_2} below the apneic threshold, results in central apnea. As a result of central apnea, P_{CO_2} rises (and P_{O_2} falls), and the cycles of hyperventilation and hypoventilation (hypopnea) or central apnea are maintained.[7] Differences in the gain of the chemoreceptors among patients with heart failure may, in part, explain why only some patients with heart failure develop periodic breathing and CSA.

The third component of the loop gain that may contribute to the development of periodic breathing in heart failure is decreased functional residual capacity, which results in underdamping.[5,33,45] This means that, for a given change in ventilation (e.g., a pause in breathing), changes in the controlled variables—namely P_{O_2} and P_{CO_2}—will be augmented (referred to as *increased plant gain*). In turn, the augmented changes in P_{O_2} and P_{CO_2} result in a pronounced compensatory ventilatory response, and overcompensation tends to destabilize breathing. Patients with heart failure may have decreased functional residual capacity for a variety of reasons, including pleural effusion, cardiomegaly, and decreased compliance of the respiratory system. Functional residual capacity may decrease further in the supine position, facilitating the development of periodic breathing in this position.

The aforementioned mechanisms that collectively increase the loop gain and the likelihood of periodic breathing are present during both sleep and wakefulness. However, in the supine position and during sleep, further changes, such as a reduction in functional residual capacity, metabolic rate (another factor in the plant gain), and cardiac output, occur that will augment the likelihood of developing periodic breathing beyond that observed during wakefulness. Furthermore, loop gain, as described previously, differs from the dynamic loop gain during sleep when periodic breathing is present and steady state is absent.[33]

Meanwhile, like obstructive apnea, central apnea usually occurs during sleep or when a subject is awake but dozing. The genesis of CSA during sleep relates specifically to the removal of the nonchemical drive of wakefulness on breathing and to the unmasking of the apneic threshold—the level of P_{CO_2} below which rhythmic breathing ceases.[33] The difference between two P_{CO_2} set points—the prevailing P_{CO_2} minus the P_{CO_2} at the apneic threshold, referred to as *P_{CO_2} reserve*—is a critical factor for the occurrence of CSA. The smaller the difference, the greater the likelihood of occurrence of apnea.

Normally, with the onset of sleep, ventilation decreases and P_{CO_2} increases. As long as the prevailing P_{CO_2} is above the apneic threshold, rhythmic breathing continues. However, in some patients with heart failure, the awake prevailing P_{CO_2} does not significantly rise with onset of sleep.[49,50] Importantly, however, heart failure patients who develop central apnea have increased CO_2 chemosensitivity below eupnea[50] while asleep, as well as above eupnea while awake, as discussed previously.[48] Because of increased CO_2 chemosensitivity below eupnea,[50] the prevailing P_{CO_2} and the apneic threshold P_{CO_2} are close together, increasing the likelihood of developing central apnea during sleep. The increased chemosensitivity above eupnea becomes particularly pathophysiological during arousals occurring following apneas, when excessive ventilatory response lowers the prevailing P_{CO_2} toward or below the apneic threshold.[48]

The reason for the lack of the normally observed rise in P_{CO_2} in some patients with heart failure is not clear. It could result from the lack of the normally observed sleep-induced decrease in ventilation. Conceivably, because of increased venous return in the supine position, and in the presence of a stiff left ventricle, pulmonary capillary pressure could rise. This results in an increase in respiratory rate and ventilation, preventing the normally observed rise in P_{CO_2}. At the same time, the increase in pulmonary capillary pressure increases the chemosensitivity below eupnea and decreases the P_{CO_2} reserve, promoting the likelihood of developing central apnea. This was demonstrated in naturally sleeping dogs in whom pulmonary capillary pressure could be increased to different levels by inflating a balloon placed in the left atrium.[51] The mechanisms remain to be fully elucidated, although vagal afferents have been shown to have significant influences on the responsiveness of both carotid bodies as well as central chemoreceptors.[32] Several studies[52-54] have shown that patients with heart failure and low arterial P_{CO_2} have a high probability of developing central apnea during sleep. Predictive value of a low steady-state arterial P_{CO_2} (<35 mm Hg) is about 80%.[54] The reason for this association lies on the fact that a low arterial P_{CO_2} is caused by increased pulmonary wedge pressure, which, per se, sensitizes carotid bodies and the central chemoreceptors promoting CSA, as noted previously.[48] Meanwhile, although an awake low arterial P_{CO_2} is highly predictive of CSA, it is not a prerequisite. Many patients with heart failure and CSA have a normal awake arterial P_{CO_2}.[55] What is important is the proximity of the apneic threshold to the arterial P_{CO_2}, and an increased CO_2 chemosensitivity below eupnea.

Mechanisms of Obstructive Sleep Apnea in Heart Failure

As noted earlier, OSA and hypopnea are also common in heart failure. The mechanisms are multifactorial.[7,56-59] First, periodic breathing resulting from heart failure predisposes the susceptible subjects to develop upper airway occlusion during the nadir of the ventilatory cycles of periodic breathing.[7,56] For this reason, we observed multiple episodes of upper airway obstruction frequently following central apneas.[7]

Second, increased venous congestion and pressure resulting from right heart failure may diminish upper airway size[57] and facilitate upper airway occlusion. Venous congestion of the upper airway may be worse in the supine (than in the erect) position, and rostral fluid shift, particularly in the presence of edema in the lower extremities and redistribution of fluid into vascular space, may further compound upper airway patency.[58] Third, patients with heart failure and OSA are commonly obese, and obesity may compromise upper airway patency. While currently 35% of patients with HFrEF are obese, almost 53% of those with HFpEF suffer from this disorder[59] for which reason the prevalence of OSA is higher in patients with HFpEF than in HFrEF.[31,32]

In summary, decreased upper airway size resulting from both venous congestion and obesity may predispose patients with heart failure to develop upper airway occlusion during the nadir of the ventilatory cycles of periodic breathing, when the tone of the dilator muscles of the upper airway decreases the most.

PATHOLOGIC CONSEQUENCES AND PROGNOSTIC SIGNIFICANCE OF SLEEP-RELATED BREATHING DISORDERS

The cycles of apnea–hypopnea and hyperpnea, both obstructive and central, are associated with three adverse

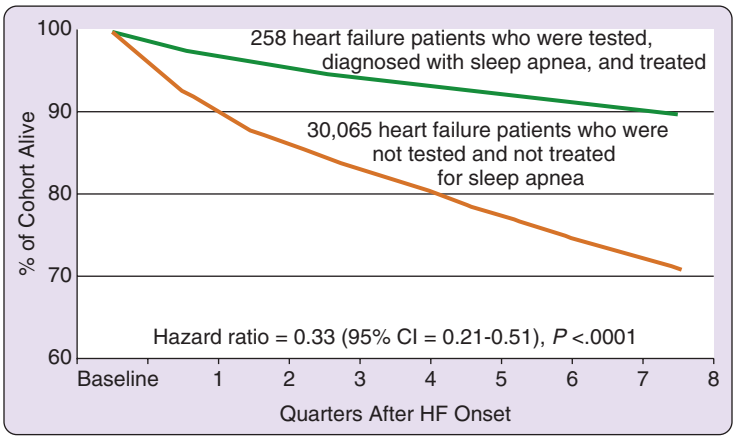


Figure 129-4 Data were from Medicare beneficiaries who were diagnosed with new heart failure. The 2-year survival of the 258 patients who were tested, diagnosed, and treated for sleep apnea was much better than the survival of the 30,065 patients who were not tested for sleep apnea. The survival was adjusted for age, gender, and Charlson Comorbidity Index. (Modified from Javaheri S, Caref B, Chen E, et al. Sleep apnea testing and outcomes in a large cohort of Medicare beneficiaries with newly diagnosed heart failure. *Am J Respir Crit Care Med* 2011;183:539–46.)

consequences. These include arterial blood gas abnormalities characterized by intermittent hypoxemia–reoxygenation and hypercapnia–hypocapnia, excessive arousals and shift to light sleep stages, and large negative swings in intrathoracic pressure (see Chapter 119). The pathophysiologic consequences of obstructive and CSAs and hypopneas are qualitatively similar (but worse in OSA than in CSA), adversely affect various cardiovascular functions, and are potentially most detrimental in the presence of established coronary artery disease and LV systolic and diastolic dysfunction. In the long run, these adverse consequences result in excess morbidity, hospital readmission, and mortality of patients with heart failure.

Effects of Obstructive Sleep Apnea on Sympathetic Activity, Cardiovascular Function, Hospital Readmission, and Mortality

In patients with heart failure, the presence of OSA is associated with increased sympathetic activity[60] and reduced LVEF, which are reversed if sleep apnea is effectively treated with nasal CPAP.[54-57] There are five randomized clinical trials[61-65] of CPAP therapy for OSA in patients with HFrEF. In three of these studies,[61-63] LVEF increased significantly (when compared with the control group) by about 10%, 5%, and 2%. In two[63,64] of these five studies, sham CPAP was used in the control group; in one,[63] LVEF increased significantly but slightly (2%), and, in the other one,[64] ejection fraction did not increase. In the latter study,[64] auto-CPAP was used, and the adherence hours to CPAP were less than in the two previous studies,[61,62] which had demonstrated 10% and 5% increases in ejection fraction. In the most recent study,[64] 45 patients with OSA (mean AHI = 27/hour of sleep) and HFrEF (mean LVEF = 36%) were randomized to CPAP (n = 22) or no CPAP (n = 23) for 6 to 8 weeks. Comparing the two groups, there were no significant changes in LVEF.

In patients with established coronary artery disease, OSA is an independent prognostic factor for recurrent cardiovascular disorders and survival.[66-67] Similarly, in heart failure, OSA is an independent predictor of mortality,[23] and two observational studies (Figure 129-4), one from Japan[68] and the other from the United States,[69] confirm this association and further

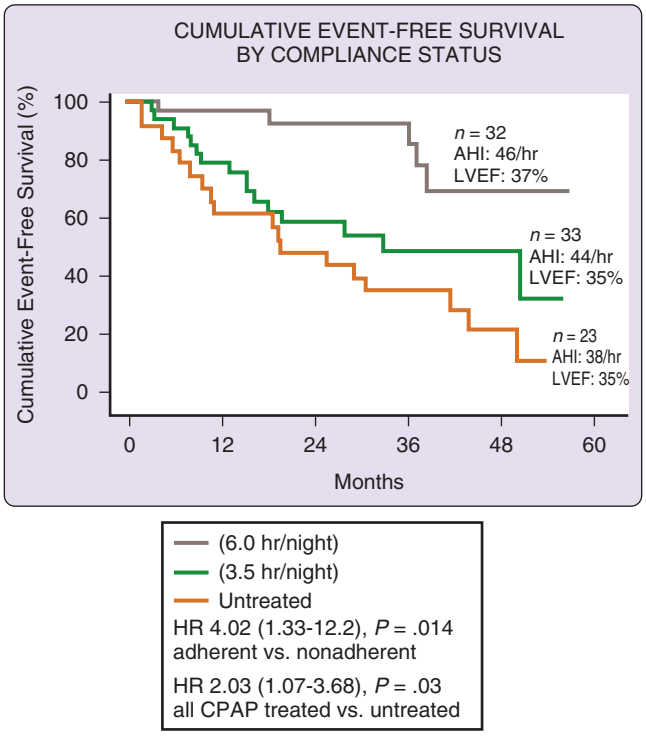


Figure 129-5 Probability for hospitalization and mortality of heart failure patients with obstructive sleep apnea decrease if they are treated with continuous positive airway pressure (CPAP) and adhere to therapy. AHI, Apnea–hypopnea index; HR, hazard ratio; LVEF, left ventricular ejection fraction. (From Kryger MH. *Atlas of clinical sleep medicine*. Philadelphia: Elsevier; 2010; modified from Kasai T, Narui K, Dohi P, et al. Prognosis of patients with heart failure and obstructive sleep apnea treated with continuous positive airway pressure. *Chest* 2008;133:690–6.)

suggest that therapy with CPAP improves survival, particularly in those who are most compliant with it (Figure 129-5).[68] This latter observation in patients with heart failure is similar to that in patients with hypertension, as studies indicate that the reduction in blood pressure is most prominent in those who are most adherent to CPAP therapy

(see Chapter 120). In the U.S. study,[69] a random sample of Medicare beneficiaries newly diagnosed with heart failure was enrolled. Of the 30,719 subjects, only 1263 (4%) were clinically suspected to have sleep apnea. Of these, 553 (2% of the total cohort) underwent sleep study, and some were treated mostly with CPAP. After adjustment for age, sex, and comorbidities, subjects with heart failure who were tested, diagnosed, and treated had a better 2-year survival rate compared with subjects with heart failure who were not tested (hazard ratio, 0.33 [95% confidence interval, 0.21–0.51]; $P = .0001$; Figure 129-4).

One important issue not previously emphasized is that, in patients with heart failure, OSA is independently associated with excess hospital readmission and that treatment of sleep apnea could lower the rate of readmissions. In the United States, Medicare began financially penalizing hospitals with excess readmissions in 2012, and, starting in October 2013, readmission penalties doubled to 2% of reimbursement. In one study from a heart hospital, Khayat and colleagues[70] demonstrated that severe OSA was independently associated with 1.5 times higher readmission when compared to heart failure patients without OSA. Importantly, these authors accounted for a large number of relevant pathologic variables, which otherwise could have associated with excess readmission. Meanwhile, two studies have shown that treatment with CPAP decreases the rate of readmission. In a study of Medicare beneficiaries with congestive heart failure, readmission costs for those treated for sleep apnea (mostly with CPAP) were much lower than for those suspected of having sleep apnea but who were not referred for sleep studies and, consequently, remained untreated.[69,71] Another retrospective study[72] of patients admitted to the hospital for a cardiac cause, including congestive heart failure who underwent polygraphy and found to have sleep apnea (AHI ≥ 5/hour, mostly OSA), hospital readmission or emergency department visit for a cardiac issue within 30 days of discharge was quite high in those who refused or were nonadherent to CPAP compared to those who adhered to CPAP therapy ($P = .025$).

Effects of Central Sleep Apnea on Sympathetic Activity, Cardiovascular Function, Hospital Readmission, and Mortality

Like OSA, CSA is associated with increased sympathetic activity and reduced LVEF, which is reversed by effective therapy with CPAP[73] and oxygen.

Several studies,[21,29,74-86] but not all,[84,85] have suggested that presence of CSA decreases survival among patients with HFrEF. In one[84] of the two studies[84,85] noted, there was a tendency for excess mortality in heart failure patients with CSA, although this was not significant, probably because of the small number of patients.

We followed 88 heart failure patients with ($n = 56$) or without ($n = 32$) CSA with a median follow-up of 51 months.[81] After controlling for 24 confounding variables, CSA was associated with excess mortality (hazard ratio, 2.14; $P = .02$; Figure 129-6). The average survival of heart failure patients without CSA was 90 months compared with 45 months for those with CSA. That CSA contributes to excess mortality in heart failure is supported by the observation that effective treatment of CSA with CPAP in HFrEF[80] and with ASV in HFpEF[87]

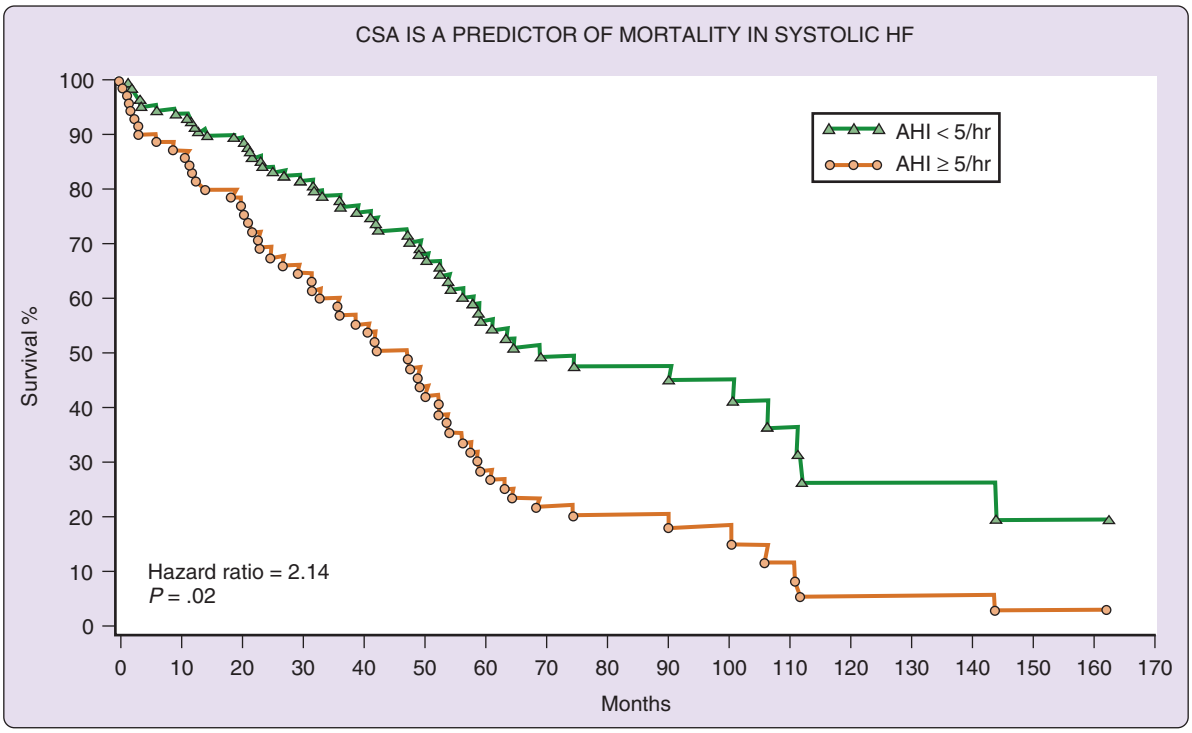


Figure 129-6 Probability of survival in patients with systolic heart failure (HF) according to the presence or absence of central sleep apnea (CSA). AHI, Apnea–hypopnea index. (From Kryger MH. *Atlas of clinical sleep medicine.* Philadelphia: Elsevier; 2010; modified from Javaheri S, Shukla R, Zeigler H, Wexler L. Central sleep apnea, right ventricular dysfunction, and low diastolic blood pressure are predictors of mortality in systolic heart failure. *J Am Coll Cardiol* 2007;49:2028–34.)

improves survival. The first study[80] was based on the post-hoc analysis of the Canadian randomized clinical trial, and the second one was a randomized trial involving the small number of patients with HFpEF.[87] The results of a relatively large observational study[29] of patients with predominantly CSA who agreed to use ASV are consistent with those of the aforementioned randomized trials.[81,87]

Similar to findings in OSA, CSA has been found to be an independent predictor of hospital readmission within 30 days after discharge. Khayat and colleagues[70] performed a prospective observational study of consecutive patients with HFrEF (LVEF = 22%), 165 with CSA and 139 without sleep apnea. In patients with CSA, the rate ratio for cardiac readmission within 1 or 6 months was 1.5 (P = .03) higher than patients without sleep apnea. The authors accounted for age, gender, body weight, blood pressure, coronary artery disease, hemoglobin, serum sodium and creatinine concentration, diabetes mellitus, and length of stay. Further, treatment of CSA has been shown to be associated with decreased hospital readmission.[88-91] Virtually all of these studies[88-91] are from Japan, are observational, the end points were a combination of premature mortality and readmission due to heart failure, and ASV devices were used to treat CSA, or when mixed with OSA.

CLINICAL PRESENTATION OF OBSTRUCTIVE AND CENTRAL SLEEP APNEAS IN PATIENTS WITH HEART FAILURE

Obesity is an important risk factor for development of OSA in patients with heart failure,[9,12,16] as it is for patients without heart failure.[29] Patients with HFrEF and OSA are significantly heavier and snore habitually (Figure 129-7). They may also have a higher systemic arterial blood pressure than subjects with CSA.[9,12] Aside from obesity and habitual snoring, it is often difficult to clinically suspect the presence of sleep apnea in patients with heart failure because (1) the prevalence of sleepiness is similar in heart failure patients with and in those without sleep apnea[9,16] (see Figure 129-7), and (2) the symptoms of heart failure and sleep apnea overlap. The overlapping symptoms of sleep apnea and heart failure include sleep-onset and maintenance insomnia, nocturia, waking up with shortness of breath (orthopnea, paroxysmal nocturnal dyspnea, hyperpnea due to periodic breathing), unrefreshed sleep, and daytime fatigue. The overlapping of the symptoms of heart failure and sleep apnea undoubtedly contributes to the underdiagnosis of sleep-related breathing disorders in patients with heart failure. CSA, in particular, is most difficult to diagnose,[8] because obesity and habitual snoring, which are the two hallmarks of OSA (see Figure 129-7), are commonly absent in heart failure patients with CSA.[8,9] However, there are some clues that, when present, should increase the probability of the presence of CSA. These include a high-numbered class in the New York Heart Association classification, low LVEF, and steady-state arterial Pco_2, atrial fibrillation, and nocturnal ventricular arrhythmias (Figure 129-8).[9]

Indications for Polysomnography in Heart Failure

As noted, patients with heart failure and sleep apnea do not generally present with symptoms that distinguish them from heart failure patients without sleep apnea. Furthermore, because heart failure is common, it is not possible to perform sleep studies on all patients with heart failure. However, there are a number of clinical and laboratory findings that, when present in patients with heart failure, should increase clinical suspicion for sleep apnea. These markers are different for obstructive and CSA.

Risk factors for OSA–hypopnea in patients with heart failure are similar to those in patients without heart failure. They include obesity, increased neck size, habitual snoring, and hypertension. These risk factors and others, such as

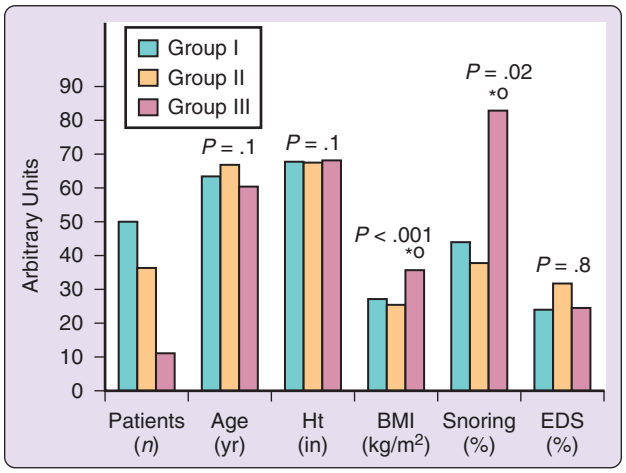


Figure 129-7 Demographics, historical data, and physical examination findings in heart failure patients without sleep apnea, with central sleep apnea (CSA), and with obstructive sleep apnea (OSA). Patients with OSA were more obese and had a higher prevalence of habitual snoring than patients with CSA. There was no difference in prevalence of excessive daytime sleepiness between the patients with heart failure and the patients without sleep apnea. BMI, Body mass index; EDS, excessive daytime sleepiness; Ht, height. (From Kryger MH. *Atlas of clinical sleep medicine*. Philadelphia: Elsevier; 2010; modified from Javaheri S. Sleep disorders in systolic heart failure: a prospective study of 100 male patients—the final report. *Int J Cardiol* 2006;106:21–8.)

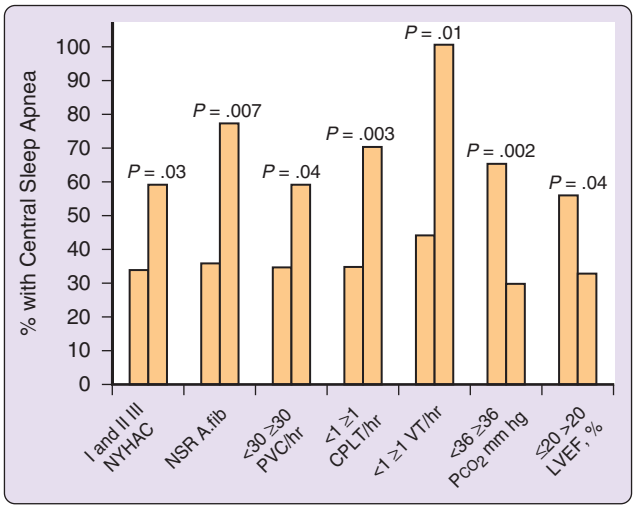


Figure 129-8 Clinical and laboratory characteristics that are more likely to be associated with central sleep apnea. A. fib, Atrial fibrillation; CPLT, couplets; LVEF, left ventricular ejection fraction; NSR, normal sinus rhythm; NYHAC, New York Heart Association Class; PVC, premature ventricular contractions; VT, ventricular tachycardia. (From Javaheri S. Sleep disorders in systolic heart failure: a prospective study of 100 male patients—the final report. *Int J Cardiol* 2006;106:21–8.)

witnessed apnea, waking up unrested, and excessive daytime sleepiness, when present, should increase the level of suspicion for the presence of OSA. The following are symptoms that should alert the clinician to the possibility of apnea in heart failure patients:

Nocturnal angina—substernal chest pain that awakens the patient—should increase suspicion for sleep apnea in the general population and for patients with coronary heart disease and heart failure.

Paroxysmal nocturnal dyspnea characteristically awakens the patient with shortness of breath, which is relieved with resumption of an erect position. However, this symptom may be a perception of shortness of breath occurring during the hyperpneic phase of periodic breathing, suggesting presence of sleep apnea.

Restless sleep, maintenance insomnia, and *leg movements* may reflect periodic arousals and movements after apneas and hypopneas. Periodic limb movement, however, is also found in patients with systolic heart failure.[16,92,93]

Patients with heart failure and progressive ventricular systolic or diastolic dysfunction or patients who remain in New York Heart Association classes III or IV, despite intensive medical therapy, should have a diagnostic sleep study.

The prevalence of sleep apnea is high in patients with an implanted cardioverter or defibrillator,[20] and those awaiting cardiac transplantation.[17] The waiting period for transplantation is long, and a large number of patients succumb to the consequences of heart failure while waiting. It is conceivable that survival of these pretransplant patients may improve if their sleep apnea is diagnosed and appropriately treated. If so, the chance of receiving cardiac transplantation may increase.

As noted earlier, several studies[52-54] have shown that heart failure patients with low arterial Pco_2 have a high prevalence of CSA. The predictive value of low Pco_2 (<35 mm Hg) is about 80%.[54] However, many patients with heart failure have CSA apnea without daytime hypocapnia.[54,55]

Several studies have shown that heart failure patients with sleep apnea have a higher prevalence of atrioventricular arrhythmias, especially atrial fibrillation[9,12,16,94] and nocturnal ventricular arrhythmias.[9,16,95] Presence of these arrhythmias should increase suspicion for the presence of CSA.

In the presence of the aforementioned risk factors for obstructive and CSA, polysomnography should be performed for diagnosis and response to therapy. Such an approach has been shown to decrease hospital readmission and improve survival, as discussed previously.

TREATMENT OF SLEEP-RELATED BREATHING DISORDERS IN PATIENTS WITH HEART FAILURE

The choice of therapy for obstructive or CSA is based on the type of sleep apnea.[96]

TREATMENT FOR OBSTRUCTIVE SLEEP APNEA

In general, treatment of OSA–hypopnea is similar in patients with and without heart failure, although there are some differences (Box 129-1). In the presence of cardiovascular disease, every attempt should be made to treat OSA with positive airway pressure devices.

Box 129-1 TREATMENT OF OBSTRUCTIVE SLEEP APNEA IN PATIENTS WITH HEART FAILURE OPTIMIZATION OF CARDIOPULMONARY FUNCTION

- To eliminate or improve periodic breathing
- To decrease right atrial and central venous pressure upper airway congestion/edema, which may increase upper airway size
- To improve functional residual capacity, which may increase upper airway size as lung volume increases
- Avoidance of benzodiazepines, opioids, alcoholic beverages, and Viagra
- Weight loss if applicable
- Nasal positive airway pressure devices:
  - Continuous positive airway pressure (CPAP)
  - Bilevel positive airway pressure (see Chapter 107)
- Oral appliances (see Chapter 109):
  - Supplemental nocturnal nasal oxygen to minimize desaturation and to decrease periodic breathing
- Upper airway procedures:
  - Uvulopalatopharyngoplasty (see Chapter 108)
  - Laser surgery (see Chapter 108)
  - Radiofrequency volume reduction (see Chapter 108)

Optimization of Cardiopulmonary Function

Optimal treatment of heart failure by improving both periodic breathing and lower extremity edema may decrease the likelihood of developing upper airway occlusion. Upper airway narrowing and occlusion may occur at the nadir of the ventilatory cycle of periodic breathing,[56] and, in some patients with heart failure, the first few breaths after central apneas are obstructed.[7] Furthermore, in biventricular heart failure, elevated right atrial and central venous pressure may result in pharyngeal congestion and edema, which along the fluid from lower extremities translocated cephalad in supine position could result in narrowing of the upper airway. Therefore, therapeutic measures to decrease the lower extremity edema and venous pressure[57,58] are advisable. Also, optimal treatment of heart failure to decrease lung water and pleural effusion could increase lung volumes, which should increase upper airway size, which is dependent on lung volume.

Weight Loss

In the general population, obesity is a major risk factor for OSA, and weight reduction improves OSA (see Chapter 106). Similarly, obesity is associated with increased risk of a new onset cardiovascular disease, including heart failure.[97,98] In spite of these aforementioned relationships between obesity, OSA, and heart failure in the general population, studies of patients with heart failure have consistently demonstrated an obesity paradox, indicating that obesity is a strong independent predictor of improved outcomes for patients with chronic heart failure. However, many patients with heart failure and OSA are obese,[9,12,16] and OSA per se has been shown to be a risk factor for the development of heart failure.[99] For this reason, we generally advise weight loss for obese patients with OSA, irrespective of heart failure. However, studies are needed to determine optimal body weight and whether purposeful weight loss in congestive heart failure comorbid with OSA improves cardiac function.

Avoidance of Alcoholic Beverages, Benzodiazepines, and Phosphodiesterase-5 Inhibitors at Bedtime, and Smoking

The use of alcoholic beverages and benzodiazepines may increase the likelihood of upper airway occlusion by promoting the relaxation of the muscles of the upper airway. We also advise patients that phosphodiesterase inhibitors used to treat erectile dysfunction (e.g., sildenafil [Viagra], vardenafil [Levitra], and tadalafil [Cialis]) may worsen OSA. In a randomized double-blind placebo-controlled study,[100] it was shown that 50 mg of sildenafil significantly increased the obstructive AHI and desaturation in a group of patients with OSA.

Smoking, via mechanisms mediated by nicotine, the active chemical in tobacco, increases efferent sympathetic activity (at least, in part, due to stimulating peripheral chemoreceptors in the carotid bodies, which contain excitatory nicotinic receptors) and plasma catecholamine resulting in increases in blood pressure, heart rate, and myocardial oxygen consumption.[101] In addition, nicotine decreases oxygen availability and causes coronary vasospasm, all promoting ventricular tachyarrhythmia. In a recent study[101] of 87 patients with HFrEF, smoking was associated with nocturnal ventricular tachycardia with an odds ratio of about 10. This is not surprising because excessive adrenergic overactivity is the underlying mechanism of arrhythmias and heart failure, in particular, when comorbid with sleep apnea is already a hyperadrenergic state.

Positive Airway Pressure Devices

Positive airway pressure devices are the treatment of choice and have been most successfully used to treat OSA in the general population and in patients with heart failure. First-night application of nasal CPAP results in a significant decrease in disordered breathing, arterial oxyhemoglobin desaturation, and arousals.[102] Short-term use of CPAP in patients with heart failure and OSA improves LVEF, blood pressure, and ventricular systolic volume,[61-63] but adherence to CPAP is a critical factor. For CPAP-noncompliant subjects who complain of a high expiratory pressure, bilevel pressure devices should be tried. As noted earlier, two recent observational studies[68,69] of patients with heart failure have shown that effective treatment of OSA with CPAP improves survival, particularly in those who are compliant with CPAP (see Figures 129-4 and 129-5).[68]

Supplemental Nasal Oxygen

For subjects with heart failure who cannot tolerate positive air pressure devices, oxygen is an alternative for treating OSA. The rationale for use of nocturnal supplemental nasal oxygen is to improve both hypoxemia and periodic breathing. Minimizing desaturation and hypoxemia–reoxygenation may have important therapeutic implications. Furthermore, as noted earlier, improvement in periodic breathing may decrease in obstructive disordered-breathing events that occur at the nadir of ventilation. We emphasize, however, that there are no systematic studies treating OSA of patients with heart failure with oxygen.

Upper Airway Surgical Procedures

Upper airway surgical procedures are performed for treatment of OSA in the general population, but there are no data in patients with heart failure.

Oral Appliances

Oral appliances are used to treat OSA, particularly in patients who cannot tolerate CPAP (see Chapter 109). Limited data are available in heart failure.[103] We speculate that efficacy of these devices in heart failure patients with OSA is similar to that in the general population. After application, a sleep study is recommended to ensure effectiveness.

TREATMENT FOR CENTRAL SLEEP APNEA

Figure 129-9 shows our approach to treatment of patients with CSA in heart failure.

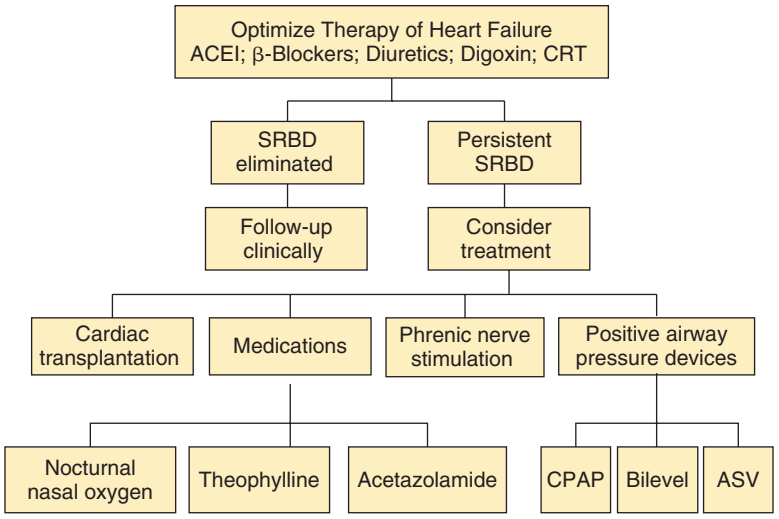


Figure 129-9 Treatment of central sleep apnea in patients with systolic heart failure. ACEI, Angiotensin-converting enzyme inhibitor; APSSV, adaptive pressure support servo-ventilation; CRT, cardiac resynchronization therapy; nCPAP, nasal continuous positive airway pressure; PAP, positive airway pressure; SRBD, sleep-related breathing disorder. (Modified from Javaheri S. Sleep-related breathing disorders in heart failure. In: Mann DL, editor. *Heart failure: a companion to Braunwald's heart disease*. Philadelphia: Saunders; 2004. p. 482.)

Optimization of Cardiopulmonary Function

Intensive therapy for heart failure with diuretics, angiotensin-converting enzyme inhibitors, beta blockers, and cardiac resynchronization therapy (CRT) can improve periodic breathing.[5] Because pulmonary congestion and edema are associated with narrowing of P_{CO_2} reserve,[51] reduction in wedge pressure should be associated with widening of P_{CO_2} reserve and improvement in CSA. Furthermore, with therapy, arterial circulation time decreases (as stroke volume increases and cardiopulmonary blood volume decreases), and functional residual capacity may increase (because of a decrease in cardiac size, pleural effusion, and intravascular and extravascular lung water). These changes contribute to the stabilization of breathing.

Beta blockers, by increasing stroke volume and decreasing pulmonary capillary pressure, should be particularly helpful in improving periodic breathing in systolic heart failure. An additional beneficial effect of beta blockers may relate to their counterbalancing of nocturnal cardiac sympathetic hyperactivity, resulting from repetitive arousals and desaturation. The reduction in cardiac sympathetic activity may have contributed to improved survival in trials of beta blockers in patients with heart failure. One particular side effect of beta blockers, however, is related to their effect on melatonin. Melatonin, a sleep-promoting chemical, is secreted via the cyclic adenosine monophosphate–mediated beta-adrenergic signal transduction system. Some beta blockers (exceptions include carvedilol), by inhibiting this process, decrease melatonin secretion[104,105] and could potentially contribute to worsening of sleep.

Regarding improvement in cardiac function and CSA, a few studies of CRT[106-110] showed some improvement in CSA, particularly noticeable in those with CRT-induced hemodynamic improvement. However, CRT devices are ineffective for OSA; although, in one study,[106] OSA improved, and improvement correlated with a decrease in circulation time.

If periodic breathing persists after cardiopulmonary function is optimized, several approaches are possible (see Figure 129-9).

Cardiac Transplantation

Preliminary studies, reviewed by Javaheri,[5] and a study of 45 patients with heart failure[111] have shown that, after cardiac transplantation, CSA is virtually eliminated. However, with time, a large number of cardiac transplant recipients develop OSA.[92] In this study, 36% had an AHI of 15 per hour or greater.[111] OSA developed in those who had gained the most weight after transplantation, and it was associated with habitual snoring, poor quality of life, and systemic hypertension. Cardiac transplantation was also associated with a high prevalence of restless legs syndrome and periodic limb movements (Figure 129-10).[111]

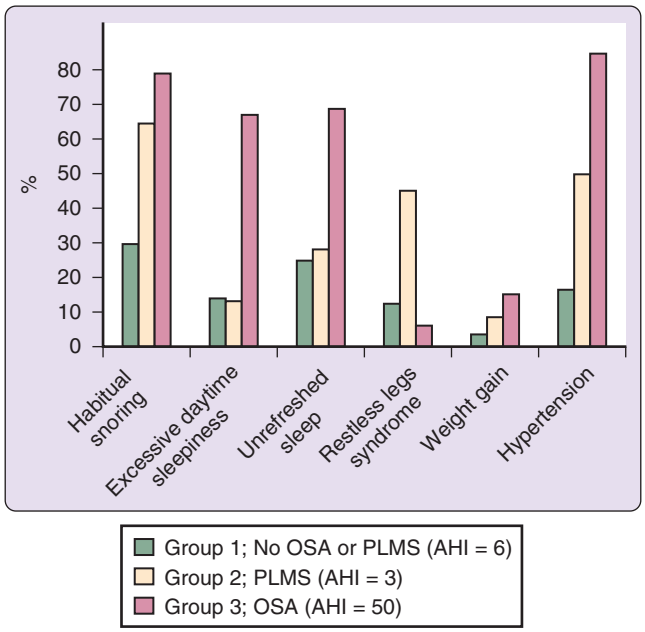


Figure 129-10 Phenotype of patients after heart transplantation. Group 1 did not have obstructive sleep apnea (OSA) or periodic limb movements during sleep (PLMS), Group 2 had PLMS, and Group 3 had OSA. AHI, Apnea–hypopnea index. Weight gain is in kilograms since the transplantation. (Modified from Javaheri S, Abraham W, Brown C, et al. Prevalence of obstructive sleep apnea and periodic limb movement in 45 subjects with heart transplantation. *Eur Heart J* 2004;25:260–6.)

Positive Airway Pressure Devices

Continuous Positive Airway Pressure

Several devices, including CPAP, bilevel pressure, and ASV, have been used to treat CSA in patients with heart failure.

In contrast to treatment of OSA, where application of nasal CPAP invariably results in virtual elimination of obstructive disordered-breathing events, treatment of CSA in patients with heart failure is difficult, and response to therapy is not uniform.[112] In our study,[102] first-night CPAP titration was effective in improving CSA in 43% of the patients (57% were considered CPAP nonresponsive). In the multicenter Canadian trial,[113] 47% of the patients were considered CPAP nonresponsive at 3 months. In this trial,[113] 132 patients were randomized to the control group and 128 to the CPAP arm. The baseline features were similar in the two randomized groups. The patients in the therapeutic arm were adapted to CPAP over 1 to 3 nights (without formal titration), and the maximum pressure was set at 10 cm H_2O or lower at whatever was tolerated.

A second polysomnography was performed at 3 months in both groups. There was no significant change in AHI in the control group. In the CPAP arm, the average AHI decreased by 50%, with considerable improvement in desaturation. Furthermore, the average plasma norepinephrine level decreased, and LVEF increased (all statistically significant). These measurements remained unchanged in the control group. However, after an interim analysis was performed, the safety-monitoring committee recommended termination of the study. This, in part, was related to worsened transplantation-free survival (primarily due to increased number of deaths from progressive heart failure and sudden death) of the CPAP-treated patients compared with the control group ($P = .02$). Although the survival curves diverged after about 3 years (favoring the CPAP arm), the difference was not statistically significant ($P = .06$).

We speculated that CPAP therapy could have resulted in excess early mortality for several reasons, including the following[114]: (1) Those who died were heart failure patients with CSA, whose periodic breathing was CPAP nonresponsive; and (2) those who died were heart failure patients whose ventricular function (according to the Frank-Starling curve) was preload dependent.

If right and LV function is preload-dependent, any reduction in venous return by the increased intrathoracic pressure, with application of CPAP, could decrease right ventricular stroke volume and return to the left ventricle, decreasing LV stroke volume and causing hypotension, diminished coronary blood flow, myocardial ischemia, and arrhythmias. Any such effect of CPAP on blood pressure is further augmented during sleep when blood pressure normally decreases.

The aforementioned assumptions that, in the Canadian trial, excess cardiovascular death due to CPAP occurred primarily in CPAP nonresponders were confirmed by a post-hoc analysis of mortality of patients with CSA (Figure 129-11).[80] In those whose CSA responded to CPAP, transplantation-free survival was significantly improved when compared with the untreated control group. In addition, the mortality of CPAP nonresponders appeared to be the worst, although the number of patients was small for statistical significance.

In the Canadian trial[113] at 3 months, 43% of heart failure patients with CSA were CPAP nonresponsive, compared with 57% in our study[102] of first-night use. Typically, CPAP-responsive patients had less severe CSA than CPAP-nonresponsive patients. In our study,[102] in CPAP-responsive patients, the average AHI decreased from 36 to 40 per hour, with elimination of desaturation. An important observation was that the number of premature ventricular contractions, couplets, and ventricular tachycardia decreased. This effect was presumed to result from decreased sympathetic activity, because arousals decreased and saturation improved. Heart failure patients with severe CSA (57% of the patients) did not respond to CPAP, and use of CPAP had no significant effect on ventricular irritability.

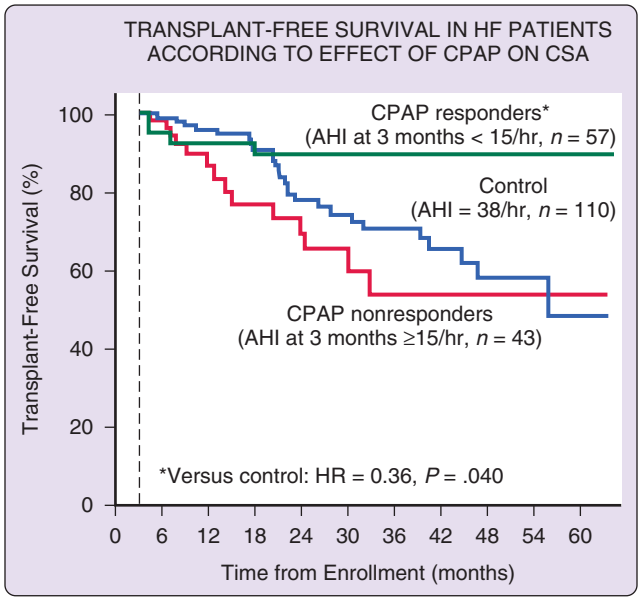


Figure 129-11 Probability of survival in patients with systolic heart failure (HF) comparing continuous positive airway pressure (CPAP) responders to a control group (patients with systolic heart failure and similar apnea–hypopnea index [AHI]) and to CPAP nonresponders. CPAP responders had a significantly increased probability of survival compared with the control group. CPAP nonresponders tended to have a poor survival when compared with the control group, although this was not significant. CSA, Central sleep apnea; HR, hazard ratio. (From Kryger MH. *Atlas of clinical sleep medicine.* Philadelphia: Elsevier; 2010; modified from Artz M, Floras JS, Logan AG, et al. Suppression of central sleep apnea by continuous positive airway pressure and transplant-free survival in heart failure. *Circulation* 2007;115:3173–80.)

The mechanisms of improvement in CSA in CPAP-responsive patients are multifactorial. One may relate to improvement in pulmonary congestion, which should widen Pco_2 reserve.[51] CPAP may improve stroke volume by decreasing LV afterload, which decreases arterial circulation time. CPAP should increase functional residual capacity, which decreases underdamping, and CPAP opens the upper airway, which is a benefit for those patients with CSA in whom upper airway closure occurs.

Adaptive Servoventilation

However, as noted earlier, 43% to 57% of heart failure patients with CSA are nonresponsive to CPAP. For these patients, and those who are intolerant to CPAP, until recently ASV devices were recommended (see end of this section).[72] These devices provide varying amounts of anticyclic inspiratory pressure support during different phases of periodic breathing, augmenting ventilation when the patient's minute ventilation decreases below a target and withdrawing support when the patient's ventilation is above the target.[112] In this way, periodic breathing is eliminated while on the device. In addition, the device initiates a breath on a timely basis, preventing development of a central apnea. Finally, the new generation of these devices is equipped with automatic end expiratory positive pressure algorithms[71,112,115] that operate to eliminate obstructive disordered breathing events. Having this virtue, these devices were expected to be advantageous for treatment of complex sleep-related breathing disorders when both CSA and OSA and hypopneas are present. Specifically in patients with heart failure, such complex breathing events are frequently observed during polysomnography. In addition, the phenotype of sleep apnea may vary during progression of heart failure and during acute decompensation when excess fluid from the lower extremities translocate to the neck area in supine position. Under such circumstances, upper airway obstruction could occur,[57,58] and a fixed end expiratory positive airway pressure with either a CPAP or bilevel device is inadequate to eliminate obstructive events.

ASV devices have been used to treat CSA in patients with congestive heart failure, generally with favorable results. In an acute (1-night) study[116] in 14 subjects with HFrEF and CSA, the ASV device decreased the AHI more than oxygen, CPAP, and bilevel devices. Indeed, in a meta-analysis of patients with congestive heart failure, in whom ASV therapy was compared to a control arm (the control arm could be medical therapy, oxygen, CPAP or bilevel), we concluded that ASV was much more effective in reducing AHI.[117] Specifically, in crossover studies, the mean AHI decreased from about 50/hour to 6/hour, compared to 20/hour. Multiple long-term studies with ASV have been reported, and these have been reviewed recently.[112,117] In general, most (but not all) of these studies show significant improvement in biomarkers of heart failure, LVEF, and reduction in primary end points, commonly a combination of mortality and readmission to the hospital, as noted previously.[83,88-91,118]

In a prospective parallel design, randomized controlled trial[119] from Norway, 51 patients with congestive heart failure and CSA, ages 57 to 81 years, were randomized to either an ASV or a control group; 30 patients completed the study (15 from each group). Three months treatment with ASV significantly improved LVEF from 32% to 36%, 6-minute walk, and a New York Heart Association classification. These

variables did not change significantly in the control group. These results were somewhat different from a similarly designed German study,[120] in which significant but equal increments in LVEF occurred after 3 months, both in the ASV arm and the control group. However, consistent with the results of the previous study,[119] reduction in N-terminal pro-brain natriuretic peptide was significantly greater in the ASV arm. However, in our meta-analysis[119] discussed previously, ASV therapy significantly improved LVEF when compared to the control. Furthermore, in a recent well-designed trial,[120] in which 23 patients with persistent CSA (in spite of using CPAP for 3 months) were randomized to either continued CPAP ($n = 11$) or ASV ($n = 12$), LVEF increased significantly (32% to 38%) in the ASV group. Furthermore, reductions in plasma B-type natriuretic peptide and urinary norepinephrine and an increase in 6-minute walk distance and quality of life as measured by short form 36 were significantly greater with ASV than CPAP.

Meanwhile, as noted previously, with automatic variable end expiratory positive pressure algorithms, ASV devices are effective in treating both CSA and OSA. In the three trials[122-124] that compared CPAP versus ASV in patients with heart failure and coexistent OSA and CSA, ASV was significantly more effective in improving LVEF than CPAP. Because both obstructive and central sleep disorders commonly occur together, and the phenotype may change with time (e.g., obstructive events becoming prominent during acute decompensation of the heart failure), ASV devices with automatic end expiratory positive pressure rhythm can be quite effective under such circumstances.

A large multi site international clinical trial (SERVE-HF)[124a] that evaluated the effect of treating central sleep apnea with an ASV device in patients with heart failure and reduced ejection fraction (HFrEF) has raised serious concerns about the safety of ASV in these patients. Not only was ASV ineffective, but post-hoc analysis found excess cardiovascular mortality in treated patients. The cause of the excess mortality is unknown; the authors hypothesized that CSA might be a compensatory mechanism with a protective effect in HFrEF.

However, there are several other (perhaps more) plausible explanations for the excess cardiovascular mortality in the trial. These include methodological issues, the use of the old generation ASV device, which is no longer manufactured by the sponsor of the trial, residual sleep-disordered breathing with significant oxygen desaturation, patient selection, data collection, and treatment adherence as well as group crossovers as potential confounding factors (Javaheri et al, unpublished). Below, only the device-related issues are briefly reviewed.

The data from the trial[124a] show that the device was ineffective in a number of patients (note the large range of AHI values downloaded from the ASV devices across the months of follow-up in Table S4 in the supplementary appendix of reference 124a). There could be multiple reasons for these residual events. Two important issues regarding the algorithm of the first generation ASV device used in SERVE-HF could have been contributory: First, it allows for only fixed expiratory positive airway pressure (EPAP), and second, there are flaws in the inspiratory pressure support algorithm that were addressed and improved considerably in the latest generation models.[122] Regarding fixed EPAP, it is known that the phenotype of sleep-disordered breathing may change over time from predominantly central to predominantly obstructive events, and data from Table 2 in the study[124a] are confirmatory. In that case, in the face of obstructive apneas and the fixed expiratory pressure of the device, the ASV device used was equipped with only one strategy for suppressing these events: progressively increasing inspiratory pressure support in an attempt to open the closed airway. Once the airway opened, the prevailing high pressures may have resulted in an excessive rise in intrathoracic pressure with consequent adverse hemodynamic effects. In addition, excess ventilation due to excessive inspiratory pressure support could result in hyperventilation and alkalemia which is arrythmogenic, and at the same time the baseline Pco_2 could be lowered excessively promoting recurrence of apneas. The current generation of ASV devices can be set to increase EPAP automatically in response to obstructive apneas and would not have been subject to this failure.

Independent of the reasons why the trial failed, manufacturers of ASV devices have declared that ASV devices are contraindicated for heart failure patients with central sleep apnea when LVEF is 45% or less. Since we no longer use ASV, our current approach to such patients is to first make sure that heart failure is maximally treated, both pharmacologically and device-wise, when indicated. Then gentle CPAP titration is performed with maximum pressure not to exceed 10 to 12 cm H_2O. If the AHI decreases below 15/hour of sleep, long term CPAP therapy is recommended. Otherwise nocturnal O_2 titration is recommended with the least amount of supplemental O_2 needed to eliminate hypoxia and maintain arterial oxyhemoglobin saturation above 92%, avoiding hyperoxia.

Meanwhile we are anxiously waiting for the results of the ADVENT trial[124b] and the remede trial.[124c] The latter uses a transvenous phrenic nerve pacemaker (see also Chapter 126).

Cardiac Pacing

In 15 subjects with predominantly mild to moderate CSA, some of whom had mild LV systolic dysfunction, atrial overdrive pacing improved periodic breathing.[125] These subjects had permanent atrial-synchronized ventricular pacemakers placed for symptomatic sinus bradycardia. Atrial overdrive (an average of 72 beats per minute versus spontaneous 57 beats per minute) moderately (but significantly) decreased the AHI from 28 to 11, improved arterial oxyhemoglobin desaturation, and decreased arousals. The mechanism remains unclear, but it could have been the improved cardiac output. However, if cardiac pacing improves CSA, biventricular pacing[126] should be more effective than atrial pacing overdrive,[106-110] as discussed earlier. In one study,[108] CRT decreased central AHI from 31/hour to 17/hour. There was no effect on obstructive disordered breathing events.

Transvenous Unilateral Phrenic Nerve Stimulation

Most recently, transvenous unilateral phrenic nerve stimulation has been used to stimulate phrenic nerve and treat CSA. In this acute study,[127] 16 patients underwent two successive nights of polysomnography—one night with and one night without phrenic nerve stimulation from either the right brachiocephalic vein or the left pericardiophrenic vein. Stimulation resulted in significant improvement in the AHI Central Apnea Index, arousal index, and oxygen desaturation index 4%. No significant changes occurred in the Obstructive Apnea

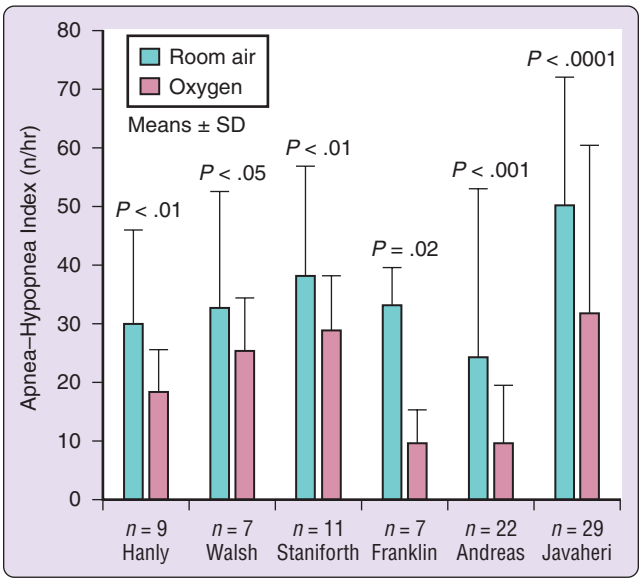


Figure 129-12 Effects of supplemental nasal oxygen on apnea–hypopnea index in patients with systolic heart failure.

Index or AHI. This approach may represent a novel therapy for CSA and warrants further study. Currently, a randomized clinical trial is in progress.

Medications

Nasal Nocturnal Oxygen. Systematic studies in patients with systolic heart failure[128-133] have shown that nocturnal therapy with supplemental nasal oxygen improves CSA (Figure 129-12). Oxygen therapy may also decrease arousals and improve the hypnogram by shifting sleep structure to deep sleep stages. In addition, randomized placebo-controlled double-blind studies have shown that short-term (1 to 4 weeks) administration of nocturnal supplemental nasal oxygen improves maximal exercise capacity[130] and decreases overnight urinary norepinephrine excretion.[131]

Three randomized clinical trials of nocturnal nasal oxygen therapy[133-136] for 9-, 12-, and 52-week periods, reported that, when compared with the control group, oxygen therapy improved CSA and desaturation and significantly increased LVEF and quality of life of patients with heart failure. In the oxygen-treated group, LVEF increased 5% (versus 1% in the control group) in the 12-week study,[136] and 5.5% (versus 1.3% in the control group) in the 52-week study.[135]

Supplemental administration of nasal oxygen may decrease periodic breathing by several mechanisms.[134] These include an increase in the difference between the prevailing P_{CO_2} and the P_{CO_2} at the apneic threshold; a reduction in the ventilatory response to CO_2 and perhaps to hypoxemia; and an increase in body stores (e.g., lung contents) of oxygen, which increases damping. Prospective placebo-controlled long-term studies, however, are necessary to determine whether nocturnal oxygen therapy has the potential to decrease mortality of patients with systolic heart failure.

Theophylline. Open[7,137] and blind studies[138] have shown the efficacy of theophylline in the treatment of CSA in heart failure. In a double-blind randomized placebo-controlled crossover study of 15 patients with treated, stable systolic heart failure, oral theophylline at therapeutic plasma concentration (11 μg/mL, range 7 to 15 μg/mL) decreased the AHI by about 50% and improved arterial oxyhemoglobin saturation.[137]

Mechanisms of action of theophylline in improving central apnea remain unclear.[123] At therapeutic serum concentrations, theophylline competes with adenosine at some of its receptor sites. In the central nervous system, adenosine is a respiratory depressant, and theophylline stimulates respiration by competing with adenosine. Conceivably, therefore, an increase in ventilation by theophylline decreasing the plant gain could decrease central apnea during sleep. Theophylline does not increase ventilatory response to CO_2.

Potential arrhythmogenic effects and phosphodiesterase inhibition are common concerns with long-term use of theophylline in patients with heart failure. Therefore, further controlled studies are necessary to ensure its safety. If theophylline is used to treat CSA, frequent and careful follow-ups are necessary.

Acetazolamide. In a double-blind placebo-controlled crossover study[139] of 12 patients with heart failure, acetazolamide, administered at about 3 mg/kg one-half hour before bedtime, decreased the central AHI significantly from about 57/hour (in the placebo arm) to 34/hour. Acetazolamide improved arterial oxyhemoglobin desaturation significantly. Furthermore, patients reported improved subjective perceptions of the following: overall sleep quality, feeling rested on awakening, falling asleep unintentionally during daytime, and fatigue. Acetazolamide, therefore, could have other advantageous effects when used in patients with heart failure and CSA, including acting as a mild diuretic and also normalizing the alkalemia (caused by loop diuretics) commonly present in patients with heart failure. In our patients, arterial blood pH decreased from 7.43 to 7.37.[139]

Acetazolamide improves CSA by decreasing the plant gain as shown in naturally sleeping canine experiments[140] and patients with heart failure with CSA.[141]

Benzodiazepines. Benzodiazepines, by decreasing arousals, may decrease CSA. However, a placebo-controlled double-blind study[142] showed a reduction in arousals but failed to show any improvement in CSA in patients with systolic heart failure. Although benzodiazepines do not increase the number of central apneas, their use may increase the likelihood of developing obstructive apneas in some heart failure patients.

Inhaled CO_2 and Addition of External Dead Space. Several studies have shown that low-level inhalation of CO_2 and addition of external dead space (by increasing P_{CO_2}) improve CSA.[145-147] However, studies[146,147] show that CO_2 inhalation increases spontaneous arousals, which are associated with increased sympathetic and decreased parasympathetic activity. One study[145] also showed that addition of dead space was associated with increased arousals. Knowing the adverse cardiovascular effects of increased sympathetic overactivity in heart failure, use of CO_2 and external dead space to treat CSA in heart failure should be avoided. However, dynamic CO_2 inhalation,[148-150] when it can be inhaled intermittently within a part of breathing cycle, could eventually prove useful.

CLINICAL PEARLS

- Because of the increased average life span and improved therapy of ischemic coronary artery disease and hypertension, the prevalence of heart failure remains high.
- Periodic breathing is common in heart failure and is characterized by apnea, hypopnea, and hyperpnea, which cause sleep disruption, arousals, hypoxemia/reoxygenation, hypercapnia/hypocapnia, and changes in intrathoracic pressure. Periodic breathing includes both obstructive and central sleep-related breathing disorders. All of these adversely affect sleep and cardiovascular function.
- Periodic breathing may contribute to the remodeling of LV dysfunction and to the progressively declining course of heart failure.
- Several studies have demonstrated that both CSA and OSA are associated with increased mortality of patients with heart failure and systolic dysfunction.
- There are only a few long-term studies on treatment of sleep apnea in systolic heart failure. These show that effective treatment of both CSA and OSA with CPAP improves mortality of patients with heart failure.
- At this time ASV is not recommended in CHF when LVEF is less than 45%.

SUMMARY

Heart failure is a common disorder that has a significant economic impact and is associated with excess morbidity and mortality. Because of increased average life spans and improved therapy for hypertension and ischemic coronary artery disease, the incidence and prevalence of heart failure remain high.

One factor that may contribute to the progressively declining course of heart failure, hospital readmission, quality of life, and premature mortality is the occurrence of periodic breathing, with repetitive episodes of apnea, hypopnea, and hyperpnea. Episodes of apnea, hypopnea, and the following hyperpnea collectively cause hypoxemia and reoxygenation, hypercapnia and hypocapnia, changes in intrathoracic pressure, and sleep disruption and arousals. These pathophysiological consequences of sleep-related breathing disorders have deleterious effects on the cardiovascular system, and they may be most pronounced in the setting of established heart failure and coronary artery disease.

Multiple studies have demonstrated increased readmission and premature mortality independently associated with obstructive and CSA comorbid with heart failure. In addition, multiple studies have also demonstrated that effective treatment of both obstructive and CSA decreases hospital readmission and improves survival, particularly in those patients who are most adherent to therapy.

ASV devices with automatic inspiratory pressure support and automatic end expiratory positive pressure algorithms, along with a backup rate, are quite effective in the treatment of hybrid sleep-related breathing disorders consisting of both central and OSA but are not recommended when LVEF is less than 45%. The best approach for patients with low LVEF is not clear, and future research will be needed to guide the management of these patients.[151]

Selected Readings

Bitter T, Faber L, Hering D, et al. Sleep-disordered breathing in heart failure with normal left ventricular ejection fraction. *Eur J Heart Fail* 2009;**11**:602–8.

Bradley TD, Flores JS. ADVENT-HF investigators. The SERVE-HF trial. *Can Respir J* 2015 Oct 16. pii: 17237. [Epub ahead of print].

Costanzo MR, Augostini R, Goldberg LR, et al. Design of the remedē System Pivotal Trial: a prospective, randomized study in the use of respiratory rhythm management to treat central sleep apnea. *J Card Fail* 2015;**21**(11):892–902.

Cowie MR, Woehrle H, Wegscheider K, et al. Adaptive servo-ventilation for central sleep apnea in systolic heart failure. *N Engl J Med* 2015;**373**:1095–105.

Galetke W, Ghassemi BM, Priegnitz C, et al. Anticyclic modulated ventilation versus continuous positive airway pressure in patients with coexisting obstructive sleep apnea and Cheyne–Stokes respiration: a randomized crossover trial. *Sleep Med* 2014;**15**:874–9.

Hetland A, Haugaa KH, Olseng M, et al. Three-month treatment with adaptive servo-ventilation improves cardiac function and physical activity in patients with chronic heart failure and Cheyne-Stokes respiration: a prospective randomized controlled trial. *Cardiology* 2013;**126**:81–96.

Javaheri S. Sleep disorders in systolic heart failure: a prospective study of 100 male patients. The final report. *Int J Cardiol* 2006;**106**:21–8.

Javaheri S, Brown L, Randerath W. Positive airway pressure therapy with adaptive servo ventilation (part 1: operational algorithms). *Chest* 2014;**146**:514–23.

Javaheri S, Brown L, Randerath W. Positive airway pressure therapy with adaptive servo-ventilation (part II: clinical applications). *Chest* 2014;**146**:855–68.

Javaheri S, Caref B, Chen E, et al. Sleep apnea testing and outcomes in a large cohort of Medicare beneficiaries with newly diagnosed heart failure. *Am J Respir Crit Care Med* 2011;**183**:539–46.

Javaheri S, Dempsey JA. Central sleep apnea. *Compr Physiol* 2013;**3**:141–63.

Jilek C, Krenn M, Sebah D, et al. Prognostic impact of sleep disordered breathing and its treatment in heart failure: an observational study. *Eur J Heart Fail* 2011;**13**:68–75.

Kasai T, Kasagi S, Maeno K-I, et al. Adaptive servo-ventilation in cardiac function and neurohormonal status in patients with heart failure and central sleep apnea nonresponsive to continuous positive airway pressure. *JACC Heart Fail* 2013;**1**:58–63.

Kasai T, Narui K, Dohi T, et al. Prognosis of patients with heart failure and obstructive sleep apnea treated with continuous positive airway pressure. *Chest* 2008;**133**:690–6.

Koyama T, Watanabe H, Igarashi G, et al. Effect of short-duration adaptive servo-ventilation therapy on cardiac function in patients with heart failure. *Circ J* 2012;**76**:2606–13.

Koyama T, Watanabe H, Igarashi G, et al. Short-term prognosis of adaptive servo-ventilation therapy in patients with heart failure. *Circ J* 2011;**75**:710–12.

Oldenburg O, Lamp B, Faber L, et al. Sleep disordered breathing in patients with symptomatic heart failure: a contemporary study of prevalence in and characteristics of 700 patients. *Eur J Heart Fail* 2007;**9**:251–7.

Priou P, d'Ortho MP, Damy T, et al. Adaptive servo-ventilation: How does it fit into the treatment of central sleep apnoea syndrome? Expert opinions. *Rev Mal Respir* 2015 Nov 20. pii: S0761-8425(15)00364-2.

Suzuki S, Yoshihisa A, Miyata M, et al. Adaptive servo-ventilation therapy improves long-term prognosis in heart failure patients with anemia and sleep-disordered breathing. *Int Heart J* 2014;**55**:342–9.

Yoshihisa A, Suzuki S, Yamaki T, et al. Impact of adaptive servo-ventilation on cardiovascular function and prognosis in heart failure patients with preserved left ventricular ejection fraction and sleep-disordered breathing. *Eur J Heart Fail* 2013;**15**:543–50.

Yumino D, Redolfi S, Ruttanaumpawan P, et al. Nocturnal rostral fluid shift: a unifying concept for the pathogenesis of obstructive and central sleep apnea in men with heart failure. *Circulation* 2010;**121**:1598–605.

Zhang J, et al. Exploring quality of life in patients with and without heart failure. *Int J Cardiol* 2016;**202**:676–84.

A complete reference list can be found online at ExpertConsult.com.

Section
16

Other Medical Disorders

Christine Won

Chapter
130

Sleep and Fatigue in Cancer Patients

Josée Savard; Marie-Hélène Savard; Sonia Ancoli-Israel

Chapter Highlights

- Insomnia and fatigue are common in cancer patients before treatment, while undergoing chemotherapy or radiation therapy, and after the completion of therapy. These symptoms appear to be part of the same cluster.
- Actigraphy data show disrupted rest-activity patterns in cancer patients, especially during chemotherapy.
- The relationship between sleep disruption and fatigue appears to be bidirectional. Fatigue is a common consequence of insomnia, but it can also increase the risk for insomnia, most likely through behavioral changes (e.g., napping) that disturb the sleep-wake cycle.
- Several cancer-specific factors may trigger the onset of sleep disturbances, including adjuvant treatments, nocturnal hot flashes, pain, opioids, and antiemetic medications. Fatigue may also be caused by a combination of factors including physical (e.g., cachexia, inflammation, hematologic and endocrine abnormalities), psychological (e.g., depression), and social factors.
- There is accumulating evidence supporting the efficacy of psychological interventions (cognitive-behavioral therapy in particular), activity-based interventions, and bright-light therapy to treat cancer-related sleep disturbances and fatigue.
- More longitudinal studies are needed to characterize the natural course of sleep complaints, circadian rhythms impairments, and fatigue during the cancer care trajectory and to better understand how these disturbances are interrelated.

This chapter is dedicated to the memory of Dr. J. Christian Gillin, dear friend and colleague, who died of cancer and was fatigued, but never let it get to him. He was an inspiration and role model to us all.

Supported by NCI CA112035, NCI CA158954, Moores UCSD Cancer Center, and the Canadian Institutes of Health Research.

Patients with cancer complain of fatigue before treatment, during chemotherapy or radiation therapy, and after the completion of therapy.[1] These patients also complain of sleep disruption.[2] Both fatigue and poor sleep likely contribute to decreased quality of life.[3,4] There is a growing body of literature on the relationship between fatigue and the quality or quantity of sleep. This chapter reviews the evidence on cancer-related sleep disruption and fatigue and their treatment as well as the interrelationships among poor sleep, desynchronized circadian rhythms, and cancer-related fatigue.

EPIDEMIOLOGY

Sleep Disruption

Early cross-sectional studies, mainly conducted in the post-treatment phase, revealed that between 30% and 50% of cancer patients report sleep difficulties and that nearly 20% meet the diagnostic criteria for an insomnia disorder (for a review of these studies, see[5-12]). However, these studies were limited by issues such as the use of small, convenience samples and of sleep measures composed of a single item or a small number of items. Moreover, sleep was often assessed several months and even years after the completion of treatment. Because of their cross-sectional nature, these studies did not provide information on the natural course of sleep impairments (incidence, remission, persistence) throughout the cancer care trajectory and beyond.

More recently, two large-scale longitudinal studies were conducted in nonmetastatic patients with heterogeneous cancer sites to assess the evolution of sleep impairments over time. Savard et al[13,14] observed 991 patients awaiting surgery for mixed cancer sites. Patients were assessed at baseline during the perioperative phase and 2, 6, 10, 14, and 18 months later. At baseline, 59% had insomnia symptoms, including 28% with an insomnia disorder, as assessed by a phone diagnostic interview and the algorithm developed by Morin and colleagues.[15] Accordingly, patients were considered to have an insomnia disorder when they met the following criteria: sleep-onset latency or wake after sleep onset of more than 30 minutes, at least 3 nights per week; sleep efficiency of less than 85%, for at least 1 month; impaired daytime functioning or marked distress; or using a hypnotic medication 3 nights/week or more for at least 1 month. Although these rates steadily decreased over time, 36% of the sample still suffered from insomnia symptoms, whereas 21% met the criteria for an insomnia disorder 18 months later, which remains much higher than in the general population. Moreover, the general persistence rate (i.e., insomnia present at two consecutive time points on 2- to 4-month intervals) was 51%, whereas 35% of patients had insomnia persisting for at least three consecutive time points. Insomnia disorder was a particularly enduring condition, with persistence rates varying from 69% to 80%.

In another large-scale prospective study, Palesh et al[16] assessed the presence of insomnia in 823 patients scheduled to receive at least four cycles of chemotherapy for various types of cancer of all stages. Sleep difficulties were assessed on day 7 of cycle 1 and cycle 2 of chemotherapy using the six sleep-related questions from the Hamilton Depression Inventory. The insomnia disorder was defined by the presence of difficulty in falling asleep, difficulty in staying asleep, and early morning awakenings (at least 30 minutes) for at least 3 nights per week for 2 weeks. This more liberal definition (not taking into account sleep efficiency and insomnia-related functioning impairments and distress), assessed with a questionnaire rather than a diagnostic interview, yielded greater prevalence rates. At cycle 1, 80% of patients exhibited insomnia symptoms, including the 43% meeting the criteria for an insomnia disorder, rates that decreased to 68% and 35%, respectively, at cycle 2. Among good sleepers at cycle 1, 35% developed insomnia symptoms at cycle 2, of whom 10% developed an insomnia disorder.

The prevalence of sleep disturbances varies as a function of cancer sites. Early studies indicated that sleep disturbances were more frequent in breast and ovarian cancer patients.[17,18] More recently, Palesh et al[16] found that the prevalence of insomnia symptoms or disorder was the highest in breast (85%) and gynecologic (83%) compared with hematologic, lung, and gastrointestinal cancers. The prevalence of the insomnia disorder alone was the highest in patients with lung cancer (51%) and the lowest in those with a gastrointestinal cancer (24%). Savard et al[13] also found the highest rates of insomnia symptoms among breast (42% to 69%) and gynecologic (33% to 68%) cancer patients, whereas the lowest rates were obtained in men with prostate cancer (25% to 39%) throughout the study. Subanalyses conducted in patients with urinary and gastrointestinal cancer, showing no difference between men and women, suggest that higher rates found in women with breast and gynecologic cancer are not solely attributable to gender.[14]

Sleep disturbances are also common in patients with advanced cancer. From about half to three quarters of outpatients attending cancer or palliative care clinics report some sleep disturbances.[19-21] In a study of patients with metastatic breast cancer, 63% reported sleep disturbances. Difficulty in falling asleep was associated with both depression and pain, whereas increased awakening during the night was associated with only depression.[22] Similarly, another study conducted among 101 patients with advanced cancer (any type) found that sleep difficulties were associated with increased pain, depressive, and anxiety symptoms and a poorer sense of well-being.[23] In a prospective study, 15% of terminally ill cancer patients had sleep disturbance (insomnia or hypersomnia) and 29% had subclinical disturbance at the moment of registration to a palliative care unit. These rates increased to 26% and 37%, respectively, at the time of admission. There was a change in sleep status in 67% of patients between the two time points; sleep deteriorated (46%) more frequently than it improved (21%).[24] Another study found that patients with advanced lung cancer reported poorer sleep and more daytime sleepiness than healthy controls and that their sleep disturbances were characterized by breathing difficulties, cough, nocturia, and frequent awakenings, all of which may be suggestive of sleep-disordered breathing.[25] Finally, a study suggested that poor sleep quality and use of sleep medications were, along with hopelessness and depression, the best predictors of desire for hastened death in 102 terminally ill patients attending a palliative care unit,[26] thus emphasizing the importance of offering appropriate sleep management to these patients.

A few studies used polysomnography (PSG) to assess sleep disturbances objectively in cancer patients. When the sleep of patients with breast or lung cancer, patients with insomnia, and volunteers with no sleep problems were

compared using PSG, insomnia patients had the shortest total sleep time, but lung cancer patients had the longest sleep-onset latency, lowest sleep efficiency, and greatest wake time during the night.[27] Savard et al[28] found an average sleep efficiency of 84% among a sample of 56 women treated with chemotherapy, radiotherapy, and hormone therapy for early-stage breast cancer. On average, wake after sleep onset was 54 minutes, whereas sleep-onset latency was 20 minutes. In that sample, sleep-onset latency, rapid eye movement (REM) sleep latency, wake after sleep onset, sleep efficiency, and distribution of sleep stages were similar to those found in healthy women of the same age. Conversely, in a sample of 114 advanced cancer patients, Parker et al[29] showed, compared with normative data, a reduced sleep quantity and quality, with a sleep efficiency of 77% on average and virtually no slow wave sleep. More research is needed to better quantify PSG alterations associated with different types of cancer and their progression.

The available literature, although limited, also suggests the presence of sleep disorders other than insomnia in cancer patients. Elevated prevalence rates of obstructive sleep apnea (OSA), ranging from 12% to 92%, were found in small-scale studies (17 to 33 patients) conducted among patients with head and neck cancer.[30,31] There is also some evidence suggesting that OSA may be associated with treatments such as surgery and radiotherapy among these patients,[32] but prospective studies are warranted to investigate to what extent OSA is caused or exacerbated by the cancer itself or by cancer treatment. Sleep disordered breathing in the form of obstructive and central apneas also appears to be frequent in patients with brain tumors, with tumor removal resulting in a significant decrease in the apnea-hypopnea index.[33] Rates of OSA have also been found to be high in women with breast cancer who had completed chemotherapy, with almost half of the patients (48%) having at least five respiratory events per hour of sleep.[34] In the same study, the prevalence of periodic limb movements was 36%. Both of these sleep disorders were substantially more frequent than in age-comparable women without cancer. These high prevalence rates of periodic limb movements and OSA may help explain some of the sleep disturbance found in this population. However, others reported no difference in the amount of sleep-disordered breathing between cancer patients and insomnia patients or healthy volunteers.[27]

Actigraphy, a noninvasive, continuous, ambulatory measure of circadian rest-activity rhythms, has also been used to objectively characterize the sleep and rhythms of patients with cancer.[35-37] Studies comparing cancer patients with healthy controls have consistently shown less contrast between daytime and nighttime activity in cancer patients, a pattern indicative of circadian disruption.[38-44] Among patients with advanced cancer, objective data from actigraphy have shown high sleep fragmentation despite normal sleep duration.[19]

Chemotherapy appears to be particularly disruptive of rest-activity circadian rhythms,[45,46,46a] although radiation therapy was also found to have a detrimental effect.[46a,47] In a study of 85 women with breast cancer, 72-hour actigraphy showed that the first administration of chemotherapy was associated with transient disruption of sleep-wake rhythm, whereas its repeated administration resulted in progressively worse and more enduring impairments.[46] A recent study, conducted among 49 patients with advanced cancer assessed with actigraphy from 3 days before to 10 days after the administration of a chemotherapy cycle, found 45% of patients showing a sustained deterioration of their rest-activity pattern after chemotherapy administration, with both increased nighttime activity level and decreased diurnal activity level. The other patients experienced some alterations but recovered at the end of the chemotherapy cycle (31%) or showed either no alteration (10%) or improvement of their rest-activity pattern throughout chemotherapy (14%).[48] Whereas there are some data suggestive of improvement of the rest-activity patterns back to precancer treatment levels once the active phase of cancer treatment is over, at least among patients with early-stage disease, the disturbance remains worse than in matched controls.[38]

Fatigue

Fatigue is one of the most frequent and disturbing complaints of patients with cancer,[1,49] with 70% to 100% of patients during active treatment and approximately 30% after the treatment phase reporting feeling weak and tired.[50] Cancer-related fatigue has been defined as a "persistent, subjective sense of tiredness related to cancer and cancer treatment that interferes with usual functioning."[51] It is believed to be distinct in nature from general fatigue, as it is unrelated to exertion level and is not relieved by rest or sleep. Fatigue may interfere with daily functioning and reduce quality of life[52] and is associated with elevated direct and indirect costs.[53] Cancer-related fatigue is one of the key reasons for discontinuation of treatment and participation in clinical trials.[54]

Overall, studies suggest that fatigue assessed subjectively with various questionnaires is highly prevalent before[55] as well as during[56] adjuvant treatments. In addition, several large-scale studies indicate that fatigue can continue for months and even years after the completion of therapy.[57-61] Fatigue is present before the start of chemotherapy,[35] and it increases significantly with the introduction of chemotherapy.[45,62-64] Different patterns have been observed for radiotherapy. Purcell et al[65] observed that fatigue increased at the end of radiotherapy, followed by a decrease 6 weeks later among 210 prostate and head and neck cancer patients. Another study, conducted in 87 oropharyngeal cancer patients, revealed that fatigue peaked about 1 to 2 weeks after the completion of radiotherapy, whereas this symptom persisted for up to 2 years after for half of the sample.[66] In the breast cancer context, some authors observed an increase in fatigue during radiotherapy, followed by a return toward baseline levels 6 to 8 weeks after treatment completion.[67,68] In contrast, others found persistent elevated levels of fatigue among women with breast cancer several weeks after radiotherapy completion.[69] In a longitudinal study conducted among 60 patients with mixed cancer sites, 22% of participants had severe persistent fatigue in the first year after cancer treatments, a rate that remained fairly stable throughout that period.[70] There are likely to be many reasons for these differences found across studies, including the differences in fatigue measurement, the types of treatments received (e.g., surgery, chemotherapy, hormone therapy), the number and duration of the treatments, and the level of preexisting fatigue.[70-72]

PATHOGENESIS

Insomnia and Sleep Disturbances

A younger age has consistently been found to be associated with an increased risk for cancer-related insomnia.[2,17,73] In addition, the cancer journey typically involves a combination of treatments that may include surgery, chemotherapy, radiation therapy, and hormone therapy, all of which have the potential to provoke or to intensify sleep disturbances because of their emotional impact, their direct physiologic effects, or their side effects.[74] Chemotherapy is thought to be particularly harmful.[3,75] Patients report more subjective sleep disturbance (lower sleep quality and duration, total sleep time) during the active phases of chemotherapy compared with rest periods.[76] Longitudinal studies using subjective measures have also shown gradual increases in sleep impairment with this treatment[77] as well as persistence of elevated insomnia rates across chemotherapy cycles.[16] Conversely, another study found no change in sleep with the introduction of chemotherapy for breast cancer as measured prospectively with a sleep diary, but most participants already had poor sleep before the initiation of this treatment.[78] Of note, studies that measured PSG data have not observed a significant deterioration of PSG sleep parameters with the introduction of chemotherapy.[79,80]

Many cancer-related somatic symptoms may affect sleep negatively. Dyspnea, urinary symptoms (e.g., due to radiation therapy in the urogenital area), gastrointestinal symptoms (e.g., chemotherapy-induced nausea), and pain (e.g., associated with use of aromatase inhibitors) in both men and women are all very likely to impair sleep.[46a,81] In addition, medications that are commonly administered with chemotherapy, such as opioids, antiemetic medications, and corticosteroids, are also known to disrupt sleep.[82-84] The occurrence or exacerbation of menopausal symptoms appears to be another important contributor to cancer-related sleep disturbances. Indeed, the estrogen deficiency induced by chemotherapy and hormone therapy, the abrupt cessation of hormone replacement therapy at cancer diagnosis, or an ovary removal may trigger or exacerbate preexisting hot flashes. Savard et al[85] showed that changes occurring in self-reported vasomotor symptoms between the end of initial adjuvant treatments and a 3-month follow-up were significantly associated with parallel changes in insomnia complaints. Two other studies from the same team using objective measures showed greater PSG sleep disturbances associated with nocturnal hot flashes detected with sternal skin conductance. In the earlier study[86] conducted among 24 breast cancer survivors, the 10-minute periods around hot flashes had more wake time and more stage changes to lighter sleep than other 10-minute periods during the night. In addition, compared with nights without, nights with hot flashes had a significantly higher percentage of wake time, a lower percentage of stage 2 sleep, and a longer REM latency. In the more recent study[28] conducted in 56 women treated for breast cancer, slower and longer hot flashes, but not increased hot flash frequency, were associated with several sleep impairments, including a greater total wake time, lower sleep efficiency, and increased number of awakenings.

The amount of insomnia in cancer patients can be as high as the amount of insomnia found in depressed patients; therefore, clinicians should not overlook the possibility that poor sleep in cancer patients may indicate some psychological distress. The study by Sharma et al[87] revealed that cancer patients with high levels of psychological distress were 4.5 times more likely to report sleep difficulties than those with low levels. In one sample of newly diagnosed breast cancer patients, insomnia was the most frequent symptom, reported by 88% of patients, and was correlated with high levels of psychological distress and anxiety.[88] However, contrary to the belief that disturbed sleep before treatment is attributable to the increased stress and anxiety resulting from a recent diagnosis of a potentially life-threatening illness, insomnia and fatigue were rated high even in those patients who rated themselves low on anxiety. Similarly, another study revealed that 46% of prostate cancer survivors with an insomnia disorder did not have clinical levels of anxiety or depressive symptoms.[89] Thus, there is evidence that although insomnia and psychological distress are interrelated, there are still a significant proportion of patients who have isolated insomnia.

Pain has often been thought to be the cause of sleep disruption, not only in patients with cancer but in patients with a multitude of other medical conditions.[90] In a sample of 2862 cancer outpatients, the risk of reporting insomnia symptoms was 2.7 times higher for those having pain,[87] which is consistent with other empiric evidence showing that cancer-related pain significantly predicts the incidence or exacerbation of sleep difficulties.[91,92] There are also data showing increased objectively assessed sleep disturbances and rest-activity rhythm impairments measured with actigraphy among men, but not women, with pain.[93]

One hypothesis explaining the association between pain and insomnia comorbid with cancer is that pain may be the initial cause of the frequent awakenings, but psychological distress prevents the patient from falling back to sleep.[94] A second hypothesis explains that whereas sleep leads to recovery and repair of tissue and may offer a temporary cessation of the psychological awareness of pain, poor sleep leads to difficulty in managing pain.[95] In this way, a cycle of pain and poor sleep may become self-perpetuating.

Fatigue

Fatigue is also likely to be caused by multiple factors including physical (e.g., cachexia; weight loss; biochemical, hematologic, and endocrine abnormalities), psychological (e.g., depression), and social factors (Figure 130-1). Anemia and other biochemical abnormalities are commonly found in cancer patients and cause fatigue,[68] although hemoglobin levels are only moderately related to fatigue and quality of life. One study examining the incremental effect of increasing hemoglobin on quality of life found that improving anemia improved quality of life only to a point, beyond which there was no further improvement.[96] Alternative possible physiologic mechanisms include inflammation, serotonin dysregulation, hypothalamus–pituitary–adrenal axis dysfunction, circadian rhythm disruption, altered muscle metabolism, and genetic dysregulation.[97,98] Among these potential mechanisms, inflammation is probably the one currently receiving the most attention and is believed to be a common pathway through which cancer and its treatment could lead to a variety of symptoms, including fatigue and sleep disturbances.[97,99]

Several studies have found significant relationships between reports of fatigue and depression,[100-102] but it is unclear to what extent these are etiologically related. A decrease in

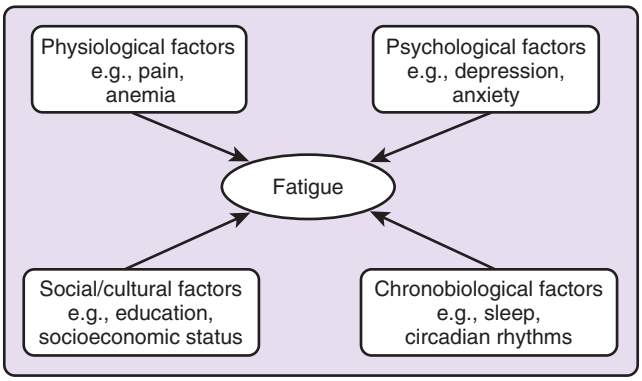


Figure 130-1 Diagrammatic representation of possible factors affecting fatigue. (From Ancoli-Israel S, Moore PJ, Jones V. The relationship between fatigue and sleep in cancer patients: a review. *Eur J Cancer Care [Engl]* 2011;10:245–55, with permission.)

activity level could constitute a common mechanism. This behavioral change, frequently observed among depressed individuals, may contribute in the long term to physical deconditioning, which may in turn exacerbate fatigue.[64] However, depression is far less common than fatigue in cancer patients, which suggests that fatigue often occurs independently. In fact, fatigue and depression show different patterns of evolution over the cancer care trajectory.[103,104] Moreover, it has been shown that an antidepressant treatment decreases depression without concurrently improving fatigue, thus suggesting differential mechanisms.[105,106] Depression has been found to significantly predict subsequent fatigue among breast cancer patients,[64] but negative results have also been found.[107-110] Overall, more research is needed to understand the pathogenesis of cancer-related fatigue.

Relationship Among Sleep Disturbances, Circadian Rhythms, and Fatigue

Studies on symptom clusters have revealed that sleep and fatigue are often part of a same cluster of three or more symptoms.[111-116] Moreover, most cross-sectional and prospective studies have found a strong correlation between sleep disturbances and fatigue,[109,111,117,118] with a similar trajectory over time.[38]

The relationship between sleep disruption and fatigue appears to be bidirectional. Patients typically report fatigue as the main consequence of their poor sleep.[119,120] Accordingly, there is evidence that sleep disturbance is a significant predictor of fatigue,[109,121-123] but the reverse may also hold true. A longitudinal study showed that fatigue significantly predicted a subsequent increase in insomnia symptoms during the cancer care trajectory, whereas insomnia was not found to predict subsequent fatigue.[110] The impact of fatigue on sleep could be due to behavioral changes occurring with fatigue. Indeed, fatigued individuals tend to nap more during the day and extend their sleep periods during the night, which may in the long run impair their circadian rhythm and make their nighttime sleep less consolidated and lighter.[124,125] This may be particularly the case during cancer treatments, when patients suffer from higher levels of fatigue and are very likely and often encouraged to rest to recuperate.[126,127]

Evidence on the relationship between circadian rhythms and subjective ratings of fatigue has been mixed, with most studies finding a significant relationship.[37,45,128] Daytime inactivity and nighttime restlessness were associated with higher subjective ratings of fatigue in one series of studies.[128] Women with breast cancer undergoing adjuvant chemotherapy reported more fatigue during treatment and less fatigue at chemotherapy cycle midpoints, in a "roller coaster" pattern. Activity levels were negatively correlated with reports of fatigue, that is, those with more fatigue showed less activity. Activity levels were reduced during the three treatment sessions compared with the cycle midpoints, thus showing the reverse "roller coaster" pattern, with inversely changing fatigue scores. Patients tended to have more nighttime restlessness at treatment times compared with cycle midpoints when higher activity during the day prevailed and there were fewer nighttime awakenings.[128,129] Others[37] found that self-reported fatigue was significantly associated with less stability of the rest-activity pattern across time but not with the proportion of time spent resting or sleeping during the day. Moreover, changes in fatigue from the second to the fourth on-study chemotherapy cycles were significantly associated with changes in the consistency of the sleep-wake pattern. Conversely, a study of breast cancer patients before chemotherapy[35] found no significant relationship between any of the rhythm variables or objective sleep variables assessed by a 72-hour actigraphy recording and subjective reports of fatigue. Another study conducted in breast cancer patients before chemotherapy found that most actigraph measures of sleep-wake, activity-rest, and circadian rhythms derived from a 48-hour recording were not significantly associated with fatigue.[130] Together, these studies suggest that fatigue becomes a significant correlate of circadian rhythms only after chemotherapy has been initiated in breast cancer patients. However, a more recent study comparing 148 women with stage I-III breast cancer scheduled to receive at least four cycles of adjuvant or neoadjuvant chemotherapy with 61 cancer-free healthy women found more severe cancer-related fatigue and disrupted circadian activity rhythms in the cancer patients both before chemotherapy and at the end of cycle 4.[43] More longitudinal studies are needed to verify to what extent the relationship between circadian rhythms and fatigue varies as a function of cancer treatments.

TREATMENT

The complaints of sleep disturbances and fatigue in cancer are often overlooked in clinical practice, and when a treatment is initiated, it is often a pharmacologic one (e.g., sedative-hypnotics for insomnia, psychostimulants for fatigue). Whereas pharmacologic therapy may be appropriate at times (e.g., short-term or occasional use), there is accumulating evidence supporting the efficacy of alternative treatments, including psychological treatments, activity-based interventions, and bright-light therapy.

Sleep

Pharmacotherapy

Hypnotic medications, particularly benzodiazepines, are by far the most commonly prescribed treatment for sleep disturbances in cancer patients. A study conducted among 1984 cancer survivors found that 41% had received a prescription for a sleeping medication since their cancer diagnosis and that 23% were currently using one.[131] The median duration of use

was 34 months, which considerably exceeds the recommendations from the 2005 National Institutes of Health state-of-the-science conference on insomnia of using hypnotic medications for no longer than 4 to 6 weeks.[132] At this conference, it was also concluded that the newer, shorter acting nonbenzodiazepines were safer and more effective than the older, longer acting benzodiazepines for the treatment of insomnia. More recently, other agents, such as a melatonin receptor agonist, have also been approved by the Food and Drug Administration for the treatment of insomnia. Although the efficacy of these medications is well established in primary insomnia, their usefulness has yet to be investigated in patients with comorbid cancer and insomnia.

Cognitive-Behavioral Therapy

The National Institutes of Health state-of-the-science conference on insomnia also concluded that cognitive-behavioral therapy (CBT) is the most effective treatment for primary insomnia.[132] There is now a large body of evidence supporting its efficacy for treating insomnia in cancer survivors.[133-141] Overall, these studies have been consistent in demonstrating that CBT (combining stimulus control, sleep restriction, cognitive restructuring, sleep hygiene, and sometimes relaxation) results in increased sleep efficiency and reduced total wake time, decreased psychological distress, and improved general quality of life. One study also showed changes in immune functioning[142] associated with CBT for insomnia, but the clinical relevance of these changes in terms of cancer prognosis or other health outcomes is unknown.

There is also some evidence showing that cancer-related fatigue may be improved after CBT for insomnia.[136,143] For instance, a study reported that 17.4% of individuals with cancer showed a remission in fatigue after CBT for insomnia compared with 2.6% of participants in the treatment-as-usual group.[113] However, nonsignificant results have also been found in other studies.[133,140] This may be due, in part, to the lack of sensitivity to change of some fatigue measures, but it still indicates that the relationship between sleep and fatigue is a complex one.

Mindfulness-Based Stress Reduction

Whereas early uncontrolled studies suggested that mindfulness-based stress reduction (MBSR) interventions could result in improved sleep quality in cancer patients,[144,145] recent studies have shown more modest results. One of them, which used a noninferiority research design, was conducted among 111 patients with various types of cancer and compared an eight-session CBT with a mindfulness-based cancer recovery intervention (MBCR, an adaptation of MBSR).[138] Short-term outcomes of MBCR on the Insomnia Severity Index were inferior to those of the CBT condition but became noninferior at the 3-month follow-up, suggesting that CBT is associated with faster improvements. Lengacher et al[146] randomly assigned 79 women who recently completed treatments for breast cancer to either a 6-week MBSR protocol or a control condition. Sleep was assessed with actigraphy, a sleep diary, and the Pittsburgh Sleep Quality Index. They found significant effects of MBSR at the 3-month follow-up on actigraphic parameters (sleep efficiency, percentage of sleep time, and number of waking bouts) but no significant effect on subjective sleep parameters. Another randomized controlled trial, conducted among 336 breast cancer patients,[147] compared an 8-week MBSR protocol with treatment as usual. Results indicated a significant and immediate benefit of the intervention on sleep as assessed with the Medical Outcomes Study Sleep Scale, but this effect was no longer significant at the 12-month follow-up.

Physical Activity

Among the cancer population, physical exercise has been found to be associated with numerous beneficial effects: improved quality of life; reduced fatigue, anxiety, pain, and depressed mood; and a possible decreased likelihood of cancer recurrence.[148,149] Only a few studies have specifically evaluated the effect of physical activity on sleep quality and quantity among cancer patients.[150-157] Although some benefits have been found, most of these trials did not select patients on the basis of minimal insomnia severity at baseline. Thus, it is not clear to what extent physical activity is potent enough to treat an insomnia disorder or chronic insomnia. In addition, questions regarding the optimal type, frequency, and dosage of exercise needed to improve sleep have yet to be answered. Comparisons with more established treatments, such as CBT, are also warranted.

Fatigue

Pharmacotherapy

There have been multiple trials of pharmacologic treatment of fatigue, reviewed by Bower.[158] Agents studied have included hematopoietic growth factors, progestational steroids, methylphenidate, and paroxetine. A meta-analysis of 27 of these randomized controlled trials concluded that treatment with hematopoietic agents led to improvements only in fatigue caused by chemotherapy-induced anemia; methylphenidate led to greater reductions in fatigue compared with placebo, whereas progestational steroids and paroxetine had no effect.[159] However, all these studies recruited patients who reported severe fatigue. Studies in cancer patients with milder levels of fatigue showed no response to pharmacotherapy.[158]

A more recent meta-analysis that included only studies of methylphenidate trials in patients with advanced-stage cancer concluded that this psychostimulant was more effective than placebo in improving cancer-related fatigue.[160] Despite this conclusion, however, only one of five studies included showed a statistically significant treatment effect. Other recent studies with larger samples of cancer patients showed no benefit for methylphenidate versus placebo for improving fatigue in general, although there was some positive effect for those with severe fatigue.[161,162]

Modafinil, a nonamphetamine-based stimulant, represents another potential treatment for cancer-related fatigue. One large multicenter trial of patients undergoing chemotherapy found beneficial effects of modafinil among patients who reported severe fatigue at baseline but not among those with mild or moderate fatigue.[163]

Overall, the results of these studies suggest that pharmacologic agents may not be the most effective first-line treatment for fatigue in cancer patients.

Nonpharmacologic Interventions

Several nonpharmacologic interventions for fatigue have been assessed in cancer patients. In 2007, a review of the literature identified a total of 41 publications, 24 assessing the efficacy of various psychological interventions, such as CBT, and 17

reporting on the efficacy of activity-based interventions.[164] Overall, the effect size obtained was of a small magnitude across all types of intervention and outcome measures (e.g., fatigue, vigor). When types of intervention were compared, a larger effect size was found for psychological interventions than for activity-based interventions. However, none of these studies selectively included severe fatigue, thus limiting the power to detect intervention effects. Moreover, no study has yet investigated the potential superior effect of an approach combining psychological and exercise-based interventions.

Possible Mechanisms of Nonpharmacologic Interventions

Not much is known about the possible mechanisms of nonpharmacologic interventions for cancer-related insomnia and fatigue. Biologic, environmental, behavioral, and cognitive factors are all potential candidates. There is some evidence showing that the effects of CBT in improving sleep would be mediated by both nonspecific (e.g., treatment expectancies) and specific (e.g., reduced maladaptive sleep habits and dysfunctional beliefs) factors.[165]

Although empiric findings are lacking, an exercise program may be beneficial for a variety of reasons, including the resynchronization of the rest-activity rhythms. A second benefit of outdoor exercises would be the increased exposure to bright light, which may promote greater daytime alertness and better sleep at night. Patients who report more fatigue tend to be exposed to less light.[166] Although the causality of sleep or fatigue and light exposure in cancer patients is not confirmed, there may be a negative feedback loop, such that less light exposure desynchronizes patients' circadian rhythms, which then causes or deteriorates fatigue and sleep impairments, and fatigue further leads to less light exposure.[166] Two small studies have shown that increased bright-light exposure, both during chemotherapy and in survivors after chemotherapy, improved or stabilized fatigue, sleep, circadian activity rhythms, and quality of life.[167-170] Larger studies are under way exploring the benefit of this easy-to-administer treatment.

DIFFERENTIAL DIAGNOSIS: IS IT SLEEPINESS, FATIGUE, OR SOMETHING ELSE?

The clinician needs to determine the cause of a patient's symptoms, recognizing that the words used by the patient to describe his or her symptoms may be vague. Is the symptom related to sleepiness (the patient may describe unintended episodes of falling asleep in the daytime or have an elevated Epworth Sleepiness Scale score) or to fatigue (complaints of muscle weakness, lack of energy, but without weakness)? Patients may also have symptoms attributable to specific effects of cancer or its treatment. When the patient has complaints related to one or more of these realms, they may become very difficult to manage.

When daytime sleepiness can be attributed to a specific sleep disorder, treatment should target that sleep disorder. If a patient has restless legs syndrome, the clinician should make sure that the patient does not have iron deficiency, which commonly occurs in gastrointestinal carcinomas. If the patient has developed movement disorders secondary to a chemotherapeutic agent, a trial of a dopaminergic agent should be initiated. If a patient has developed OSA, for example, secondary to enlarged lymph nodes in the pharynx, as might occur with lymphoma or with nasopharyngeal carcinoma, continuous positive airway pressure treatment as well as specific treatment directed at these areas should be initiated. If the patient has developed clinical depression along with insomnia, concurrent therapy for the mood disorder as well as for insomnia should be initiated. If the cancer is causing pain that is disturbing sleep, the pain needs to be treated concurrently with the insomnia. Hypoxemia caused by spread of cancer to the lung or the development of lung fibrosis in response to chemotherapy or radiation therapy may require treatment, as patients with hypoxemia are known to have disturbed sleep.

As described before, fatigue, weakness, and loss of energy are all hallmarks of cancer. Although the pathophysiology is still poorly understood, the clinician should try to determine if the fatigue is caused in part by a correctable factor, such as an electrolyte imbalance (as might occur in a patient receiving chemotherapy with severe nausea) or vomiting, an underlying infection, or an undiagnosed metabolic disorder (such as thyroid disease or diabetes mellitus). Diabetes mellitus may develop as a result of some types of therapy, such as large doses of corticosteroids. Fatigue might also be exacerbated by poor sleep, in which case treating sleep disturbances might result in improvements in fatigue.

PITFALLS AND CONTROVERSIES

Although the numbers of research studies in the last few years have increased, there is still much that remains unknown about the cause, consequences, and cures of sleeping difficulties and fatigue in patients with cancer. In particular, more longitudinal studies are needed to characterize the natural course of sleep complaints, circadian rhythms impairments, and fatigue during the cancer trajectory and to better understand how these disturbances are interrelated. Areas for future research also include a better characterization of those disorders across cancer sites. More research is also warranted on patients with advanced cancer, including clinical studies investigating the efficacy of nonpharmacologic interventions for sleep disturbances and fatigue, as it is unclear whether the same treatment modalities can be offered to these patients. Finally, mechanisms through which these interventions are effective also deserve investigation.

CLINICAL PEARL

Fatigue and sleep disturbances are among the most common and most distressing complaints of patients with cancer. When left untreated, these symptoms can significantly impair patients' quality of life. Clinicians should screen routinely for these disturbances and administer evidenced-based treatment strategies to help patients coping with them.

SUMMARY

Significant progress has been made in the last decades in the study of sleep and fatigue in cancer. Although causes and mechanisms of the sleep disruption or the fatigue experienced

by these patients remain to be fully elucidated, the available evidence suggests that both are common. Sleep disruption and fatigue are major complaints in patients with cancer, before treatment, while undergoing chemotherapy or radiation therapy, and after the completion of therapy. The relationship between fatigue and the quality or quantity of sleep or between fatigue and the sleep-wake circadian rhythm cycle is still unclear. Some of the cancer-related fatigue may be related to disturbed sleep or to disturbed sleep-wake rhythms. In cancer patients, as in other medically ill individuals, disturbed sleep may be important not only to the expression of fatigue but to the patients' quality of life, to their tolerance to treatment, and to the development of mood disorders, particularly clinical depression. Disruptions in circadian rhythms themselves affect sleep quality as well as many other physiologic mechanisms that pertain to fatigue. The degree of sleep disruption found in patients with cancer is not trivial. Objectively recorded sleep and biologic rhythms have not been well investigated in patients with cancer, but because it appears that a large proportion of them may in fact not be getting a good night's sleep, more research is needed to better characterize the sleep disruption and to establish the efficacy of various treatment approaches to improve sleep in this population.

Selected Readings

Bower JE. Cancer-related fatigue—mechanisms, risk factors, and treatments. *Nat Rev Clin Oncol* 2014;**11**:597–609.

Casault L, Savard J, Ivers H, Savard MH. A randomized-controlled trial of an early minimal cognitive-behavioural therapy for insomnia comorbid with cancer. *Behav Res Ther* 2015;**67**:45–54.

Howell D, Oliver TK, Keller-Olaman S, et al. Sleep disturbance in adults with cancer: a systematic review of evidence for best practices in assessment and management for clinical practice. *Ann Oncol* 2014;**25**:791–800.

Liu L, Rissling M, Natarajan L, et al. The longitudinal relationship between fatigue and sleep in breast cancer patients undergoing chemotherapy. *Sleep* 2012;**35**:237–45.

Miller AH, Ancoli-Israel S, Bower JE, et al. Neuroendocrine-immune mechanisms of behavioral comorbidities in patients with cancer. *J Clin Oncol* 2008;**26**:971–82.

Palesh OG, Peppone L, Innominato PF, et al. Prevalence, putative mechanisms, and current management of sleep problems during chemotherapy for cancer. *Nat Sci Sleep* 2012;**4**:151–62.

Savard J, Ivers H, Savard MH, Morin CM. Cancer treatments and their side effects are associated with aggravation of insomnia: results of a longitudinal study. *Cancer* 2015;doi:10.1002/cncr.29244.

Savard J, Ivers H, Villa J, et al. Natural course of insomnia comorbid with cancer: an 18-month longitudinal study. *J Clin Oncol* 2011;**29**:3580–6.

Savard J, Savard M-H. Insomnia and cancer: prevalence, nature, and non-pharmacological treatment strategies. *Sleep Med Clin* 2013;**8**:373–87.

A complete reference list can be found online at ExpertConsult.com.

Chapter
131

Fibromyalgia and Chronic Fatigue Syndromes

Christine Won; Douglas Kirsch

Chapter Highlights

- Sleep-related complaints and sleep disturbances are principal features of fibromyalgia and chronic fatigue syndrome. (There is some debate about the relationship of nomenclature between myalgic encephalomyelitis and chronic fatigue syndrome; this chapter will refer to it solely as chronic fatigue syndrome.) Sleep quality affects patients' pain and fatigue symptoms as well as quality of life. The pathophysiologic role of sleep disturbances in fibromyalgia and chronic fatigue syndrome is currently unknown despite the recognition of this as a frequent symptom.
- Several polysomnographic findings have been described in patients with fibromyalgia and chronic fatigue syndrome. The relevance of sleep electroencephalogram findings to daytime symptoms of fibromyalgia and chronic fatigue syndrome are not well understood. Conventional methods for sleep stage scoring or quantifying electroencephalogram may not be sensitive enough to detect important sleep changes in patients with fibromyalgia or chronic fatigue syndrome.
- Comorbid sleep disorders may affect sleep quality of patients with fibromyalgia and chronic fatigue syndrome, and treating primary sleep disorders may augment therapy for these syndromes. Therapy includes pharmacologic and nonpharmacologic approaches that directly target sleep disorders or sleep complaints.
- This chapter illustrates the complexities of identifying, measuring, and treating sleep disturbances in syndromes defined by chronic pain and fatigue.

FIBROMYALGIA AND SLEEP

Fibromyalgia (FM) is characterized by generalized pain, chronic fatigue, and nonrestorative sleep. The 1990 American College of Rheumatology criteria defined FM as the presence of more than 11 of 18 tender points, which are bilateral, occur above and below the waist, involve axial sites, are distributed widely, and occur for at least 3 months.[1] An alternative criteria published in 2011 uses a patient-reported survey asking about locations of pain and the presence and severity of fatigue, sleep disturbances, memory deficits, headaches, irritable bowel, and mood problems. The new criteria do not require tender-point examination; however, they identify most of those individuals who would meet 1990 criteria, while including many others.[2] The global prevalence of FM is approximately 3% to 8%. Approximately 75% of patients with FM are women, and most patients are aged 30 to 50 years.[3]

Although current diagnostic criteria for FM do not include sleep-related abnormalities, sleep disturbances and complaints of poor sleep quality and daytime fatigue are common. In fact, sleep-related complaints are second only to pain.[4] Nonrestorative sleep correlates closely with somatic symptoms and pain severity. A specific etiology for FM is unknown, although it is speculated that disturbances in central nervous system (CNS) function likely contribute. The pathophysiologic role of sleep disturbances in FM is currently unknown despite the recognition of this as a frequent symptom.

Polysomnographic Features of Fibromyalgia

Patients with FM compared with healthy individuals demonstrate greater disturbances during sleep on polysomnography. Sleep-onset latency and stage N1 tend to be increased, whereas total sleep time and sleep efficiency are reduced. There are more frequent arousals and awakenings as well as longer awakening periods. Patients with FM have three times more arousals per hour than healthy subjects.[5] They have an increased number of sleep-stage shifts and a shorter duration of N2 sleep stage.[6] In addition, polysomnography data suggest that patients with FM have reduced total rapid eye movement (REM) and slow wave sleep time corrected for age, and these changes are associated with increased musculoskeletal and mood symptoms.

Several findings related to sleep electroencephalography (EEG) microstructure have been described using power spectral and frequency domain analysis. Beta frequency (14 to 38 Hz) is generally considered to reflect arousal; it is associated with lightened sleep perception and is linked with depression. When muscle and joint pain are experimentally provoked during sleep, beta power and alpha (8 to 13 Hz) frequency bands are increased on spectral analysis, while delta (0.5 to 4 Hz) and sigma (12 to 14 Hz) power is reduced.[7] Sigma frequency, likely reflecting sleep spindles, is generated from corticothalamic networks, appears to be responsible for lack of perceptual awareness and unresponsiveness during

sleep, and is associated with perception of greater sleep depth. Thus a reduction in sigma frequency suggests a higher level of alertness during sleep in patients with FM. Patients with FM also demonstrate fewer sleep spindles per minute of stage N2 sleep and reduced spindle frequency even when controlling for age, depression, and psychiatric diagnosis. Reduced spindle number and spindle frequency activity in stage N2 sleep may affect sensory processing in the thalamus and is associated with a lower pain threshold.[8]

Occipital alpha frequency (8 to 13 Hz), which is associated with relaxed wakefulness with eyes closed in healthy individuals, has been found to intrude on the sleep EEG in patients with FM.[9,10] Alpha intrusion has been identified in 70% of patients with FM compared with 16% of control subjects.[10] Two patterns of alpha activity during sleep have been described: (1) a phasic alpha pattern or alpha-delta sleep, which describes a pattern of alpha activity superimposed on delta waves of slow wave sleep (found in 71% of patients with FM with alpha intrusion); (2) a tonic alpha pattern, in which the alpha frequency occurs throughout NREM sleep (found in 29% of patients with FM with alpha intrusion). Alpha-delta sleep is associated with worse sleep efficiency, perception of lighter sleep, longer pain duration, and increased morning stiffness and pain. In patients with FM, but not in healthy controls, the alpha-delta ratio increases exponentially through the night.[11] Nonrestorative sleep has been described in 100% of patients with FM with the alpha-delta pattern, 25% with the tonic alpha pattern, and 58% of patients with FM without significant alpha intrusion. Alpha intrusion is also associated with symptoms of vigilance during sleep manifested as an increased tendency to wake to an external response and the perception of shallow or unrefreshing sleep.[12] There is no association between the occurrence of depressed mood or memory complaints in patients with FM and any of the alpha frequency patterns.[13]

Whether alpha intrusion is causative or results from chronic pain is unclear. When painful stimuli to muscles and joints are applied during slow wave sleep, an arousal effect with decreased delta waves and increased alpha activity is observed. Stimulation of superficial pain to the skin, on the other hand, does not elicit the same EEG response during sleep, suggesting that deep but not superficial pain alters sleep architecture.[7] Alternatively, alpha sleep may predispose the individual to increased arousability because of pain or other stimuli.

Not all investigators agree that the alpha EEG findings during sleep are particularly significant because they do not occur in all patients with FM. Additionally, this signal pattern also occurs in normal individuals and in patients with other pain syndromes such as rheumatoid arthritis.[14] The alpha intrusion pattern may be observed in 15% of normal subjects. Approximately 40% of subjects with alpha intrusion have pain syndromes, and most subjects exhibiting this pattern have a non-pain-related medical or psychiatric condition.[15] The occurrence of alpha intrusion in normal subjects is associated with subjective complaints of disturbed sleep.[16] Furthermore, alpha intrusion may be elicited by auditory or deep pain stimuli during slow wave sleep in normal individuals.

Therefore the alpha EEG findings, although not specific to FM, correlate with the subjective feeling of nonrestorative sleep and are observed in most patients with FM. The amount of this rhythm correlates with objective measurements of pain, and decreasing the amount of alpha intrusion with medications results in perceived improvement in sleep and pain. The biochemical and cellular processes that occur with this EEG finding are unknown and warrant further study.

Cyclic alternating pattern (CAP) describes a phenomenon whereby the EEG activity is periodic within NREM sleep. CAP is characterized by sequences of transient electrocortical events that are distinct from background EEG activity and recur at up to 1-minute intervals. This periodic activity, originally thought to be arousals, is now theorized to be the process of sleep maintenance and sleep fragmentation. CAP phase A1 pattern is considered to be an index of sleep stability, whereas CAP phases A2 and A3 are markers of sleep instability or poor sleep quality.[17] CAP A2 and A3 are increased in patients with FM compared with controls. Greater frequency of these subtypes are associated with increased number of tender points in patients with FM.[18] Furthermore, when A2 and A3 subtypes are decreased pharmacologically, patients with FM describe improvements in fatigue and in Hospital and Anxiety and Depression Scale (HADS) score.[19] Some scientists suspect that excessive phasic EEG activity may reflect enhanced or abnormal activity within the insula, anterior cingulate cortex, and ventromedial prefrontal cortex, which are in some way responsible for modulating abnormal nociceptive processing. However, it remains unclear whether CAP EEG findings reflect disturbed sleep processes and are pathogenic in FM, or whether they are consequences of pain and other syndromes found in patients with FM.

Pathophysiologic Correlates of Altered Sleep in Fibromyalgia

Neurologic mechanisms that regulate sleep and are in turn affected by sleep disturbances may play a role in the etiology and maintenance of FM. Several studies have shown that patients with FM have decreased levels of serotonin and the serotonin metabolite, 5-hydroxyindole-3-acetic acid, in the cerebrospinal fluid and blood serum.[20,21] Serotonin modulates pain transmission, and decreased serotonin synthesis induces a hyperalgesic state as well as insomnia. The precursor to serotonin, 5-hydroxytryptamine, is released at axonal nerve endings in the basal hypothalamus as a neurotransmitter during wakefulness and is important in the sleep state–dependent modulation of the nociceptive process of pain. A reduction in serotonin may be partially responsible for reducing delta sleep and predisposing to an alpha EEG rhythm. High levels of substance P are found in the cerebrospinal fluid of patients with FM and may contribute to serotoninergic deficiency.[22] Substance P appears to be a biologic marker for the presence of chronic pain.[23,24] Studies suggest that some individuals have a genetic predisposition for FM through the elevated frequency of polymorphisms in the serotonin transporter, dopamine D4 receptor, and catechol-*O*-methyltransferase genes.[25,26] Together, these findings suggest serotonin plays a role in promoting both analgesia and sleep.

Serotonin is transformed into melatonin in the pineal gland. The decrease in serotonin in patients with FM has been speculated to result in decrease melatonin synthesis, leading to abnormal sleep patterns. Some studies have described low nocturnal melatonin peak levels and a decrease in overall melatonin secretion in patients with FM.[27] In addition, melatonin has been used successfully to treat sleep-wake cycle

disturbances in some patients with FM and has resulted in improved pain.[28] However, other studies have found no difference in melatonin levels between patients with FM and healthy controls and have not found an association between melatonin levels of patients with FM and disease duration, sleep disturbances, fatigue, and pain scores.[29,30]

There is substantial evidence indicating that sensitization of the central nervous system pain pathway plays a crucial role in the pathophysiology of FM.[31] Neuroimaging studies show that patients with FM may exhibit abnormal perception of painful stimuli through central changes that result in increased pronociception or through decreased activity in descending pain inhibitory pathways such as the anterior cingulate cortex and thalamus.[32] There is altered regional cerebral blood flow to the thalamic areas where encoding and inhibiting pain transmission occurs and to areas where ablation has been shown to result in persistent insomnia in cats.[33,34] Altered thalamic regional cerebral blood flow may cause loss of growth hormone (GH) secretion during slow wave sleep and result in sleep alterations in patients with FM. Somatomedin C or insulin-like growth factor-1, which regulates GH production, is low in about one third of patients with FM,[35] and exogenous GH may improve sleep quality in patients with FM.[36]

Fibromyalgia may be considered a stress-related disorder. Stressors such as physical trauma, hormonal alterations, and emotional stress play an important role in triggering the development of FM and its somatic symptoms, including widespread chronic pain. Stress may directly affect the hypothalamic-pituitary-adrenal axis and the autonomic system, which induces central alterations that blunt cortisol, lead to abnormal regulation of GH, and cause heart rate variation and sensitivity to orthostatic changes on tilt-table tests.[37,38] Stress may directly affect sleep, causing sleep disruption especially during slow wave sleep, and further modulate the hypothalamic-pituitary-adrenal axis and exacerbate musculoskeletal pain and fatigue symptoms. Sleep disturbances are also related to disturbances in mood and cognition and may affect coping mechanisms to stress.

Common Sleep Complaints and Sleep Disorders in Fibromyalgia

One of the most prevalent and clinically challenging complaints in this patient population is unrefreshing or nonrestorative sleep, which is present in more than 75% of patients with FM.[39] Patients often report feeling unrested or worse after a night of sleep. In addition, patients with FM commonly report sleep fragmentation, early morning awakenings, and insomnia.[40] Despite reporting significant subjective sleepiness and fatigue, however, patients with FM demonstrate less objective daytime sleepiness on Multiple Sleep Latency Tests (MSLT) than healthy controls.[41]

Primary sleep disorders, such as sleep apnea, restless legs syndrome (RLS), and periodic limb movements of sleep, may be found in patients with FM. It has been reported that approximately 2% of women presenting to a rheumatology clinic with a new diagnosis of FM will have sleep apnea.[42] Women with FM may be more likely to have other sleep-related breathing disorders such as oxygen desaturation, inspiratory airflow limitation, and upper airway resistance syndrome, rather than frank sleep apnea. Treatment of sleep disordered breathing with positive airway pressure (PAP) therapy improves functional outcomes in female patients with FM.[43] In contrast, 44% of men presenting to a rheumatology clinic with a new diagnosis of FM have been purported to have sleep apnea,[44] although this finding has not been consistently supported. FM was found only in approximately 3% of patients presenting at a pulmonary-based sleep disorders clinic.[45] Thus sleep apnea and other types of sleep-disordered breathing may be found in select patients with FM, but current evidence is not convincing for a direct association between sleep apnea and FM.

RLS has been described to be more frequent in female patients with FM (20% to 64%) compared with women in the general population (2% to 15%).[46-48] Genetic studies suggest a common genetic background and coheritability of FM and RLS.[49,50] The occurrence of RLS in patients with FM leads to worse sleep quality and quality of life.[51] The reason for the high prevalence of RLS in this population is unknown but may result from common central processes such as dysfunction of the dopaminergic system. There are emerging data suggesting that the experience of pain among patients with FM may be in part due to dysfunction in the release of dopamine within the CNS. When subjected to deep muscle pain, normal subjects release dopamine in the basal ganglia, whereas fibromyalgia patients do not. In normal subjects, the amount of dopamine release correlates with the amount of perceived pain however there is no such correlation in fibromyalgia patients.[52]

Approach to the Management of Sleep Disturbances in Fibromyalgia

The general goal of FM treatment is to develop an individualized therapeutic program that involves various approaches. Fibromyalgia symptoms of pain and mood disturbances and sleep disturbances affect quality of life, including the patient's ability to work, to participate in everyday activities, and to maintain relationships. For these reasons, therapies directed at alleviating sleep symptoms are integral to the treatment plan of patients with FM.

Actigraphy may aid in identifying issues related to sleep-wake schedules and sleep maintenance issues. Sleep quality reported by patients with FM is directly related to total sleep time and sleep fragmentation, and fatigue in patients with FM is directly related to wake time after sleep onset and sleep efficiency measured by actigraphy.[53] Addressing sleep management issues, such as a regular sleep-wake and light-dark schedule, sufficient sleep time, a sleep-conducive bedroom environment, and avoidance of alcohol, caffeine, nicotine, and other substances that might interfere with sleep, is essential in the management of sleeping difficulties in patients with FM. Other lifestyle recommendations may include daily morning or early afternoon exercise, avoiding heavy meals before bedtime, and nighttime relaxation exercises. Finally, the patient's medications should be reviewed to determine whether medications are contributing to sleep fragmentation or daytime fatigue.

At present, only three drugs—pregabalin, duloxetine, and milnacipran–are approved by the U.S. Food and Drug Administration to reduce FM symptoms.[54] Pregabalin is an anticonvulsant that downregulates presynaptic excitatory neurotransmitter release through its action on an $\alpha_2\beta$ receptor that regulates flow of calcium channels.[55] It is found to be effective in reducing the severity of the two major FM symptom domains: pain and sleep disturbance.[56] Duloxetine

and milnacipran, serotonin-norepinephrine reuptake inhibitor antidepressants, increase the availability of serotonin and norepinephrine at CNS synaptic clefts. They have the potential to reduce pain by correcting the functional deficit of 5-hydroxytryptamine and norepinephrine neurotransmission in the descending inhibitory pain pathway.[57]

In addition, several pharmacologic agents have been tested specifically for their potential to improve sleep in patients with FM. These medications include chlorpromazine, amitriptyline, zopiclone, fluoxetine, mirtazapine, trazodone, and cyclobenzaprine. Chlorpromazine has been shown to decrease alpha-delta EEG frequencies, while reducing pain and trigger point tenderness.[58] Zopiclone has been shown to marginally increase delta sleep, improve subjective sleep quality, and decrease awakenings during the night.[59] Tricyclic antidepressant medications such as amitriptyline, which affects serotonin metabolism in the CNS, reduce alpha-delta sleep time and promote restorative sleep in patients with FM, although effects appear to decline after 6 months of therapy.[60,61] Fluoxetine has been shown to help sleep and depressed mood in women with FM.[62] Amitriptyline and fluoxetine combined have resulted in significant improvement among patients with FM in general sleep scores and appeared to be superior to either drug alone.[63] Mirtazapine has a pharmacologic profile that includes serotonergic and antihistaminic properties, which lead to antidepressive as well as sedative effects. Mirtazapine has been shown to reduce pain and fatigue and to improve sleep quality in most patients with FM.[64] Trazodone has been shown to markedly improve sleep duration, sleep efficiency, and sleep quality based on the Pittsburgh Sleep Quality Index in patients with FM. Improvements with the use of trazodone were also observed on the Fibromyalgia Impact Questionnaire, the Beck Depression Inventory, the HADS, and pain interference with daily activities.[65] Trazodone plus pregabalin administered over 12 weeks showed significant improvement in sleep quality, depression, pain, and global fibromyalgia severity.[66] Very low-dose cyclobenzaprine (1 to 4 mg) at bedtime resulted in improvement in CAP sleep EEG and improved FM symptoms. Responding patients showed lower frequency of CAP A2 and A3 and reported improved fatigue and HADS depression score. CAP changes appeared to be a useful biomarker that predicted treatment benefit.[19] These findings underscore the clinical benefit of treating sleep complaints and normalizing sleep architecture in FM.

Sodium oxybate has been evaluated in double-blind randomized placebo-controlled trials for treatment of FM. Sodium oxybate has been noted to reduce tender points and fatigue symptoms and to improve sleep quality in patients with FM. Six of seven pain and fatigue scores (overall pain, pain at rest, pain during movement, end-of-day fatigue, overall fatigue, and morning fatigue) were relieved for approximately one third of subjects treated with sodium oxybate compared with 6% to 10% in the placebo group.[67,68] Sodium oxybate also led to improved scores on the Epworth Sleepiness Scale, Jenkins scale for sleep, Functional Outcomes of Sleep Questionnaire, and Short Form-36 health questionnaire. Doses of both 4.5 and 6 g split into two doses (bedtime and 2.5 to 4 hours later) improved outcomes, although only the 6-g dose improved sleep efficiency, stage N2 sleep, and slow wave sleep.[69] This pharmacologic therapy also significantly decreased alpha intrusion, sleep latency, and REM sleep.

Nonpharmacologic therapies for improving sleep and FM symptoms have been explored. Spinal and extremity joint manipulation or mobilization, massage, and various soft tissue techniques improve pain intensity, FM symptom impact, depressive symptoms, and sleep quality. Sex differences are observed in response to these treatments, with women having a greater reduction in pain and FM symptoms than men, although sleep quality improve equally in men and women.[70]

Screening for primary sleep disorders, such as obstructive sleep apnea, restless legs syndrome, and circadian disorders, is important in patients with FM. Even mild inspiratory flow limitation is associated with sleep fragmentation, arousals, fatigue, and excessive daytime sleepiness, notably in women with FM. Nasal PAP therapy in patients with sleep-disordered breathing may improve fatigue, pain, sleep fragmentation, disability, and Rheumatology Distress Index scores.[43] When patients with FM and RLS are treated with pramipexole, a dopamine-3 receptor agonist, they experience significant improvement in measures of pain, fatigue, and functional and global status.[71]

In conclusion, screening for and treating primary sleep disorders, implementing good sleep hygiene recommendations, and pharmacologic management of FM may improve sleep quality and FM symptoms. Treatment strategies should be individualized to improve the patient's quality of life.

CHRONIC FATIGUE SYNDROME AND SLEEP

Chronic fatigue syndrome (CFS) is a disabling condition characterized by severe fatigue lasting for more than 6 months and the presence of at least four out of eight minor criteria. CFS may overlap with other chronic somatic syndromes, such as fibromyalgia syndrome and irritable bowel disorder. Diffuse muscular pain, fatigue, and sleep disturbances are part of the definitions of both CFS and FM. As a result, 20% to 70% of patients with FM meet criteria for CFS, and 35% to 70% of those with CFS have coexistent FM.[72,73] These conditions may constitute continuums of the same physiologic and psychosocial processes, with varying degrees of pain, fatigue, and disturbed sleep.

Complaints of unrefreshing or nonrestorative sleep are very common in patients with CFS. Disturbed sleep is a known cause of fatigue and may play a role in the pathogenesis of CFS. However, the nature of the sleep impairment in CFS remains unknown. Although complaints of unrefreshing sleep are pervasive, there are no apparent neurophysiologic correlates of dysfunctional sleep in CFS, and there are no polysomnographic findings that discriminate between subjects with CFS and normal controls. These observations suggest that psychosocial factors may affect the perception of poor sleep quality. Primary sleep disorders occur often in patients with CFS and may contribute to the presence of daytime dysfunction. However, there is currently little evidence to indicate that treating primary sleep disorders improves fatigue associated with CFS. Therefore primary sleep disorders are likely comorbid rather than an associated condition of CFS.

Polysomnographic Findings in Chronic Fatigue Syndrome

Despite the fact that sleep complaints are frequently reported in patients with CFS, there are no specific or consistent polysomnographic findings that characterize sleep in patients with

CFS. Polysomnographic EEG may show increased number and duration of awakenings.[74,75] Reduced total sleep time and sleep efficiency, both in single night in-laboratory PSG recording and in-home studies, have been described.[76,77] Sleep-onset latencies are often longer in patients with CFS compared with healthy control subjects.[78] It has been reported that patients with CFS may have altered parasympathetic activity at sleep onset, which may contribute to difficulty with sleep initiation.[79] This finding may also reflect poor sleep hygiene or daytime napping, which may decrease nocturnal sleep drive. Based on limited data, there appear to be very few differences or mixed findings in sleep architecture between CFS patients and healthy individuals. Shorter durations of NREM stage 2 have been shown in those with CFS and FM, but not in those with CFS alone. Although total slow wave sleep time has been shown to be unchanged in monozygotic twins discordant for CFS, when the twin with CFS is intentionally sleep delayed by 4 hours, there is decreased slow wave activity during the first NREM period, which suggests a potential deficiency in sleep homeostasis.[80] Some have reported reduced REM sleep in patients with CFS,[76] whereas others have observed a higher percentage of REM sleep in patients with CFS compared with controls.[81-83] Probabilities and rates of sleep-state transitions between wake, N1, and REM into N2 sleep, and transitions from N3 into wake or N1, may also be greater in those with CFS and comorbid FM. In those with CFS without FM, sleep-state transitions from REM to wake may be greater.[84] CAP EEG has not been explored extensively in CFS; existing limited data suggest possibly an increase in CAP during NREM sleep in patients with CFS compared with healthy controls.[77] Therefore a single-night polysomnography performed with traditional technology does not appear useful for discriminating individuals with CFS from healthy controls or from other chronic pain or fatigue conditions, nor do these subtle differences in sleep architecture and EEG findings provide a correlate for sleep complaints in patients with CFS.[85]

Power spectrum analysis of the EEG does not seem to provide strong evidence for abnormal alpha intrusion in NREM sleep in subjects with CFS. No significant differences in spectral power in any frequency band have been found between cotwins with and without CFS.[86] In fact, some studies showed a reduction in alpha power by spectrum analysis during NREM stages 2 and slow wave sleep and during REM sleep in CFS subjects.[77,87] Analysis of other spectral bands (theta, delta, beta) shows inconsistent results. For example, delta power has been reported to be increased, decreased, or no different in CFS groups compared with normal controls.[86-88] Therefore, conventional methods for sleep-stage scoring or quantifying EEG have failed to find specific findings or may not be sensitive enough to detect sleep changes in patients with CFS.

Biochemical Correlates of Altered Sleep in Chronic Fatigue Syndrome

The pathogenesis of CFS remains unknown. Abnormalities of the central and autonomic nervous systems, and in some cases infections, have been hypothesized.[89] Adolescent patients with CFS often have findings of postural orthostatic tachycardia syndrome, suggesting autonomic dysfunction and a potential neutrally mediated cause of chronic fatigue.[90] The cognitive behavior model for CFS suggests predisposing and perpetuating factors. The onset may be related to an infection or other CNS abnormalities, but the perpetuation of the condition may be determined by psychosocial factors such as maladaptive behavior or negative conditioning.[91] Maladaptive responses include reduced physical activity, leading to autonomic dysfunction, increased nervous system sensitivity, and increased CFS symptom burden.

Common Sleep Complaints and Sleep Disorders in Chronic Fatigue Syndrome

Unrefreshing or nonrestorative sleep even after a night of sufficient sleep duration is a frequent complaint occurring in 87% to 95% of CFS patients.[92,93] Nocturnal neurophysiologic disturbances that result in the nonrestorative sensation following sleep in CFS patients are not detected by traditional sleep monitoring, suggesting current measurements may not be sufficiently sensitive to detect subtle disturbances in sleep in this population.

Subjects with CFS with complaints of nonrestorative sleep have a very high co-occurrence of complaints of daytime dysfunction. Objective testing has shown cognitive impairments in attention, motor functioning, information processing, and executive functioning.[94] Pathologic sleepiness, however, is not objectively demonstrated on MSLT despite CFS patients reporting greater subjective sleepiness and poorer sleep quality than healthy controls.[95,96] One potential reason for this may be that patients with CFS may suffer sleep quality misperception. Poor self-rated health and depressive symptoms have been found to be associated with overreporting of sleep difficulties and underestimation of sleep efficiency.[97] Alternatively, current measures may not reliably differentiate between fatigue and sleepiness, resulting in discordances between subjective and objective sleep-related measurements.[98]

Actigraphy studies in children with CFS report a sleep pattern consisting of more than 10 hours of continuous sleep.[99] Impaired daily sleep-wake rhythms and disturbed sleep were observed only in some children with CFS. Most actigraphy studies to date have shown mixed findings relating to circadian rhythm disturbances.[100-102]

Polysomnography studies on population-based subjects have shown the presence of a primary sleep disorder in approximately 18% of patients with CFS compared with 7% of controls.[83] In the clinical setting, 46% to 81% of patients with CFS have comorbid primary sleep disorders such as primary insomnia, obstructive sleep apnea, periodic limb movement disorder, or narcolepsy.[74,103] Organic sleep disorders, such as sleep apnea and narcolepsy, were once considered exclusionary for the diagnosis of CFS. This concept is challenged; it has been found that there are no significant differences between patients with CFS in the presence or absence of OSA with regard to subjective sleep variables, CFS symptoms, indexes of anxiety and depression, and Short Form-36 quality-of-life measures.[104] Although there is evidence to indicate that treatment of sleep-related breathing disorders with PAP therapy may reduce fatigue and increase vigor in otherwise healthy subjects, it is not clear whether primary fatigue due to CFS is similarly responsive to treatment of comorbid sleep disorders.[105] However, because of their high frequency and their potential compounding of symptoms in patients with CFS, a thorough evaluation for underlying organic sleep disorders and aggressive treatment are warranted.

CLINICAL PEARLS

- Sleep-related complaints such as poor sleep quality and nonrestorative sleep are pervasive in fibromyalgia and chronic fatigue syndrome and may be associated with complaints of pain and fatigue.
- Patients with fibromyalgia demonstrate greater sleep disturbances on polysomnography such as reduced total sleep time, sleep efficiency, REM and slow wave sleep, and increased arousals and alpha intrusion.
- Decreased serotonin levels in the CNS may contribute to altered nociception and sleep disturbances in fibromyalgia.
- Therapies targeted at improving sleep, including treating comorbid sleep disorders, are important adjuncts to treatment of fibromyalgia and chronic fatigue syndrome.

SUMMARY

Sleep disturbances, sleep complaints, and comorbid sleep disorders are common in patients with fibromyalgia and chronic fatigue syndrome. Many characteristic daytime symptoms, such as chronic pain and fatigue, may be related to nonrestorative sleep patterns associated with the diseases. Pain and sleep disturbances interact reciprocally—pain affects sleep processes, and sleep disturbances affect pain threshold—to influence disease severity and chronicity. Sleep disturbances therefore may contribute to fatigue, psychological disturbances, and impaired quality of life. To the extent that nonrestorative sleep may exacerbate fibromyalgia or chronic fatigue symptoms, treatments that directly improve sleep quality may improve daytime symptoms of these disorders.

Selected Readings

Ahmed M, Aamir R, Jishi Z, Scharf MB. The effects of milnacipran on sleep disturbance in fibromyalgia: a randomized, double-blind, placebo-controlled, two-way crossover study. *J Clin Sleep Med* 2015 Sep 14. [Epub ahead of print].

Anderson RJ, McCrae CS, Staud R, et al. Predictors of clinical pain in fibromyalgia: examining the role of sleep. *J Pain* 2012;**13**(4):350–8.

Bigatti SM, Hernandez AM, Cronan TA, Rand KL. Sleep disturbances in fibromyalgia syndrome: relationship to pain and depression. *Arthritis Rheum* 2008;**59**(7):961–7.

Falk Hvidberg M, Brinth LS, Olesen AV, et al. The health-related quality of life for patients with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *PLoS One* 2015;**10**(7):e0132421.

Haney E, Smith ME, McDonagh M, et al. Diagnostic methods for myalgic encephalomyelitis/chronic fatigue syndrome: a systematic review for a National Institutes of Health Pathways to Prevention workshop. *Ann Intern Med* 2015;**162**(12):834–40.

Jason LA, Kot B, Sunnquist M, et al. Chronic fatigue syndrome and myalgic encephalomyelitis: toward an empirical case definition. *Health Psychol Behav Med* 2015;**3**(1):82–93.

Liedberg GM, Björk M, Börsbo B. Self-reported nonrestorative sleep in fibromyalgia—relationship to impairments of body functions, personal function factors, and quality of life. *J Pain Res* 2015;**8**:499–505.

Reeves WC, Heim C, Maloney EM, et al. Sleep characteristics of persons with chronic fatigue syndrome and non-fatigued controls: results from a population-based study. *BMC Neurol* 2006;**6**:41.

Roth T, Lankford DA, Bhadra P, et al. Effect of pregabalin on sleep in patients with fibromyalgia and sleep maintenance disturbance: a randomized, placebo-controlled, 2-way crossover polysomnography study. *Arthritis Care Res (Hoboken)* 2012;**64**:597–606.

Russell IJ, Larson AA. Neurophysiopathogenesis of fibromyalgia syndrome: a unified hypothesis. *Rheum Dis Clin North Am* 2009;**35**:421–35.

Scharf MB, Baumann M, Berkowitz DV. The effects of sodium oxybate on clinical symptoms and sleep patterns in patients with fibromyalgia. *J Rheumatol* 2003;**30**:1070–4.

Shah MA, Feinberg S, Krishnan E. Sleep-disordered breathing among women with fibromyalgia syndrome. *J Clin Rheumatol* 2006;**12**:277–81.

Smith ME, Haney E, McDonagh M, et al. Treatment of myalgic encephalomyelitis/chronic fatigue syndrome: a systematic review for a National Institutes of Health Pathways to Prevention workshop. *Ann Intern Med* 2015;**162**(12):841–50.

Unger ER, Nisenbaum R, Moldofsky H, et al. Sleep assessment in a population-based study of chronic fatigue syndrome. *BMC Neurol* 2004;**4**:6.

Vijayan S, Klerman EB, Adler GK, Kopell NJ. Thalamic mechanisms underlying alpha-delta sleep with implications for fibromyalgia. *J Neurophysiol* 2015;**114**(3):1923–30.

Wagner JS, DiBonaventura MD, Chandran AB, Cappelleri JC. The association of sleep difficulties with health-related quality of life among patients with fibromyalgia. *BMC Musculoskelet Disord* 2012;**13**:199.

A complete reference list can be found online at ExpertConsult.com.

Chapter
132

Endocrine Disorders

Adriana G. Ioachimescu; Octavian C. Ioachimescu

Chapter Highlights

- The hypothalamus plays a major role in integrating sleep and various metabolic functions.
- Most hypothalamic and pituitary hormones exert important effects on sleep-wake homeostasis.
- Sleep-disordered breathing is very common in acromegaly (both obstructive and central sleep apnea).
- The association between hypothyroidism and obstructive sleep apnea remains controversial.
- The hypothalamic-pituitary-adrenal axis plays a significant role in stress coping, with major influences on sleep homeostasis.
- Low testosterone levels and obesity play important roles in obstructive sleep apnea.

The endocrine system can be conceptualized as an array of hypothalamus-pituitary-target organ axes, with the target organs being represented by the "effector" endocrine glands (e.g., thyroid, adrenal, gonads, adipose tissue; Figure 132-1). These effects are important in the sleep neurobiology because some of these hormones are neuropeptides and may have significant effects. We discuss mainly the endocrine effects of these hormones, whereas the paracrine, autocrine, and local neuromodulating effects of these factors are beyond the scope of this chapter. We will not discuss here in great detail the endocrine physiology in relationship to sleep or sleep deprivation, which are reviewed in detail in Chapter 20.

HYPOTHALAMIC DISORDERS AND SLEEP

Fine coordination between metabolic processes and sleep-wake states involves at least one major player: the hypothalamus, which is considered the grand regulator of the master endocrine gland, that is, the hypophysis (or pituitary gland). The hypothalamus is involved in regulating several metabolic processes: thirst, hunger and satiety, thermoregulation, autonomic nervous system, response to stress, and modulation of emotional behaviors. The hypothalamus controls sleep and metabolism in two different ways: directly (using neural connections that control arousal and metabolic pathways) and indirectly (using appetite and behavioral controls, stress-coping mechanisms, and several reward systems).

The homeostatic role of the hypothalamus is exerted by integrating inputs from peripheral and central circadian oscillators, data on current sleep debt (or "homeostatic sleep pressure," process S; see Chapter 36), autonomic nervous system, caloric intake, and body temperature changes and by orchestrating the metabolic rate, arousal mechanisms, and water and food intake. Although sleep is generally coupled with energy saving and wakefulness with energy intake and consumption, it is not surprising that there is common integration and regulation of these processes at the hypothalamic level.

Two brain areas are of crucial importance in the neuroendocrine and metabolic effects in discussion: (1) the anterior hypothalamus–suprachiasmatic nucleus (SCN, which generates oscillations or rhythmicities by translating day-night information relayed from the retina into transcriptional and translational feedback loops of various clock genes; see Chapters 32 to 40) and (2) the lateral hypothalamus (e.g., the area secreting hypocretin). The SCN orchestrates global and peripheral circadian rhythms through both endocrine and neural pathways. Hormonal systems that are under overt circadian control include the pineal gland (secreting melatonin), the hypothalamic-pituitary-adrenal (HPA) axis (active hormone: cortisol) and the hypothalamic-pituitary-thyroid (HPT) axis (active hormones: thyroxine [T_4] and triiodothyronine [T_3]). In addition, hormones under SCN control (e.g., cortisol) may act to entrain other circadian oscillators, located in peripheral tissues, organs, and systems.

Since the initial description of encephalitis lethargica or von Economo's "sleepy sickness" due to lesions of the dorsal part of the hypothalamus and our understanding that lesions of certain anterior hypothalamic areas lead to severe insomnia, the role of the hypothalamus in sleep-wake regulation has become more evident. The hypothalamic regions could be damaged by locoregional tumors or their treatment and by inflammatory or infiltrative conditions. Given the plethora of functions of the hypothalamus, its damage may be associated with a wide variety of sleep symptoms, including sleepiness, fatigue, and disturbances in the sleep-wake regulatory system and circadian rhythms.

Many case reports or short series illustrate that pathology of the lateral hypothalamus can lead to hypersomnia or narcolepsy, whereas involvement by infiltrative, inflammatory, or neoplastic conditions of the anterior hypothalamus may disrupt the circadian clock of these patients or lead to severe insomnia. Inflammatory infiltrates were observed in the hypothalamus of two patients who died with Kleine-Levin syndrome.[1,2] Although the exact biochemical changes or hormonal

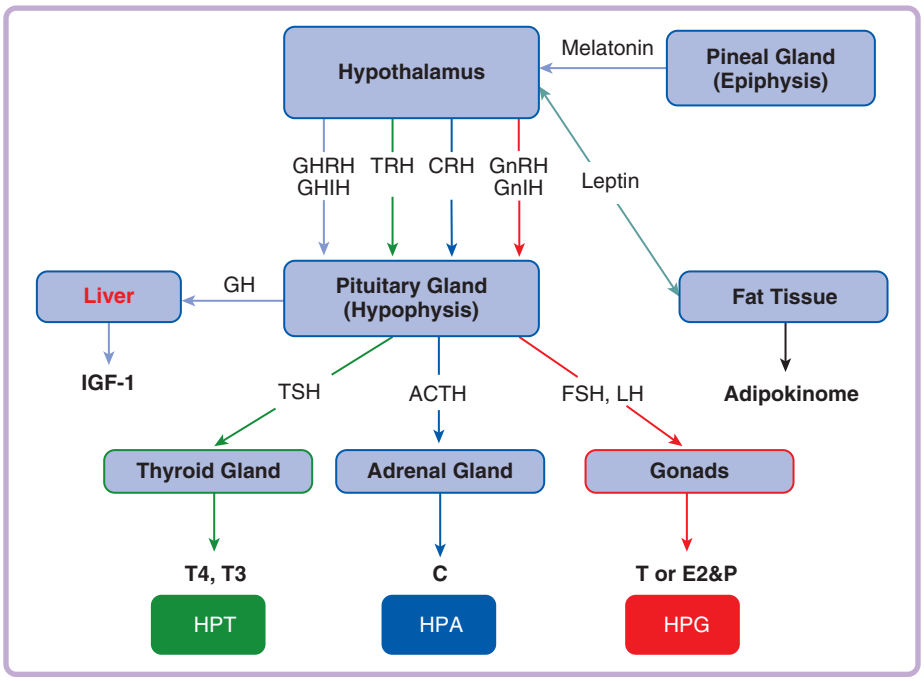


Figure 132-1 Diagram representing the main hypothalamic-pituitary–target gland axes: somatotropic axis (on the left, in *light blue*, via growth hormone [GH], growth hormone–releasing hormone [GHRH], growth hormone–inhibiting hormone [GHIH], and insulin-like growth factor-1 [IGF-1]), hypothalamic-pituitary-thyroid axis (HPT, in *green*, via thyrotropin-releasing hormone [TRH], thyroid-stimulating hormone [TSH], thyroxine [T_4], and triiodothyronine [T_3]), hypothalamic-pituitary-adrenal axis (HPA, in *dark blue*, via corticotropin-releasing hormone [CRH], adrenocorticotropic hormone [ACTH], and cortisol [C]), and hypothalamic-pituitary-gonadal axis (on the right, in *red*, via gonadotropin-releasing hormone [GnRH], gonadotropin-inhibiting hormone [GnIH]; follicle-stimulating hormone [FSH]; luteinizing hormone [LH], and testosterone [T] or estradiol [E2]/progesterone [P]). Additional interactions illustrated here are the interactions between the fat tissue, which secretes adipokines or so-called adipokinome, and the interaction between pineal gland and hypothalamus, via melatonin. Not illustrated here are the negative feedback loops exerted by the effector hormones (IGF-1, T_3, C, T, E2/P) on the hypothalamus and the pituitary gland; additionally, for purposes of simplicity, prolactin pathways were omitted.

changes occurring in this disorder are still unclear, it is conceivable that an autoimmune insult at the level of the thalamus or the hypothalamus is involved in the pathogenesis.

One special note needs to be made about hypothalamic obesity (see also Adipose Tissue). Most syndromes of morbid obesity result from single gene mutations expressed in the hypothalamus. Hypothalamic obesity is generally the result of impairments in the hypothalamic regulatory centers of body weight and energy expenditure and can be caused by anatomic destruction by tumors or other infiltrative processes, radiotherapy, Prader-Willi syndrome, and other disorders. Mutations in the *LEP*, *LEPR*, *POMC*, *MC4R*, or *CART* genes have been described in association with hypothalamic obesity.[4] Pathophysiologic mechanisms include loss of sensitivity to afferent peripheral humoral signals such as leptin, dysregulated insulin secretion, and impaired sympathetic nervous system activity. Dysregulation of 11β-hydroxysteroid dehydrogenase-1 activity and melatonin may also have a role in the development of hypothalamic obesity.

PITUITARY DISORDERS AND SLEEP

The anterior pituitary hormones are part of the following endocrine axes: HPA, HPT, and hypothalamic-pituitary-gonadal (HPG) as well as somatotropic (hypothalamic-pituitary–growth hormone) and lactotrophic (hypothalamus-pituitary-prolactin) systems. Vasopressin and oxytocin are synthesized and released from posterior hypophysis; they exert many autocrine, paracrine, or neuropeptidergic actions but have less known endocrine effects relevant to sleep or involvement in sleep disorders. Similarly, less is known about the physiologic roles of pituitary adenylate cyclase–activating polypeptide (PACAP), which may be involved in the homeostatic regulation of sleep.[5,6]

Growth Hormone–Secreting Tumors (Acromegaly and Gigantism) and Sleep Apnea

The neuroendocrine condition of growth hormone (GH) excess leads to acromegaly in adults and gigantism in children. The GH and insulin-like growth factor-1 (IGF-1) excess lead to acral enlargement, coarse facial features, and growth of the synovial tissues and articular cartilages (with acromegalic arthropathy in up to 75% of cases[7]). Sleep-disordered breathing (SDB) is very common in acromegaly. Studies have shown that up to 70% of patients diagnosed with acromegaly have sleep apnea, even after correction for obesity.[8-10] In a landmark study by Grunstein and colleagues performed on 53 consecutive referrals of patients with acromegaly, almost all patients with acromegaly had severe snoring.[11] Thirty-one patients (93%) referred because of suspected SDB had sleep apnea, compared with 12 patients (60%) referred without suspected sleep apnea. Central sleep apnea (CSA) was the predominant type of apnea in 33% of patients.

Excessive daytime sleepiness and fatigue have been reported as common symptoms in patients with acromegaly. The currently accepted explanations are represented by coexistent obstructive sleep apnea (OSA; most of the cases), direct effects of GH in promoting sleep (still unclear in humans), effect of prior cranial radiotherapy, coexistent hypogonadism or hypothyroidism, and hypothalamic GH-secreting tumors affecting somnogenic hypothalamic areas (rarely). Recognizing that SDB is underassessed in patients with acromegaly, current guidelines recommend that every patient diagnosed with this condition should be evaluated upfront for daytime symptoms or assessed by polysomnography (PSG) in collaboration with a sleep specialist.

The pathophysiologic mechanisms of SDB in acromegaly are still unclear. Several hypotheses have been proposed. OSA may result from macroglossia, hypertrophy of the submaxillary salivary glands, soft tissue thickening (especially of the soft palate and uvula), changes in the bone architecture of the upper airway (dorsocaudal rotation of the mandibular angle or reduced lingual "enclosure"), leading to smaller cross-sectional area and greater collapsibility of the hypopharynx. OSA may also result from altered neuromuscular control of the upper airway, dysfunction of the upper airway dilators such as sternohyoid myopathy, or obesity (if present). CSA in patients with acromegaly may be due to abnormalities in the control of breathing that result from left ventricular (systolic or diastolic) dysfunction, central pathways disinhibiting respiratory control, increased response to hypercapnic hypoxic drive, or direct or indirect effects exerted by GH on respiratory control centers.

The presence of OSA may confer additional cardiovascular risk for patients with acromegaly. GH and IGF-1 excess lead to left ventricular hypertrophy, systolic and diastolic myocardial dysfunction and cardiac dysrhythmias, and glucose intolerance and diabetes mellitus. Hypertension is highly prevalent in patients with acromegaly (up to 50% of cases), whereas cardiomyopathy is present in most patients. Hypertension is related to chronic hypervolemia (GH and IGF-1 increase sodium reabsorption by acting directly on the distal tubules epithelial cells' sodium channels[12]), endothelial dysfunction, insulin resistance, or coexistent sleep apnea that could further increase the blood pressure.[8] Most acromegaly patients die of cardiovascular disease.

OSA improves in many patients with acromegaly after GH tumor resection surgery. Nevertheless, prospective studies show that OSA persists in at least 40% of cases cured of acromegaly, requiring periodic reassessment.[8,13,14] Indeed, current guidelines state that, despite successful treatment of acromegaly, SDB does not consistently resolve, so posttherapeutic assessments with yearly evaluations (e.g., Epworth Sleepiness Scale, PSG, positive airway pressure use, and effectiveness data) are suggested.

Growth Hormone–Secreting Tumors and Restless Legs Syndrome

A group of investigators[15] looked recently into a possible association between acromegaly and restless legs syndrome (RLS), or Willis-Ekbom disease. Using both questionnaires and PSG, they found an increased prevalence of RLS in patients with active acromegaly.[15] Overall, approximately 21% of patients with acromegaly had RLS (36% in active disease group, 12% in controlled disease group, and 4% in healthy controls). Patients with active disease had worse acromegaly-related quality of life (QoL), higher severity of RLS symptoms by the International RLS Rating Scale, longer sleep latencies and wake after sleep onset, higher arousal and periodic limb movements during sleep indexes, and worse sleep efficiency.[15] Moreover, prevalence of RLS was not correlated with serum GH or IGF-1 levels, and the relationship stood even after exclusion of patients with OSA.[15] Interestingly, Taylor-Gjevre and colleagues[16] showed that the prevalence of RLS was almost as high (24.4%) in patients with osteoarthropathy, a condition frequently associated with acromegaly, so it would be interesting to parse out how many patients with acromegaly with and without osteoarthropathy have RLS.

Prolactin-Secreting Tumors (Prolactinoma)

Prolactinomas are prolactin (PRL)-secreting tumors and account for approximately 40% of pituitary adenomas. In patients with untreated prolactinoma, secondary hypogonadism is the most frequent endocrine abnormality. Patients with prolactinoma spend more time in slow wave sleep (SWS) than their healthy controls (on average, 79.4 vs. 36.6 minutes in controls), whereas rapid eye movement (REM) sleep does not seem to differ between the two groups. The current working hypothesis is that PRL stimulates both SWS and REM sleep, although the exact mechanism is still unclear. These findings are concordant with reports of good sleep quality in patients with prolactinoma, and discordant with the abnormal sleep of patients with other endocrinopathies.[17] An increased prevalence of obesity in male patients with prolactinoma has been described,[18,19] and because obesity is a risk factor for OSA, this morbid association should not be surprising. One study showed that elevated PRL levels may represent a modifiable and reversible cause of overweight state, especially in men.[20] Potential mechanisms explaining weight gain in patients with PRL-secreting pituitary adenomas are (1) associated hypogonadism; (2) increased hypothalamic pressure; (3) reduced dopaminergic tone; (4) leptin resistance; (5) adipokine imbalance such as hypoadiponectinemia; and (6) visceral fat deposition as a direct effect of PRL.

Adrenocorticotropic Hormone–Secreting Pituitary Tumors (Cushing Disease)

Excess adrenocorticotropic hormone (ACTH) secretion by pituitary tumors (Cushing disease, most frequently caused by pituitary microadenomas) or ectopically (by extrapituitary ACTH- or corticotropin-releasing hormone [CRH]-secreting tumors such as small cell lung cancer and atypical bronchial carcinoids) presents with manifestations due to the systemic effects of hypercortisolism.[21] When nocturnal cortisol secretory profile was analyzed in a group of patients with Cushing disease, adrenal activity was found to start predominantly during periods of non–rapid eye movement (NREM) sleep, similarly to matched healthy controls. Thus, even when the typical pituitary-adrenal axis nocturnal oscillation pattern is lost or blunted (such as in Cushing disease), the link between the endocrine activity and the ultradian rhythmicity of NREM and REM sleep seems to be preserved.[22]

Nonfunctioning and Gonadotropin-Secreting Pituitary Adenomas

Nonfunctioning macroadenomas (NFMAs) are pituitary tumors without systemic hormonal hypersecretion. Several

authors reported that patients with resected NFMAs have sleep complaints and circadian alterations as well as objective PSG or actigraphy changes.[23] In one study, patients with NFMAs had reduced sleep efficiency, less REM sleep, more N1 sleep, and more awakenings in the absence of associated apneas or periodic limb movements (compared with controls). Actigraphy revealed longer rest durations, more awakenings at night, and less activity during the day. Patients with treated NFMAs reported more fatigue and impaired QoL.[23] In another study, Joustra and colleagues[24] found melatonin secretion abnormalities in a significant percentage of patients with NFMAs and pituitary craniopharyngiomas, likely owing to either suprachiasmatic clock abnormalities induced by suprasellar extension or the treatment instituted for the pituitary tumor. Patients with NFMAs have also been found to have increased risk for developing metabolic syndrome (reduced high-density lipoprotein cholesterol and increased triglycerides)[25]; interestingly, identified factors were preoperative visual field defects and hypopituitarism. This association may be explained by hypothalamic dysfunction leading to the well-known hypothalamic obesity,[26] intrinsic imperfections of hormone replacement therapy,[25] direct effect of associated sleep disturbances,[27] or other, unidentified causes.

Craniopharyngioma

So far, little is known about the circadian and sleep-wake abnormalities of patients with craniopharyngiomas. Tumor growth and its treatment may lead to hypopituitarism, visual field defects, obesity, and sometimes increased daytime sleepiness.[28] Interestingly, the latter was associated with decreased nocturnal melatonin levels and higher body mass index (BMI); as such, further studies on the possible beneficial effects of melatonin substitution on daytime sleepiness and weight control in these patients are needed. Compared with healthy controls, patients successfully treated for craniopharyngiomas may report impaired QoL, excessive fatigue,[29] increased daytime sleepiness,[21,30,31] or severe sleepiness despite normal sleep patterns.[31,32] In one recent study,[33] craniopharyngioma patients presented with significantly decreased area-under-the curve melatonin concentrations compared with healthy controls and with a strong association between low midnight melatonin and reduced total sleep time, reduced time of night sleep, impaired sleep efficiency, and reduced physical activity.

Growth Hormone Deficiency

Adults with growth hormone deficiency (GHD) often complain of fatigue, fatigability, or impaired overall QoL.[34] Numerous animal[35,36] and human[37-39] studies have showed that growth hormone–releasing hormone (GHRH) has soporific effects, increasing total sleep time (TST) and SWS, even in the absence of GH.[35,40] As such, enhanced hypothalamic GHRH activity may lead to increased sleep pressure and excessive daytime fatigue, even in untreated GHD of primary pituitary origin because of the GH negative feedback loop (i.e., lack of GH leads to compensatory increase in GHRH). Furthermore, findings in rodents[36] and even in humans[41,42] indicate that GH itself may stimulate REM sleep. Thus there is enough evidence to go along with the hypothesis that patients with GHD may have sleep disturbances explaining daytime symptoms such as fatigue and impaired QoL. These sleep disturbances may differ according to the origin of GHD, that is, either pituitary (with overactive hypothalamic GHRH neurons) or hypothalamic (with reduced GHRH activity).

To date, few studies have objectively characterized sleep in GHD.[43-45] In one of the more recent studies,[46] authors evaluated sleep both subjectively and objectively in 30 adult patients with GHD (primary pituitary defects in 26 patients, hypothalamic causes in 4 patients, none on GH supplementation) and 30 matched healthy controls. The study found that GHD patients had worse scores on the QoL questionnaire, with tiredness being the most affected domain, irrespective of etiology. Patients with pituitary GHD spent more time in SWS and had a higher intensity of SWS than their controls. In contrast to pituitary GHD, patients with hypothalamic GHD had lower intensity of SWS than their controls. In the same study, older patients with pituitary GHD had more fragmented sleep and overall less REM sleep[46]; because insufficient REM sleep may be related to the emergence of memory disturbances,[47,48] it will be interesting to investigate to what extent GHD contributes to memory disturbances in elderly people.

OSA patients appear to have relative GH deficiency, which is likely related to hypoxia and abnormal sleep architecture from OSA and is reversible after continuous positive airway pressure (CPAP) therapy.[49,50] Contrary to earlier observations, more recent data suggest that GH therapy administered to patients with GHD does not induce or aggravate OSA. On the contrary, GH therapy seems to improve in these patients.[51] Twelve weeks, but not 6 weeks, of CPAP therapy for OSA increases IGF-1, with a further increase after 24 weeks. Total and pulsatile GH secretion, secretory burst mass, and pulse frequency are also increased by 12 weeks. CPAP therapy seems to improve specific elements of the GH–IGF-1 axis in a time-dependent manner.[52]

PINEAL DISORDERS AND SLEEP

The pineal gland is constituted by pinealocytes, which secrete melatonin (derived from serotonin), with important regulatory roles of the circadian and infradian rhythms. The pineal gland melatonin acts on numerous peripheral tissues, in concert with the melatonin secreted in the retina, bone marrow, platelets, and gastrointestinal tract. Although the complete surgical excision of the pineal gland leads to very few endocrine disturbances (which could point toward a vestigial aspect or a nonessential role), pineal pathology may occasionally have some consequences on sleep-wake homeostasis (Table 132-1). Pineal gland tumors are relatively rare but are the most frequent pineal pathology seen in clinical practice. These tumors may lead to signs and symptoms attributable to local mass effect, involvement of adjacent anatomic structures, and rarely endocrine effects. There are inconsistent data about the effects of hypermelatoninemia or hypomelatoninemia in certain types of tumors. Recent studies show that serum melatonin levels are generally very low or absent (rather than elevated) in both treated and untreated pineal tumors.[53] Nevertheless, two clinical presentations that have been reported deserve to be briefly mentioned here. First, rare cases of pineal tumors may present with clinical manifestations indistinguishable from narcolepsy ("secondary narcolepsy"), although they may also be due to concurrent hypothalamic involvements.[54,55] Second, cases of circadian-timed headaches, excessive daytime fatigue, and severe insomnia have been

Table 132-1 Main Neuropeptide Hormones in the Central Nervous System with Important Actions on Sleep-Wake Homeostasis

Source	Neuropeptide Hormone	Role in Sleep-Wake Homeostasis	Other Regulatory Roles
Hypothalamus	Hypocretin	Consolidates sleep Suppresses REM sleep	Feeding (+); mood; thermoregulation and energy expenditures; reward
	Melanin-concentrating hormone (MCH)	Promotes sleep	Feeding (+); memory
	Galanin	Promotes NREM sleep Promotes REM sleep	Anxiety; neuroregeneration; pain
	Growth hormone–releasing hormone (GHRH)	Promotes NREM sleep	Growth hormone (GH) release (+)
	Thyrotropin-releasing hormone (TRH)	Promotes wakefulness Inhibits NREM sleep	Feeding (–); energy homeostasis; locomotion
	Corticotropin-releasing hormone (CRH)	Promotes wakefulness Modulates REM sleep	Anxiety; depression; stress
	Cholecystokinin (CCK)	Promotes wakefulness Promotes NREM sleep	Anxiety; feeding (–); pain
	Somatostatin (SST)	Promotes wakefulness Promotes REM sleep Inhibits NREM sleep	GH release (–)
	Corticostatin (CST)	Promotes NREM sleep Inhibits REM sleep	GH release (–); learning; memory; pain (–)
	Brain-derived neurotrophic factor (BDNF)	Promotes NREM sleep (?) Inhibits wakefulness	Feeding (–); synaptic plasticity (+)
	Dynorphin	Promotes NREM sleep	Pain
	Melanocyte-stimulating hormone (MSH)	Promotes NREM sleep (?)	Stress regulation; feeding (–); motivation; pain; reward
	Neuropeptide Y	Modulates both NREM and wakefulness	Feeding (+); anxiety; addiction
	Vasoactive intestinal peptide (VIP)	Promotes REM sleep	Circadian regulation; feeding (–)
Anterior pituitary	GH	Promotes REM sleep Inhibits NREM sleep (?)	Cell growth (anabolic effects)
	Adrenocorticotropic hormone (ACTH)	Promotes wakefulness	Feeding (–); motivation; pain; reward
	Prolactin (PRL)	Promotes REM sleep	Lactation; stress modulation
	Pituitary adenylate cyclase–activating polypeptide (PACAP)	Promotes REM sleep	Circadian regulation; feeding (+); memory; stress; pain
Posterior pituitary	Vasopressin (AVP)	Circadian regulation (present also in the hypothalamus, including suprachiasmatic nucleus)	ACTH secretion (+); thirst and water excretion
	Oxytocin	Circadian regulation (?)	Lactation; anxiety; mood
Pineal gland	Melatonin	Promotes sleep	Circadian regulation; reproduction
Stomach (and hypothalamus)	Ghrelin	Promotes wakefulness	Feeding (+)
Adipose tissue (and hypothalamus)	Leptin	Promotes NREM sleep Promotes wakefulness (?)	Feeding (–); energy expenditure (+); thermogenesis (+)

reported and posited to be related to a pineal endocrine insufficiency because exogenous melatonin administration improved or resolved the presenting clinical symptoms.[56,57]

THYROID DISORDERS

Thyroid hormones influence many central nervous system (CNS) neurochemical pathways and networks involved in the sleep-wake homeostasis. Disturbances of the HPT axis such as altered thyroid function or abnormal thyroid-stimulating hormone (TSH) levels may affect the arousal-promoting system. Overall, the effects of thyroid hormones are more likely activating than soporific. So far, there have been only a few studies that shed light on the association between HPT axis and neural pathways involved in generating or maintaining alertness. Thyrotropin-releasing hormone (TRH) and its receptors are widely distributed in the CNS. Aside from its TSH-releasing effect, TRH exerts many other neuromodulatory actions, including stimulative, antidepressant, and neurotrophic effects.

A recent open-label study[58] examined the effects of a small dose of oral levothyroxine (25 mcg) on sleep and daytime somnolence of nine patients with idiopathic hypersomnia. The authors found that after 4 weeks of levothyroxine administration, TST and hypersomnolence improved significantly, effects that were maintained at 8 weeks. Unfortunately, posttreatment thyroid function testing was not available in most of the subjects. Mechanistically, it is possible that small doses of levothyroxine reduce TSH production or that patients with idiopathic hypersomnia may have intrinsic alteration of the HPT axis.

Hyperthyroidism and Restless Legs Syndrome

Reduced levels of dopamine (DA) in certain areas of the brain seem to play an important role in the pathophysiology of RLS, either as a primary biochemical disturbance or as a secondary condition such as in iron deficiency (because iron is a co-factor for tyrosine hydroxylase, a key enzyme in the synthesis of DA). DA agonists effectively alleviate RLS symptoms, and they are the first line of therapy. Currently, there is an accumulating body of evidence showing that DA may suppress the activity of the HPT axis.[59-61] First, the 24-hour profile of TSH resembles the typical circadian variation in clinical symptoms for RLS patients; levels of TSH increase in the evening, as does the severity of RLS symptoms (interestingly, dopaminergic tone seems also to increase in the evening[60]). Second, DA also modulates thyroid hormones by enhancing the biochemical functions of the cytochrome P-450 (CYP-450, heme and iron containing) liver enzymes that metabolize them and by inhibiting directly the pituitary TSH secretion. In addition, low iron levels diminish the available catalytic units of CYP-450 capable of degrading thyroid hormones.

One interesting recent theory[62] posits that the main cause of RLS is an imbalance between the dopaminergic system and HPT axis. This theory stems from several observations: (1) thyroid hormone levels seem to follow DA levels[59,60]; (2) TRH regulates TSH synthesis by stimulating transcription and translation of the TSH β-subunit gene, while DA inhibits this process[63]; (3) antidopaminergic agents (e.g., metoclopramide, neuroleptics) tend to worsen RLS complaints,[64] while DA agonists diminish RLS symptoms; (4) reduced levels of iron in the brain diminish the DA system activity, which in turn aggravates RLS; (5) the HPT axis increases its activity during sleep restriction,[65] which could also aggravate RLS; (6) some of the effects of elevated thyroxine levels, as seen in hyperthyroidism, resemble some symptoms of RLS; (7) opioids, the first class of drugs used to treat RLS, are also known to depress the HPT axis[66,67]; (8) several drugs known to alleviate RLS symptoms, such as carbamazepine, phenobarbital, and valproic acid, are inducers of CYP-450 system; and (9) conversely, tricyclic antidepressants, antihistaminic agents, selective serotonin reuptake inhibitors, and neuroleptics are CYP-450 inhibitors, and they all have the potential to worsen RLS symptoms.[62] Support of this theory comes from an earlier report of a patient with RLS and hypothyroidism on thyroid replacement therapy, who had a low serum ferritin level (iron deficiency). On successive challenge and withdrawal of L-thyroxine, there were significant changes in the International Restless Legs Syndrome Study Group (IRLSSG) severity score, periodic limb movement index, and sleep efficiency (all worse while taking thyroxine).[68] A study by Tan and colleagues[69] found a strong association between hyperthyroidism and RLS symptoms, despite the fact that the overall prevalence of RLS was very low (0.2%). The study examined 125 hyperthyroid patients, 21 hypothyroid individuals, and 434 normal healthy controls. None of the patients with thyroid disease had a diagnosis of RLS, as defined by the IRLSSG (i.e., none met all four IRLSSG criteria). In normal subjects, 4 out of 434 (0.9%) met the first three IRLSSG diagnostic criteria, as did 1 out of 21 hypothyroid individuals (4.8%) and 11 out of 125 hyperthyroid patients (8.8%). A recently published German study[70] analyzed the impact of comorbidities on prevalent and incident RLS in two distinct cohorts of patients (Dortmund Health Study [DHS], $n = 1312$, median follow-up of 2 years; and Study of Health in Pomerania [SHIP], $n = 4308$, median follow-up of 5 years). At baseline, 10.1% and 6.6% of patients had thyroid disease, and 7.3% and 10.1% had RLS in DHS and SHIP, respectively. Interestingly, odds ratios for incident RLS associated with thyroid disease (as defined by patient taking thyroid medications, either suppressive or replacement therapy) were statistically significant only in SHIP and found to be 1.63 (95% confidence interval, 1.04 to 2.58). Overall, the relationship of multimorbidity with RLS was stronger with incident than with prevalent RLS, suggesting that RLS developed subsequently to preexisting chronic conditions. This was an important observation because prior studies were mostly cross-sectional and thus unable to assess the temporal relationship between RLS and comorbidities.

Hypothyroidism and Obstructive Sleep Apnea

Hypothyroidism is one of the most common endocrine disorders; its prevalence tends to increase with age and is more common in women. It was estimated that approximately 2% of women between 70 and 80 years of age have overt hypothyroidism, whereas 5 to 10% of women older than 50 years of age have subclinical hypothyroidism.[71] Subclinical primary hypothyroidism generally refers to normal thyroid hormone levels with increased serum TSH in asymptomatic subjects.

In the general population, OSA is also a frequent condition.[72,73] Interestingly, OSA and hypothyroidism share a few similarities: (1) they are very prevalent disorders, (2) both have subclinical phenotypes that can go on undetected for years, (3) their exact prevalence in a particular population is often

unknown and inviting for wrongful extrapolations, (4) both are defined with somewhat fluid diagnostic criteria (e.g., hypopnea definition, apnea-hypopnea index [AHI] threshold, technology employed for diagnosis, TSH assays, threshold for subclinical hypothyroidism), and (5) they have common clinical manifestations and major comorbidities; for all these reasons, clinically they could be very easily confused with each other. For example, OSA is characterized by snoring, excessive daytime sleepiness, fatigue, apathy, frequent headaches, and memory impairments and is often associated with obesity or depression; these manifestations are quite nonspecific and also frequently seen in hypothyroidism. Furthermore, large goiters can also cause upper airway compression, ventilatory impairment, and possibly OSA.[74] Given these considerations, it is not surprising that a number of investigators found these conditions to coexist in 1.2% to 11% of the population studied.[75-77]

In a study on 50 consecutive patients with hypothyroidism, 30% were reported to have some degree of OSA.[78] In contrast, Bahammam and colleagues[79] failed to detect a significant difference in the prevalence of hypothyroidism in patients with or without a diagnosis of OSA. They found subclinical hypothyroidism in 11.1% of patients with OSA and in 4% of patients without OSA and concluded that subclinical hypothyroidism was more frequent in OSA.[79] In a cross-sectional study from Turkey,[80] from among 150 patients with OSA and 32 controls, thyroid ultrasonographic examination did not reveal any significant differences, and Hashimoto thyroiditis was found to have similar prevalence in OSA and control groups. In contrast, Bozkurt and colleagues[81] found that obese women with OSA had higher prevalence of Hashimoto thyroiditis, in parallel with OSA severity. Furthermore, Carratu and colleagues[82] found that subclinical hypothyroidism and thyroid replacement therapy did not change prevalence and severity of OSA. In another study, Mickelson and colleagues[77] evaluated the thyroid status of 834 patients with SDB and concluded that routine thyroid function testing was not necessary for these patients. Several authors found similar results and conclusions,[76,83,84] whereas others disagreed.[85,86]

Sleepiness, obtundation, and ultimately coma seen in advanced myxedema could be the result of either (1) a very reduced rate of basal metabolism, or (2) undiagnosed and untreated hypoxic (and possibly hypercapnic) OSA. Several mechanisms explaining the association between OSA and hypothyroidism have been posited: reduced cross-sectional area of the upper airway due to infiltration with mucopolysaccharides and water (e.g., macroglossia), myxedema-related upper airway myopathy, or reduced respiratory drive (as hypothyroidism blunts both hypoxic and hypercapnic ventilatory responses).[87,88] The downstream effects of the respiratory drive blunting on OSA likely vary by disease phenotype. As such, in patients with high loop gain or a tendency for ventilatory overshoot, blunted sensitivity may lead to relative stability of the airway and a decrease in OSA severity[89,90]; similarly, the impact of thyroid replacement therapy on SDB may also vary by phenotype.

The effect of thyroid hormone replacement therapy on OSA coexistent with hypothyroidism has been variable and based on rather small cohorts.[87] A more recent study of 50 patients suggested that OSA was a secondary phenomenon in many of the hypothyroid patients and that it may be reversible.[78] Thyroxine replacement therapy was associated with improvement in macroglossia, myxedema, and facial puffiness, suggesting that changes in upper airway anatomy seen in hypothyroidism may contribute to the development of OSA in these patients. Furthermore, several studies suggested the possibility of a stronger connection between OSA and cardiovascular complications in the initial phase of thyroid hormone replacement therapy because rapid restoration of the euthyroid state in these patients may confer additional cardiovascular risks.[87,91]

In summary, the association between OSA and hypothyroidism remains highly controversial and debatable, possibly because the hypothyroidism diagnosed nowadays is much milder than the severe myxedema seen in earlier investigations.

Hypothyroidism and Restless Legs Syndrome

Back in 1985, Schlienger[92] reported a case of RLS due to moderate hypothyroidism. Banno and colleagues[93] analyzed the comorbidities of RLS and found a trend toward a prior diagnosis of acquired hypothyroidism in female patients with RLS (within 5 years before PSG). Interestingly, women were twice as likely to be on thyroid replacement therapy at the time of evaluation. For some reason, patients with pulmonary hypertension were found to have a very high prevalence of RLS (43.6%), with moderate to severe symptoms. Among these patients, those with a history of hypothyroidism (67%) and those on opioids for relief of leg pain (69%) were more likely to have RLS.[94] Again, of interest is the fact that most of these patients were on thyroid replacement therapy.

Sleep Quality in Hypothyroidism

Excessive daytime sleepiness and fatigue are frequent symptoms in hypothyroidism. Although at times other sleep disorders (e.g., OSA, RLS) may be the cause of these symptoms, a primary CNS effect is also possible. Marked reductions in SWS have been described in patients with hypothyroidism, which are reversible with thyroid replacement therapy. Increased movements in sleep and reduced REM sleep have been described in patients with congenital hypothyroidism.

A recent study by Akatsu and colleagues[95] assessed the relationship between thyroid status and sleep quality in 5994 community-dwelling men older than 65 years; among the 682 men examined, 15 (2.2%) had subclinical hyperthyroidism (TSH <0.55 mIU/L) and 38 (4.9%) had subclinical hypothyroidism (TSH >4.78 mIU/L). Overall, there were no sleep qualitative differences between subclinical hypothyroidism and euthyroid status. Compared with euthyroid men, subjects with subclinical hyperthyroidism had lower mean TST (–27.4 minutes), lower mean sleep efficiency (–4.5%), and higher mean wake after sleep onset (+13.5 minutes), whereas 41% had increased risk for actigraphy-measured TST of less than 6 hours, and 83% had increased risk for sleep latency of 60 minutes or more (statistically nonsignificant). The study concluded that neither subclinical hypothyroidism nor hyperthyroidism was significantly associated with decreased sleep quality in elderly men.

ADRENAL DISORDERS AND SLEEP

The existing literature abounds in largely conflicting articles on the relationship between OSA and HPA activity in

humans. A recent study[96] looked at cortisol awakening response (CAR) in obese male subjects newly diagnosed with severe OSA compared with obese nonapneic matched controls and found the following: (1) a flattening of the CAR in OSA, (2) lower levels of cortisol at awakening in patients with OSA, (3) preserved circadian activity of the HPA axis, and (4) that 3 to 6 months of CPAP therapy led to significant restoration of the sleep architecture and the pattern of CAR (Figure 132-2)[96]; additionally, CPAP reduced the morning cortisol level differences between OSA patients and controls. In summary, the authors found a significant dysregulation of HPA axis activity in adult OSA patients, as shown by the flattening of the diurnal pattern of cortisol production, seen mostly in the first hour after awakening, which was restored after 3 or 6 months of CPAP therapy.[96]

Cushing Syndrome

Patients with Cushing syndrome typically have truncal or central obesity, diabetes mellitus, hypertension, acne, androgenic hirsutism, violaceus skin striae, psychiatric manifestations such as depression, and cognitive impairments. Most of these patients had sleep complaints and sometimes frank sleep disorders as comorbidities. In one study of 17 patients with pituitary ACTH–dependent Cushing disease and 5 patients with Cushing syndrome due to an adrenal tumor, 32% of patients had at least mild sleep apnea.[97] "Considerable" snoring and obesity were found in both apneic and nonapneic patients. Nonapneic patients with Cushing disease differed strikingly from healthy volunteers, having lighter and more fragmented sleep.[97] In another analysis by the same authors of patients with pituitary ACTH–dependent Cushing disease or ACTH-independent Cushing syndrome, patients with major depressive disorder, and normal healthy controls, there were substantial PSG similarities: All three patient groups demonstrated poorer sleep continuity, shorter REM latency, and increased first REM period density compared with normal subjects. Furthermore, ACTH-independent Cushing syndrome and major depressive disorder patients had elevated REM activity and density.[98] In another study, patients with Cushing disease *without* OSA exhibited poorer sleep continuity, shorter REM latency, and increased first-REM period density compared with normal subjects with reduced SWS.[99]

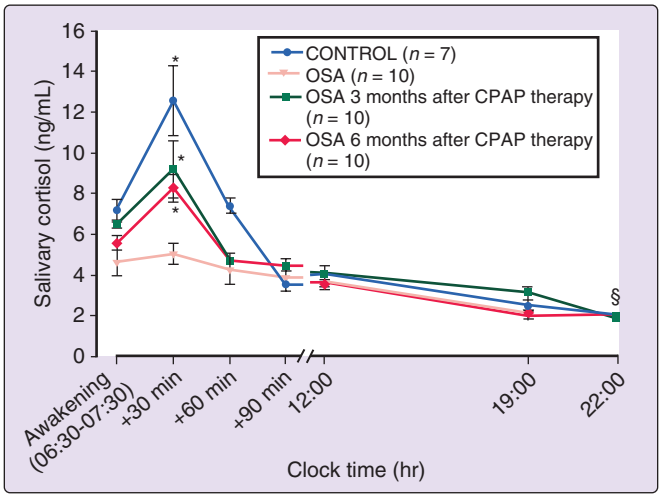


Figure 132-2 Cortisol awakening response before and after 3 and 6 months of continuous positive airway pressure therapy. CPAP, Continuous positive airway pressure; OSA, obstructive sleep apnea. (From Ghiciuc CM, Dima Cozma LC, Bercea RM, et al. Restoring the salivary cortisol awakening response through nasal continuous positive airway pressure therapy in obstructive sleep apnea. *Chronobiol Int* 2013;30[8]:1024–31. Copyright Informa Healthcare USA, Inc.)

Adrenal Insufficiency

Adrenal insufficiency is caused by primary adrenal dysfunction (Addison disease, characterized by both mineralocorticoid and glucocorticoid secretion insufficiency and increased ACTH) or secondary adrenal dysfunction (i.e., a hypothalamic or pituitary dysfunction resulting in reduced cortisol levels but normal levels of mineralocorticoids). Very few systematic assessments of sleep have been done in patients with untreated Addison disease. One recent, single-center, cross-sectional study[100] found that fatigue was a frequent complaint in patients with adrenal insufficiency (between 41% and 50% of patients, depending on the etiology) and that fatigue is influenced by the cortisol levels; additionally, salivary cortisol levels were not correlated with instantaneous fatigue. Acute administration of exogenous cortisol can improve daytime fatigue significantly not only in patients with Addison disease but also in healthy women.[101]

SEX HORMONE DISORDERS AND SLEEP

Testosterone in Sleep Restriction

Impaired sleep quality, reduced TST, circadian rhythm disruptions, and SDB may be associated with reduced testosterone levels. (See Chapter 20 for discussion on the role of sex hormones on sleep in men). Although studies confirm the effect of total sleep deprivation[102,103] on lower testosterone, data on the effect of sleep restriction on the HPG axis remain somewhat contradictory. The effect of sleep deprivation on testosterone levels seems to be age dependent[104]; other factors explaining discrepancies in the literature may be confounding factors, such as time of the day spent asleep,[102,105] circadian shifts, changes in sex hormone–binding globulin (SHBG) levels, circadian rhythm disruptions, stress, depression, medications, and methodologies used (e.g., self-reported vs. objective TST). For example, although the physiology described previously suggests that it is the first 3 to 4 hours of sleep that are critical to determining the increase in testosterone, a recent study has shown that partial sleep restriction to 4.5 hours was associated with a lower morning testosterone level when sleep was permitted in the first half rather than the second half of the night.[102] This was not unexpected because testosterone levels have been shown to decrease with increasing time awake.[106] A recent study showed a marked reduction (10% to 15%) in circulating testosterone levels in young healthy men after 8 days of partial sleep restriction of 5 hours per day (00:30 to 05:30 hours)[107]; in this study, SHBG levels were not measured. In a subsequent study in which sleep was restricted during the first part of the night and permitted from 04:00 to 08:00 hours for 5 nights, there was no significant change in testosterone levels, whereas SHBG decreased.[108] Similarly, serum concentrations of testosterone, luteinizing hormone, and PRL were reduced after only 24 to 48 hours of total sleep deprivation.[109-111]

Sleep Apnea Effects on Testosterone

As discussed earlier, disturbed sleep quality, reduced TST, chronodisruption, and SDB may be associated with reduced testosterone levels, although the direct causality is sometimes challenged. This association may be in fact mediated by effects on SHBG or by comorbid conditions.

Low testosterone levels have been reported in men with OSA,[49] and this appeared to be independent of age, degree of obesity, and presence of awake hypoxemia or hypercapnia. Subsequently, in a study on 89 obese men (BMI ≥35 kg/m^2), the severity of OSA was inversely correlated with the free testosterone levels, even after correction for age and BMI.[112] In a case-control study of 15 men with OSA and 15 matched controls, both total and free testosterone levels were lower in OSA patients, and testosterone was inversely correlated with the OSA severity, as defined by the oxygen desaturation index (ODI), and independent of obesity.[113] Triglyceride and uric acid levels were also significantly higher in OSA patients. A negative correlation between testosterone and uric acid levels and a positive correlation between testosterone and high-density lipoprotein cholesterol levels were also found, independent of BMI and waist circumference. This study suggested that, in patients with obesity and OSA, the severity of hypoxia during sleep might be an additional contributing factor to reduced testosterone levels, regardless of BMI or central obesity. In addition to hypoxia, sleep fragmentation may contribute to reduced testosterone levels.[114]

Conversely, there is a body of literature that supports the idea that low testosterone in men with OSA is primarily related to obesity.[114-116] For example, in a cohort of 1312 men 65 years or older followed for 3.4 years, total testosterone levels were unrelated to age or TST but were inversely correlated with AHI and ODI. After adjusting for BMI and waist circumference, significant associations were either absent or markedly weakened.[117]

Although older data suggest that treatment of OSA by CPAP therapy or uvulopalatopharyngoplasty results in increases in morning plasma testosterone levels at 3 months,[49,118] most of the studies[50,119,120] showed that CPAP therapy used for a single night[121] up to an average of 10 months[122] is generally without effects on follicle-stimulating hormone, luteinizing hormone, or testosterone, even when good compliance is ensured. The only significant outliers are the older studies by Grunstein and colleagues[49] and Santamaria and colleagues.[118]

In contrast to the effect of CPAP therapy, there seems to be a linear relationship between the degree of weight loss and an increase in plasma total testosterone levels in obese men[123,124] (Figure 132-3). Given the increasing evidence that obesity is a major determinant of the age-related decline in testosterone, it is thus encouraging that this decline can be at least partially prevented, or reversed, by successful lifestyle modifications such as diet, exercise, or even surgery.

In conclusion, OSA may not have a direct effect on testosterone levels after adjusting for age and obesity, although sleep fragmentation and intermittent hypoxia may be significant contributors. Furthermore, CPAP therapy instituted for moderate or severe OSA does not consistently increase testosterone levels in all studies, whereas weight loss does so in a linear fashion.

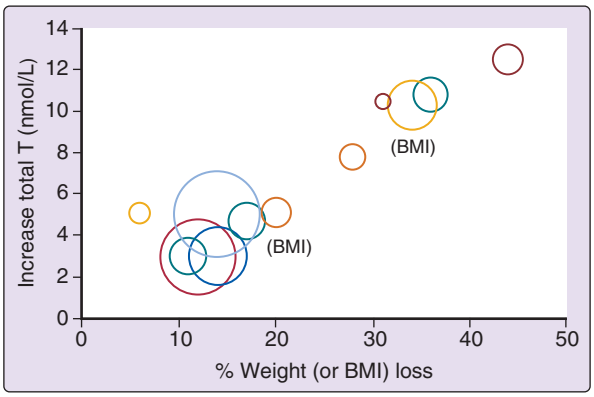


Figure 132-3 Effect of weight loss on testosterone (T) levels, as shown in different studies. Each *circle* represents a study, and its diameter is directly proportional to the study size. BMI, Body mass index. (From Fui MN, Dupuis P, Grossmann M. Lowered testosterone in male obesity: mechanisms, morbidity and management. *Asian J Androl* 2014;16[2]:223–31, which was modified after Grossmann M. Low testosterone in men with type 2 diabetes: significance and treatment. *J Clin Endocrinol Metab* 2011;96[8]:2341–53, Figure 1. Copyright of Endocrine Society, *The Journal of Clinical Endocrinology & Metabolism*.)

Testosterone Effects on Sleep Apnea

A number of reports have described the development of OSA after testosterone therapy in both genders, although these reports or small studies used mainly supraphysiologic doses of intramuscular testosterone.[125-127] Although the evidence is still rather anecdotal, exogenous testosterone administration has been considered deleterious enough in OSA that the current guidelines contraindicate testosterone therapy in the presence of untreated OSA.[128,129] In a recent study of obese men with OSA, testosterone undecanoate versus placebo was administered intramuscularly at baseline, 6 weeks, and 12 weeks. The authors used the modified Duffin rebreathing method in isohyperoxic and isohypoxic conditions and assessed the changes in minute ventilation versus CO_2 concentrations or $Paco_2$.[130,131] Ventilatory recruitment threshold is defined as the breakpoint between the horizontal and the linear ascending part of the graph. The slope of the latter is represented by the central and mixed central-peripheral chemosensitivity to CO_2 in hyperoxic and hypoxic conditions, respectively. Testosterone administration was associated with a slightly worse ODI at 7 weeks, but not at 18 weeks. There were no correlations between ODI and testosterone levels,[132] but positive correlations were noted between changes in serum testosterone and hyperoxic ventilatory recruitment threshold and between changes in hyperoxic ventilatory recruitment threshold and hypoxic burden at 6 to 7 weeks, but not at 18 weeks.[133] The authors suggested that time-dependent alterations in ventilatory recruitment threshold may mediate the changes in respiration during sleep observed with testosterone.

In summary, according to the current evidence and apart from transient adverse effects, testosterone treatment at normal doses may not lead to SDB.

Testosterone and Sleep Quality

Both insufficient and excessive testosterone levels have been shown to affect sleep and sleep architecture. In a cohort study of men aged 65 years or older, those with lower testosterone levels had reduced sleep efficiency, increased nocturnal

awakenings, and less time in SWS.[117] In mice, the loss of testosterone following gonadectomy resulted in a very small reduction in the amount of SWS, which was reversed by testosterone replacement, even after 6 hours of sleep deprivation.[134] In humans, the administration of testosterone and the abuse of androgenic or anabolic steroids have been reported to be associated with reduced TST, insomnia, and increased awakenings.[125,135] Administered acutely, methyltestosterone increases arousal and diminishes sleep changes attributed to activation of the brain serotonergic system.[136] In a study of patients with drug-induced hypogonadism[137] with and without gonadal steroid replacement, hypogonadal males had reduced 24-hour PRL levels and a reduced percentage of deep sleep in the hypogonadal state compared with those receiving testosterone replacement. Melatonin secretion was found to be increased in male patients with GnRH deficiency and in low-testosterone hypergonadotropic hypogonadal patients.[138] The observation of reduced TST without worsening alertness on task with high-dose testosterone therapy was also intriguing in this context.[125] In summary, low testosterone levels may affect overall sleep quality, and this may be improved by replacement-range doses. It seems that higher doses of exogenous testosterone or anabolic steroids are associated with more significant abnormalities of sleep duration and architecture.

Sex Hormones and Sleep-Disordered Breathing in Women

Lower estradiol levels have been associated with poor sleep quality in 45- to 49-year-old women.[139] Sleep disruption during pregnancy and after birth is well recognized, and menopause is often associated with insomnia (see Chapter 159). The latter is likely related to several factors, including hot flashes, mood disorders, and development or worsening of SDB, in parallel with a higher propensity to have central or android obesity (vs. gynoid) type.[140]

Estrogens and progesterone seem to be protective factors against development of OSA in women, thus explaining the gender-based differences in disease prevalence and severity. The evidence comes from studies in which sleep was evaluated during different phases of the menstrual phase,[141,142] menopause,[143] hormone replacement therapy (postmenopausally),[143] or pregnancy.[144,145] Some authors found that upper airway resistance is lower during the luteal compared with follicular phase.[141] Progesterone is thought to promote its effects through direct stimulation of respiratory drive by increased ventilatory responses to hypercapnea and hypoxia[146,147] and by enhancing upper airway dilator muscle activity[148] (and consequently, reduced upper airway resistance).

Lower estradiol levels have been associated not only with abnormal sleep architecture but also with a higher severity of OSA across a broad age spectrum (24- to 72-year-old women).[142] Conversely, postmenopausal subjects who received estrogen replacement therapy seem to have less severe SDB compared with those taking placebo.[143,149] Currently it is still unclear whether hormonal therapy has an advantageous risk-to-benefit ratio for SDB in women.[150,151] Similarly, although progesterone levels fall after menopause, no consistent therapeutic effect of progesterone administration in OSA has been demonstrated[150] (Table 132-2).

Androgens may also explain the observed gender-based dimorphisms in sleep architecture and SDB severity. O'Connor and colleagues[152] performed a retrospective analysis of 830 patients with PSG-diagnosed OSA and found that the total AHI was significantly higher in men compared with women; the REM difference (REM AHI minus NREM AHI) was higher in women than men across all levels of SDB severity, even after adjustment for age, weight, or duration of apnea. The study showed that women with SDB tend to have more frequent respiratory events during REM sleep compared with men and to have a higher prevalence of OSA that occurs mostly in REM sleep. Koo and colleagues[153] also analyzed the sexual dimorphism of OSA by assessing REM-related SDB among 1540 patients with OSA. Among 221 subjects (14.4%) who met the criteria for REM-related SDB (i.e., AHI ≥5 events/hour, NREM AHI <15, and REM AHI/NREM AHI >2), female patients had higher prevalence than men (24.5% vs. 7.9%), whereas younger women had a significantly higher prevalence than older women (27.2% vs. 18.6%); similarly, younger men had a significantly higher prevalence than older men (9.9% vs. 4.5%). In conclusion, REM-related SDB was found to be more prevalent in women than men and more prevalent in younger individuals of both genders, suggesting that differences may depend also on age (or duration of OSA).

Several studies showed that testosterone influences both neural control of breathing[154] and upper airway mechanics.[126] For example, Zhou and colleagues[155] examined the effect of transdermal testosterone (5 mg/day, administered during the follicular phase of the menstrual cycle) on apneic threshold during sleep in eight healthy premenopausal women. The authors concluded that testosterone increases apneic threshold in premenopausal women, thus leading to breathing instability during sleep.

Sleep and Polycystic Ovary Syndrome

Polycystic ovary syndrome (PCOS) affects up to 10% of reproductive-age women of all ethnic groups, making it the most common endocrine disorder of women in this age range. PCOS is defined by four components: oligomenorrhea, clinical and biologic features of hyperandrogenism, polycystic ovaries by ultrasonography, and exclusion of other causes of androgen excess.

Several studies showed that PCOS is strongly associated with increased central adiposity and with OSA, in direct relationship with the degree of androgen excess.[156-160] Patients with PCOS are oligoovulatory or anovulatory (by definition); thus they have low circulating progesterone levels, which may contribute to the high prevalence of SDB in this condition. Furthermore, studies in adults with PCOS found various strengths of association between PSG measurements and serum androgens and parameters of glucose metabolism. For example, Fogel and colleagues[156] revealed that the AHI of obese PCOS patients correlated with total and free testosterone levels, whereas others[159] found a correlation between the risk for or severity of SDB and insulin levels or other measures of glucose tolerance. Vgontzas and colleagues[157] showed that the strongest risk factor for OSA in women with PCOS were fasting insulin levels and glucose-to-insulin ratio, a measure of insulin resistance.

ADIPOSE TISSUE

The adipose tissue represents one of the largest and most active endocrine organs, known to be secreting a plethora of

Table 132-2 Known Associations between Neuroendocrine Abnormalities and Major Sleep Conditions

Sleep Disorder/ Endocrine Abnormality Associated	Somatotropic Axis	HPT Axis	HPA Axis	HPG Axis	Other
OSA	Reduced GH and IGF-1 levels (correlated with severity of SDB)	Controversial (hypothyroidism is a risk factor for OSA)	Unclear (exaggerated response of corticotrophs to CRH, not explained by obesity alone)	Low testosterone (males), elevated androgens and reduced progesterone (females) are risk factors for OSA (obesity is a confounder)	Increased sympathetic activity, changes in adipokinome (increased leptin, leptin resistance, decreased adiponectin), decreased hypocretin, higher incidence of diabetes mellitus, hyperinsulinemia, and metabolic syndrome Activated RAAS
Insomnia	Unclear	Hyperthyroidism	Hypercortisolism	In perimenopausal women: reduced sleep quality, sleep maintenance insomnia ("hot flashes") Abnormal CAR?	Unclear
RLS	Acromegaly patients have higher rates of RLS	Hyperthyroidism and hormone replacement for hypothyroidism are associated with RLS	Unclear	Unclear	Unclear

CAR, Cortisol awakening response; CRH, corticotropin-releasing hormone; GH, growth hormone; HPA, hypothalamic-pituitary-adrenal; HPG, hypothalamic-pituitary-gonadal; HPT, hypothalamic-pituitary-thyroid; IGF-1, insulin-like growth factor-1; OSA, obstructive sleep apnea; RAAS, renin-angiotensin-aldosterone system; RLS, restless legs syndrome; SDB, sleep-disordered breathing.

hormones. White adipocytes are secretory cells, releasing lipid products such as fatty acids resulting from lipolysis, cholesterol, prostaglandins, endocannabinoids, fat-soluble vitamins such as α-tocopherol and active form of vitamin D_3, glucocorticoids (converting cortisone to cortisol), estrogens, and others. Several major protein hormones are synthesized and secreted by adipocytes, the most prominent of which are leptin and adiponectin, which are produced primarily (not exclusively) in the fat cells. The circulating level of leptin is generally directly related to the body mass or the amount of fat; by contrast, circulating adiponectin levels are reduced in obesity. Leptin is involved in regulation of appetite, angiogenesis, and insulin secretion, whereas adiponectin is mainly insulin sensitizing, antiinflammatory and proangiogenic.

Obesity and Sleep

In obesity, secretion of a number of adipokines is dysregulated and associated with adipose tissue inflammation and, possibly, with the development of obesity-associated complications (vs. pure surrogate, "innocent" biomarkers). There is a solid body of literature showing that, as obesity worsens and adipose tissue grows disproportionately to its vascular supply, local hypoxia develops, and other downstream biologic changes converge to develop more robust local and systemic inflammation (Figure 132-4).

The relationship between quantity of sleep (TST) and obesity has been extensively examined in the past decade.[161-168] Most of the studies found a significant association between obesity and short sleep in children and some even in adults (although the latter was weaker). For example, Magee and Hale[166] reviewed 20 longitudinal studies published before 2010 and found an inconsistent relationship between short sleep and obesity in adults and a positive relationship in children. Nielson and colleagues[167] analyzed another set of 23 original studies and found similar conclusions. Before this, a meta-analysis on this topic published in 2008,[164] which included 45 cross-sectional or prospective studies of adults or children, found that short sleep conferred an odds ratio (OR) for developing obesity of 1.6 (1.4 to 1.7) in adults and 1.9 (1.5 to 2.4) in children. Patel and Hu[165] also considered cross-sectional and prospective studies of children and adults separately. The results from both study designs were consistent in children, whereas studies of adults were less so.

As far the relationship between sleep quality and obesity goes, this was less studied and reported. One recent study[168] examined in a family medicine setting the relationship between obesity and three sleep characteristics (duration, quality, and stability). Among 225 consecutive patients seen over a period of 10 weeks who completed a battery of questionnaires, 78% reported poor quality sleep, 59% had high Berlin questionnaire–based apnea risk scores, 12% reported RLS symptoms, and 9% reported a prior diagnosis of sleep apnea; 62% were obese. The authors found significant associations between sleep quality, duration or bedtime stability, and

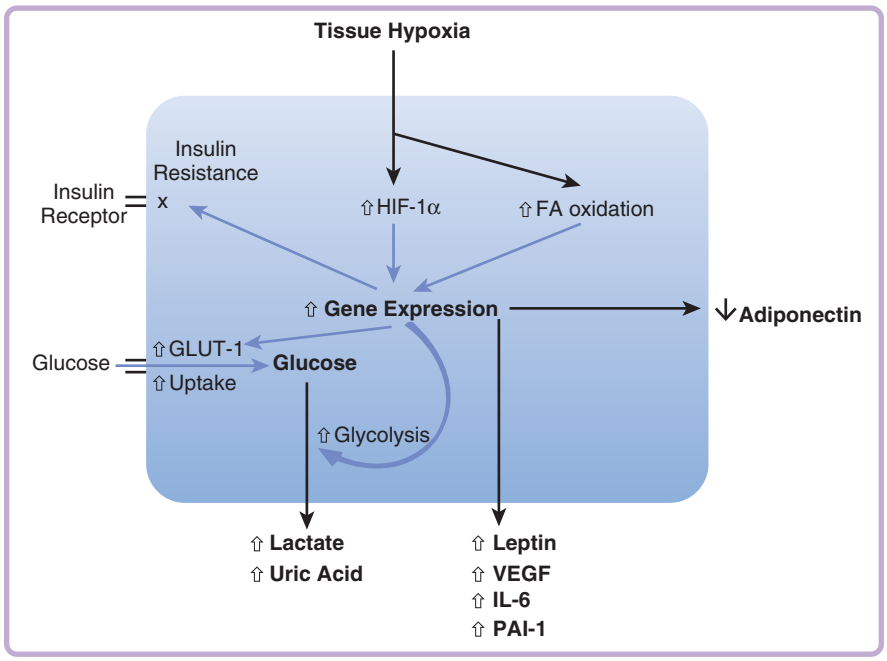


Figure 132-4 Diagram illustrates the effects of intermittent hypoxia at the level of a white adipocyte: activated transcription factor (hypoxia-induced factor-1α [HIF-1α]) leads to expression of more than 1000 genes; fatty acid (FA) oxidation in the mitochondria; increased glucose uptake through the cell membrane, mainly due to increased availability of glucose transporter-1 (GLUT-1); blocking of the insulin receptor, with subsequent insulin resistance; and activation of glycolysis, which leads to increased production of lactate, which will be released into the circulation. In terms of adipokines, low oxygen tension at the level of the adipose cell leads to reduced adiponectin and increased leptin, vascular endothelial growth factor (VEGF), and plasminogen activator inhibitor-1 (PAI-1).

obesity. The association between sleep quality and obesity was negative and linear, while the association between sleep duration and obesity was U shaped (as seen in multiple other publications and without a clear explanation to date). Less stable bedtimes during the week (OR = 2.3) or on the weekend (OR = 1.8) were also associated with obesity. The association between sleep quality and obesity was not explained by patient demographics or snoring.[168]

What is the pathogenic connection between obesity and poor or short sleep? First, short sleep, poor quality sleep, or unstable bedtimes may provide more opportunities to eat and could launch a metabolic cascade that increases appetite, reduces satiety, and worsens the fatigue, leading to adverse behavioral changes (e.g., curtailed levels of exercise, increased intake of hypercaloric, concentrated sweets) and more fat mass deposition. Conversely, obesity is an inflammatory condition that leads to release of biologically active cytokines, which could lead to activation of the HPA axis, with subsequent reduced TST or impaired sleep quality; in turn, this could perpetuate the vicious cycle.

Indeed, both TST reduction (with preserved SWS) and SWS suppression (with preserved TST) have been shown to be associated with insulin resistance (without compensatory hyperinsulinemia), resulting in impaired glucose tolerance and increased risk for type 2 diabetes mellitus. Furthermore, sleep restriction is also associated with decreased serum levels of leptin (an anorexigenic hormone) and increased levels of ghrelin (an orexigenic hormone), leading to hunger, exacerbated appetite, and craving for hypercaloric foods. Further evidence suggests that reduced TST may represent a permissive environment for the activation of multiple genes that may contribute to the development of obesity. Indeed, the heritability of BMI is higher in short sleepers. Therefore the chronic and progressive sleep curtailment seen nowadays in modern Western society conceivably may contribute to the current obesity, diabetes, and metabolic syndrome epidemics seen in both adults and children.

Besides the relationship between obesity and sleep, central, android-type obesity is also a powerful and consistent epidemiologic predictor of OSA.[169,170] Conversely, significant weight loss leads to improvement in SDB severity and associated metabolic abnormalities.[170] Because association does not mean causality, the reverse is also possible, that is, that OSA promotes the development of obesity. Although the data are limited for the latter directionality,[171] several mechanisms are conceivable. For example, chronic intermittent hypoxia and sleep fragmentation of SDB could alter the central control of energy regulation and general metabolism (e.g., trough-altered serotonergic activity in the hypothalamus[171] through leptin, insulin resistance, or hyperinsulinemia), and increase appetite and oral intake. As such, selective serotoninergic agonists and antagonists have been tried as treatments for central obesity[170] and for OSA but with limited effects.[172] The low circulating testosterone and GH levels noted in OSA also seem to occur in some patients with central obesity.[49,170] In obesity, recombinant GH administration appears to reduce central body fat, whereas CPAP therapy for OSA may restore the GH secretion during sleep.[171] If indeed OSA causes central obesity, it would be expected that treatments that reverse SDB would result in weight loss. Some studies of patients with OSA suggested that CPAP therapy reduces visceral fat mass, even without significant changes in BMI,[173,174] but large, controlled,

and randomized studies specifically addressing this issue are still lacking. As mentioned previously, central obesity is also associated with hyperinsulinemia and insulin resistance, but their exact links with OSA are still under investigation. Hyperinsulinemia has been found in some patients with OSA independent of weight, BMI, or central type of obesity.[175] Nevertheless, the questions of whether and to what extent CPAP improves insulin resistance or diabetes control remain controversial. Further discussion on the interrelationships among obesity, diabetes, and metabolic syndrome and SDB can be found in Chapter 118.

CLINICAL PEARLS

- Endocrine derangements may lead to or are associated with hypersomnia, OSA, RLS, insomnia, circadian disorders, and other sleep disorders.
- Endogenous or exogenous excess of growth hormone, cortisol, or androgens may predispose to development of OSA.
- Excess of catecholamines, cortisol, or thyroxine is associated with insomnia.
- OSA may be associated with low testosterone (direct effect or mediated by aging, obesity, fragmented sleep, and intermittent hypoxia).
- Adipose tissue is one of the largest and most active endocrine gland and secretes a very complex set of factors that constitute the "adipokinome."

SUMMARY

In a general conceptual framework, sleep is under both neural and hormonal control. Neural control is exerted through a complex network of nuclei or neuronal groups, highly interconnected and multifunctional, whereas hormones exert many autocrine, paracrine, or endocrine effects. There is a strong connection between sleep and metabolic processes. Disorders of the hypothalamus-pituitary-target organ axes frequently present with sleep symptoms and, occasionally, with frank comorbid sleep disorders. Accumulating evidence suggests an important role played by sleep curtailment or specific disorders in the pathogenesis of metabolic disturbances associated with obesity. Furthermore, sleep disorders such as OSA are very common in several endocrine disorders, such as acromegaly, obesity, and diabetes. Most patients with acromegaly have some degree of SDB, either OSA or CSA. OSA seems to be associated with low testosterone levels, whereas androgen administration appears to exacerbate OSA. Furthermore, there is some evidence of a protective effect of female sex steroids on upper airway patency during sleep. It is still controversial whether hypothyroidism is a distinct risk factor for SDB.

Selected Readings

Cuesta M, Cermakian N, Boivin DB. Glucocorticoids entrain molecular clock components in human peripheral cells. *FASEB J* 2015;**29**(4):1360–70.

Klibanski A. Clinical practice: prolactinomas. *N Engl J Med* 2010;**362**(13):1219–26.

Massumi RA, Winnacker JL. Severe depression of the respiratory center in myxedema. *Am J Med* 1964;**36**:876–82.

McGinty DJ, Sterman MB. Sleep suppression after basal forebrain lesions in the cat. *Science* 1968;**160**(3833):1253–5.

Melmed S, Casanueva FF, Hoffman AR, et al. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2011;**96**(2):273–88.

Melmed S, Casanueva FF, Klibanski A, et al. A consensus on the diagnosis and treatment of acromegaly complications. *Pituitary* 2013;**16**(3):294–302.

Mirer AG, Peppard PE, Palta M, et al. Menopausal hormone therapy and sleep-disordered breathing: evidence for a healthy user bias. *Ann Epidemiol* 2015;**25**(10):779–84.

Rao MN, Neylan TC, Grunfeld C, et al. Subchronic sleep restriction causes tissue-specific insulin resistance. *J Clin Endocrinol Metab* 2015;**100**(4):1664–71.

A complete reference list can be found online at ExpertConsult.com.

Chapter 133

Pain and Sleep

Anthony G. Doufas

Chapter Highlights

- Chronic pain and poor sleep hygiene are major public health challenges with great economic and societal impact. Pain disturbs sleep, and recent evidence from longitudinal cohorts indicates that impaired sleep may in turn promote or exacerbate chronic pain.
- Experimental and epidemiologic evidence supports a bidirectional influence between pain and common pain-related comorbidities, such as insomnia and mood disorders, in patients with chronic pain. This chapter presents current evidence regarding the theories and potential mechanisms mediating the development and maintenance of this interactive nexus of comorbidities in patients suffering from chronic pain.
- A diligent diagnostic approach is considered essential to comprehensively manage patients with chronic pain and comorbid insomnia. This chapter describes sleep assessment instruments that are in wide use in clinical and research practice in both pain and sleep medicine.
- Evidence from randomized controlled trials provides support for comprehensive, multimodal management of comorbidities associated with chronic pain, including insomnia and depression. This chapter examines recent randomized controlled trial evidence for the effectiveness of modern pharmacologic and behavioral interventions in the treatment of chronic pain patients.

Sleep and pain are both critical processes of life with great biologic and evolutionary value. In the normal state, pain and sleep operate together through a sensitive and well-orchestrated physiologic balance, which primarily serves to protect the sleep function. When pain is excessive or its function becomes deregulated, pain can alter sleep and diminish its capacity to provide the necessary physiologic and mental recovery for the individual. Although the physiologic link between sleep and pain remains elusive, clinical and experimental evidence supports a bidirectional, dynamic influence between the two. Steadily accumulating evidence from observational micro- and macro-longitudinal studies, as well as from randomized controlled interventions, suggests that inadequate or disrupted sleep may exacerbate an existing pain condition or mediate the development of a new pain condition, which in turn may augment sleep dysfunction.

EPIDEMIOLOGY OF PAIN AND COMORBID INSOMNIA

Because of their sheer prevalence and significant impact, chronic pain and impaired sleep qualify as two major, yet unmet, public health challenges that are associated with enormous economic and societal cost.[1-5] According to the National Center for Health Statistics, 43% of Americans are affected by a chronic pain condition, with low back pain being the most prevalent (25% in average) pain-related diagnosis, whereas 42 million people (National Sleep Foundation; 2000 Sleep in America poll) report frequent sleep disruptions (more than few nights a week) as a consequence of pain or physical discomfort.[6] Moreover, the 2006 Voices of Chronic Pain Survey indicates that, apart from sleep, other important aspects of quality of life, such as mood function and the ability to concentrate during normal daily activities, are most commonly and severely affected in patients with chronic pain.[6] International large population surveys have estimated similar prevalence rates for chronic pain conditions (1-year prevalence up to 40%) in both developed and developing countries, with joint-related (17.5%) and low back (18.5%) pain ranking at the top.[7] Large epidemiologic cross-sectional assessments in general and clinical populations suggest a dense connection between impaired sleep, physical pain, and mood disturbances, primarily expressed as anxiety and depression disorder.

Community-based studies estimate that more than 40% of people who suffer from chronic insomnia (the most common sleep disorder) do so in the context of a pain condition,[8] whereas moderate to severe sleep disturbances and comorbid depression occur in up to 80% of patients with chronic pain.[6,9-11] Although cross-sectional population assessments are important in characterizing the various morbidity phenotypes in patients suffering from chronic pain and disturbed sleep, they cannot provide evidence regarding the causal pathophysiologic mechanisms underlying the commonly observed associations. Instead, longitudinal observational cohorts are better fit to evaluate potential causal factors or modifiers of the relationship among sleep, pain, and mood dysfunction in patients suffering from chronic pain. Table 133-1 presents recent evidence from such longitudinal cohorts that evaluated the change in pain- and sleep-related outcomes in patients with primary insomnia or pain at the beginning of the assessment period.

Table 133-1 Prospective Longitudinal Studies Evaluating the Relationship between Impaired Sleep and Pain*

Baseline Condition	Study	Population	Follow-Up Period	Sleep Measures	Pain Type	Pain Measures	Outcome	Results
Impaired sleep	Cavinet et al., 2008[13]	Middle-aged healthy population (n = 3767)	1 yr	Insomnia questionnaire (4-item)	Musculoskeletal pain	Frequency of pain symptoms	Risk for developing pain	Increase in the odds for developing pain for both men (adjusted OR = 1.83; 95% CI, 1.16–2.87) and women (adjusted OR = 1.92; 95% CI, 1.34–2.75)
	Odegard et al., 2011[16]	Headache-free population (HUNT-2 and HUNT-3, n = 15,060)	11 yr	Sleep-onset and terminal insomnia (composite score)	Headache (all types)	ICHD-2 criteria	New-onset headache	Increase the risk for new-onset headache (adjusted RR = 1.4; 95% CI,1.2–1.7)
	Mork et al., 2012[136]	Unselected CWP-free women (n = 12,350)	11 yr	Sleep problems (ordinal variable)		CWP and musculoskeletal pain (yes/no)	New-onset CWP	Increase in the risk for new-onset CWS (adjusted RR = 3.4; 95% CI, 2.3–5.2)
	Jansson-Fojmark et al., 2012[14]	Unselected CWP-free women (n = 1551)	1 yr	Sleep problems by questionnaire (yes/no)	Musculoskeletal pain	Frequency of pain symptoms	Persistence of pain	Increase in the odds for persistence of pain (OR = 1.49; 95% CI, 1.15–1.97)
	Sanders et al., 2013[17]	Healthy TMD-free (OPPERA, n = 3263)	3 yr	PSQI (7-item, composite score)	TMD	Questionnaire and clinical examination	First-onset TMD	Increase in the risk for first-onset TMD (adjusted hazard ratio = 1.32; 95% CI, 1.18–1.47)
	McBeth et al., 2014[15]	Middle-aged free of CWP (n = 4326)	3 yr	Sleep problems (4-item ordinal scale)	CWP	CWP (ACR criteria)	Prevalence of CWP	Increase in the odds for CWP (adjusted OR = 1.8; 95% CI, 1.2–2.8)
Pain	Odegard et al., 2013[19]	Insomnia-free population (HUNT-2 and HUNT-3, n = 19,279)	11 yr	Sleep-onset and terminal insomnia (composite score)	Headache and chronic musculoskeletal pain	ICHD-2 criteria; Nordic questionnaire and ACR criteria	New-onset insomnia	Increase in the odds for headache (adjusted OR = 2.2; 95% CI, 1.9–2.6) and chronic musculoskeletal pain (adjusted OR = 1.8; 95% CI, 1.6–1.9)
	Tang et al., 2014[20]	Older adults (n = 6676)	3 yr	Jenkins sleep scale	Musculoskeletal pain	ACR criteria	New-onset insomnia	Increase in the risk for new-onset insomnia (adjusted† OR = 1.46; 95% CI, 1.21–1.75 for some pain, and 1.80; 95% CI, 1.47–2.22 for CWP)

*This list presents in chronologic order major prospective longitudinal cohorts that were published in the past 5 years.
†Statistical model adjustment included the presence of sleep disturbances at baseline.
ACR, American College of Rheumatology; CWP, chronic widespread pain (a newer term for fibromyalgia); HUNT, Nord-Trøndelag Health Study, Norway; ICHD, International Classification of Headache Disorders; OPPERA, Orofacial Pain: Prospective Evaluation and Risk Assessment Study; PACE, Paracetamol for Low-Back Pain Study; PSQI, Pittsburgh Sleep Quality Index; TMD, temporomandibular disorder.

Insomnia as a Risk Factor for Chronic Pain

Subjects who suffer from primary insomnia (i.e., insomnia with no obvious medical cause) have been found to be more sensitive to experimental pain than normal sleepers,[12] suggesting the presence of a reciprocal action between disturbed sleep and pain. This relationship is also supported by recent longitudinal studies showing the development of new pain or worsening of a pain condition in patients suffering from chronic sleep loss.

Evidence from large prospective longitudinal studies (see Table 133-1) with follow-up periods ranging from 1 to 11 years have demonstrated that disturbed sleep in pain-free subjects could significantly increase the risk (odds and relative risk ratios ranging from 1.3 to 3.4) for different chronic pain conditions, including musculoskeletal pain,[13,14] chronic widespread pain,[15] headaches,[16] and temporomandibular joint disorder.[17]

Pain as a Risk Factor for Insomnia

In the acute experimental setting, electroencephalography has shown that pain invades sleep and disturbs its structured continuity and architecture. Among other factors, the magnitude of these acute effects of pain on the sleep function depends on the severity of nociceptive stimuli as well as on the vulnerability of sleep to the disruptive influence of such stimuli.[18] Currently it remains unknown whether chronic infliction of pain can alter the sensitivity of sleep to painful or other stimuli, thus resulting in the development of an insomnia-like sleep disturbance.

Evidence from a large-scale prospective longitudinal Norwegian cohort (Nord-Trøndelag Health Study, n = 19,279) has shown that, compared with pain-free controls, insomnia-free subjects who were suffering from headache or chronic musculoskeletal pain at baseline were twice as likely to develop insomnia 11 years later (odds ratio [OR] = 1.8; 95% confidence interval [CI], 1.8 to 2.2),[19] Similarly, various types of chronic pain complaints among older adults were associated with an increased risk for insomnia of up to four times for the next 3 years.[20] Interestingly, the combination of physical limitations and reduced social participation explained more than 65% of the effect of pain on insomnia onset.

Depression as a Comorbid Link Between Chronic Pain and Impaired Sleep

Abnormal sleep is a primary symptom of affective diseases and has also been incriminated as a risk factor for the development of depression.[21] In addition, mood disorders are quite common among patients suffering from chronic pain.[22] Evidence from large international (World Mental Health Surveys, n = 42,249)[7,23,24] and national (National Comorbidity Survey, n = 5877)[25] cross-sectional surveys showed that chronic pain conditions increased the odds for developing depression and anxiety disorders by two to four times, with approximately 20% of pain patients reporting symptoms of depression and 35% of them reporting an anxiety disorder in the past year.[25] The close relationship between mood disorders and insomnia in the context of chronic pain has also been demonstrated in a large community-based sample of 18,980 European participants, where pain conditions like backache and joint or articular disease were as strongly associated with symptoms of insomnia as they were with depression and bipolar disorders (OR ≈ 5).[8] Interestingly, although pain-related insomnia shares several mood, cognitive, and behavioral characteristics with primary insomnia,[11,26,27] insomnia due to pain seems to last longer and is associated with more daytime dysfunction[28] compared with primary insomnia.[8] This points to a possible interactive effect of impaired sleep and pain on daily functioning, a hypothesis that has recently received support by evidence from two separate cohorts of Norwegian (n = 6892) and Finnish (n = 6060) employees, in which pain and insomnia at baseline synergistically predicted objective health outcomes and disability retirement in these populations.[29]

A recent longitudinal investigation found that anxiety and depression in pain patients increased the likelihood for persistence of insomnia over the course of a year,[14] and a possible bidirectional association between anxiety and chronic pain has recently been suggested by other longitudinal cohorts examining these outcomes in patients suffering from acute and chronic pain.[30,31] Among the 614 participants in the Netherlands Study of Depression and Anxiety who were free of depression and anxiety at baseline, those with multiple locations of joint-related pain were at higher risk for first-onset mood disorder during the 4-year-long follow-up period (adjusted hazard ratio of 2.9; 95% CI, 1.7 to 4.8).[31] Alternatively, feelings of anxiety experienced in the early phase after lower extremity trauma predicted the presence of pain up to 24 months later.[30]

It needs to be recognized that the morbid nexus between pain, abnormal sleep, and mood dysfunction is dense, and cause-and-effect relationships cannot be inferred without the application of well-designed targeted interventions.

BIOLOGIC AND BEHAVIORAL MECHANISMS FOR THE PAIN-SLEEP ASSOCIATION

In the past decades, several human studies involving both healthy and clinical populations have attempted to elucidate the nature of the association between abnormal sleep and pain processing. Investigations focusing on the experimental curtailing or disruption of sleep, and others that have examined pain processing in individuals suffering from primary sleep or chronic pain disorders, suggest that deterioration of sleep function might enhance pain responses, possibly through multiple biologic and behavioral mechanisms.[32-34] Although most of these investigations involve a small number of subjects and suboptimal control conditions, an informal meta-analytic review of this research appears to support the pain-enhancing effect of fragmented or heavily curtailed sleep.

Pain Disturbs Sleep

The central nervous system (CNS) is organized to protect the integrity and preserve the continuity of sleep. For this purpose, neural filters at the brainstem level and descending inhibitory pathways that are activated during sleep regulate the arousing effect of ascending sensory information.[35,36] Thus, although the application of nociceptive stimuli during sleep may alter sleep microstructure and cause sleep-stage shifts associated with arousal patterns in humans,[37] deeper stages of sleep seem to be more resistant to such effects compared with lighter ones.[38,39] The overall response of sleeping humans to nociceptive stimuli, which may even invade consciousness and be incorporated into dreams,[39] is determined by the magnitude and type of these stimuli.

This line of research is important because (1) interindividual variations in the vulnerability of patients to pain-related insomnia might partly be due to differences in the capacity of the CNS to regulate the input of sensory information during sleep, and (2) chronic engagement of CNS sensory filters by nociceptive information might ultimately alter their function (neuroplasticity), with potential implications in the natural course of both pain disease and insomnia.

Experimental Sleep Disturbance Enhances Sensitivity to Pain

Over the years, several human studies using various experimental pain paradigms induced by thermal or mechanical stimuli have focused on elucidating the effect of abnormally short sleep on pain processing.[33] Experimental conditions involving total sleep deprivation were associated with the development of hyperalgesia and spontaneous pain symptoms in healthy volunteers, whereas targeted elimination of specific sleep stages, like rapid eye movement (REM) or slow wave sleep (i.e., stage 3 non–rapid eye movement [NREM] sleep), has not demonstrated conclusive findings.

The "forced awakenings" method[40] has been developed as a means to disrupt sleep continuity without completely eliminating a particular stage of sleep. This has provided researchers with an experimental paradigm more relevant to chronic pain patients, whose sleep is commonly disrupted by pain. Thus, when sleep was experimentally curtailed by forced awakenings in healthy volunteers, central pain-inhibitory mechanisms (namely, conditioned pain modulation) were impaired and spontaneous pain was increased.[40] This finding is of potential clinical relevance because impaired conditioned pain modulation regulation has been demonstrated in several chronic pain conditions, including chronic widespread pain,[41] temporomandibular joint disorder,[42,43] and low back pain,[44] in which sleep disturbances are also highly prevalent. In an effort to further elucidate the association between pain and sleep, a recent experimental study showed that fragmented, but not shortened, sleep was observed in mice with musculoskeletal sensitization (a "chronic pain" analogy),[45] and this sleep disturbance phenotype resulted in further aggravation of both nociceptive and sleep-wake outcomes.[46]

General musculoskeletal sensitization with spontaneous pain symptoms has been described in healthy humans following experimentally generated total and partial sleep deprivation for one or more consequent nights. Compared with normal sleepers, these experimental interventions led to activation of major inflammatory pathways,[47] with increased concentrations of proinflammatory cytokines like tumor necrosis factor-α, interleukin-1β, and interleukin-6 in the peripheral blood.[48] Although these experimental interventions were extreme, they demonstrate that peripheral inflammatory response may be another potential mechanism for the hyperalgesic effect of sleep loss.[49]

Pain Responses in the Context of Chronic Sleep Loss and Primary Sleep Disorders

The experimental application of painful stimuli in volunteers indicated that physiologic sleepiness (as determined by the Maintenance of Wakefulness Test) due to chronic sleep loss was associated with increased sensitivity to experimental pain[50] and decreased codeine analgesia.[51] This increased sensitivity to painful stimuli was reduced when subjects recovering from chronic sleep loss were given the opportunity to sleep for an extended period.[52]

Hyperalgesic responses to experimental pain in patients with primary sleep disorders like insomnia[53] and restless legs syndrome[54] were associated with a significant weakening of central pain-inhibitory[53] mechanisms or an amplification of descending pain-facilitatory[54] signals. Moreover, treatment with continuous positive airway pressure (CPAP) reversed the observed hyperalgesia in subjects suffering from sleep-disordered breathing,[55] although it is not clear whether the effect of CPAP on hyperalgesia[55] was due to CPAP restoration of normal sleep architecture or correction of the associated nocturnal hypoxemia. A recent analysis of data from the Cleveland Family Study demonstrated a significant positive association between recurrent nocturnal hypoxemia and pain complaints in subjects with sleep-disordered breathing, independent of sleep fragmentation.[56] This, along with other findings from human[57,58] and in vitro[59] experiments, may incriminate intermittent hypoxia as another potential mechanism, in addition to sleep disruption, involved in altering pain processing in humans.[60]

Behavioral, Mood, and Neurochemical Alterations in Pain and Comorbid Insomnia

The observed cognitive and affective changes encountered in patients suffering from pain or insomnia could also act as potential confounders of the relationship between sleep and pain. For example, the ability of attention to modulate pain perception is one of the main mechanisms people employ to cope with pain.[61] Insomnia with objectively determined short sleep duration has recently been associated with deficits in the executive attention control function,[62] which is also found to be impaired in pain-free volunteers after one night of partial sleep restriction.[63] Similarly, application of painful stimuli in volunteers showed that attentional bias toward pain and pain-related information reduced distraction from the painful stimuli,[64] undermined sleep,[26] and strengthened the association between daily pain severity and disability in patients with chronic pain.[65]

Pain catastrophizing[66] has been shown to magnify pain, possibly because it has detrimental effects on sleep.[67] Presleep cognitive arousal in patients with chronic pain was found to be a more reliable predictor of subsequent sleep quality than was presleep pain.[68] These intriguing findings point to the hyperarousal hypothesis for insomnia[69] as a potential underlying mechanism linking insomnia to depression and pain. According to this concept, in individuals with a neurocognitive state of intense focus and rumination about sleep, an acute stressful event that could otherwise only transiently affect sleep may trigger and perpetuate a more permanent process of sleep disturbance associated with changes in mood. This interplay between a vulnerable hyperfocusing psychology and a biology characterized by the preponderance of arousal-promoting activity in the CNS might be responsible for a set of maladaptive responses, ultimately resulting in morbid sleep- and pain-related-outcomes. Similarly, the "dopamine hypothesis"[70] posits that dysregulation of central dopaminergic transmission (a major arousal-promoting pathway involved in sleep-wake regulation) that diminishes mesocorticolimbic dopaminergic activity may act as a common precursor for several of the ill phenotypes encountered in chronic pain patients, including hyperalgesia,

insomnia, and depression.[33,71,72] Although the dopamine hypothesis is intriguing in conception, it is yet unclear to what extent the emerging nociception phenotypes could be the result of a dopaminergic dysfunction alone or an interaction between dopaminergic and other neurotransmission or neurophysiologic processes.[73]

Numerous neurochemical pathways of sleep-wake regulation,[74] including cholinergic,[75] adenosinergic,[76,77] gamma-aminobutyric acid (GABA)–ergic,[78] serotonergic,[77] and orexinergic[79,80] transmission, have been found to intersect with pain-related functions, making them attractive candidates for establishing mechanistic foundations for the association between sleep, pain, and relevant changes in mood. Nevertheless, experimental evidence shows that the relationship between arousal-promoting actions in the CNS and sensitivity to nociceptive stimuli is not monotonic; that is, the derived nociception phenotypes strongly depend on various factors. These include the type (e.g., various subtypes of dopamine, orexin, or serotonin receptors) and location (i.e., central versus peripheral nervous system) of activated receptors, the employed experimental model (e.g., acute versus chronic neuropathy pain), and, most important, the complex interaction between several neurotransmission paths.[75,80,81] The success of multimodal therapy strategies for chronic pain also emphasizes the experimental evidence that multiple brain regions and neurochemicals (including genetic and epigenetic molecules) modulate the psychophysiology of human pain.

Although primary and pain-related insomnia share several behavioral and phenomenologic characteristics,[11,26,27] neuroanatomic and clinical evidence support that chronic pain comorbidities like insomnia and depression might arise through a unique, pain-evoked pathophysiology and should be treated differently.[82] For example, human functional imaging studies have demonstrated that pain perception engages limbic and cortical areas of the brain, which are important for the regulation of mood and cognition.[82,83] In addition, patients with chronic pain and comorbid depression have been shown to respond less favorably to antidepressants compared with those who were pain free.[84] This evidence may indicate the existence of a deeper, physiology-based dissimilarity between primary insomnia and insomnia or depression as comorbidities of chronic pain.

Epidemiologic studies of pain and theoretical models for chronic disease[82,85] suggest that the transition from an acute pain episode to a chronic state of persistent pain is characterized by molecular and anatomic alterations (neuroplasticity). These alterations may contribute both to the chronicity and irreversibility of pain disease[86,87] and to the development of comorbidities like insomnia and depression.[88] Interestingly, individuals who suffer from chronic insomnia demonstrate an increased potential for neuroplasticity changes in response to training or other triggering events.[89] In addition, an abnormally high plasticity background in these patients might be also relevant to the development of mood disorders[90] or persistent insomnia syndrome.[91]

ASSESSMENT OF SLEEP IN THE CONTEXT OF CHRONIC PAIN

To accurately characterize the relationship between disturbed sleep and pain and effectively manage chronic pain patients, it is essential to use reliable instruments to assess both sleep- and pain-related phenotypes. The 2008 Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) consensus on interpreting the clinical importance of treatment outcomes in chronic pain trials found that the interference of pain with sleep, as assessed by the Brief Pain Inventory Interference Scale,[92] is one of the main factors affecting quality of life in this patient population. Thus it is among the primary therapeutic outcomes of interest.[93,94] A deeper and better understanding of the physiologic interaction between impaired sleep and pain would directly depend on our ability to reliably translate sleep behavior phenotypes to easily obtainable clinical surrogates.

In contrast to pain, sleep can be evaluated both by objective and subjective measures. Overnight, attended polysomnography (i.e., the evaluation of quantitative and qualitative parameters of sleep with the use of electroencephalographic, respiratory, and electrocardiographic monitoring modalities) is considered a gold standard for the objective assessment of sleep. Actigraphy monitoring, another method for the objective evaluation of sleep, has also been used in clinical trials, although it has less specificity than polysomnography in detecting wakefulness after sleep onset.[95] However, the ability to apply objective tests in large clinical trials has been limited by the associated cost and cumbersome preparation logistics. Despite the fact that patients tend to misperceive their sleep and objective tests demonstrate superior performance or measure different sleep qualities[96] than do self-reported sleep assessment methods,[97] the latter demonstrate sufficient validity[98] to be used in large chronic pain cohorts with longitudinal trial designs. Table 133-2 presents a list of scoring instruments that have been employed to assess pain-related sleep disturbances.

These scoring instruments consist of either a simple, one-item assessment scale for sleep quality or more complex and extended questionnaires assessing several aspects of sleep behavior; that is, sleep initiation and maintenance, number of awakenings, and the restorative quality of sleep. Validation studies have demonstrated that these tests can discriminate between subjects with insomnia and those with normal sleep. Certain assessment instruments, such as the Medical Outcomes Study sleep scale,[99-101] the Chronic Pain Sleep Inventory,[102] and the Insomnia Severity Index,[103,104] were able to detect changes in sleep behavior associated with the use of analgesia in pain patients. It is still not clear, however, how such tests should be constructed in order for them to have clinically meaningful effects in the management of chronic pain patients. Are more elaborate scoring instruments and those containing pain-specific components (e.g., Brief Pain Inventory Interference Scale,[92] Chronic Pain Sleep Inventory,[102] and Daily Sleep Diary[105]) better equipped to characterize sleep disturbances in the context of pain compared with simple single-item scales? A recent validation study of the single-item Sleep Quality Scale (11-point; 0 for *best possible sleep* to 10 for *worst possible sleep*) showed that in patients treated for chronic widespread pain, the simple instrument correlated closely with the Medical Outcomes Study sleep scale and was also sensitive enough to detect improvement in the quality of sleep as a result of pain treatment.[106]

Sleep is lately receiving more attention inside the overall evaluation framework of the chronic pain patient.[3,94] An outcomes-based approach is certainly required to guide future research to determine the type of sleep assessment that could

Table 133-2 Scales and Questionnaires that Are Commonly Used to Evaluate Sleep in Patients Suffering from Chronic Pain

Measure/Scale	Content	Item Format	Scoring Format	Method of Administration	Score Output	Validation Populations
Brief Pain Inventory (BPI)[92]	Location, intensity, and pain interference with several functions, including sleep	15 items composed from 2 multiple-item scales	2 open-ended and 1 yes/no question and 12 Likert scales (0–10)	Self-report	Mean of items 3–6 for pain severity score and mean of the 7 components for pain interference score	Arthritis and low back pain[137]
Chronic Pain Sleep Inventory (CPSI)[102]	Sleep initiation, need for sleep medications, midsleep and morning awakenings because of pain, and overall sleep quality	5 items; assessment of various sleep behaviors in patients with chronic pain	Each of the 5 items is scored using a 100-mm visual analogue scale (0–100)	Self-report	Items are scored independently; a single index of sleep problems due to pain is scored from 3 out of the 5 items	Chronic pain due to osteoarthritis
Daily Sleep Diary (DSD)[105]	Sleep onset, sleep duration, nighttime and early-morning awakenings, nonrestorative sleep, and overall sleep quality	7 items; frequency and severity of distinct sleep behaviors	Free entry, categorical, and 5- and 6-point ordinal scales	Self-report	Independent scoring for each individual DSD item	Chronic pain patients
Functional Outcome of Sleep Questionnaire (FOSQ)[138]	Impact of sleep disturbance on daily function	30 items; rating all evaluated functional outcomes	4- and 5-point rating scales	Self-report	Summation of the scores from all items; total score: 0–120	Patients with sleep-disordered breathing
Insomnia Severity Index (ISI)[103]	Nature, severity, and impact of insomnia	7 items; severity of insomnia and impact on mental health and daily function	5-point ordinal and Likert scales (0–4)	Self-report	Summation of the responses to each of the items; total score: 0–28	Insomnia patients[104]
Jenkins Sleep Scale (JSS)[139]	Symptoms related to sleep onset, maintenance, and nonrestorative sleep	4 items; rating the severity of various insomnia symptoms	6-point ordinal scale (0–5)	Self-report	Summation of the responses to each of the items; total score: 0–20	Air traffic controllers
Medical Outcomes Study (MOS) sleep scale[99]	Sleep initiation, maintenance, and quantity; nonrestorative sleep	12 items; rating various aspects of sleep behavior	5- and 6-point ordinal scales and free entry for the quantity of sleep	Self-report	Scores aggregated in 7 subscale domains and the composite sleep problems index; possible score for each subscale: 0–100	Patients with neuropathic[99] and chronic widespread[101] pain
Pittsburgh Sleep Quality Index (PSQI)[140]	Sleep-wake patterns, sleep duration and latency, sleep disturbances, sleep medications, nonrestorative sleep, and overall sleep quality	19 items; severity and frequency of disturbances in sleep continuity, various insomnia symptoms, and global quality of sleep	Free entry and 4-point ordinal scales	Self-report	7 component scores each rated from 0–3; total sleep quality score: 0–21	Patients with sleep disorders and depression
Sleep-EVAL Expert System[8]	Difficulties in sleep onset and maintenance; nonrestorative sleep	4 items; rating the severity of various insomnia symptoms	Dichotomous (yes/no) response	Clinician interview	Independent scoring of each of the 4 items	Psychiatric populations and patients with sleep disorders

add clinical value to the management of patients with chronic pain (e.g., objective versus subjective tests, extended and more complex questionnaires versus shorter and simpler ones, combination of objective and subjective methods).

MANAGEMENT OF PATIENTS WITH CHRONIC PAIN AND COMORBID INSOMNIA

Because of the physiologic and phenotypic complexities associated with the interaction of pain, pain-related insomnia, and comorbid mental illness in chronic pain, the therapeutic management of this patient population is a challenging task. The significance of these comorbid relationships is further emphasized by the adoption of multimodal therapeutic interventions that combine analgesic, hypnotic, and antidepressant effects. Thus it is encouraging that sleep-related outcomes are being incorporated into randomized controlled clinical trials designed to evaluate treatments for chronic pain. The inclusion of sleep assessments in such trials might also provide indirect evidence for the physiologic basis of the association between pain and disturbed sleep. Table 133-3 presents a list of important recent randomized placebo-controlled trials evaluating treatments for various types of chronic pain conditions, in which both pain and sleep-related outcomes have been assessed.

Interventions Targeting Pain

Pharmacologic analgesic interventions in patients suffering from chronic pain of various etiologies include acetaminophen, nonsteroidal antiinflammatory drugs, and opioids.[107-109] In patients suffering from chronic widespread or neuropathic pain, tricyclic and nontricyclic antidepressants and anticonvulsant drugs, prescribed as adjuvant or even monotherapy therapy, have also shown relative success in improving pain-related outcomes.[110-112] With the exceptions of acetaminophen and newer antidepressants (e.g., trazodone, mirtazapine, and buproprion) and anticonvulsants (e.g., gabapentin, pregabalin), most analgesics seem to alter sleep in a manner quite similar to that of pain (i.e., by disrupting sleep continuity or suppressing both REM and slow wave sleep).[113] It is thus not surprising that effective analgesia with agents like opioids and nonsteroid antiinflammatory drugs may not be associated with a complete restoration of normal sleep architecture, despite improving subjective sleep assessment outcomes.[114-116] Because systematic objective and longitudinal assessment for the sleep-disrupting effects of analgesics is missing, it is currently unclear whether a sole disturbance of sleep architecture (i.e., not associated with subjective complaints) constitutes a physiologic insult with the potential of altering pain outcomes or changing the natural course of pain disease in the long-term. This issue directly relates to the yet unanswered question about the clinical meaning of the various sleep phenotypes evaluated by different assessment methods.

Interventions Targeting Insomnia

Several recent randomized placebo-controlled, double-blinded trials have demonstrated the efficacy of novel CNS depressants and nonbenzodiazepine hypnotics that specifically target sleep function in reducing pain and improving function in patients suffering from chronic widespread and low back pain.

Sodium oxybate is the sodium salt of gamma-hydroxybutyrate, an endogenous metabolite of GABA with CNS-depressant properties.[117] Administration of sodium oxybate in patients suffering from chronic widespread pain significantly improved subjective sleep, pain, and functional outcomes, whereas it also increased physiologic sleep stability and the amount of slow wave sleep in polysomnography recordings.[118,119] Similarly, the administration of eszopiclone, a nonbenzodiazepine hypnotic,[120] in patients suffering from chronic low back pain, was associated with a significant increase in total sleep time, a reduction in pain, and an improvement in the ratings of depression.[121]

Although rigorous consensus criteria[93] may question the clinical meaningfulness of the improvement in pain-related and other functional outcomes demonstrated by the small analgesic effects in some of these interventions,[118,121] these findings are the first important randomized controlled trial (RCT) evidence that interventions targeting sleep-related physiology may alter pain perception and improve associated functional outcomes in patients suffering from chronic pain.

Interventions Targeting Both Pain and Disturbed Sleep

Melatonin is pineal gland hormone with a circadian pattern of secretion and well-known sleep-wake regulation, antiinflammatory, and antihyperalgesic effects. Although melatonin has demonstrated efficacy both in treating pain and improving sleep quality when administered in patients with temporomandibular joint disorder[122] or endometriosis-associated chronic pelvic pain,[123] the small sample size of these trials does not allow for robust conclusions regarding possible interactions between the pain-relieving and sleep-restoring effects of the drug.

Pregabalin, a novel anticonvulsant agent with calcium channel α_2-δ ligand properties,[124] has been shown to be effective both in treating pain and restoring the quality of sleep in patients with neuropathic disease, and it is one of first-line therapies recommended[111,125] for neuropathic pain. Studies in volunteers[126] and patients suffering from chronic pain[98,127] have demonstrated the beneficial effects of pregabalin on sleep function. The fraction of slow wave sleep is increased, and there is an associated improvement in self-assessed sleep quality. Both the analgesic and sleep-enhancing effects of pregabalin are presumably mediated by the same molecular mechanism, that is, a reduction in the influx of calcium ions into oversensitized neurons, followed by a subsequent reduction in the synaptic activity of the latter.[124] A recent retrospective analysis has shown that patients with neuropathic pain and comorbid sleep problems tend to respond better to the pain-relieving effects of pregabalin, suggesting that the analgesic action of the latter might be mediated through its sleep-restoring effects.[128] However, the nonspecific character of pregabalin's mechanism of action makes it difficult to test of this hypothesis prospectively.

Cognitive behavior therapy for the treatment of sleep disturbances has been a successful psychological intervention in patients suffering from primary[129] or comorbid (e.g., pain-related) insomnia.[130] However, the pain-relieving effects of this treatment modality have been weak, at best.[129,131,132] A recent RCT evaluating both pain- and sleep-related behaviors in older patients with osteoarthritis demonstrated that 9 months of cognitive behavior therapy failed to significantly

Table 133-3 Randomized Controlled Trials Evaluating the Effects of Various Interventions on Pain- and Sleep-Related Outcomes in Patients with Comorbid Chronic Pain and Impaired Sleep*†

Disease	Study	Population	Blinding	Control	Follow-up	Intervention	Intervention Focus	Pain Measures	Sleep Measures	Improved Outcome		Sleep-Pain Agreement
										Pain	Sleep	
Chronic widespread pain	Arnold et al., 2010[110]	CWP ($n = 507$)	Double	Placebo	12 wk	Duloxetine (SNRI)	Pain	BPI, SF-36 bodily pain	BPI and 11-point Likert scale (0–10)	Yes	Yes	Yes
	Moldofsky et al., 2010[118]	CWP ($n = 151$)	Double	Placebo	8 wk	Sodium oxybate (γ-hydroxybutyric acid)	Sleep	VAS, FIQ pain, SF-36 bodily pain	PSG, ESS, JSS, FOSQ	Yes	Yes	Yes
	Russell et al., 2011[119]	CWP ($n = 334$)	Double	Placebo	14 wk	Sodium oxybate (γ-hydroxybutyric acid)	Sleep	VAS, FIQ pain, SF-36 bodily pain	JSS	Yes	Yes	Yes
	Roth et al., 2012[98]	CWP ($n = 206$)	Double	Placebo	4 wk	Pregabalin (anticonvulsant)	Pain and sleep	NRS (0–10)	PSG, self-reported sleep assessments	Yes	Yes	Yes
	Kashikar-Zuck et al., 2012[141]	Juvenile CWS, ($n = 112$)	Single	CWP education	6 mo	CBT	Pain	VAS (0–10)	VAS (0–10)	No	No	Yes
Low back pain	Steiner et al., 2011[142]	CLBP ($n = 539$)	Double	Placebo	84 days	Buprenorphine (partial agonist of the μ-opioid receptor)	Pain	NRS (0–10)	MOS sleep scale (0–100)	Yes	Yes	Yes
	Williams et al., 2014[143]	Acute low back pain ($n = 1596$)	Double	Placebo	3 mo	Paracetamol (mild analgesic)	Pain	Days until recovery from pain, NRS (0–10)	Sleep quality from PSQI (item No 6)	No	No	Yes
	Goforth et al., 2014[121]	CLBP ($n = 52$)	Double	Placebo	1 mo	Eszopiclone (nonbenzodiazepine hypnotic)	Sleep	VAS (0–100)	Total sleep time (sleep diary)	Yes	Yes	Yes
Neuropathic pain	Richter et al., 2005[144]	Painful diabetic neuropathy ($n = 219$)	Double	Placebo	6 wk	Pregabalin (anticonvulsant)	Pain	NRS (0–10, daily), SF-MPQ (weekly)	Sleep interference (NRS, 0–10, daily)	Yes	Yes	Yes

	Siddall et al., 2006[145]	Central neuropathic pain (n = 137)	Double	Placebo	12 wk	Pregabalin (anticonvulsant)	Pain	NRS (0–10, daily), SF-MPQ	Sleep interference (NRS, 0–10, daily), MOS sleep scale	Yes	Yes	Yes
	Kalliomäki et al., 2013[146]	Posttraumatic neuralgia (n = 133)	Double	Placebo	1 mo	Chemokine receptor 2 (CCR2) antagonist (antiinflammatory)	Pain	NRS (0–10, every 12h), NPSI	Sleep interference (NRS, 0–10)	No	No	Yes
	Serpell et al., 2014[147]	Peripheral neuropathic pain (n = 173)	Double	Placebo	15 wk	Tetrahydrocannabinol/ cannabidiol (THC/ CBD)	Pain	Peripheral neuropathic pain (NRS, 0–10, daily), BPI	Sleep quality (NRS, 0–10, daily)	Yes	Yes	Yes
Temporomandibular disorder	Vidor et al., 2013[122]	Myofacial TMD pain (n = 32)	Double	Placebo	4 wk	Melatonin	Pain and sleep	VAS (0–10, daily)	Sleep quality (VAS, 0–10, daily)	Yes	Yes	Yes
Osteoarthritis	Vitiello et al., 2013[130]	Elderly (>60 yr) patients with osteoarthritis (n = 367)	Double	Education only	9 mo	CBT for pain and insomnia	Pain and sleep	Chronic pain scale (6 items; 0–10)	Insomnia severity index, sleep efficiency (actigraphy)	No	Yes	No
Postherpeutic neuralgia	Dworkin et al., 2003[148]	Postherpeutic neuralgia (n = 173)	Double	Placebo	9 wk	Pregabalin (anticonvulsant)	Pain	NRS (0–10, daily), SF-MPQ	Sleep interference (NRS, 0–10, daily), MOS sleep scale	Yes	Yes	Yes
	Apalla et al., 2013[149]	Postherpetic neuralgia (n = 30)	Double	Placebo	24 wk	Botulinum toxin A	Pain	NRS (0–10)	Sleep quality scale (5 items)	Yes	Yes	Yes
Chronic pelvic pain	Schwertner et al., 2013[123]	Endometriosis (n = 36)	Double	Placebo	8 wk	Melatonin	Pain and sleep	VAS (0–10, daily)	Sleep quality (VAS, 0–10, daily)	Yes	Yes	Yes

*Emphasis is given to the agreement between the changes in the pain- and sleep-related outcomes as a result of interventions targeting sleep, pain, or both sleep and pain functions.

†This list presents in chronologic order main prospective trials that were published in the past decade. Investigations with no clear prospectively defined hypotheses regarding the examined outcomes, as well as studies with conclusions based on exploratory and/or subgroup unplanned analyses, were not included.

ACR, American College of Rheumatology; BPI, Brief Pain Inventory; CBT, cognitive behavior therapy; CLBP, chronic low back pain; CWP, chronic widespread pain (a newer term for fibromyalgia); ESS, Epworth Sleepiness Scale; FIQ, Fibromyalgia Impact Questionnaire; FOSQ, Functional Outcomes of Sleep Questionnaire; IBS, irritable bowel syndrome; JSS, Jenkins Scale for Sleep; MOS, Medical Outcomes Study; NPSI, Neuropathic Pain Symptom Inventory; NRS, Numeric Rating Scale; PSG, polysomnography; PSQI, Pittsburgh Sleep Quality Index; TMD, temporomandibular disorder; SF-36, The 36-Item Medical Outcomes Study Short-Form Health Survey; SF-MPQ, Short Form McGill Pain Questionnaire.

improve pain compared with only education-based management.[130] Nonetheless, a recent secondary analysis of this RCT evidence by the same investigators showed that, independent of the received treatment, patients who experienced sleep improvement in the first 2 months of the treatment were more likely to demonstrate sustained improvement in their pain severity scores during the remaining 16-month follow-up period.[133]

As shown in Table 133-3, recent randomized placebo-controlled blinded interventions for a variety of chronic pain conditions have demonstrated that agreement between treatment-induced changes in pain- and sleep-related outcomes is common, regardless of the effectiveness of the therapy under evaluation. This observation confirms previous findings from a systematic review of Cochrane meta-analyses, showing that changes in insomnia- and pain-related outcomes of various RCT interventions tend to be in the same direction.[134] Although these findings support a strong link between sleep and pain functions, it is evident that, when employed alone and independent of the targeted function (i.e., sleep or pain), no current pharmacologic or behavioral modality is both sufficiently safe and effective to improve chronic pain in the long term. A combination of treatments, treatment modalities, and treatment targets (i.e., pain, insomnia, depression) may result, through already determined therapeutic paths, in larger analgesic and safer effects. Moreover, more accurate characterization of the various patient morbidity phenotypes may reveal new markers for identifying those who are responsive to each particular treatment or combination of treatments.[133,135]

CLINICAL PEARL

Pain disturbs sleep, and impaired sleep might exacerbate responses to pain. Insomnia and depression are highly prevalent among patients with chronic pain, and primary sleep disorders such as insomnia and sleep apnea have also been associated with various types of pain complaints. The sleep medicine physician should be aware of the complex epidemiology of chronic pain disease and follow a diagnostic approach that includes a comprehensive evaluation of both pain- and sleep-related phenotypes as well as a thorough assessment of the patient's mental health. Different types of chronic pain benefit from different pharmacologic approaches (e.g., modern antidepressants, sodium oxybate, and anticonvulsants for chronic widespread pain, and anticonvulsants for neuropathic pain), but all interventions are uniformly associated with improvement in both sleep and pain functions independent of their therapeutic target. Cognitive behavior therapies, despite their effectiveness in treating insomnia, have yet to show an important benefit as a holistic therapy for the chronic pain patient with comorbid insomnia.

SUMMARY

Chronic pain and poor sleep hygiene are major public health challenges with great economic and societal impact. An interaction between the two is slowly being established based on micro- and macro-longitudinal studies in healthy humans and in patients with chronic pain. Both the activation of major inflammatory pathways and inhibition of central descending analgesic signals seem to be candidate mechanisms for the hyperalgesic effect of acute sleep impairment. In patients suffering from chronic pain, the mood disorders and cognitive changes associated with preoccupation with pain- or sleep-related information seem to potentiate the bidirectional influence between pain and comorbid insomnia. Self-administered sleep assessment instruments have demonstrated that they can reasonably discriminate between patients with insomnia and normal sleepers. In addition, when applied to pain patients, they can detect improvement in sleep quality associated with the use of analgesia. Pharmacologic approaches that target either pain- or sleep-related outcomes in patients with chronic pain have resulted in a significant improvement in both outcomes. Evidence suggests that the effect sizes of such interventions on sleep, or pain, may be further enhanced by combining pharmacologic and behavioral therapies for both sleep- and pain-related symptoms.

Selected Readings

Buenaver LF, Quartana PJ, Grace EG, et al. Evidence for indirect effects of pain catastrophizing on clinical pain among myofascial temporomandibular disorder participants: the mediating role of sleep disturbance. *Pain* 2012;**153**:1159–66.

Cairns BE. Alteration of sleep quality by pain medication: an overview. In: Lavigne G, Sessle BJ, Choiniere M, Soja P, editors. *Sleep and pain*. Seattle, WA: International Association for the Study of Pain Press; 2007.

Doufas AG, Tian L, Davies MF, et al. Nocturnal intermittent hypoxia is independently associated with pain in subjects suffering from sleep-disordered breathing. *Anesthesiology* 2013;**119**:1149–62.

Institute of Medicine (US) Forum on Advancing Pain Research Care, and Education. *Relieving pain in america: a blueprint for transforming prevention, care, education, and research*. Washington, DC: National Academies Press; 2011.

Odegard SS, Sand T, Engstrom M, et al. The long-term effect of insomnia on primary headaches: a prospective population-based cohort study (HUNT-2 and HUNT-3). *Headache* 2011;**51**:570–80.

Ohayon MM. Relationship between chronic painful physical condition and insomnia. *J Psychiatr Res* 2005;**39**:151–9.

Russell IJ, Holman AJ, Swick TJ, et al. Sodium oxybate reduces pain, fatigue, and sleep disturbance and improves functionality in fibromyalgia: results from a 14-week, randomized, double-blind, placebo-controlled study. *Pain* 2011;**152**:1007–17.

Schrimpf M, Liegl G, Boeckle M, et al. The effect of sleep deprivation on pain perception in healthy subjects: a meta-analysis. *Sleep Med* 2015;**16**(11):1313–20.

Tang NK, Goodchild CE, Hester J, et al. Pain-related insomnia versus primary insomnia: a comparison study of sleep pattern, psychological characteristics, and cognitive-behavioral processes. *Clin J Pain* 2012;**28**:428–36.

Tang NK, McBeth J, Jordan KP, et al. Impact of musculoskeletal pain on insomnia onset: a prospective cohort study. *Rheumatology (Oxford)* 2015;**54**(2):248–56.

Tsang A, Von Korff M, Lee S, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. *J Pain* 2008;**9**:883–91.

Turk DC, Dworkin RH, Revicki D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. *Pain* 2008;**137**:276–85.

A complete reference list can be found online at ExpertConsult.com.

Chapter 134

Sleep and Chronic Kidney Disease

John G. Park; Kannan Ramar

Chapter Highlights

- Sleep disturbances such as insomnia, restless legs syndrome, periodic leg movements during sleep, and obstructive sleep apnea are common across the wide spectrum of chronic kidney disease (CKD) patients. CKD is linked to an increased risk for sleep disorders through various pathophysiologic mechanisms. The occurrence of CKD continues to rise, and as a result the prevalence of sleep disorders is likely to grow. Similarly, sleep disorders may increase risks associated with CKD and contribute to the rising prevalence of CKD.
- Comorbid sleep disorders add significant burden to the quality of life, health care costs, and morbidity and mortality of patients with CKD.
- This chapter provides an up-to-date review of the current available literature regarding sleep disturbances in patients with CKD and discusses the associated clinical implications as well as treatment options for the various sleep disorders common to this population.

Chronic kidney disease (CKD) is defined as abnormalities of kidney structure or function (defined by estimated glomerular filtration rate [eGFR] <60 mL/minute per 1.73 m^2) that is present for 3 or more months, with negative implications on health.[1] CKD is a common but serious condition associated with an increasing incidence and prevalence (16.8%) in the United States, particularly with the increasing prevalence of diabetes, metabolic syndrome, and hypertension, all being significant risk factors for CKD.[2,3] CKD is associated with increased mortality, decreased quality of life, and increased health care costs. CKD is also linked to an increased risk for sleep disorders. This chapter discusses these various sleep disorders across the spectrum of CKD and reviews the consequences and treatment options in this specific population.

EPIDEMIOLOGY

The prevalence of sleep disorders in patients with end-stage renal disease (ESRD) requiring dialysis is well established. Poor sleep, as defined as sleep-wake complaints, sleep-disordered breathing, and excessive daytime sleepiness, occurs in 45% to 80% of dialysis patients. This prevalence is much higher than that of the general population or CKD patients not requiring dialysis.[4-6] In contrast, the prevalence of sleep disorders in non–dialysis-dependent CKD patients varies widely between 14% and 85%. The reasons for this wide variation are multifactorial and may be related to the use of subjective versus objective data to assess sleep, the type of sleep disorder that is being evaluated, and the stage of CKD.

Some studies have not shown significant differences in the Pittsburg Sleep Quality Index (PSQI) between CKD patients and general medical outpatients,[7] whereas others have found progressive worsening of sleep quality with progression of renal disease based on CKD stages.[8] Poor sleep quality on the PSQI has been noted in 84.6% of CKD patients compared with 59.5% of patients with hepatitis C.[9] Inadequate sleep (defined as ≤6 hours/night) has shown to differ by CKD stages, with 37.4%, 43.0%, and 30.9% reported for no CKD, CKD stages 1 and 2 (eGFR ≥60 mL/minute per 1.73 m^2 with albuminuria), and CKD stages 3 and 4 (eGFR 15 to 59 mL/minute per 1.73 m^2), respectively.[10] Leg symptoms and sleeping pill use were more frequent in CKD stages 3 and 4 compared with stages 1 and 2 or no CKD.[10] Poor sleep quality, when measured by the Kidney Disease Quality of Life (KDQOL) survey, has been reported in approximately 57% of patients with CKD and was more common in CKD stages 3 to 5 (mean eGFR 24.9 ± 10.6 mL/minute per 1.73 m^2).[11]

Reduced sleep efficiency and greater sleep fragmentation as measured by wrist actigraphy have been noted in CKD compared with non-CKD patients.[12] Similarly, patients with CKD stages 4 and 5 had short and fragmented sleep resulting in total sleep time and sleep efficiency being worse than patients on hemodialysis (HD).[13] Patients with non-HD CKD had a higher prevalence (54.3%) of sleep-disordered breathing and periodic limb movement disorder (30%) compared with the general population.[14]

Although further studies are needed to clarify the heterogeneity of the results, the overall evidence points toward higher prevalence of sleep disorders in CKD patients compared with the general population. This association may also be bidirectional. CKD may promote sleep disturbances with anemia, resulting in restless legs syndrome (RLS) and periodic limb movements during sleep (PLMS), or in fluid gain that is noted in CKD, resulting in sleep-disordered breathing. Similarly, sleep-disordered breathing may result in hypertension, which may then worsen CKD.

INSOMNIA

Insomnia is defined as a persistent difficulty with sleep initiation, duration, consolidation, or quality that occurs despite adequate opportunity and circumstances for sleep, and results in some form of daytime impairment (*International Classification of Sleep Disorders*, third edition [ICSD3]). (See Chapter 80 for comprehensive discussion of insomnia.) Sleep complaints most typically include difficulty initiating or maintaining sleep.

Insomnia symptoms are more common in patients with CKD compared with the general population. About 60% of subjects on HD have insomnia, ranging from difficulty falling asleep in nearly one half of patients to difficulty maintaining sleep or having early morning awakening in one fourth of patients.[15] In another multicenter trial involving more than 11,000 patients on maintenance HD, one half of them reported poor sleep quality.[16] Not surprisingly, the prevalence of routine use of sleep hypnotics was 25.8%,[15] which not only might interact with the patients' other medications but also may adversely affect some physiologic functions such as daytime sleepiness. Although the mechanism is not entirely understood, poor sleep quality appears to increase the mortality risk among patients with CKD.[16]

It is likely that psychological and physiologic factors are responsible for the high prevalence of insomnia complaints in patients with CKD. Higher prevalence of anxiety, stress, and depression is reported in patients with CKD, and these are well known risk factors for insomnia.[17] Rapid fluid, electrolyte, and acid-base changes that occur are often associated with central nervous system symptoms such as changes in arousal and fatigue during or immediately after treatment. A fall in cerebral spinal fluid pH during dialysis and slow movement of bicarbonate across the blood–brain barrier may also be contributing factors to daytime somnolence and subsequent decreased nocturnal sleep quality and ventilatory instability. In addition, several studies have reported an increase in cytokine production secondary to blood interactions with the bioincompatible equipment used in the HD procedure.[18] These substances have both somnogenic and pyrogenic properties and have been linked to a number of postdialysis symptoms, including daytime sleepiness and sleep disturbances.[19] In particular, the somnogenic effects of cytokines are most pronounced during the day when binding of these substances to brain receptor sites is typically much lower than during the night. Dialysis-associated changes in melatonin levels and pattern of secretion may play a role in disrupting the circadian cycle.[20] Timing of HD session may also affect circadian systems by altering exposure to zeitgebers such as wake-up and bed times, activity patterns, meal times, light exposure, and social activity (Figure 134-1). In addition, HD induces a heat load, and patients often respond with an increase in body temperature of approximately 0.5° to 1.0° C; this increase in body temperature may persist for several hours after treatment.[21] These changes may disrupt the circadian regulation of sleep and explain the increased sleep onset and rapid eye movement (REM) latencies and decreased percentage of REM sleep, measures that are specifically linked to body temperature rhythms, in subjects on HD.

There is very little evidence to guide treatment of insomnia among CKD patients. The main goal of treatment is to improve sleep quality to reduce daytime impairment such as daytime fatigue and sleepiness. This can be accomplished by either pharmacologic, nonpharmacologic, or a combination of both of these approaches. The evidence, efficacy, and safety profile of both of these approaches for insomnia are well outlined in Chapters 80 to 88. However, studies using these approaches specifically in insomnia patients with CKD are very limited.

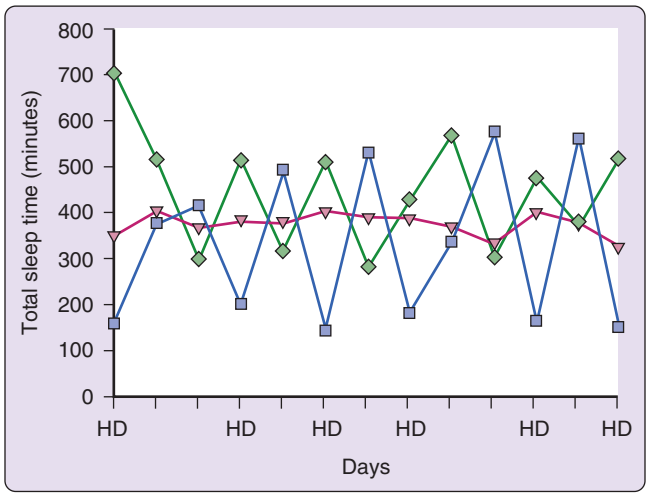


Figure 134-1 Total sleep time (TST) on hemodialysis (HD) treatment and non-HD nights. *Diamonds* represent subjects with longer TST on post-HD nights. *Squares* represent subjects with shorter TST on post-HD nights. *Triangles* represent subjects with a stable sleep pattern on pre- and post-HD nights.

Although nonpharmacologic approaches such as sleep hygiene, stimulus control, and cognitive behavior therapy (CBT) have been well studied in insomnia patients in the general population, there are limited data for CKD patients with insomnia. In a study of patients on peritoneal dialysis, CBT has shown to improve fatigue as measured by the Global Fatigue Severity Scale (GFSC) and decrease inflammatory cytokines.[22] In patients on HD, CBT results in improved PSQI, Global Fatigue Severity Scale, and Beck Depression and Anxiety Inventory scores along with reduced inflammatory markers and oxidative stress.[23] When the benefit of exercise during dialysis was evaluated, it was noted that aerobic activity with moderate intensity during the first 2 hours of a dialysis session could improve sleep quality as measured by the PSQI.[24] The effect of acupressure to improve sleep quality and fatigue in renal patients on HD was evaluated in a randomized controlled trial of 1 month. Patients who received acupressure had significantly lower levels of fatigue, better sleep quality, and less depression compared with controls. Despite the overall limited data of nonpharmacologic approaches to treat insomnia in CKD patients, it is reasonable to consider these approaches because their side effects and complications are relatively very few.

Caution needs to be exercised when considering a pharmacologic approach in CKD patients because of the requirement of dose changes, drug interactions (particularly in the setting of polypharmacy in this patient population), and side effects. In a small randomized study comparing zaleplon with placebo using the PSQI to assess for sleep quality, zaleplon reduced sleep latency and improved sleep efficiency and sleep quality without significant side effects in HD patients.[25] In another randomized crossover study comparing the effects of

clonazepam with zolpidem, both drugs improved PSQI in HD patients; clonazepam was more effective in decreasing PSQI scores than zolpidem, whereas the latter was better tolerated.[26] HD causes physiologic changes that affect the ability to dissipate heat after HD, which may result in insomnia on the nights following treatment. The use of cool diasylate during HD has been shown to improve nocturnal sleep (i.e., shorter sleep latency, longer sleep duration, and longer REM latencies) by decreasing sympathetic activation.[27]

In summary, nonpharmacologic methods such as behavioral techniques and cognitive therapy as well as pharmacologic approaches and combinations of these methods may be used for the treatment of insomnia in patients with CKD, noting the previous caveats. There remains a need for studies assessing the effect of long-term therapies for chronic insomnia in CKD patients.

RESTLESS LEGS SYNDROME AND PERIODIC LIMB MOVEMENTS DURING SLEEP

Although the pathogenesis is not well understood, CKD is one of the secondary causes of RLS and PLMS (see Chapter 95 for comprehensive discussion of RLS and PLMD). Prevalence of RLS in CKD patients not yet on dialysis is reported to be 11% to 26% based on structured interview studies rather than questionnaires alone.[28,29] In nondialyzed CKD patients, it has been reported that approximately 11% fulfilled the RLS criteria of the International Restless Legs Syndrome Study Group[30] compared with 3% in healthy controls.[28] Looking at RLS across different stages of renal dysfunction, the prevalence among CKD patients may be as high as 26%[29] compared with what has been described in the general population (3% to 15%).[31-33] An increased prevalence of RLS has also been described in Japanese patients with CKD; 3.5% of Japanese patients with CKD had RLS compared with 1.5% of healthy controls.[34] The occurrence of RLS is also greater in children with CKD (15.3%) compared with healthy controls (5.9%) and appears to be underdiagnosed.[35,36] In children, there is no association between RLS and CKD stage, etiology, duration, dialysis, or transplant status. Children with RLS are more likely to rate their sleep quality as poor and report using sleep medications.

Among patients with ESRD requiring HD, RLS is present in 14% to 58%.[37-39] Some of the earlier studies suggested even higher prevalence rates, but they may have been limited by inconsistent criteria used for RLS or use of surveys without a structured clinical interview to validate symptoms.[40] The wide range of prevalence may also be due to small sample size in some studies and the location of patient recruitment (dialysis centers vs. sleep centers). Most of the studies suggest prevalence rates of 20% to 30% among ESRD patients requiring HD.

PLMS is reported in 42% of CKD patients on the transplant waitlist compared with 27% of patients after kidney transplantation.[41] A study using 48-hour seven-channel ambulatory polysomnogram (PSG) that included electromyogram signals from bilateral anterior tibialis muscles, found PLMS (defined by periodic limb movement index [PLMI] ≥5/hour of sleep) to be present in 85.4% of HD subjects.[39] Defining "clinically relevant" as PLMI ≥25/hour, 71% met that criterion. They also noted that those who also had RLS had much higher PLMI (median PLMI of 87/hour in those with coexisting RLS versus 16/hour in those without RLS).[39]

These data suggest that the overall rate of RLS is much more frequent in patients with CKD and ESRD than the general population. Similar to patients without CKD, PLMS are much more frequent in those with coexisting RLS. The exact mechanism of how CKD results in increased RLS remains elusive, but some proposed theories associate the sleep-related movement disorder to uremia[42] and low intact parathyroid hormone,[43] whereas the association with reduced ferritin has been much less robust compared with subjects without CKD. There also appears to be a strong genetic influence in single-nucleotide polymorphisms that may still influence the development of RLS even among CKD patients.[44]

Clinical Characteristics

Similar to in the general population, RLS appears to be more prevalent among females with CKD, but the mean ages between those with and without RLS do not differ.[28,29,37] Furthermore, the prevalence of RLS appears to increase with worsening eGFR.[34,45] Molnar and colleagues looked at a group of 176 CKD patients on a waitlist for renal transplantation and found the prevalences of RLS to be 1.8%, 5.1%, 6.5%, and 23.5% in patients with eGFRs greater than 60, 30 to 59, 15 to 29, and less than 15 mL/minute per 1.73 m^2, respectively[45] (Figure 134-2). In HD patients, there are inconsistent findings regarding RLS association with sex, age, iron, hemoglobin, and ferritin.[37,38,46] One study concluded the only difference in laboratory values between HD patients with or without RLS was the level of C-reactive protein.[47] Compared with idiopathic RLS subjects, those with RLS and ESRD had much higher PLMI (103.6 ± 74.4 vs. 22.0 ± 18.9; P <.001) and much longer suggested immobilization test index (PLMS per hour of immobility tested during awake) (127.9 ± 82.3 vs. 13.8 ± 29.0; P <.001).[48] These findings suggest some physiologic differences between uremic RLS and idiopathic RLS.

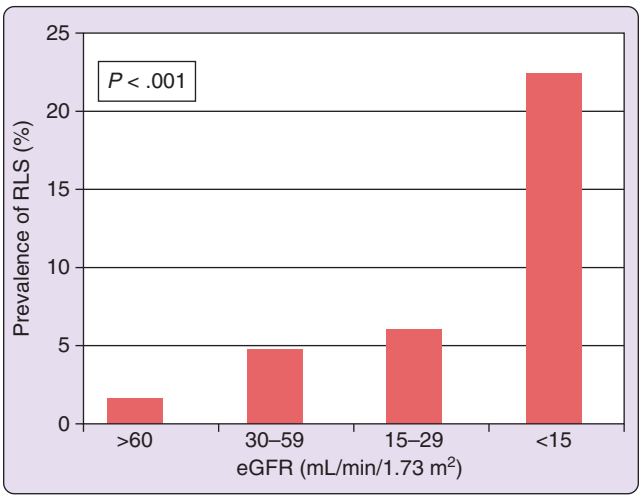


Figure 134-2 Association between presence of restless legs syndrome (RLS) and severity of chronic kidney disease based on estimated glomerular filtration rate (eGFR). (From Molnar MZ, et al. Restless legs syndrome in patients after renal transplantation. *Am J Kidney Dis* 2005;45[2]:388–96, used with permission.)

Morbidity and Mortality

Not unexpectedly, RLS is independently associated with a multitude of sleep-related complaints, worse quality of life, and increased mortality. In patients with CKD and ESRD, RLS is independently associated with greater fatigue.[49] Part of the reason for this increased fatigue may be the increased risk for sleep apnea (see Sleep-Related Breathing Disorder) and insomnia in CKD patients with RLS.[39,50,51] RLS is also independently associated with depression in patients on HD and after renal transplantation, even after adjusting for the presence of insomnia or use of antidepressants (because antidepressants may cause RLS symptoms).[52,53] These findings likely explain the worse quality of life reported by patients with CKD and RLS compared with those with idiopathic RLS.[54] RLS is independently associated with lower quality-of-life assessments even after adjusting for insomnia, age, gender, and comorbid conditions.[51,55,56]

Although the data on RLS and PLMS contributing to cardiovascular morbidity in the general population are inconclusive, there are some data suggesting that RLS in ESRD patients may increase cardiac morbidity. The incidence of new cardiovascular events (myocardial infarction, cerebral stroke, or peripheral artery occlusion) in patients with CKD and RLS is 64.5% compared with 39.1% in patients without RLS and likely contribute to a higher 18-month mortality in this group[57] (Figure 134-3). In a follow-up study of nearly 30 months, ESRD patients with RLS have increased adjusted hazard ratio (HR) of mortality (adjusted for age, sex, comorbid conditions, functional status, and clinic location) of 1.39 (HR = 1.08 to 1.79) compared with ESRD patients without RLS.[55] Increased PLMS seem to predict mortality. Survival rate at 20 months among ESRD patients with PLMI of more than 20/hour is 50% compared with 90% in ESRD patients with PLMI of less than 20/hour.[58]

Treatment

Based on a limited number of studies addressing the needs of CKD patients, the overall treatment approach to RLS does not appear to be different from that of non-CKD patients.[59]

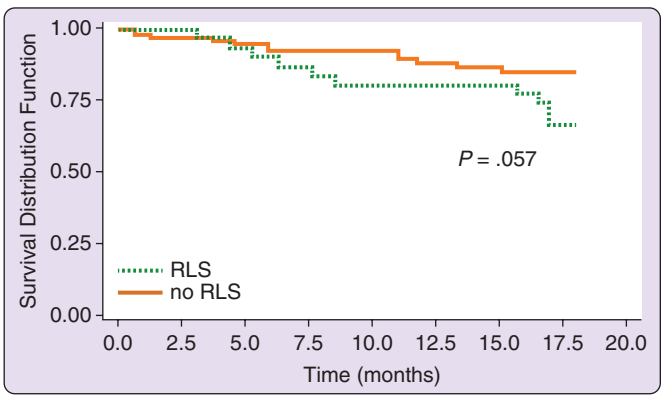


Figure 134-3 Kaplan-Meier estimates of all-cause mortality at 18 months in patients with and without restless legs syndrome (RLS). (From La Manna G, et al. Restless legs syndrome enhances cardiovascular risk and mortality in patients with end-stage kidney disease undergoing long-term haemodialysis treatment. *Nephrol Dial Transplant* 2011;26[6]:1976–83, used with permission.)

Few small trials have shown the benefit of exercise in reducing RLS symptoms, including exercise during dialysis.[60,61] In CKD patients, the treating clinician must keep in mind that most of the medications currently used to treat RLS are renally excreted. Therefore it would be advisable to start at the lowest dose possible and uptitrate the dose at a slower rate than for those with normal renal function. In patients who are on dialysis, it would be advisable to give a dose after dialysis. There are few comparative studies of different agents, and those studies are significantly limited by small sample size, such that one class of medication cannot be recommended over another as the initial agent. Medications such as iron supplementation, gabapentin, dopamine agonists, benzodiazepines, and narcotics could be considered based on severity of symptoms.[59,62]

Kidney transplantation is reported to have some benefit in improving RLS symptoms. Resolution of RLS symptoms is reported to occur between 1 and 38 days after undergoing renal transplantation, and symptoms may recur in 10 to 60 days in those whose transplantation fails.[63,64] In a follow-up period of up to 3 years, patients whose transplanted kidneys are still functioning have continued resolution of RLS.[64] These findings suggest that normalizing renal function can improve RLS symptoms, but the specifics of this mechanism remain elusive.

SLEEP-RELATED BREATHING DISORDER

There appears to be a strong association between CKD and sleep apnea, of which the most common form is obstructive sleep apnea (OSA) (see Section 14 for a comprehensive discussion of sleep breathing disorders). Its prevalence in CKD and HD patients ranges from 30% to 73%.[14,65-68] This wide range of reported prevalence is due to the variable cutoff value of the apnea-hypopnea index (AHI) or respiratory disturbance index (RDI) that was used to define OSA as well as the varying degree of renal dysfunction. For example, in a study of 1624 subjects who were diagnosed with OSA (based on AHI ≥5 events/hour), 30.5% had CKD.[65] This prevalence of CKD was threefold higher in the OSA group than the control group (prevalence of 9.1%).[65] In a much smaller study of 63 subjects on HD, 51% had OSA, defined as RDI of 15 or more.[69] It is however worthwhile to note that a cross-sectional study of patients enrolled in the Kaiser Permanente health care system found that among 85,376 patients with CKD, only 3.3% had sleep apnea (based solely on diagnostic codes or positive airway pressure device prescriptions).[70] This low prevalence rate seems inconsistent with most of the other literature, and in fact the 3.3% prevalence rate is even lower than the rate among the non-CKD population.[71,72]

Despite the variability in the prevalence studies mentioned previously, a more consistent finding is the increasing prevalence of OSA with worsening renal function.[73-75] Using an RDI of 15 or greater, 38% of subjects with CKD were shown to have OSA, whereas 51% of subjects with ESRD had OSA.[69] It has been demonstrated that for each 10 mL/minute per 1.73 m^2 decrease in eGFR, the odds ratio for OSA is 1.42, even after adjustment for age, body mass index (BMI), and presence of diabetes.[76] Similarly, it has been shown that patients on HD have an odds ratio of 4.14 for having an AHI of more than 15, compared with 2.19 in CKD patients (eGFR ≤40 mL/minute per 1.73 m^2).[75]

A few studies specifically address central sleep apnea (CSA) in CKD. The number of CSA events may be six times higher, with greater resultant desaturations, in those with CKD than those with normal renal function.[66] Compared with those with eGFR of 90 mL/minute per 1.73 m^2 or greater, the percentage of total RDI that is due to CSA events is three times greater in those with eGFR of 60 mL/minute per 1.73 m^2 or less (14.9% of total RDI compared with 4.9%).[13] Patients on HD with sleep apnea have CSA-predominant sleep apnea (defined by CSA index >5).[77] In these CSA-predominant subjects, there is a greater incidence of atrial fibrillation. Interestingly, the central apnea index is lower on the night of HD, suggesting that volume overload may contribute to the development of CSA in susceptible patients.[77]

Identifying CKD patients with OSA requires a high index of suspicion because these patients seem to present with less severe or atypical sleep-related complaints.[78-80] A history of snoring and witnessed apneas are reported less often in ESRD subjects than in controls.[80] Similarly, complaints of unrefreshing sleep and morning headaches are less frequent in ESRD subjects.[80] The mean maximal snoring intensity during the PSG is less in ESRD subjects compared with controls.[80] Furthermore, subjects with ESRD have lower BMI and neck circumference compared with controls with comparable AHI.[80] When compared with CKD subjects without OSA, CKD patients with OSA do not report worse sleepiness as measured by the Epworth Sleepiness Scale.[79]

Pathophysiology

CKD may potentiate the likelihood of sleep apnea through mechanisms of volume overload, fluid redistribution, and altered chemoresponsiveness. Internal jugular vein volume and upper airway mucosal water content, and not upper airway cross-sectional area, correlate with AHI in subjects undergoing HD.[81] This correlation remains significant even after adjusting for age, sex, height, BMI, and percent reduction of urea. This is consistent with the finding that even a 0.5-L rostral fluid shift during recumbency can lead to significant increase in neck circumference and upper airway resistance.[82,83] The findings that nocturnal HD can reduce AHI and that AHI may increase back to baseline when off of nocturnal HD support the fluid-shift theory because nocturnal HD can mitigate nocturnal fluid shifts by extracellular fluid removal at night and by improved ultrafiltration owing to longer dialysis periods.[4]

Another contributing mechanism may be altered chemosensitivity. Unstable chemosensitivity leads to destabilization of respiratory control and may contribute to the severity of OSA.[84] Subjects with ESRD and coexisting OSA have higher ventilatory sensitivity to arterial partial pressure of carbon dioxide than those without OSA.[68] This relationship is independent of age, sex, or BMI. By switching from conventional HD to nocturnal HD, chemoresponsiveness to hypercapnia may be reduced and lead to a decrease in AHI.[85]

Although chronic uremia has also been proposed to result in upper airway muscle dysfunction, there have not been convincing data to confirm this theory. Rather, it is more likely that multiple factors are ultimately responsible for the increased likelihood of OSA in CKD. For any given individual, one factor may have a greater influence over another, leading to the variable expression of OSA in these patients.

Influence of Sleep Apnea on Outcomes

Patients with OSA have pathophysiologically shared comorbidities, including hypertension, diabetes mellitus, and obesity, with CKD patients, and therefore it is difficult to attribute worsening kidney function to OSA alone. However, there is some evidence to suggest that the coexistence of OSA in patients with CKD may worsen microalbuminuria and eGFR and be associated with increased levels of cystatin C (a sensitive biomarker that reflects impaired renal function and is associated with latent CKD).[86,87] Furthermore, being at high risk for OSA (as defined by scoring positively in two of three main domains in the Berlin Sleep Apnea Questionnaire) may be an independent risk factor for kidney graft loss, particularly in female transplant recipients.[52] Potential mechanisms by which OSA contributes to worsening renal function may include intermittent hypoxemia, sympathetic surges associated with arousals, and systemic inflammation, which may lead to endothelial dysfunction and tubulointerstitial injury.[88,89]

OSA also contributes to worse quality of life, as it does in those without CKD. Those with OSA and CKD score worse in vitality, social functioning, and mental health measures on the Short Form (36) Health Survey.[90] Patients with CKD and moderate or severe OSA (AHI >15) have greater complaints of excessive daytime sleepiness and impairment in verbal memory, working memory, attention, and psychomotor speed compared with those without OSA.[91]

Treatment

With OSA and CKD independently increasing cardiovascular morbidity and mortality, it is crucial to treat OSA to mitigate these consequences. Unfortunately, although there is an exhaustive literature confirming the benefit of continuous positive airway pressure (CPAP) therapy in treatment of OSA, there is a limited literature regarding CPAP therapy specifically in CKD patients.[92] Some studies have shown the efficacy of CPAP in improving glomerular hyperfiltration and eGFR in subjects with OSA but with normal baseline renal function.[53,93] Adaptive servoventilation therapy in CKD patients with moderate to severe obstructive and central sleep apnea, and heart failure (mean ejection fraction of around 45%), may result in improvement in eGFR, New York Heart Association functional class, and serum levels of brain natriuretic peptide, creatinine, cystatin C, C-reactive protein, and noradrenaline.[94,95]

Beyond positive airway pressure devices, changing the type of dialysis in renal failure patients has been shown to improve sleep apnea. As mentioned earlier, nocturnal HD compared with conventional HD may improve AHI severity[4] (Figure 134-4). Nocturnal HD may also lead to a decrease in sleep percent time spent at oxygen saturation less than 90%, a decrease in heart rate, and an increase in vagal tone.[96] Nocturnal HD may also result in a decrease in chemoreflex response and improve AHI.[85]

Nocturnal peritoneal dialysis (NPD) may also result in improvement in OSA compared with continuous ambulatory peritoneal dialysis (CAPD). When subjects matched for demographics, BMI, comorbidities, adequacy of dialysis, and peritoneal transport properties are converted from NPD to CAPD, AHI worsens markedly.[97] It appears that NPD is more effective in treating OSA because of its greater efficacy in fluid removal based on greater reduction in total body water,

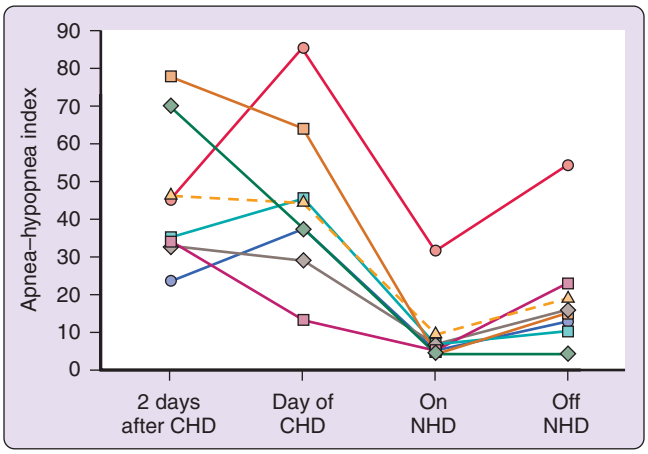


Figure 134-4 Apnea-hypopnea index (AHI) in seven patients with a baseline AHI >15 events/hour. Mean values are represented by the *broken line*. CHD, Conventional hemodialysis; ND, nocturnal hemodialysis.

as measured by multifrequency bioelectrical impedance analysis, and in percent reduction in hydration fraction, despite similar targeted dialysis regimen adjusted to achieve euvolemia and a weekly Kt/V_{urea} (measure of dialysis adequacy) of 1.8 to 2.1 in both groups.[97] Magnetic resonance imaging assessment of the upper airways confirms significant reductions in nasopharyngeal and oropharyngeal volumes, minimal pharyngeal cross-sectional area, and tongue volume enlargement after converting to NPD from CAPD.[98]

Several studies report improvements in OSA after renal transplantation. Since earlier reports of significant improvement in AHI after transplantation,[99] there have been conflicting results reported by others. Several reports suggest that not all patients undergoing transplantation had resolution of their OSA, with some suggesting that less than half of their study patients had resolution.[100-102] The small sample sizes of these reports and varying durations between the pretransplantation and posttransplantation PSG (which may result in BMI increase due to immunosuppressives) may account for the conflicting results. However, taken together, these reports suggest that some patients with OSA may significantly improve after transplantation.

CPAP should still be considered the initial therapy of choice for patients with OSA and CKD because of its proven efficacy. For CKD patients, there appears to be other means by which OSA can be controlled. These additional considerations include converting from conventional to nocturnal HD or from continuous ambulatory peritoneal dialysis to nocturnal peritoneal dialysis. Furthermore, transplantation may potentially improve OSA, but this effect may be counterbalanced by the potential weight gain associated with medications required as part of the antirejection regimen.

CLINICAL PEARL

Concurrent sleep disorders are common in patients with CKD, and the prevalence of various sleep disorders seems to increase with worsening kidney function. Because of coexisting comorbidities, their presenting symptoms may be atypical or initially attributed to other conditions. Thus it is important for the clinician to specifically inquire about coexisting sleep disorders such as insomnia, RLS, or sleep apnea. When employing pharmacologic therapies, clinicians must carefully consider renal metabolism of some medications and appropriately adjust the dose or timing of the medication. Treating these coexisting disorders can substantially improve patients' quality of life and, particularly in the case of sleep apnea, may also mitigate worsening morbidities.

SUMMARY

Sleep disturbances such as insomnia, RLS, PLMS, and OSA are common across the spectrum of CKD. The pathophysiologic mechanisms that drive these sleep disorders in CKD patients are complex and incompletely understood. What is appreciated is that sleep disorders have significant effects on morbidity, quality of life, functional status, and possibly mortality in this population. Therefore an assessment of sleep-related complaints and sleep disorders should be pursued in all CKD patients. Treatment of sleep disorders may be complex and unique to this population because of the nature of their underlying medical disease.

Selected Readings

Abuyassin B, et al. Obstructive sleep apnea and kidney disease: a potential bidirectional relationship? *J Clin Sleep Med* 2015;**11**(8):915–24.

Al Mawed S, Unruh M. Diabetic kidney disease and obstructive sleep apnea: a new frontier? *Curr Opin Pulm Med* 2016;**22**(1):80–8.

Elder SJ, et al. Sleep quality predicts quality of life and mortality risk in haemodialysis patients: results from the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Nephrol Dial Transplant* 2008;**23**(3):998–1004.

Lee JJ, et al. Improvement of sleep-related breathing disorder in patients with end-stage renal disease after kidney transplantation. *Clin Transplant* 2011;**25**(1):126–30.

Lee J, et al. The prevalence of restless legs syndrome across the full spectrum of kidney disease. *J Clin Sleep Med* 2013;**9**(5):455–9.

Lindner AV, Novak M, Bohra M, Mucsi I. Insomnia in patients with chronic kidney disease. *Semin Nephrol* 2015;**35**(4):359–72.

Molnar MZ, et al. Restless legs syndrome in patients after renal transplantation. *Am J Kidney Dis* 2005;**45**(2):388–96.

Owada T, et al. Adaptive servoventilation improves cardiorenal function and prognosis in heart failure patients with chronic kidney disease and sleep-disordered breathing. *J Card Fail* 2013;**19**(4):225–32.

Puckrin R, Iqbal S, Zidulka A, et al. Renoprotective effects of continuous positive airway pressure in chronic kidney disease patients with sleep apnea. *Int Urol Nephrol* 2015;**47**(11):1839–45.

Sabbatini M, et al. Zaleplon improves sleep quality in maintenance hemodialysis patients. *Nephron Clin Pract* 2003;**94**(4):c99–103.

Szentkiralyi A, et al. High risk of obstructive sleep apnea is a risk factor of death censored graft loss in kidney transplant recipients: an observational cohort study. *Sleep Med* 2011;**12**(3):267–73.

A complete reference list can be found online at ExpertConsult.com.

Chapter 135

Sleep in the Critically Ill Patient

Siavash Farshidpanah; Margaret A. Pisani; E. Wesley Ely; Paula L. Watson

Chapter Highlights

- Sleep disruption was for many years regarded as an unfortunate, unavoidable, and relatively unimportant consequence in the course of critical illness management. Improved understanding of the potential importance of sleep in recovery has made sleep in the intensive care unit a new area of research and clinical focus.
- Poor sleep quality is common in critically ill patients, and efforts to understand the causes and promote improved sleep may result in a positive impact on patient outcomes such as delirium, time on mechanical ventilation, and post-illness mood disorders.
- Efforts to study sleep in critically ill patients are currently labor-intensive, expensive, and limited by electroencephalographic abnormalities caused by neuropathology and frequently used psychoactive medications. However, clinicians and researchers should be excited about more cost-effective technology and new proposed staging criteria based on the most recent data.
- Improving the sleep of critically ill patients should begin with optimizing environmental factors and minimizing use of medications known to disrupt sleep. If necessary, pharmacologic sleep aids can be used.

INTRODUCTION

More than four million adults are cared for in U.S. intensive care units (ICUs) each year.[1,2] With advancements in care, more patients are surviving their critical care illness, and significant efforts are now being directed toward optimizing modifiable risk factors that can further improve ICU outcomes. Nearly 60% of patients report sleep disruptions as a major cause of distress during their ICU stay. Poor sleep quality during critical care illness has been described for the past four decades[3-6] and was previously perceived as an unavoidable effect of ICU management. In addition to emotional distress, sleep deprivation in this group of patients has been hypothesized to contribute to impaired immune function, prolongation of mechanical ventilation, delirium, and cognitive dysfunction. In this chapter, we will review the methods available to measure sleep, the characteristics of sleep in ICU patients, clinical outcomes that might be adversely affected by poor sleep quality, and, last, methods to improve sleep in critically ill patients.

MEASUREMENT OF SLEEP IN THE INTENSIVE CARE UNIT

Sleep can be evaluated in three domains: quantity, quality, and pattern of distribution. Nursing logs and systematic observation by health care staff would be an obvious, cost-effective tool to evaluate sleep. Despite a few early efforts that found good correlation between nursing observation and polysomnographic (PSG) measures of sleep,[7,8] other studies show a typical overestimation on the part of health care staff and an underestimation by patients.[4,5,9,10] PSG is the gold standard method for measuring sleep in ICU patients, although other modalities have been studied, and alternative approaches are being explored.

Polysomnography

PSG is regarded as the most reliable method for measuring sleep in critically ill patients. However, several important factors hinder its widespread use to evaluate sleep in this population. Trained personnel are required to conduct the initial setup and maintain electrode signal quality throughout the study. The use of electroencephalographic (EEG) electrodes and leads may interfere with routine patient care needs, and the quality of PSG recording may be compromised by electrical interference from the many patient care devices found in the typical ICU room.[11] The measurement of sleep is also complicated by the atypical sleep patterns frequently present in ICU patients. Critically ill patients frequently experience as much sleep time during the day as during the night[12] (Figure 135-1). Therefore, if a full assessment of a patient's sleep is to be obtained, 24-hour recordings are necessary.

Although PSG is the gold standard for sleep measurement, there is now a widespread acceptance by researchers in the field that the standard Rechtschaffen and Kales (R&K) and even the newer American Academy of Sleep Medicine (AASM) guidelines for scoring are not reliable in critically ill patients.[13,14] The interpretation of EEG data in this population is complicated by the effects of both the underlying illness and the multiple psychotropic medications that these patients receive during their ICU stay.[15-17] Cooper and colleagues[12] were the first to note the difficulty in applying

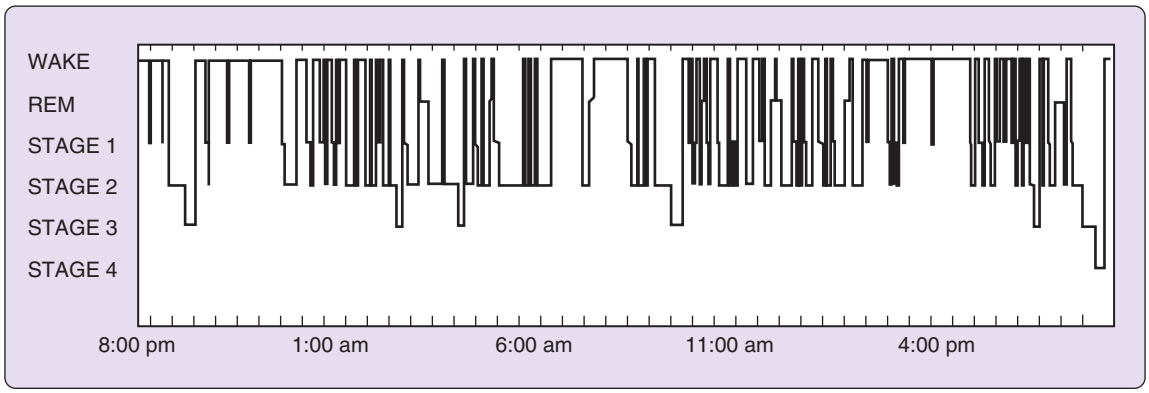


Figure 135-1 A 24-hr hypnogram of an ICU patient. Sleep during critical care illness is severely fragmented and distributed evenly across 24 hours. Oftentimes, a clear circadian rhythm is not seen in these patients. REM, Rapid-eye movement sleep. (From Cooper AB, Thornley KS, Young GB, et al. Sleep in critically ill patients requiring mechanical ventilation. *Chest* 2000;117:809–18, with permission.)

standard scoring criteria to this population, finding that electrophysiologic sleep was not identifiable in 12 of 20 ICU patients in their study. These 12 patients either lacked the normal transitions of sleep stages or had slowing of EEG activity—sometimes without evidence of EEG activation to painful stimuli and, thus, more consistent with coma than sleep. These patients were noted to have received higher doses of sedative medications than patients with EEG-identifiable sleep.

Further evidence of the limitations of R&K criteria was noted by Ambrogio and colleagues[18] who compared manual sleep assessment to spectral analysis. They found that the interobserver reliability of R&K methodology was poor ($\kappa = 0.19$). Two other research groups have also evaluated the limitations of R&K criteria in critically ill patients and have proposed new scoring criteria. Drouot and colleagues,[19] in a study of 57 ICU patients who had been free of sedative and other psychotropic medications for at least 48 hours, found that R&K sleep scoring was not feasible in 16 (28%) patients. They recommend adding two new states: atypical sleep and pathologic wakefulness (Figure 135-2). Watson and colleagues[20] encountered similar limitations of the standard scoring criteria in their study of 37 mechanically ventilated patients. As previously described by Cooper and colleagues and by Drouot and colleagues, patients in this study also had atypical sleep characterized by a lack of K-complexes and sleep spindles as well as pathologic wakefulness. As opposed to the study by Drouot, patients receiving psychotropic medications were included in this investigation. Watson and colleagues proposed a comprehensive scoring criteria that incorporated both stages of encephalopathy (based on previous neurology literature[21]) and typical sleep (Figure 135-3, *A* and *B*). This approach allowed the co-evaluation of sleep state and consciousness in critically ill patients throughout the course of their illness even while pathologic brain activity was present. While there is controversy on whether this approach should be called a *sleep* staging method,[22,23] there is evidence that tracking the pathologic brain waves can lead to discoveries regarding prognosis and outcomes (Figure 135-4).[24,25] At this time, there are no scoring criteria that have been adequately validated for measurement of sleep in critically ill patients, and more research is needed in this area.

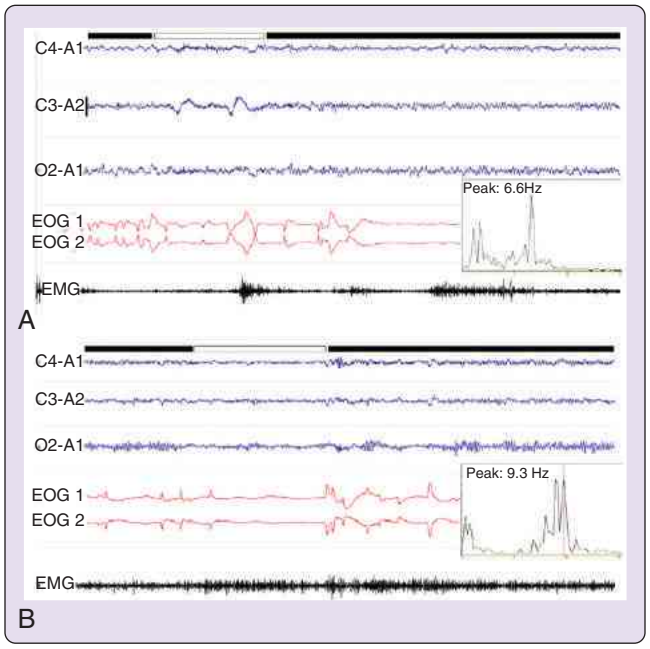
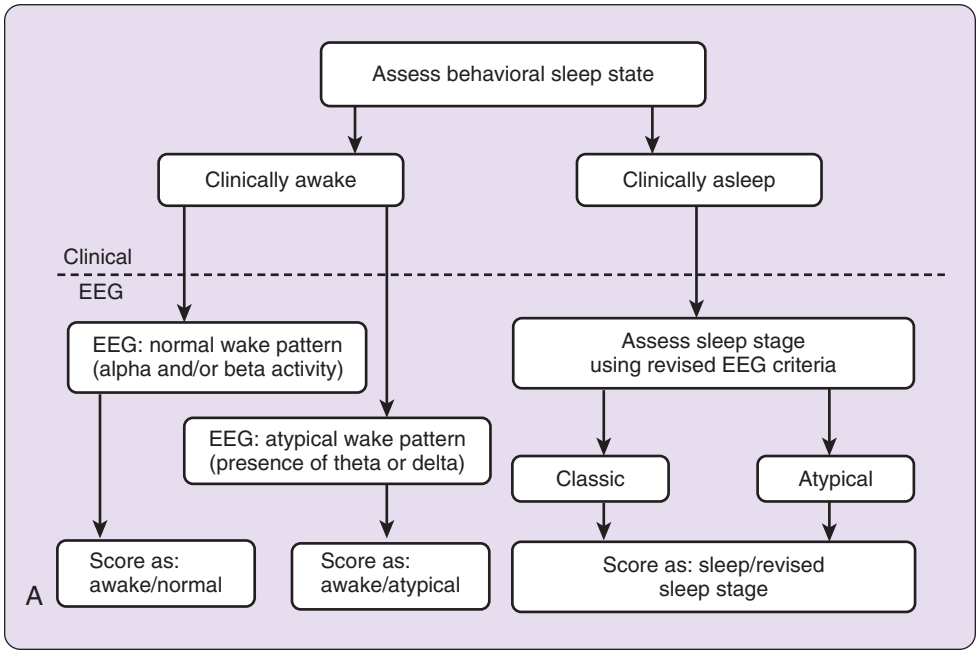


Figure 135-2 Polysomnographic features of critically ill patients showing pathologic wakefulness. **A,** EEG shows slow background frequency with numerous rapid eye movements typical of wakefulness and high submental EMG activity. These findings were obtained while the patients were fully awake. *Inset* shows spectral power with peak EEG frequency. **B,** Alpha waves in a normal wake EEG occurring only during periods with closed eyes. *Inset* shows spectral power with peak EEG frequency. The *white horizontal bars* indicate periods with open eyes, and the *black bars* indicate periods with closed eyes. EEG, Electroencephalogram; EMG, electromyogram. (From Drouot X, Roche-Campo F, Thille AW, et al. A new classification for sleep analysis in critically ill patients. *Sleep Med* 2012;13:7–14, with permission.)

Processed Electroencephalographic Monitors

Given the difficulty and expense of performing PSG, there has been interest in the use of processed EEG devices to monitor sleep in the ICU. Processed EEG monitoring tools, such as the bispectral index (BIS) monitor, were introduced into clinical care in the 1990s to monitor levels of sedation in patients undergoing anesthesia. These devices record brain



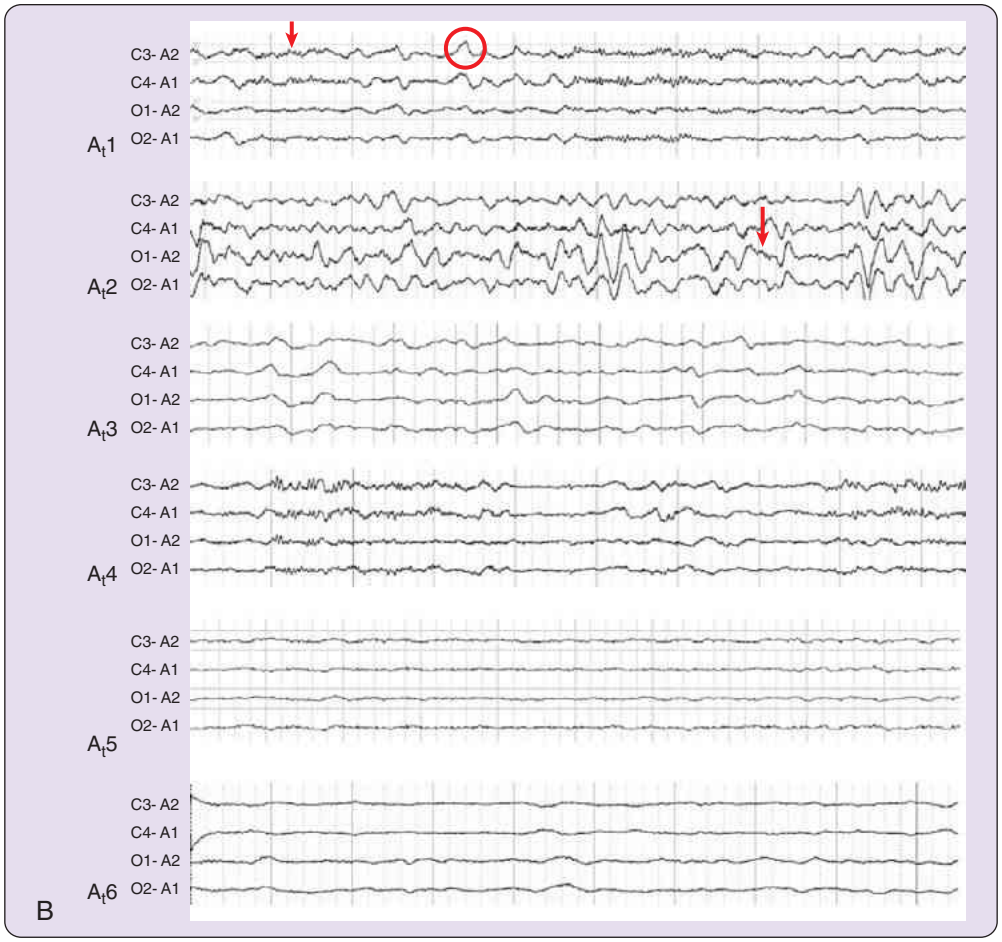


Figure 135-3 **A,** Proposed approach to scoring sleep in critically ill patients. Due to the presence of pathologic wakefulness, a clinical assessment of the behavioral sleep state is first necessary. It should then be determined whether the EEG pattern is normal. If normal, classic sleep stages can be used. If the EEG pattern is abnormal, atypical stages as demonstrated in **B** should be used. **B,** Six proposed atypical sleep stages noted on EEG obtained during 24-hr illness. **A_t1** (atypical stage 1), characterized by having at least 10% alpha and/or theta activity (indicated by the *arrow*) but may also include delta activity (indicated by the *circle*). **A_t2** (atypical stage 2), characterized by the presence of polymorphic delta but with the presence of background beta, alpha, or theta activity (indicated by the *arrow*). **A_t3** (atypical stage 3), characterized by a polymorphic delta activity without the presence of background beta, alpha, or theta activity. **A_t4** (atypical stage 4), defined by a burst-suppression pattern, intermittent EEG activity alternating with periods of isoelectric EEG activity is classified as atypical stage 4. **A_t5** (atypical stage 5), defined by a suppression pattern EEG, a very low-voltage EEG activity (<20 μV amplitude) is classified as atypical stage 5. **A_t6** (atypical stage 6), characterized by a complete lack of EEG/cortical activity as shown. EEG, Electroencephalogram. (From Watson PL, Pandharipande P, Gehlbach BK, et al. Atypical sleep in ventilated patients: empirical electroencephalography findings and the path toward revised ICU sleep scoring criteria. *Crit Care Med* 2013;41:1958–67, with permission.)

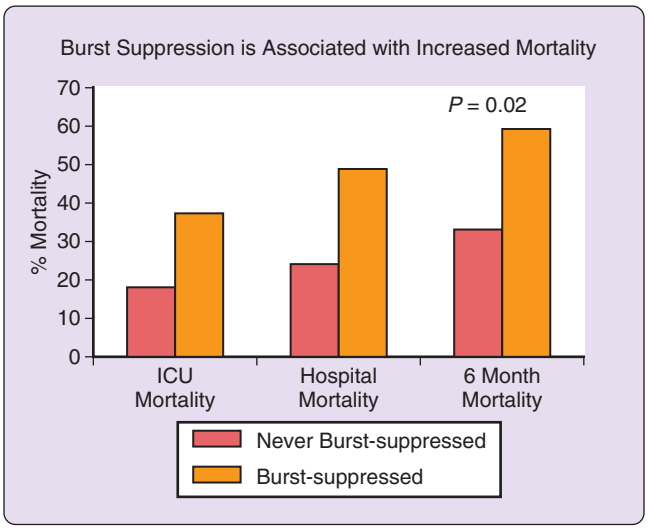


Figure 135-4 Presence of burst suppression is associated with increased mortality. (From Watson PL, Shintani AK, Tyson R, et al. Presence of electroencephalogram burst suppression in sedated, critically ill patients is associated with increased mortality. *Crit Care Med* 2008;36:3171–7, with permission.)

activity via two to four leads placed on the forehead and report a numerical value calculated from analysis of the EEG waveforms that correlates with clinical depth of anesthesia in healthy volunteers.[26] Reference values of 90 to 100 infer an awake or conscious clinical state, whereas a value of zero reflects isoelectric EEG activity.[27,28] Effort has been made to evaluate the utility of BIS in assessing the level of sedation outside of the operating room, including various ICU settings.[29-36]

At the time of this printing, two studies have aimed to use BIS in the evaluation of sleep rather than sedation or consciousness levels in critically ill patients. Nicholson and colleagues[37] observed sleep in 27 medical ICU patients during the recovery phase of critical illness and used BIS values to classify patients as awake (if more than 85), light sleep (if 60 to 85), slow wave sleep (SWS) (if less than 60), or as rapid eye movement (REM) sleep (if BIS>60 with reduced electromyogram [EMG]). Bourne and colleagues[38] used BIS monitors in a study aimed at evaluating the effects of exogenous melatonin on sleep (which they defined as a BIS level below 80) in the ICU. However, both studies relied on previously published BIS values for sleep identification that had been established in healthy volunteers,[39,40] and neither study used PSG as a reference to stage sleep. No study to date has validated BIS measurement of sleep in ICU patients by comparing BIS to PSG as a standard.

The EEG in critical illness is frequently altered, showing a slowing in EEG frequency even in the wake state, which will skew the BIS value to lower than expected for a patient's level of consciousness. Conversely, the BIS signal is subject to electrical interference, which can lead to spuriously high BIS values and, thus, overestimating a patient's level of consciousness. Additionally, the overlap of the BIS values for a given stage of sleep likely prevents its use as a depth of sleep monitor.[41] These issues currently decrease the reliability of processed EEG monitors when used alone for measurement of ICU sleep.

Actigraphy

Actigraphy has been widely used in research for nearly three decades as a surrogate measure for sleep in clinical practice.[42] Although not officially recommended for monitoring of sleep in the ICU setting, many groups have attempted to validate its use in this population with varying degrees of success.[43] Actigraphy in critically ill patients does not correlate well with PSG[44] and lacks specificity for identifying sleep transitions.[45,46] This is not surprising, given that the basis of actigraphic "sleep" is lack of limb movement, which may occur for various other reasons (neuromuscular blockade, physical restraint, acute neurologic deficit, ICU myopathy, etc.) in critically ill patients. Nonetheless, the relative unobtrusive nature and cost-effectiveness of this tool make it appealing for use in long-term follow-up studies on the effects of sleep-promoting interventions.

Devices Based on Physiologic Changes during Sleep

As discussed elsewhere in this volume, there are well-documented autonomic changes that occur during the different stages of sleep. Investigators have tried to capitalize on these physiologic variations to develop automatic scoring algorithms that can identify sleep and its divided stages. These devices are unlikely to be useful in ICU patients. By virtue of their critical illness, patients' physiology is unlikely to be at baseline, and their autonomic response is often artificially manipulated via medications to acutely stabilize the patient in the ICU. Therefore, despite their validation in healthy subjects,[47,48] it is extremely doubtful that such devices can be reliably used for monitoring sleep in critically ill patients.

It is out of the scope of this chapter to provide a comprehensive list of devices or techniques and review the evidence. Currently, the gold standard to assess sleep in the ICU remains to be PSG. As a whole, the atypical EEG of ICU patients coupled with changes in guidelines from R&K to AASM have created a heterogeneous landscape of reported data over the past 40 years. However, despite these limitations, noticeably persistent patterns are observed when evaluating sleep in the ICU. In the next section, we will review findings from currently reported studies to review characteristics of sleep in the critically ill.

CHARACTERISTICS OF SLEEP IN THE INTENSIVE CARE UNIT

The characteristics of sleep can be broadly divided into those reported subjectively by patients themselves and those measured objectively by a range of assessment tools used by health care staff or clinical researchers.

Subjectively Reported Quality of Sleep

The overall results from studies to date reveal a consistent perception of poor sleep quality.[5,10,49,50] Patients specifically complain about trouble initiating sleep, experiencing lighter sleep, and noticing frequent awakenings with significant difficulty returning to sleep.

Subjective sleep quality reports are based on survey tools that patients complete at variable time points during their ICU stay. A few traditional sleep surveys, such as the Pittsburgh Sleep Quality Index (PSQI), have been applied in ICU

populations. Other tools, such as the Sleep in Intensive Care Questionnaire and the Richards Campbell Sleep Questionnaire (RCSQ), have been specifically developed for and validated in critically ill patients.[5,51] These tools are cost-effective, readily available, and easily applied by novice personnel without significant need for training. Although they have been compared even to PSG findings in ICU patients,[7,8,51,52] the chief limitation of self-report instruments for ICU patients is that sufficient cognitive function and memory recall function is required to render the survey reliable. For example, a study by Bourne and colleagues[46] reported that 20% of patients were unable to complete the RCSQ mainly due to delirium. A study led by Frisk[53] found an even higher failure rate, estimating that 50% of ICU patients were neither conscious nor orientated enough to complete the RCSQ. Although patient self-reports are helpful in providing an intimate look into the psyche and perception of critically ill patients,[54] a quantitative analysis of sleep may be more reliable.

Objectively Reported Sleep

The inherent limitations of objectively assessing sleep in the ICU were discussed earlier in this chapter. Despite these known difficulties, investigators have remained persistent in their efforts, and the existing sleep characteristics data will be reviewed next.

Critically ill patients appear to spend a prolonged time in non–rapid eye movement sleep stages N1 and N2.[7,44,55,56] Mild variations are reported among different ICU cohorts, but this difference is unlikely to be clinically significant. For example, in surgical ICU patients, N1 and N2 sleep comprised 96% to 99% of total sleep time (TST).[4,57,58] Meanwhile, medical and trauma ICU patients had 76%, 83%, and 80% to 100% of TST in N1 or N2 sleep, in three separate studies.[10,12,59]

The majority of data in critically ill patients show a decrease in SWS ranging from less than 1% [58] to 9 ± 18% of TST.[60] However, a study by Broughton and Baron[61] reported prolonged SWS up to 25% in patients immediately after myocardial infarction. More recently, Cabello[62] found SWS to be near the upper limits of normal (19% TST) in long-term ventilated medical ICU patients.

In the critically ill population, REM sleep is reduced across all ICU settings. Studies have shown a moderate decrease (10% to 15% of TST),[10,61,62] dramatic reduction (0.02% to 0.36% of TST),[20,58,63,64] including absence of REM sleep[60] in critically ill patients. Although no formal hypothesis exists, several explanations for the diminution of REM are worth considering. Some of the most commonly used medications in ICU (e.g., benzodiazepines, opioids) are known to suppress REM sleep.[65] Unconventional transitions through sleep stages in ICU patients have been reported,[12,55,58] which may mean that the normative data in the healthy cohort do not directly apply to the ICU. Another explanation may involve the discontinuous nature of sleep in the ICU precluding the natural progression toward REM throughout the night.

Reports on TST vary between studies that monitored only overnight sleep or an entire 24-hour cycle. If only overnight sleep is recorded, there is a trend toward decreased sleep ranging from 1 to 6 hours.[3,4,10,66] However, TST is surprisingly within the normal range of 6 to 8 hours when all sleep during a 24-hour cycle is captured.[12,57,58,60] Also, as might be expected, in patients receiving sedative mediations, the total "sleep" time is prolonged (10.9 to 18.9 hours).[67] The caveat is that sleep is scattered over 24 hours and markedly fragmented (Figure 135-1).

Arousal indices are difficult to assess in this population as inter-rater variability and scarcity of concurrent nursing documentation of extrinsic disruptions [68] complicate the reproducibility of these data. Nonetheless, indices of fragmentation are generally high during the sleep of ICU patients.[69] Cabello and colleagues[62] reported a median of 29 arousals per hour (interquartile range: 19–41), whereas Edell-Gustafsson and colleagues[57] noted even higher median hourly arousals of 56.6 ± 32.6 per hour. Viewed in more practical terms, other investigators reported the mean sleep duration before an arousal occurred ranging from 3 minutes[52] to 15 ± 9 minutes.[60]

In summary, ICU cohorts report subjectively poor quality sleep in the setting of normal TST measured over 24 hours. The proportions of individual stages are deranged with an increase in amounts of N1 and N2 possibly due to frequent arousals and awakenings. A corresponding decrease in SWS as well as stage REM is also noted. Sleep appears to be severely fragmented and evenly split between night and day. This means that roughly 50% of sleep is occurring during daytime even though external disrupting factors may be at a maximal at that time.

FACTORS CONTRIBUTING TO POOR SLEEP

The disrupting factors in the ICU are likely multifactorial and interrelated; however, for the purpose of discussion, these factors can be divided into two broad categories. *Extrinsic factors* include ambient conditions in which the patient is immersed (noise level, light intensity, etc.) as well as therapeutic applications (mechanical ventilation, medications, care events, etc.). *Intrinsic factors,* such as pre-illness natural sleep tendencies, emotional state, pain, and physiologic disruptions due to critical illness itself, also play a significant role in determining sleep quality.

Extrinsic Factors

Noise in the Intensive Care Unit

The majority of studies published in this realm have implicated noise as a bothersome environmental factor. Noise can produce physiologic changes similar to a generalized stress reaction, and these effects may prevent patients from initiating or maintaining sleep.[70] Whereas some authors have implicated noise from equipment and hardware in the room,[71,72] patients have identified hearing people (i.e., health care staff) talking as the most "annoying" impedance to their sleep.[73]

The World Health Organization Guidelines for Community Noise includes specific recommendations on noise levels in hospitals.[74] Based on a review of scientific studies, they concluded that hospitalized patients are more susceptible to excess stress levels generated by intrusive environmental noise. They recommend that sound level in hospitals should not exceed 35 decibels (dB) in patient care areas. Similar guidelines are also proposed by the U.S. Environmental Protection Agency; however, in critical care areas, noise levels often exceed these recommendations.[75] To add perspective, a recording of a vacuum cleaner yields 55 to 65 dB. Although individual tolerances vary, stress reaction and quantifiable effects from sleep disturbance are seen at time-averaged sound levels as low as 30 dB and peak sound levels above 70 dB.[76,77]

Increased sound levels account for approximately one-third of arousals and awakenings in ICU studies.[10,52] In particular, increased *peak* sound levels demonstrate unfavorable effects on sleep in ICU patients.[10,76,78] Elliot and colleagues[52] showed that, despite a slight decrease in noise level at night, the ICU remained above recommended levels both during the day (53.95 dB) and at night (50.20 dB). Two studies did not find that noise was a major factor,[5,49] but both of these studies relied on subjective retrospective recall, which could have been affected if patients did not reach conscious arousal long enough to form a memory of the disturbance. Freedman and colleagues[60] subsequently evaluated sleep disruption objectively using PSG. Although they noted noise was responsible for 11.5% of arousals and 17% of awakenings, they concluded that noise is not the main disturbing factor compared to other spontaneous (non-noise) arousals.[60] In a recent study, Tegnestedt and colleagues[79] noted that 64% of disruptive sounds were caused by monitor alarms and "[staff] conversations not related to patient care," which are avoidable. Overall, noise is persistently reported as a sleep-disruptive factor in the ICU environment.

Light in the Intensive Care Unit

Patients do not identify light intensity as a major disrupting force for their sleep.[5,49] Once again, caution should be taken with data obtained retrospectively from patients, because recall bias and the intensity of other disrupting factors may overshadow the impact that they felt from lighting differences. Furthermore, patients may not readily identify the potential effect on circadian rhythm changes. Light is the most potent "zeitgeber" for regulating our circadian rhythm,[80] and the study of endogenous melatonin secretion in healthy participants indicates that luminance levels of <100 lux (present in the ICU setting) may not be sufficiently bright to suppress melatonin secretion in some individuals.[81] In a study of 57 ICU patients, Elliot and colleagues[52] noted appropriately low light levels at night (median <2 lux), but the observed daytime luminance levels were only 74 lux and, therefore, too low to entrain a 24-hour circadian rhythm. Despite the lack of recognition by the patients, a non-circadian light exposure may contribute to sleep disturbance in ICU patients.

Mechanical Ventilation

The usual accessories required for mechanical ventilation include an endotracheal tube, suctioning apparatus, bite blocks, and usually a nasogastric tube, in addition to restraints. Although this setup alone can be a source of discomfort and makes sleep more difficult, during the past decade, much attention has been paid to how various modes of mechanical ventilation affect sleep in critically ill patients. Unfortunately, most studies are limited by small numbers of subjects, confounding factors (e.g., medications, severity of illness), and varying proprietary ventilator algorithms.

Early studies showed better sleep quality with assist-control ventilation (ACV) compared to pressure-support ventilation (PSV),[56,82] whereas later investigations indicated that proportional-assist ventilation resulted in better sleep than PSV.[83] Parthasarathy and colleagues[56] noted increased fragmentation of sleep during PSV compared to ACV because of the development of central apneas; this effect was more prominent in patients with heart failure. Cabello and colleagues[62] found no difference in sleep quality between three modes of ventilation (ACV, clinically adjusted PSV, or automatically adjusted PSV), noting equally abnormal sleep in the entire study population of conscious mechanically ventilated patients regardless of the assigned mode.

More recent studies continue to show conflicting results. Andrejak[84] found better sleep quantity and quality in patients receiving pressure-control ventilation compared to those on low-PSV. Roche-Campo and colleagues[85] did not find a difference in sleep quality on PSG recordings of nonsedated tracheostomized patients when comparing low PSV to spontaneous breathing in a cross-over study design. Although the number of arousals and awakenings were similar, they noted longer TST when patients were on mechanical ventilation.[85] The same group also compared a traditional ICU ventilator to a standalone noninvasive positive pressure ventilator (NIPPV); both devices were used with an oronasal mask in hypercapnic patients in the ICU. They found that patients experienced better sleep while on noninvasive mode of ventilation (regardless of the type of device used) compared to the spontaneous breathing periods.[86]

In summary, the true effect of mechanical ventilation or individual modes of ventilation on sleep is not yet fully clear. One hypothesis is that ventilator settings that reduce the likelihood of hypocapnia lead to fewer arousals from central apneas and thus improve sleep.[56,87] Intuitively, any mode that brings about better patient-ventilator synchrony will be less likely to disrupt sleep. A confounding factor is that most patients who are mechanically ventilated in the ICU are likely to receive some degree of sedation and analgesia.

Medications

Both the use and the withdrawal of common ICU medications are often implicated in sleep disruption because they act on shared receptor pathways used to regulate sleep. Benzodiazepines and opioids are among the most commonly administered medications in the ICU. Even in healthy patients and when administered at standard doses, these medications have well-described deleterious effects on sleep architecture.[88,89] Several authors have previously reviewed the effect of medications commonly administered (Table 135-1) in the ICU on sleep architecture.[90-92]

Although sedatives and opioids are sometimes used in an attempt to improve sleep in ICU patients, these medications are unlikely to provide the same restorative benefits as natural sleep. Benzodiazepines and opioids have been shown to decrease SWS, in favor of increasing N1 and N2 sleep, whereas propofol increases SWS and decreases REM sleep.[93] Additionally, benzodiazepines decrease sleep latency and REM sleep while increasing N2 sleep and possibly extending total sleep time.[89,94] Other classes of medications often used in the ICU, such as beta-antagonists and vasoactive agents, have also been shown to negatively affect sleep in healthy volunteers. Beta-antagonists prolong sleep latency and decrease the amount of REM sleep.[95] Norepinephrine and epinephrine are believed to reduce SWS and REM due to their stimulation of alpha-1 receptors as well as increase arousals by stimulating the reticular activating system.[96]

Sleep induced by sedatives is believed to differ in important aspects from normal sleep.[90] For example, a loss of transition through the sleep cycles and changes in sleep patterns are generally present regardless of the sedative administered. Dexmedetomidine sedation has been theorized to offer a

Table 135-1 Common Medications Used for Patients in the Intensive Care Unit

Drug Class	Examples of Drugs	Effect on Sleep Architecture	Potential Mechanism
CNS			
AED	Phenobarbital, carbamazepine, phenytoin	Very sedating. AEDs tend to ↑ TST, ↓ sleep latency. May ↑ SWS	Action on neuronal sodium influx in glutamate channels or GABA type A
TCA	Amitriptyline, imipramine, nortriptyline, desipramine, doxepin, clomipramine	Very sedating; suppresses REM sleep, ↑ TST, ↑ stage 2 sleep	Antimuscarinic activity, α1-receptior stimulation
Anxiolytic BzRA	Alprazolam, lorazepam, diazepam, oxazepam, propofol	Very sedating, ↑ TST, ↓ sleep latency ↓ SWS duration, ↑ stage 2 sleep same	GABA type A receptor stimulation; may also effect endocannabinoid
SSRI	Sedating: paroxetine, fluvoxamine; "activating": fluoxetine, sertraline, citalopram	In general, SSRIs tend to ↑ TST; less sedating than TCAs and MAOIs; ↓ REM, ↓ SWS ↑ TST, ↓ SE	↑5HT activity
SNRI	Venlafaxine	TST	↑5HT activity and NE activity
Mood stabilizer	Lithium	↑ TST, ↑ SWS, ↑ stage 2 sleep, ↓ REM,	Neuronal sodium channels
Antiparkinson	Bromocriptine, levodopa	Sedating; nightmares, ↓ SWS	Dopamine
cardiovascular stimulant	Norepinephrine, epinephrine dopamine		α1-, α2-, β-receptor stimulation; D2 α1-receptor stimulation
Lipophilic β-blocker	Propranolol, pindolol, metoprolol, timolol	Activating; ↑ awakenings, ↑TWT, ↓ REM, nightmares	CNS β-blockade
α^2-receptor agonist	Clonidine, DEX	↑ Stage 1, ↓ REM, nightmares ↑ TST	α2-receptor stimulation
α1-receptor blocker	Doxazosin, prazosin, terazosin	↑ TST	α1-receptor inhibition
Analgesic	Codeine, morphine,	Sedating; ↓ SWS, ↓ REM	μ-receptor stimulation
Opioid	hydrocodone	↓TST, ↓ SE	Prostaglandin synthesis
NSAID	Ibuprofen, indomethacin, celecoxib	Activating; ↓ TST, ↓ SE, ↑ stage 1, ↓ REM	inhibition
Other			Inhibits adenosine
Methylxanthine	Theophylline	Sedating	
Antihistamine	Diphenhydramine, promethazine	Sedating	Histamine 1 receptor blockade and can have Ach effect
Corticosteroid	Dexamethasone, prednisone	Activating; ↓ REM, ↓ SWS, nightmares	↓ Melatonin secretion

Although individual reactions may vary, some of the known effects on sleep architecture should be taken into account whenever possible to decrease the incidence of sleep disturbance during critical care illness.
Ach, Acetylcholine; AED, antiepileptic drug; BzRA, benzodiazepine; CNS, central nervous system; D or DOPA, dopamine; DEX, dexmedetomidine; GABA, gamma-aminobutyric acid; MAOI, monoamine oxidase inhibitor; NE, norepinephrine; NSAID, non-steroidal antiinflammatory drug; REM, rapid eye movement sleep; SE, sleep efficiency; SNRI, serotonin norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; SWS, slow wave sleep; TCA, tricyclic and tetracyclic antidepressant; TWT, total wake time; 5HT, serotonin, serotonergic.
From Hardin KA. Sleep in the ICU: potential mechanisms and clinical implications. *Chest* 2009;136:284–94.

more restorative—perhaps natural sleep-like—state than gamma aminobutyric acid (GABA)-ergic sedatives because alpha-2 agonists act on the sleep pathway at the brainstem level, whereas GABAergic agents act at the level of the hypothalamus.[97] Patients are more arousable when sedated with alpha-2 agonists compared to GABAergic agents such as benzodiazepines and propofol. This property is desired to facilitate weaning from mechanical ventilation, but it may not be conducive to the maintenance of sleep in a noisy ICU.

Acute medication withdrawal in ICU patients, the incidence of which is likely underestimated, may be an important issue affecting the sleep physiology of patients.[90] For example, in cases when REM-suppressing medications (such as benzodiazepines and opiates) are abruptly discontinued, the recovery sleep may have a disproportionately high percentage of REM sleep.[98] In the critically ill, this may play a more important role because cardiac and respiratory instability are greatest during REM sleep. Therefore, a gradual withdrawal of these agents whenever possible may reduce disruption of sleep neurophysiology.

Patient Care Events

Hands-on assessment at frequent intervals is a characteristic and necessary part of ICUs. In the reported literature, the number of direct patient contact occurrences ranges from 40 to 60 per night.[99,100] Although efforts should be made to coordinate care in such a way that allows for consolidated sleep times, only about 10% of arousals and awakenings have been shown to be attributable to patient-care activities.[10]

Intrinsic Factors

Sleep disruptions are not always clearly attributable to external causes and may be due to intrinsic factors that directly interfere with sleep. Some studies have implicated age because older patients may have a lower arousal threshold.[101] However, findings have been inconsistent with some showing older patients slept better in the ICU,[49] whereas others found age not to be a factor in determining quality of sleep in the ICU.[5] Pre-illness factors, such as psychological disorders (e.g., anxiety, posttraumatic stress disorder, panic attacks) as well as poor sleep habits and inconsistent bedtime routines, have been associated with disturbed sleep quality during acute hospitalization.[49,102] Additionally, circadian dysregulation and factors inherent to critical illness itself have been suggested as possible sleep-disruptive forces.

Disrupted Circadian Rhythm

The role of melatonin in the sleep of critically ill patients is the subject of renewed and increased attention. Although clinically not very useful, serum melatonin and urine 6-sulfatoxymelatonin (6-SMT, a melatonin metabolite) levels are accepted biologic markers for circadian rhythm.[103] In ICU cohorts, several studies show that melatonin secretion is decreased and does not follow a typical circadian pattern.[6,63,104-107] In a recent study of sedated and mechanically ventilated ICU patients, the authors noted a pronounced temporal disorganization of circadian rhythm. They report that most subjects exhibited preserved, but phase delayed, excretion of 6-SMT, implying a free-running circadian pacemaker.[63] Variations in the body core temperature acrophase in critically ill patients also suggest that the circadian rhythm is abnormal.[108] However, in contrast to these studies, Mundigler and colleagues[109] demonstrated an opposite effect in septic patients, showing persistently elevated levels of melatonin, which they hypothesized play a role in immune modulation and possibly contribute to the increased sleep noted during sepsis.

Importantly, some commonly used continuous infusions in the ICU (e.g., catecholamines and opioids) also affect melatonin secretion.[110] As reviewed earlier, sleep distribution and efficiency are abnormal in ICU patients; however, the role of decreased, non-circadian melatonin secretion as a contributing factor is still unclear. The use of exogenous melatonin to treat sleep disruption in the ICU is discussed later in the chapter.

Critical Illness

Acute illness can exacerbate sleep disruption and increase the need for sleep.[91] Data suggest that critically ill patients with higher severity of illness scores have greater sleep disturbance, as well as abnormal electrophysiologic sleep patterns showing absent sleep stages and increased low amplitude slow waves.[12,67] Freedman and colleagues[60] noted low-voltage mixed frequency or slow waves in nonsedated critically ill patients that had sepsis or positive blood cultures. These EEG changes were recorded both while the patient's eyes were open and closed, which was thought to reflect a dissociative state of consciousness reflecting neither normal sleep nor wakefulness.[60,111] This is similar to the pathologic wakefulness described by Drouot and colleagues[19] and Watson and colleagues.[64] One theory is that chemokines, such as interleukin-1 (IL-1) and tumor necrosis factor, are released during sepsis and induce a centrally mediated somnogenic effect[112,113] (Figure 135-5). Extending the hypothesis, Hardin[91] cites a possible protective effect because REM sleep, a vulnerable period associated with more cardiovascular instability, is significantly decreased in critical care illness (and sepsis, specifically).

All in all, multiple factors disturb sleep in critically ill patients. Some elements of extrinsic and intrinsic disrupting forces may be unavoidable in the setting of acute treatment. More research is necessary to determine which disrupting factors are associated with undesirable sleep-related ICU outcomes.

SLEEP-RELATED INTENSIVE CARE UNIT OUTCOMES

As a research field, sleep during critical care illness is still in its infancy, and, therefore, robust long-term ICU outcomes are not yet available. There is, however, growing evidence that disrupted sleep during the ICU negatively affects outcomes both during and following critical illness. Data from non-ICU subjects and animal studies indicate that sleep deprivation is adversely associated with immune and respiratory dysfunction. Specific data in ICU survivors associate poor sleep with worse psychological, neurocognitive, and quality of life (QoL) measures. Even after discharge, ICU patients report long-term sleep disruption. Overall, outcomes research in critically ill cohorts is challenging due to the many confounding factors, the presence of recall bias, and the absence of a true control group for comparison. Therefore, in this section, we will also consider hypotheses based on animal data, healthy volunteers, and non-ICU subjects.

Immune System and Healing

The effects of sleep deprivation on the immune system have not yet been directly studied in ICU populations. Given the vital role of host defense during critical illness, this is an important area for future research. See Figure 135-5 for proposed immune pathways involved in critically ill patients. In addition to cellular and humoral immunity, altered cytokine production is also implicated in wound healing.[114,115] Protein synthesis and metabolism have well described diurnal patterns and likely play a role in recovery from acute illness.[116-118] Patients in the ICU suffer from circadian dysregulation that may alter this diurnal pattern and further complicate their recovery. It is unclear whether the negative effects on immune function are associated with chronic versus acute sleep loss or sleep fragmentation. In a small study of healthy young males, oxygen consumption and carbon dioxide (CO_2) production increased during experimentally fragmented sleep but returned back to baseline during the recovery sleep period.[119] Furthermore, abnormalities in cellular metabolism, nitrogen balance, and hormonal secretions have been described in both sleep-deprived and critically ill patients; see Figure 135-6 for the suggested overlapping features.[69,120]

Respiratory Dysfunction

Sleep disruption in critically ill patients may increase the risk for respiratory muscle weakness, upper airway dysfunction, and sleep-related breathing disorder. In a classic article involving 30 healthy male volunteers, Chen and Tang[121] reported a decrease in respiratory muscle endurance after 30 hours of

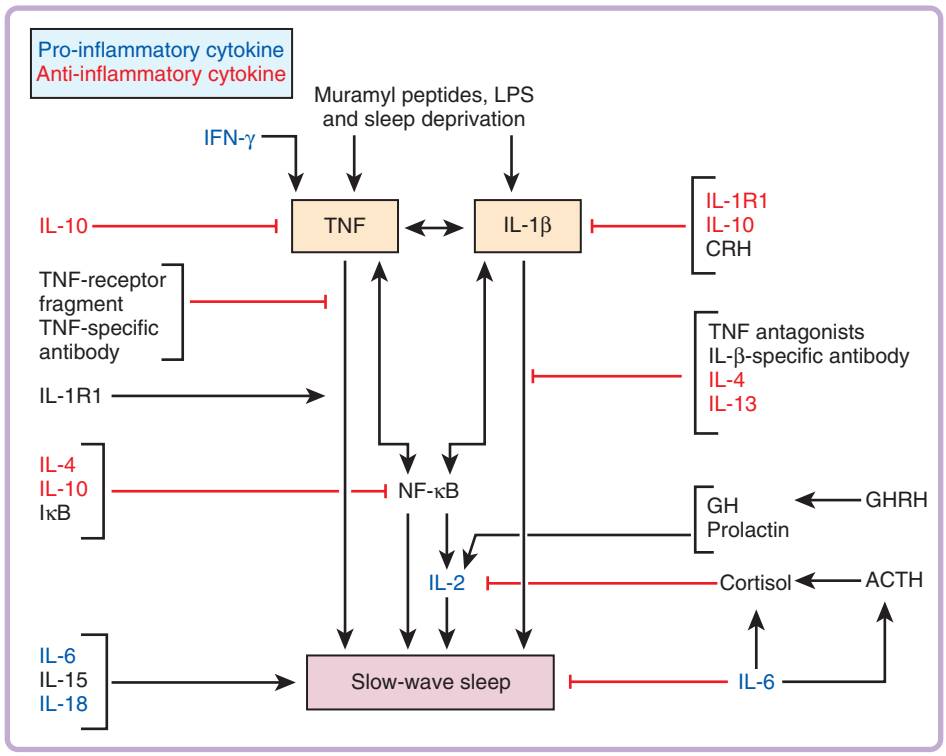


Figure 135-5 An overview of proposed immune mechanisms involved in sleep. This is not intended to be comprehensive but rather to illustrate some of the reciprocal effects of sleep and cytokines. In general, pro-inflammatory cytokines induce sleep, whereas anti-inflammatory cytokines inhibit sleep. ACTH, Adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; GH, growth hormone; GHRH, growth-hormone-releasing hormone; IκB, inhibitor of NF-κB; IL, interleukin; IL-1R1, type 1 interleukin-1 receptor; LPS, lipopolysaccharide; NF-κB, nuclear factor-κB; TNF, tumor necrosis factor. (From Bryant PA, Trinder J, Curtis N. Sick and tired: does sleep have a vital role in the immune system? *Nat Rev Immunol* 2004;4:457–67.)

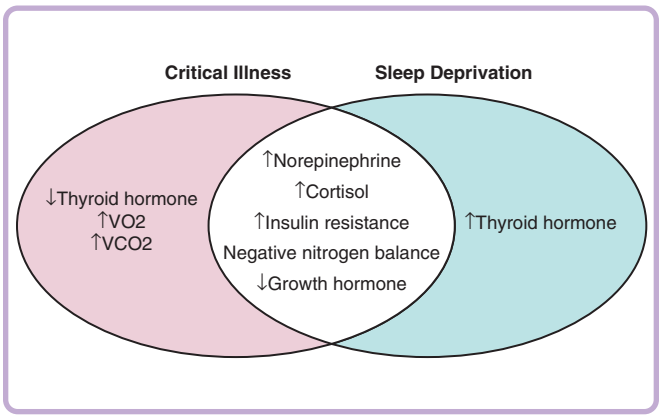


Figure 135-6 The overlapping metabolic and hormonal changes seen in sleep deprivation and critical illness. Although definitive studies in ICU patients are still lacking, several investigators have alluded to the similarities between these two states.

sleep deprivation. Leiter and colleagues[122] observed blunting of genioglossus EMG response to CO_2 as a result of sleep deprivation (which was worse with advanced age). In non-ICU patients with chronic obstructive pulmonary disease, Phillips and colleagues[123] noted a decrease in pulmonary function after only one night of sleep loss. In a small study of healthy subjects, Series and colleagues[124] demonstrated that increased upper airway collapsibility occurred primarily with sleep fragmentation rather than sleep deprivation per se. Taken together, these factors are hypothesized to either precipitate or exacerbate obstructive sleep apnea (OSA) in susceptible ICU patients. Therefore, some authors have argued in favor of early evaluation for sleep-related breathing disorder in suspected patients.[125,126] This may be particularly important for at-risk spontaneously breathing ICU patients for whom initiating NIPPV should perhaps be considered.

The data in animal models and non-ICU patients support the need for further investigation in order to optimize respiratory outcomes and prevent unnecessarily prolonged mechanical ventilation.

Neurocognitive Effects

The modifiable risk factors that can lead to neurocognitive dysfunction and delirium in the ICU have been extensively explored in the past decade and now include sleep. In other populations such as those with OSA, sleep fragmentation has been shown to result in a broad range of neurocognitive deficits that improve with treatment. Chronic partial sleep deprivation also causes neurocognitive deficits.[127] Because both sleep fragmentation and partial sleep deprivation are common in ICU patients, it is likely that poor sleep quality plays a role in the neurocognitive dysfunction frequently seen in these patients.

Delirium, a form of acute brain dysfunction commonly seen in ICU patients, has been implicated as both a cause and a consequence of sleep disruption. However, until the exact

mechanism is identified, it is impossible to differentiate whether sleep disruption leads to brain dysfunction or vice versa.[128,129] Delirium occurs in 20% to 50% of ICU patients with low severity of illness and 80% of those receiving mechanical ventilation.[130] Roche-Campo and colleagues[129] were the first to report evidence of a possible association between sleep disturbance noted on PSG and incidence of delirium in ICU. Subsequently, Trompeo and colleagues[94] found a 73.3% incidence of delirium in surgical ICU patients and noted that delirium was independently associated with sleep disturbance. More research is needed to investigate the specific role and possible dose–response relations between delirium and sleep in ICU patients. However, there may be a benefit in an integrated approach to address both sleep disruption and delirium because each has been independently associated with worse long-term outcomes.[131]

Psychological Effects and Quality of Life

Loss of sleep in a critical care setting is associated with decreased subjective QoL.[132,133] The negative effect on QoL lasts for variable periods of time following ICU discharge.[5,134-137] Survivors of ICU often develop depression, anxiety, and posttraumatic stress disorder, any of which can lead to sleep disruption and decreased quality of life.[138-140] It is unclear whether depression, anxiety, and delirium lead to poor sleep or vice versa. Most studies to date were performed by interviewing subjects retrospectively and, therefore, may be biased by the patients' poor recall.

Sleep During Post–Intensive Care Unit

Many patients continue to report sleep disruption long after their ICU stay. Sleep patterns may take days to months to normalize.[61,137,141-143] For example, survivors of the acute respiratory distress syndrome who reported sleep disturbance during their ICU admission noted persistent changes in their sleep quality 6 months or more after hospital discharge.[144] A study of cardiothoracic and neurosurgical ICUs enrolled 222 patients and used surveys (RCSQ) to retrospectively inquire regarding their sleep pre-ICU, while they were in the ICU and on the hospital ward.[137] Eighteen percent of participants reported insomnia prior to ICU admission, 73% reported poor sleep in the ICU, and 68% reported poor sleep on the hospital ward. Using the PSQI, 62% of participants noted poor sleep at 2 months, and 57% had poor sleep at 6 months. Examination of sleep quality over time showed that 7% of participants had poor sleep quality at all five time points, 10% had poor sleep that began in the ICU and continued out to 6 months, and 16% reported poor sleep after hospital discharge only. Choi et al[145] examined self-reported physical symptoms in 28 adult medical ICU survivors at 2 weeks, 2 months, and 4 months post-ICU discharge. Across all time points, sleep disturbance was one of the top complaints reported.[145] In a randomized controlled trial evaluating health status and QoL in 195 survivors of critical illness the authors assessed sleep at one, eight and 26 weeks after hospital discharge. Fifty percent of patients reported moderate to severe sleep disturbance at one week and one-third of patients were still reporting sleep disturbance at 26 weeks.[136] In addition, there were associations with poor sleep and worse scores on all measures of psychological outcomes at all three time points.

Patients' perceptions of the ICU vary, and it is recognized that, while some patients have clear memories of their ICU stay, others often have no memory of false memories. Further study is needed to clarify the impact of sleep quality during the ICU and hospital course on post-ICU sleep and to determine the best ways to prevent and treat post-ICU sleep disturbances.

MANAGEMENT OF SLEEP

In order to promote better sleep in the ICU, a multifaceted approach is necessary. The first step should include optimizing the environment in which the critically ill patient is expected to sleep. When possible, psychological and physical comfort should be addressed prior to initiating pharmacologic therapy for sleep.

Among the modifiable environmental factors, noise and light have been the most frequently studied. Applying earplugs, eye masks, and "white noise" sound machines improved sleep quality in various cohorts of critically ill patients[9,146-149] as well as in healthy volunteers exposed to recorded ICU sounds.[150-152] Other common-sense approaches to controlling these disrupting factors include opening blinds during the day and decreasing light intensity at night as well as reducing noise from avoidable sources (e.g., television, bedside phone, staff conversations). If possible, staff pagers should be set to vibrate, and bedside monitor alarm thresholds should be set more liberally when possible. Investigators have demonstrated improved sleep in critically ill patients by implementing two "quiet time" periods from 2:00 to 4:00 PM and 2:00 to 4:00 AM, when the physiologic propensity to sleep is highest.[153,154] After implementing a multifaceted paradigm shift to promote sleep in the ICU, Kamdar and colleagues[155] noted significant improvements in nighttime noise (as perceived by patients), a decrease in the incidence of delirium/coma, and an improvement in daily delirium/coma-free status. Coordinating routines and implementing non-disturbance periods are feasible,[57,156,157] and ICU nurses report an overall interest in sleep-promoting interventions.[158]

Other alterations to routine ICU care can also be considered. Non-critical care events should be minimized as a part of an integrated strategy for sleep promotion in the ICU. For example, routine medication delivery, laboratory blood draws, bathing, and bedding changes should be clustered outside of designated sleeping periods (especially at night). Individualizing settings to improve patient-ventilator synchrony will likely decrease patient effort and arousals. The patient's baseline apnea threshold, arterial partial pressure of CO_2 during sleep, and known sleep-related breathing disorders (i.e., OSA) might possibly influence ventilator settings. A list of the patient's home medications and chemical dependencies should be reviewed for mechanisms of action and withdrawal effects that may negatively impact sleep. Patients may experience nightmares, hallucinations, and parasomnias while in the ICU. It is helpful to place calendars, clocks (that designate AM and PM), and staff names within view of the patient to facilitate frequent reorientation. Last, some holistic techniques, such as guided-imagery relaxation,[159,160] back massage,[55] and even acupressure,[161] have resulted in improved sleep in randomized controlled trials in the ICU setting.

Despite optimization of modifiable disrupting factors, carefully selected pharmacotherapy to help treat pain, anxiety, and promote sleep may be necessary. Opioids should be used to target pain control rather than sedation or "sleep." If

Table 135-2 Clinical Trials Evaluating the Effects of Exogenous Melatonin Replacement in Critically Ill Patients

Author (Year)	Study Design	Reported Results	Comments
Shilo (2000)[171] *n* = 8	• Double-blind placebo controlled • 3 mg controlled-release melatonin or placebo given at 22:00	• Treatment dramatically improved both the duration and quality of sleep • Melatonin in ICU patients may help sleep induction and resynchronization of the "biologic clock"	• Wrist actigraphy was used to measure "sleep" and disruptions • Half of the ICU patients were on a ventilator • Control group included general medicine ward patients
Ibrahim (2006)[172] *n* = 32	• Double-blind, randomized, placebo-controlled • 3 mg oral melatonin or placebo given at 20:00	• Melatonin was well absorbed, and a standard dose increased blood levels approximately 1000-fold • Failed to increase observed nocturnal sleep	• Sleep was assessed by bedside nurse observation only • Use of "extra sedation" and haloperidol was nonsignificantly higher in control group
Bourne (2008)[38] *n* = 24	• Double-blind, randomized, placebo-controlled • 10 mg oral melatonin or placebo at 21:00 for four nights	• Nocturnal sleep quantity was severely compromised in both groups at the start of the study • Melatonin use was associated with increased nocturnal sleep efficiency	• "Sleep" monitoring was done by bispectral index (BIS) as well as actigraphy, nurse and patient survey • Proposed that 10-mg dose is too high and may lead to carryover of effects; suggested 1–2 mg immediate-release dose

All patients in these studies had undergone tracheostomy placement and were not receiving continuous sedation at the time of the study.

benzodiazepines are needed to treat anxiety, short-acting or medium-acting agents are preferred over formulations with longer half-lives. Although no specific efficacy data exist for using sedating antidepressants as hypnotics agents in the ICU, trazodone is frequently used in the inpatient setting due to its desirable pharmacokinetic properties (i.e., quick onset of action with an elimination half-life of 7 to 8 hours).[162] Sedating antihistamines (e.g., promethazine and diphenhydramine) should be avoided due to possible antimuscarinic effects associated with ICU delirium and daytime "hangover" effect.[163] Haloperidol and the new "atypical" antipsychotics (e.g., quetiapine, risperidone, and olanzapine) are often used as hypnotic agents due to their sedating properties. However, when administered in the absence of documented mental disorder, antipsychotics were associated with higher ICU and hospital length of stay as well as hospital mortality.[164] For intubated patients, continuous infusions of sedatives and analgesics should be titrated for comfort rather than to induce sleep. Dexmedetomidine has been suggested as a suitable agent to help with sleep as well as sedation, given its inhibition of norepinephrine release from the locus ceruleus, although further investigation is needed.[165] Gabapentin deserves consideration in critically ill patients because it has a large therapeutic window, a favorable side-effect profile, and it has been shown to increase SWS in healthy subjects.[166] Nonbenzodiazepine, nonopioid sleeping aids, such as zolpidem and eszopiclone have not been studied in critically ill patients. The recent concerns of parasomnias and increased falls during hospital admission in patients taking zolpidem[167] render these medications less desirable for routine management of sleep disruption in the ICU.

Melatonin and ramelteon are receiving renewed interest for use in the ICU population. In addition to possible effects on sleep, studies suggest that low endogenous melatonin levels during critical illness may be associated with delirium.[168] Three investigators have evaluated the efficacy of exogenous melatonin as a sleep aid in critically ill patients with conflicting results; the findings are summarized in Table 135-2. Ramelteon is an agonist of the MT1 and MT2 melatonin receptors and is used to treat insomnia. However, despite recently reported decreased delirium incidence in the ICU, the clinical impact of ramelteon on sleep has been noted to be small in critically ill patients.[169,170]

CONCLUSION

Sleep plays an integral role in our overall health and is vital to our ability to recover from illness. More research is needed to identify the exact bidirectional relationship between sleep and outcomes from critical care illness. The duration, quality, and distribution of sleep should be considered as a routine part of patient assessment—especially in critically ill patients who are more susceptible to sleep dysregulation.

CLINICAL PEARL

Sleep in critically ill patients is extremely fragmented and distributed across a 24-hour period without an apparent circadian pattern. Patients experience a significant decrease in the restorative sleep stages—SWS and REM. Multiple factors can disrupt sleep in the ICU; they include anxiety, noise, mechanical ventilation, medication effect, and critical illness itself. The loss of circadian melatonin secretion may also play a role in ICU sleep disruption. Sleep deprivation during critical care illness is an independent risk factor for worse clinical outcomes. A team-based and interdisciplinary approach toward implementation of nonpharmacologic measures to improve sleep should be considered whenever possible. Currently, no pharmacologic agent has been approved specifically for treating sleep disturbances in critically ill patients. However, small clinical trials have reported success with melatonin.

SUMMARY

Millions of patients are treated in ICUs each year and are known to suffer poor sleep during their critical care illness. Recently, increased focus on the intersection of sleep and critical care medicine has elucidated better understanding of the importance of sleep quality in this population. Sleep is best evaluated with PSG, although some alternative techniques have been used. During critical illness, patients experience increased N1 and N2 sleep and decreased SWS and REM sleep. In the ICU, sleep is likely to be fragmented and distributed in brief interrupted periods over 24 hours, not following a circadian pattern. Common sleep-disrupting factors in the ICU are noise, deleterious medication effects, and critical illness itself. The management of ICU patients should include a team-based and interdisciplinary approach to minimize disruptive factors that prevent patients from achieving consolidated sleep. Efforts should be made to address their anxiety and to reduce avoidable noise while trying to cluster the delivery of care outside of dedicated sleep periods. Currently, no pharmacologic agent is approved for use as a sleep aid in the ICU.

Selected Readings

Bourne RS, Mills GH. Sleep disruption in critically ill patients—pharmacological considerations. *Anaesthesia* 2004;**59**:374–84.

Cabello B, Thille AW, Drouot X, et al. Sleep quality in mechanically ventilated patients: comparison of three ventilatory modes. *Crit Care Med* 2008;**36**:1749–55.

Cooper AB, Thornley KS, Young GB. Sleep in critically ill patients requiring mechanical ventilation. *Chest* 2000;**117**:809–18.

DAS-Taskforce 2015; Baron R, Binder A, Biniek R, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015)-short version. *Ger Med Sci* 2015;**13**:Doc19.

Freedman NS, Gazendam J, Levan L. Abnormal sleep/wake cycles and the effect of environmental noise on sleep disruption in the intensive care unit. *Am J Respir Crit Care Med* 2001;**163**:451–7.

Guo Y, Sun L, Li L, et al. Impact of multicomponent, nonpharmacologic interventions on perioperative cortisol and melatonin levels and postoperative delirium in elderly oral cancer patients. *Arch Gerontol Geriatr* 2016;**62**:112–17.

Hu RF, Jiang XY, Chen J, et al. Non-pharmacological interventions for sleep promotion in the intensive care unit. *Cochrane Database Syst Rev* 2015;**10**:CD008808.

Kamdar BB, King LM, Collop NA, et al. The effect of a quality improvement intervention on perceived sleep quality and cognition in a medical ICU. *Crit Care Med* 2013;**41**:800–9.

Kaplan PW. The EEG in metabolic encephalopathy and coma. *J Clin Neurophysiol* 2004;**21**:307–18.

McKinley S, Fien M, Elliott R, Elliott D. Sleep and psychological health during early recovery from critical illness: an observational study. *J Psychosom Res* 2013;**75**:539–45.

Roche Campo F, Drouot X, Thille AW, et al. Poor sleep quality is associated with late noninvasive ventilation failure in patients with acute hypercapnic respiratory failure. *Crit Care Med* 2010;**38**:477–85.

Watson PL, Pandharipande P, Gehlbach BK. Atypical sleep in ventilated patients: empirical electroencephalography findings and the path toward revised ICU sleep scoring criteria. *Crit Care Med* 2013;**41**:1958–67.

Weinhouse GL, Schwab RJ, Watson PL. Bench-to-bedside review: delirium in ICU patients—importance of sleep deprivation. *Crit Care* 2009;**13**(6):234–41.

A complete reference list can be found online at ExpertConsult.com.

Section 17

Psychiatric Disorders

Chapter 136

Anxiety Disorders and Posttraumatic Stress Disorder

Andrew D. Krystal; Murray B. Stein; Steven T. Szabo

Chapter Highlights

- Anxiety and trauma-related disorders are associated with complaints of disturbed sleep. This chapter provides a review of the diagnostic criteria, sleep features, and treatment of sleep-related problems in patients with panic disorder, generalized anxiety disorder, social anxiety disorder (social phobia), and posttraumatic stress disorder.
- Abrupt nocturnal awakenings are often encountered in anxiety and trauma-related disorders such as panic disorder and posttraumatic stress disorder, in which non–rapid eye movement and rapid eye movement sleep abnormalities have been implicated, respectively.
- Treatments that target sleep can improve outcomes in patients with anxiety and trauma-related disorders and will be discussed where possible.

Anxiety and trauma-related disorders are the most common group of mental disorders, affecting more than 18% of the general population in a 1-year period.[1] The National Comorbidity Survey Replication, a large, cross-national epidemiologic survey conducted in 2001 to 2003, found that 28.8% of adults 18 years and older had a diagnosis of an anxiety disorder at some point in their lifetime.[2] Up to one third of the population experiences insomnia at any given time,[3-5] and insomnia is often comorbid with mental disorders.[3,6] Although insomnia is often a preexisting condition in individuals who experience both insomnia and an anxiety disorder and some evidence suggests that disturbed sleep is likely a risk factor for the development of anxiety disorders, the most typical pattern is for insomnia to begin concurrent with or following the onset of an anxiety disorder.[3,7-10] Thus onset of an anxiety disorder often heralds the onset of a sleep problem. which suggests that a sizable portion of the burden of insomnia in the general population is associated with—and perhaps even etiologically attributable to—anxiety disorders.

Anxiety disorders are commonly seen in primary care settings where patient complaints of sleep problems are often prominent.[11] Sleep disturbances are included among the diagnostic features of generalized anxiety disorder (GAD), separation anxiety disorder, and posttraumatic stress disorder (PTSD). Treatments targeted to sleep problems and worry, tension, and other manifestations of anxiety are often similar (e.g., cognitive or behavioral techniques, benzodiazepine medications, relaxation). The antihypertensive prazosin, which is effective for treating nightmares and sleep disturbance in patients with PTSD, is a relatively recent addition to this set of therapies.[12] Thus it is important to consider the diagnosis and treatment of anxiety disorders when caring for a patient with prominent sleep complaints. The converse is equally true: Attention to sleep problems is integral to the management of patients with anxiety disorders. Whereas sleep dysregulation has been extensively studied in depressive disorders, the polysomnographic (PSG) study of anxiety disorders is less well developed, but given the comorbidity of anxiety disorders and

depression there are likely to be some common features. PSG studies of anxiety disorders indicate that there are abnormalities in initiating and maintaining sleep and in sleep stage distribution. Such studies are one of the primary foci of the material reviewed in this chapter.

There has been relatively little work on the origin of these dysregulations of sleep. However, there are well-documented neurobiologic links between anxiety and sleep dysfunction. This includes circadian gene abnormalities in individuals with anxiety disorders.[13] There is also elevated cortical and peripheral arousal in both anxiety and insomnia patients. In both groups there appears to activation of arousal systems.[14,15] In anxiety, limbic structures such as the amygdala and hippocampus are activated in response to emotion-provoking stimuli. These structures in turn stimulate systems mediating arousal, including lateral hypothalamic hypocretin-orexin neurons, noradrenergic neurons in the locus coeruleus, and serotoninergic neurons in the raphe nuclei, which promote wakefulness.

As sleep disturbances related to specific phobias have been seldom investigated and obsessive-compulsive disorders are now a distinct category apart from anxiety disorders in DSM-5, they are not included here. In the cases of PTSD (which has also been moved to a separate chapter in the *Diagnostic and Statistical Manual of Mental* Disorders, fifth edition [DSM-5] but will nonetheless be considered here) and panic disorder, an additional focus relates to the core paroxysmal events that can manifest in relation to sleep (e.g., nightmares, panic attacks). This chapter also reviews current pathophysiologic concepts for sleep disturbances in anxiety and related disorders, and it reviews the types of treatments (and their outcomes) that target or indirectly influence sleep in these conditions.

PANIC DISORDER

Epidemiology and Clinical Features

Panic disorder has a 12-month general population prevalence of 2% to 3%, is more common in women than men, and has a typical age of onset in late teens or early 20s (although it can start earlier in life).[1,2] Onset is rare in older adulthood.[2] The characteristic feature of panic disorder is recurrent unexpected panic attacks (Box 136-1), which are acute episodes of severe anxiety associated with a wide array of somatic symptoms, such as chest pain, tachycardia, shortness of breath, psychosensory disturbances (i.e., changes in sound or light intensity, alterations in the perception of time, derealization), and lightheadedness. Classic panic attacks reach peak severity quickly and last only seconds to minutes in most cases. Panic attacks are well documented to be able to occur during sleep; in this chapter, the terms *nocturnal panic* and *sleep panic* are synonymous and refer to the same phenomenon.

Panic disorder is diagnosed when a person experiences recurrent unexpected panic attacks (Box 136-2). Some persons experience infrequent panic attacks for years without conspicuous changes in their health. More often, however, panic attacks are complicated by anticipatory anxiety (i.e., apprehension about future attacks) or worry about possible underlying medical disorders (e.g., heart disease). Such concerns, if lasting for 1 month or longer after the panic attack or attacks, satisfy criteria for the diagnosis of panic disorder. Alternatively, behavioral change (e.g., frequent visits to the emergency department; avoidance of places where attacks have occurred in the past) subsequent to the attacks also justifies this diagnosis. Patients also may become frightened of places or situations in which unexpected panic attacks have occurred.

Marked distress in—or the actual avoidance of—places (e.g., bridges, tunnels, airplanes) or situations (e.g., driving, shopping, traveling) in which unexpected panic attacks or panic-like symptoms have occurred in the past is referred to as *agoraphobia*. Although agoraphobia is frequently a consequence of panic disorder, it can also occur independently of

Box 136-1 DSM-5 CRITERIA FOR PANIC DISORDER

Diagnostic Criteria: 300.01 (F41.0)

Recurrent unexpected panic attacks. A panic attack is an abrupt surge of intense fear or intense discomfort that reaches a peak within minutes, and during which time four (or more) of the following symptoms occur:

Note: The abrupt surge can occur from a calm state or an anxious state.

- Palpitations, pounding heart, or accelerated heart rate
- Sweating
- Trembling or shaking
- Sensations of shortness of breath or smothering
- Feelings of choking
- Chest pain or discomfort
- Nausea or abdominal distress
- Feeling dizzy, unsteady, lightheaded, or faint
- Chills or heat sensations
- Paresthesias (numbness or tingling sensations)
- Derealization (feelings of unreality) or depersonalization (being detached from oneself)
- Fear of losing control or "going crazy"
- Fear of dying

Note: Culture-specific symptoms (e.g., tinnitus, neck soreness, headache, uncontrollable screaming or crying) may be seen. Such symptoms should not count as one of the four required symptoms.

- At least one of the attacks has been followed by 1 month (or more) of one or both of the following:
  1. Persistent concern or worry about additional panic attacks or their consequences (e.g., losing control, having a heart attack, "going crazy")
  2. A significant maladaptive change in behavior related to the attacks (e.g., behaviors designed to avoid having panic attacks, such as avoidance of exercise or unfamiliar situations)
- The disturbance is not attributable to the physiologic effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism, cardiopulmonary disorders).
- The disturbance is not better explained by another mental disorder (e.g., the panic attacks do not occur only in response to feared social situations, as in social anxiety disorder; in response to circumscribed phobic objects or situations, as in specific phobia; in response to obsessions, as in obsessive-compulsive disorder; in response to reminders of traumatic events, as in posttraumatic stress disorder; or in response to separation from attachment figures, as in separation anxiety disorder).

Box 136-2 PANIC ATTACK SPECIFIER

Note: Symptoms are presented for the purpose of identifying a panic attack; however, panic attack is not a mental disorder and cannot be coded. Panic attacks can occur in the context of any anxiety disorder as well as other mental disorders (e.g., depressive disorders, posttraumatic stress disorder, substance use disorders) and some medical conditions (e.g., cardiac, respiratory, vestibular, gastrointestinal). When the presence of a panic attack is identified, it should be noted as a specifier (e.g., "posttraumatic stress disorder with panic attacks"). For panic disorder, the presence of panic attack is contained within the criteria for the disorder, and panic attack is not used as a specifier.

An abrupt surge of intense fear or intense discomfort that reaches a peak within minutes and during which time four (or more) of the following symptoms occur:

Note: The abrupt surge can occur from a calm state or an anxious state.

- Palpitations, pounding heart, or accelerated heart rate
- Sweating
- Trembling or shaking
- Sensations of shortness of breath or smothering
- Feelings of choking
- Chest pain or discomfort
- Nausea or abdominal distress
- Feeling dizzy, unsteady, lightheaded, or faint
- Chills or heat sensations
- Paresthesias (numbness or tingling sensations)
- Derealization (feelings of unreality) or depersonalization (being detached from oneself)
- Fear of losing control or "going crazy"
- Fear of dying

Note: Culture-specific symptoms (e.g., tinnitus, neck soreness, headache, uncontrollable screaming or crying) may be seen. Such symptoms should not count as one of the four required symptoms.

DSM-5, *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition. Modified with permission from the American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Press; 2013. Copyright 2013 American Psychiatric Association.

panic disorder and is separately diagnosable in DSM-5. When not tied to panic disorder, agoraphobia (i.e., fear and avoidance of particular situations owing to fear of incapacitation or embarrassment) may be a complication of illness (and the repercussions thereof) such as vertigo or other forms of physical incapacity.

Sleep Features

At least two thirds of patients with panic disorder report moderate to severe sleep difficulties, including difficulty initiating and maintaining sleep, nonrestorative sleep, and nocturnal panic attacks.[16-18] Sleep difficulties and sleep deprivation can lead to worsening of anxiety symptoms, including panic attacks, in patients with panic disorder.[19] Most PSG studies have found decreased sleep efficiency and sleep duration in panic disorder patients,[20-24] although some studies have not found such disturbances.[21] Because panic disorder and major depression are often comorbid, it is possible that comorbid depression may be partially responsible for sleep disturbances in panic disorder; however, some of these studies have excluded subjects with comorbid depression.[22,24,25] Of those excluding comorbid depression, some still found evidence for sleep disturbance in persons with panic disorder.[22,24]

One type of sleep disturbance reported to occur in panic disorder is isolated sleep paralysis, a transient gross motor paralysis that can occur at sleep onset or offset. Although common in those with narcolepsy, it can also occasionally occur in those without this condition. It appears to arise when the involuntary immobility characteristic of rapid eye movement (REM) sleep intrudes into the waking state. This may relate to brainstem norepinephrine neurons in the locus coeruleus being quiescent in REM sleep and failing to revert to spontaneous pacemaker-like firing activity on waking, which is thought to lead to cortical arousal.[26,27] In addition to being unable to move during isolated sleep paralysis, some patients report anxiety, chest pressure, and other somatic sensations. Isolated sleep paralysis has not only been reported in association with panic disorder, it can also occur in other anxiety and trauma-related disorders such as PTSD.[28-32] One study investigating the prevalence of isolated sleep paralysis found that the prevalence of isolated sleep paralysis in subjects with a primary diagnosis of panic disorder (20.8%) did not seem differentially higher than that in PTSD (22.2%) or GAD (15.8%).[33]

Several surveys and studies of populations with panic disorder have documented the occurrence of panic attacks emerging from sleep as a not uncommon feature of the disorder. These episodes are often described as being awakened abruptly from sleep, usually with physical symptoms such as shortness of breath that also characterize the person's panic attacks in awake states. Sleep panic attack episodes appear to be non–rapid eye movement (NREM) sleep phenomena that occur at the transition between sleep stages N2 and N3 (NREM sleep stages 2 and 3) and thus they are not associated with dream mentation.[20,34,35] Approximately one half of patients with panic disorder report experiencing sleep panic attacks at some point during the course of their illness.[16,36] Some studies estimate that up to one third of patients experience recurrent nocturnal panic.[16,17,36,37] Although it has been suggested that nocturnal panic may itself be a marker of more severe panic disorder,[36,38] this has not consistently been found across studies.[37,39,40] Nocturnal panic appears to be associated with states of diminished arousal such as sleep and relaxation.[16,40-42] Although greater motor activity during sleep as evidenced by increased epochs of movement time has been reported with panic disorder, patients might actually move *less* on the nights when they experience sleep panic attacks,[43] leading the authors to suggest that movement during sleep may serve as a temporary protective mechanism against the episodes.

Several authors have suggested that the occurrence of panic attacks during NREM sleep implicates a more endogenous, physiologic (rather than cognitive or attributional) explanatory mechanism. Specific mechanisms that have been proposed include sensitivity to subtle increases in blood carbon dioxide levels,[44] irregular breathing during slow wave sleep,[45] and abnormalities in autonomic activity.[20,21,42] Additionally, these symptoms may also be mediated through activation of locus coeruleus activity.[35] Further, sodium lactate administered during sleep was found to be associated with increased cardiac and respiratory reactivity in panic disorder patients relative to controls,[46] and pentagastrin administered during sleep resulted in abrupt awakenings accompanied by panic symptoms in

patients with panic disorder.[48] These findings have been cited as evidence for a physiologic explanation for nocturnal panic because during sleep the influence of cognitive factors is purportedly less.

However, several studies suggest that cognitive factors also play a role. In an investigation employing caffeine administration during sleep, more fully elaborated panic attacks were preceded by a period of lighter sleep before awakening, providing support for a mixture of physiologic and cognitive influences on sleep panic.[49] In another study, participants with recurrent nocturnal panic attacks who were primed to expect intense physiologic changes during sleep (as indicated by an auditory signal) were less likely to awaken with panic symptoms than those for whom such a signal was unexpected, highlighting a role for presleep attributions.[50] Physiologic differences in those with nocturnal panic have also been found to normalize with cognitive behavior therapy (CBT).[51] Based on this evidence, it has been argued that although physiologic differences exist for those with nocturnal panic, they should be seen as a function of panic disorder psychopathology rather than as an explanatory mechanism.[42]

Treatment

The aim of drug treatment is to block panic attacks (waking and nocturnal panic attacks) and to eliminate secondary fears and avoidance activities (e.g., sleep phobias). The removal of exogenous stimulants (e.g., caffeine, amphetamine, catecholamine enhancers) and the correction of maladaptive behavior (e.g., sleep deprivation) that often exacerbate panic disorder should also be an integral part of the drug treatment program. If a thorough medical assessment has not recently been performed, this should be conducted (including, routinely, thyroid-stimulating hormone measurement to rule out most thyroid problems) before proceeding with specific antianxiety treatments.

Tricyclic antidepressants (e.g., imipramine) and monoamine oxidase inhibitors (e.g., phenelzine) were considered mainstays in the treatment of panic disorder as well as associated depressive symptoms that are commonly encountered. For reasons of tolerability and safety, these have been largely supplanted in the past two decades by the selective serotonin reuptake inhibitors (SSRIs), such as fluoxetine, sertraline, paroxetine, and controlled-release paroxetine, which are approved by the U.S. Food and Drug Administration (FDA) for this purpose. High-potency benzodiazepines (alprazolam, extended-release alprazolam, and clonazepam are approved by the FDA for this purpose) have also been widely used to treat panic disorder. Some medications (e.g., propranolol, buspirone, hydroxyzine) used often in the management of other forms of anxiety have been shown to be ineffective in the treatment of panic disorder.[52,53] Although there has been little pharmacologic research on the treatment of sleep disturbances associated with panic disorder, preliminary observations suggest that sleep-panic attacks are responsive to antidepressant-antipanic medications.[54,55]

Research has shown CBT to be at least as beneficial as first-line drug treatments.[56] CBT involves challenging irrational thoughts about panic symptoms and their consequences, the elimination of avoidance behavior, and gradual exposure to feared interoceptive sensations and agoraphobic situations. CBT also has the benefit of yielding long-lasting effects.[56] It is unclear whether standard CBT for panic disorder is beneficial in improving sleep, although one study suggests that combined pharmacologic treatment and CBT was insufficient in eliminating objective and self-reported sleep disturbances.[24] One CBT study included modifications targeted to nocturnal panic, such as psychoeducation about normal physiologic changes during sleep, challenging of catastrophic thoughts about nocturnal panic, interoceptive exposure to relaxation conditions, and sleep hygiene.[51] Compared with waitlist controls, participants who received this intervention fared better on measures of physiologic and self-reported anxiety and sleep quality, including nocturnal panic specifically. There is otherwise little information to guide the specific treatment of patients with panic disorder who have nocturnal panic attacks.[57]

In the absence of empiric data in this regard, it is recommended that patients with significant sleep disturbance including nocturnal panic attacks be treated with an antipanic agent or with CBT and that they undergo cognitive behavioral therapy for insomnia or sleep-targeted pharmacotherapy. Human subjects experiencing panic anxiety have been found to have elevated levels of the peptide hypocretin-orexin in their cerebrospinal fluid compared with subjects not experiencing it.[58] Based on the large body of evidence indicating that hypocretin-orexin activity promotes wakefulness, there is reason to believe that panic anxiety-related sleep disturbance might be mediated by this system.[59] As such, agents that block hypocretin-orexin receptors, which have recently become available in the United States (e.g., suvorexant) may be particularly useful in disorders of sleep associated with a panic anxiety component. Future research is needed to test this hypothesis.

GENERALIZED ANXIETY DISORDER

Epidemiology and Clinical Features

GAD is typified by chronic anxiety and excessive, pervasive worry (Box 136-3). In community surveys, the 12-month prevalence of GAD is approximately 3%,[1] with lifetime rates being higher (approximately 6%).[2] As is the case for all of the anxiety disorders the prevalence is higher in women than in men, with GAD showing an approximate 2:1 female-to-male ratio. The natural course of GAD can be characterized as chronic, with few complete remissions, a waxing and waning course of symptoms, and the substantial depressive comorbidity.

GAD, it might be argued, has been poorly named. This has led to a tendency to consider it a generic form of anxiety and to make the diagnosis inappropriately and pervasively. Many anxiety (and depressive) disorders are characterized by chronic anxiety and tension; these features alone are therefore insufficient to make this diagnosis. It is the presence of excessive and uncontrollable worry about multiple factors—such as work, health, and the well-being and safety of family members—that defines GAD. Patients with GAD often present to their primary care practitioner, where somatic complaints (e.g., headache, back or shoulder pain due to chronic muscle tension, chronic gastrointestinal distress) might predominate.

Sleep Features

Insomnia and GAD are highly overlapping, frequently comorbid disorders. Sleep disturbance, defined as difficulty

Box 136-3 DSM-5 CRITERIA FOR GENERALIZED ANXIETY DISORDER

Diagnostic Criteria: 300.02 (F41.1)

1. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance)
2. The individual finds it difficult to control the worry.
3. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms having been present for more days than not for the past 6 months):
   Note: Only one item is required in children.
   - Restlessness or feeling keyed up or on edge
   - Being easily fatigued
   - Difficulty concentrating or mind going blank
   - Irritability
   - Muscle tension
   - Sleep disturbance (difficulty falling or staying asleep, or restless, unsatisfying sleep)
4. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
5. The disturbance is not attributable to the physiologic effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism).
6. The disturbance is not better explained by another mental disorder (e.g., anxiety or worry about having panic attacks in panic disorder, negative evaluation in social anxiety disorder [social phobia], contamination or other obsessions in obsessive-compulsive disorder, separation from attachment figures in separation anxiety disorder, reminders of traumatic events in posttraumatic stress disorder, gaining weight in anorexia nervosa, physical complaints in somatic symptom disorder, perceived appearance flaws in body dysmorphic disorder, having a serious illness in illness anxiety disorder, or the content of delusional beliefs in schizophrenia or delusional disorder).

DSM-5, *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition. Modified with permission from the American Psychiatric Association: Diagnostic and statistical manual of mental disorders, 5th ed. Washington, DC: American Psychiatric Press; 2013. Copyright 2013 American Psychiatric Association.

initiating or maintaining sleep, or sleep that is restless and unsatisfying, is one of the six features (a minimum of three of which are needed to establish the diagnosis) associated with chronic worry in the DSM-5 criteria for GAD. Three of the other five features—fatigue, irritability, and difficulty concentrating—are also possible consequences of sleep loss. The core cognitive feature of GAD, excessive worry ("apprehensive expectation"), is commonly implicated in the genesis and maintenance of insomnia problems, in that patients often report their worry as most uncontrollable and bothersome at bedtime, interfering with their ability to fall asleep. In a study of comorbid psychiatric disorders in an insomnia sample, GAD was the most commonly diagnosed anxiety disorder.[60] Conversely, difficulty sleeping has been reported in 56% to 75% of persons with GAD,[61,62] although empiric data are largely lacking on the prevalence of sleep disturbance in GAD samples. One difference between GAD and primary insomnia may be the foci of worry at night; in primary insomnia the focus of worry is typically the insomnia itself, whereas in GAD, the worry is focused on areas that are also preoccupations during the day (e.g., career, finances, relationships).

The PSG sleep of insomniac patients with GAD has been compared with that of healthy control subjects in a handful of studies.[63] Patients with GAD have increased sleep latency, increased wake time after sleep onset, lower sleep efficiency, and reduced total sleep time relative to controls. The sleep architecture findings in GAD are unremarkable, and the conclusion is that GAD is characterized by a nonspecific sleep-onset and sleep-maintenance insomnia that compromises sleep quality. Notably, these studies provide evidence that GAD can be differentiated from major depression: The classic reduction in REM sleep latency seen in endogenous major depression is usually not seen in nondepressed patients with GAD.[64-66] However, given that most patients with GAD, particularly those encountered in general medical settings, also suffer from major depression, it should be expected that more classic depression-related sleep problems (e.g., early morning awakening) also will be seen. It is doubtful that differentiation of GAD from other anxiety or depressive disorders can be made on the basis of differences in sleep symptoms or PSG findings.

Treatment

Treatment for this chronic condition is, not surprisingly, often prolonged (i.e., several years). Benzodiazepines (e.g., alprazolam, clonazepam) are used extensively in the management of GAD. The available evidence suggests that the anti-anxiety effects of these compounds persist indefinitely in most cases and are not associated with dosage escalation in the long term.[67] In addition, strong evidence from double-blind, placebo-controlled studies indicates that certain classes of antidepressants, such as the SSRIs[68] and the dual reuptake inhibitors (also known as norepinephrine and serotonin reuptake inhibitors [NSRIs]; e.g., venlafaxine extended-release)[69] are efficacious.[70] A substantial advantage of antidepressants over benzodiazepines is the fact that the former treat comorbid depression, whereas the latter do not. Tricyclic antidepressants are also effective, although their use has largely been supplanted by the SSRIs and NSRIs. Recently, effects of extended-release quetiapine fumarate (50 to 300 mg/daily) on long-term functioning and sleep quality in patients with GAD using a randomized-withdrawal, placebo-controlled maintenance study in 432 participants showed that quetiapine XR monotherapy was effective at maintaining improvements in functioning and sleep quality.[71] However, at this writing, neither quetiapine XR or any of the other "atypical antipsychotic" medications is FDA approved for the treatment of GAD (or any other anxiety disorder).

There is evidence to suggest that antihistamines may also be helpful for treating core GAD symptoms,[72,73] although few studies have been conducted to date. Clinical experience suggests that improvement in insomnia parallels the overall benefits associated with pharmacologic treatment of GAD; however, studies often do not report sleep findings in GAD. One placebo-controlled study found that adding the hypnotic eszopiclone to the SSRI escitalopram was beneficial for both insomnia symptoms and daytime anxiety associated with GAD.[74] However, when this study was repeated with zolpidem CR, improvement with sleep but not anxiety symptoms

was observed, suggesting that eszopiclone may have direct anxiolytic effects.[75] Pregabalin has been demonstrated to have efficacy for GAD in several placebo controlled trials and has been reported to alleviate sleep disturbances associated with GAD; however, it is not FDA approved for the treatment of GAD or insomnia.[76-78]

CBT is highly effective in treating GAD,[79] and such treatments have also been shown to be effective for insomnia complaints.[80] In older adults, in whom benzodiazepines may be relatively contraindicated because of concerns about adverse effects (e.g., falls leading to fractures), psychosocial treatments may be particularly appealing.[81] The potential efficiency of integrated psychotherapeutic interventions that target both excessive generalized worries and worries about sleep warrants further exploration[82]; however, empiric investigation of such treatment approaches is not yet available.

SOCIAL ANXIETY DISORDER (SOCIAL PHOBIA)

Epidemiology and Clinical Features

The term *social anxiety disorder* is synonymous with *social phobia,* and the two terms may be used interchangeably. Social anxiety disorder can be described as excessive fear of being judged negatively, embarrassed, or humiliated in one or more social situations (Box 136-4). Anxiety in social situations can take the form of a panic attack, marked by extreme discomfort and physical symptoms such as heart racing, shaking, sweating, blushing, and other symptoms. In other cases, the symptoms are less acute but last longer, especially in anticipation of or before an upcoming social situation (e.g., worrying for days or weeks ahead of time about having to attend a dinner party).

Overt signs of discomfort (blushing, tremor in voice, sweating, motor tics) that might be apparent to another person are extremely distressing to the social phobic patient. One of these symptoms ("I sweat too much"; "I have the shakes") might be the chief and exclusive complaint of the social phobic patient when seeking treatment from his or her family physician. Public speaking may also be endorsed as a prominent source of anxiety, although further inquiry often reveals social fears and avoidance in many more socially routine situations such as speaking to small groups, interacting with authority figures, and relating to peers. DSM-5 encourages the use of the specifier "performance-only type" for the subset of individuals whose social fears and anxiety truly are limited to public speaking or other performance situations.

Social anxiety disorder and the associated avoidance can significantly interfere with daily functioning and reduce quality of life.[81] Many patients report that they always have been "shy"; onset of the disorder is in early childhood (i.e., it has "always been there") in approximately half of cases, whereas in the other half it appears to develop de novo in adolescence among those without a background of pathologic shyness. Social anxiety disorder affects slightly more women than men, although men more commonly present to mental health treatment settings for its treatment.

Box 136-4 DSM-5 CRITERIA FOR SOCIAL ANXIETY DISORDER (SOCIAL PHOBIA)

Diagnostic Criteria: 300.23 (F40.10)

1. Marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others. Examples include social interactions (e.g., having a conversation, meeting unfamiliar people), being observed (e.g., eating or drinking), and performing in front of others (e.g., giving a speech).

Note: In children, the anxiety must occur in peer settings and not just during interactions with adults.

2. The individual fears that he or she will act in a way or show anxiety symptoms that will be negatively evaluated (i.e., will be humiliating or embarrassing; will lead to rejection or offend others).
3. The social situations almost always provoke fear or anxiety.

Note: In children, the fear or anxiety may be expressed by crying, tantrums, freezing, clinging, shrinking, or failing to speak in social situations.

4. The social situations are avoided or endured with intense fear or anxiety.
5. The fear or anxiety is out of proportion to the actual threat posed by the social situation and to the sociocultural context.
6. The fear, anxiety, or avoidance is persistent, typically lasting for 6 months or more.
7. The fear, anxiety, or avoidance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
8. The fear, anxiety, or avoidance is not attributable to the physiologic effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.
9. The fear, anxiety, or avoidance is not better explained by the symptoms of another mental disorder, such as panic disorder, body dysmorphic disorder, or autism spectrum disorder.
10. If another medical condition (e.g., Parkinson disease, obesity, disfigurement from burns or injury) is present, the fear, anxiety, or avoidance is clearly unrelated or is excessive.

Specify if:

- **Performance only:** If the fear is restricted to speaking or performing in public.

Specifiers

Individuals with the performance-only type of social anxiety disorder have performance fears that are typically most impairing in their professional lives (e.g., musicians, dancers, performers, athletes) or in roles that require regular public speaking. Performance fears may also manifest in work, school, or academic settings in which regular public presentations are required. Individuals with performance only social anxiety disorder do not fear or avoid nonperformance social situations.

DSM-5, *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition. Modified with permission from the American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Press; 2013. Copyright 2013 American Psychiatric Association.

Sleep Features

Complaints of insomnia are not uncommon in patients with social anxiety disorder if the clinician specifically elicits them, although it is rare for a patient with social anxiety disorder to present with sleep disturbances as his or her chief complaint. Anecdotally, persons with social anxiety disorder are somewhat more likely to experience sleep difficulties, particularly increased sleep latency, when experiencing anticipatory anxiety in advance of a highly feared social event (e.g., job interview, oral presentation).

In one of only two published studies that specifically focus on subjective sleep in patients with social phobia, 60% of participants could be categorized as poor sleepers compared with only 7% of healthy controls.[83] A more recent investigation found that 18% of participants with social phobia had moderate to severe insomnia versus only 5% of controls, and an additional 40% had subthreshold insomnia symptoms.[84] Furthermore, this study found that the relationship between social phobia and insomnia was partially mediated by depressive symptoms.

The results of PSG are largely normal in social anxiety disorder, with sleep latency and sleep efficiency being similar in patients and healthy control subjects. REM sleep latency, REM sleep distribution, and REM sleep density are also normal in social anxiety disorder.[85]

Treatment

There is a solid evidence base for managing social anxiety disorder with either pharmacotherapy or CBT. Five classes of medications are effective in the treatment of social phobia: high-potency benzodiazepines (alprazolam, clonazepam), monoamine oxidase inhibitor antidepressants (i.e., phenelzine, tranylcypromine), SSRIs (paroxetine and sertraline have FDA approval for the treatment of social phobia), SNRIs (venlafaxine extended-release has FDA approval for the treatment of social phobia), and, in limited circumstances, beta blockers (e.g., atenolol, propranolol).[86] Clinical experience suggests that some patients with the performance-only specifier of social anxiety disorder (e.g., public speaking or anxiety limited to other performance situations such as playing a musical instrument in public) may respond to beta blockers on an as-needed basis, whereas individuals with more generalized, pervasive social anxiety symptoms require regular dosing—to *prevent* the occurrence of symptoms in situations that occur so commonly that as-needed medication cannot be used—with either a benzodiazepine or one of the aforementioned antidepressants.

Individual and group CBT are also effective, and their duration of effects is longer than that for medication treatments.[87-89] A combination of psychosocial and drug therapies may be considered to achieve an optimal response; often, rather than starting CBT and medications together, a stepped approach can be employed, starting with one modality and adding the other if needed.[90] Whereas there had been initial excitement about the possibility of speeding or otherwise facilitating the effects of CBT with D-cycloserine for social anxiety disorder,[91] a recent study failed to show benefits of such an approach over CBT alone..

Interestingly, poorer pre-CBT sleep quality has been noted as a predictor of poorer response to subsequent CBT, with the concomitant suggestion that sleep problems should be addressed before starting CBT. The impact of successful social anxiety disorder treatment on sleep symptoms has not been ascertained. For transient insomnia symptoms, a short course of hypnotics may be beneficial.

POSTTRAUMATIC STRESS DISORDER

Epidemiology and Clinical Features

PTSD, previously included within the anxiety disorders, in now included in a separate trauma-related disorders chapter in DSM-5. PTSD is characterized by the recurrent, unwanted mental and emotional reexperiencing of a previous traumatic event. The trauma is one that would be experienced by almost anyone as profoundly disturbing, usually falling into the category of an event that was life-threatening (e.g., violent attack with physical injury, sexual assault, serious motor vehicle collision) or profoundly and abruptly life-altering (e.g., sudden death of a loved one from accidental or unanticipated medical causes). Traumatic experiences that yield PTSD, such as military combat or domestic violence (i.e., intimate partner abuse), often occur as repeated episodes and not single events.

After exposure to severe traumatic experiences, it is the norm to experience brief (lasting several days) periods of anxiety, recurrent thinking about the event, and insomnia. In cases in which these symptoms persist more than a few days and cause functional impairment, accompanied by prominent feelings of unreality or memory problems (i.e., dissociative symptoms), a diagnosis of acute stress disorder may be appropriate. In most such cases, the symptoms wane over the ensuing weeks, but when they do not (which is the pattern in 10% to 30% of cases, depending on the nature and severity of the traumatic event) and the symptoms interfere with functioning or cause great distress, a diagnosis of PTSD may be applied. To qualify for a PTSD diagnosis, characteristic symptoms must be noted in four domains: reexperiencing symptoms (e.g., nightmares or daytime intrusive thoughts or images, including flashbacks), negative cognitions and mood, avoidance and numbing symptoms (e.g., avoidance of reminders of the trauma), and hyperarousal symptoms (e.g., insomnia or increased startle response) (Box 136-5).

Lifetime prevalence of PTSD is reported to be as high as 7% to 8%, with a 12-month prevalence in the range of 2% to 3%. PTSD is approximately twice as prevalent in women as men in the general population. Common sources of PTSD in men (although increasingly in women) are military combat and physical assault; in women, intimate partner violence and violent injury (often associated with sexual assault) are common sources. Like many other anxiety and related disorders, PTSD is often encountered comorbid with major depression, both in general population samples and in clinical settings.

Sleep complaints are virtually ubiquitous in patients with PTSD.[92] Patients sometimes report not having slept well for decades, with these reports corroborated by their bed partners. Extreme hypervigilance (sometimes bordering on paranoia, but evidently linked to habits learned during combat experiences) may take the form of a combat veteran with PTSD spending several hours each evening "patrolling" the perimeter of his home to ensure that it is protected from intruders. Nightmares, often accompanied by vivid recall on wakening, are commonplace, as is extreme motor activity, according to companions' reports. It is not uncommon to encounter patients who have tried numerous over-the-counter and prescription medications to help with their sleep, to no avail. Many patients also have alcohol abuse problems, which can complicate the clinical picture and make determining the nature of the sleep disorder even more difficult. In some cases, other risk factors for disturbed sleep—such as obstructive sleep apnea—are present, making an informed evaluation by a sleep expert (with a particular focus on possible sleep-disordered breathing) of considerable value.

Individuals suffering with PTSD are afflicted with hyperactivity of the norepinephrine system during waking and

Box 136-5 DSM-5 CRITERIA FOR POSTTRAUMATIC STRESS DISORDER

In posttraumatic stress disorder, the person has been exposed to a traumatic event in which both of the following were present:

- The person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury or a threat to the physical integrity of self or others.
- The person's response involved intense fear, helplessness, or horror. In children, this may be expressed instead by disorganized or agitated behavior.
- The traumatic event is persistently reexperienced in one (or more) of the following ways:
  - Recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions. In young children, repetitive play can occur in which themes or aspects of the trauma are expressed.
  - Recurrent distressing dreams of the event. Children can have frightening dreams without recognizable content.
  - Acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur on awakening or when intoxicated).
  - Intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.
  - Physiologic reactivity on exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.

The person persistently avoids stimuli associated with the trauma and has numbing of general responsiveness (not present before the trauma), as indicated by three (or more) of the following:

- Efforts to avoid thoughts, feelings, or conversations associated with the trauma
- Efforts to avoid activities, places, or people that arouse recollections of the trauma
- Inability to recall an important aspect of the trauma
- Markedly diminished interest or participation in significant activities
- Feeling of detachment or estrangement from others
- Restricted range of affect (e.g., unable to have loving feelings)
- Sense of foreshortened future (e.g., does not expect to have a career, marriage, children, or a normal life span)

The person has persistent symptoms of increased arousal (not present before the trauma), as indicated by two (or more) of the following:

- Difficulty falling or staying asleep
- Irritability or outbursts of anger
- Difficulty concentrating
- Hypervigilance
- Exaggerated startle response

Duration of the symptoms is longer than 1 month.

The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify if the symptoms are acute (duration less than 3 months) or chronic (duration 3 months or longer).

Specify if the symptoms have delayed onset (onset at least 6 months after the stressor).

DSM-5, *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition. Modified with permission from the American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Press; 2013. Copyright 2013 American Psychiatric Association.

sleep.[93,94] Preclinical work has implicated changes in the amygdala, which leads to activation of the locus coeruleus through corticotropin-releasing hormone receptors as a key mechanism mediating hyperarousal, hypervigilance, and sleep disturbances in PTSD.[95] These considerations support the use of agents that enhance serotonin tone, such as most of the antidepressants because this inhibits locus coeruleus norepinephrine neurons and is also thought to inhibit amygdala activity.[96-98] They also support the use of α-adrenergic antagonists such as prazosin, which block locus coeruleus outputs that have hyperarousing effects (see later).

Sleep Features

Subjective sleep complaints in PTSD often include the two sleep symptoms listed in the diagnostic criteria: nightmares (which are viewed as reexperiencing phenomena) and insomnia (i.e., impairment in initiating and maintaining sleep). A survey of male Vietnam War veterans with PTSD confirms that insomnia complaints are very common symptoms of PTSD, although nightmares are more specific to the disorder.[99] In a community survey, PTSD was often (in approximately 70% of cases) associated with sleep complaints including insomnia as well as other types of sleep disturbance such as violent or injurious activities during sleep, sleeptalking, or hypnagogic and hypnopompic hallucinations.[100] These rates seem very high and need to be replicated, but they do point to the usefulness of a thorough sleep assessment in conjunction with diagnostic assessment of PTSD and other trauma-related disorders. Heightened arousal during sleep has also been reported by patients with PTSD, including excessive motor activity and awakenings with somatic anxiety symptoms.[101] Isolated sleep paralysis is also associated with PTSD, and with trauma exposure even in the absence of PTSD, although its specificity to PTSD relative to other anxiety disorders is unclear (see also Panic Disorder, earlier).[99,102,103] It has been reported that the severity of sleep complaints mediates the relationship between PTSD and functional disability.[104]

Findings on objective sleep disturbances in PTSD are inconsistent. Subjective reports of sleep disturbance are often discrepant with the findings of PSG. In early PSG studies of PTSD, mostly featuring male combat veterans during a chronic phase of the disorder, findings are mixed with respect to the presence of impaired sleep initiation and maintenance. Several studies report reduced sleep time or efficiency or increased awakenings in the patients with PTSD,[105] whereas in other studies measures of sleep maintenance did not differ from that of control subjects.[106-108] In the largest PSG study of PTSD conducted to date, investigators conducted sleep studies for two consecutive nights with 292 trauma-exposed young adults from the community, 71 of whom had lifetime PTSD.[109] On standard measures of sleep disturbance, no differences were detected between subjects with PTSD and controls, irrespective of trauma history or major depression in the control subjects. The investigators concluded that they found no objective evidence for clinically relevant sleep disturbances in PTSD. This study raises the possibility that the presence of sleep disturbance in PTSD may be a prevailing factor in determining treatment seeking and disability, so that patient samples are enriched for sleep disturbance.

However, the authors of a meta-analysis of 20 PSG studies found that overall, PTSD patients had increased

stage 1 sleep, decreased slow wave sleep, and increased frequency of eye movements during REM sleep periods (REM sleep density).[110] Moreover, they found that several moderating variables—age, sex, and comorbid depression and substance use—affected the relationship between PTSD and sleep disturbances. Another factor that can contribute to the lack of significant findings in individual studies, as raised in a review,[111] is that some persons with PTSD (and other psychiatric conditions, such as primary insomnia) might sleep better in the laboratory environment because of the perception that it is a "safe" setting, thus making evidence of objective sleep disturbance less discernible. Numerous studies have contributed data which indicate that individuals with PTSD have greater apnea-hypopnea index, more transitions from REM sleep to waking, and greater REM density,[112-114] but more work is needed to determine whether these features play a role in pathophysiology. Other sleep disturbances in PTSD include movement abnormalities and sleep-disordered breathing. Self-reports of excessive motor activity have been corroborated by studies documenting frequent periodic limb or gross body movement during sleep,[115-118] although one report found reduced movement time during sleep in patients with PTSD compared with controls.[119] The latter study also found reduced sleep movement time to be associated with self-report of trauma-related nightmares in the past month.[120]

The prominence of nightmares in PTSD (which most typically arise from REM sleep) has focused interest on REM sleep variables. Abnormalities in the timing or amount of REM sleep in PTSD have not been consistently found, but other subtle differences in REM sleep in PTSD are emerging. First, increases in REM sleep density have been reported.[120,121] Second, there is evidence for greater REM sleep fragmentation in PTSD. For example, increased frequency of REM sleep to wakefulness[103] and REM sleep to stage 1 transitions[121] has been reported. Third, abnormalities in REM sleep activity after trauma exposure might predict subsequent development of PTSD.[108]

In a study of physically injured subjects who were admitted to hospital and who had at least one night of PSG recording within a month of injury, development of PTSD symptoms was associated with a shorter average duration of REM sleep before a stage change and with more periods of REM sleep.[122] More recently, in a subsequent report from a subset of this sample, sympathetic activation indexes were higher during the REM sleep of the 9 subjects who were positive for PTSD symptoms, compared with the 10 subjects who were PTSD negative.[123] Furthermore, the development of PTSD symptoms after traumatic injury is associated with a more fragmented pattern of REM sleep and increased sympathetic activation during REM sleep. Taken together, these studies suggest that sleep in PTSD is characterized by REM sleep abnormalities, suggesting that increased noradrenergic activity during REM sleep may contribute to the development of PTSD. These hypotheses deserve to be explored further because they might suggest avenues for pharmacologic intervention to prevent the development of PTSD in persons exposed to traumatic events.

Treatment

As noted previously, thorough assessment of sleep complaints is integral to the overall management of PTSD. Although the absolute rates of sleep disturbances such as obstructive sleep apnea or parasomnias in patients with PTSD remain to be determined, clinical experience suggests that these are commonly encountered in this clinical population and should be seriously considered in most cases. In cases in which the suspicion is high, patients should be referred to a sleep physician for assessment. Patients with PTSD are also likely to have comorbid mental disorders such as major depression and other anxiety disorders such as panic disorder. Treatment therefore needs to encompass these clinical entities, as well as comorbid alcohol or other substance abuse or dependence, which are also often encountered in patients with PTSD.

There is a strong and growing evidence base for particular pharmacologic and psychological treatments for PTSD. Pharmacologic treatments with support for efficacy in the treatment of PTSD include the SSRIs (sertraline and paroxetine are approved by the FDA for this indication) and, to a lesser extent in terms of strength of evidence, tricyclic antidepressants and monoamine oxidase inhibitors.[124] The dual-reuptake inhibitor venlafaxine has also been recommended as a first-line treatment for PTSD. However, none of these agents has been demonstrated to have therapeutic effects on sleep in PTSD patients. An agent that has been found to have a particular beneficial effect on sleep in PTSD patients is the α_1-adrenergic receptor antagonist prazosin.[125-128] There is also some support for the adjunctive (or monotherapeutic) use of atypical antipsychotics (e.g., olanzapine, risperidone, or quetiapine), although not all trials have shown benefits.[129] Early results for the anticonvulsant tiagabine showed promise,[130] although findings from a large double-blind study found no differences in PTSD treatment efficacy from placebo.[131] A recent study indicating that changes in sleep occurring on the first night of tiagabine administration can predict subsequent PTSD response suggests the possibility that sleep changes may be mediators or early markers of subsequent therapeutic PTSD effects.[132] The use of prazosin in PTSD has become mainstream in many settings for the treatment of patients with PTSD. The benefits of prazosin for PTSD are likely a property of α_1-receptor antagonists as a class and not a prazosin-specific effect because terazosin has shown benefit in a case report.[133,134]

More than in any of the other anxiety and related disorders, the use of evidence-based psychotherapy is critical either as primary or adjunctive treatment for PTSD.[135-137] There is accumulating evidence that psychosocial treatments that focus on sleep aspects of PTSD such as nightmares can have strong therapeutic effects across the full spectrum of PTSD symptoms.[138,139] Practice guidelines based on evidence are available (Box 136-6).[140] Few studies have used psychotherapeutic interventions to treat PTSD and examined their effects on sleep quality. In one small study 5-week treatment using eye movement desensitization reprocessing therapy improved sleep consolidation and reduced wake time after sleep onset.[141] Another study reported the safe and effective use of CBT for insomnia in patients with comorbid insomnia and PTSD and improved sleep was reported to facilitate entry into exposure therapy for PTSD.[142] Additional studies that incorporate psychotherapeutic techniques with medication strategies are needed to formally evaluate their benefits.[143]

Box 136-6 TREATMENT RECOMMENDATIONS OF PTSD-ASSOCIATED NIGHTMARES

Level A: Supported by a substantial amount of high-grade evidence and/or based on a consensus of clinical judgment	Prazosin
Level B: Supported by a sparse amount of high-grade evidence or a substantial amount of low-grade data and/or clinical consensus by task force	Clonidine Venlafaxine is *not* suggested.
Level C: Supported by low-grade data without volume to recommend more highly and likely subject to revision with further studies	*Medications:* trazodone, atypical antipsychotic medications, topiramate, low-dose cortisol, fluvoxamine, triazolam and nitrazepam, phenelzine, gabapentin, cyproheptadine, and tricyclic antidepressants. Nefazodone is not recommended as first-line therapy for nightmare disorder because of the increased risk for hepatotoxicity. *Behavioral therapies:* Exposure, Relaxation, and Rescripting Therapy (ERRT); Sleep Dynamic Therapy; Hypnosis; Eye-Movement Desensitization and Reprocessing (EMDR); and the Testimony Method.

Modified from Aurora RN, Zak RS, Auerbach SH, et al. AASM Standards of Practice Committee. Best practice guidelines for the treatment of nightmare disorders in adults. *J Clin Sleep Med* 2010;6:389–401.

CLINICAL PEARLS

- Effective treatment of anxiety and trauma-related disorders includes assessing and managing sleep symptoms.
- Clinical experience suggests that education and encouragement about behavioral factors that disturb sleep can be helpful as an adjunct to treatment in nearly all cases.
- If sleep disturbances persist after the successful treatment of the primary anxiety disorder, the patient should be reevaluated for other possible medical or sleep disorders.
- In the case of PTSD, where there is reason to suspect that comorbid medical, psychiatric, and sleep disorders (e.g., alcohol or substance abuse, obstructive sleep apnea) are especially common, a high index of suspicion should be maintained and a thorough medical and sleep evaluation may be considered early in the course of assessment and treatment.

SUMMARY

Anxiety and trauma-related disorders are extremely common and often associated with disturbances of sleep. These disturbances vary across the anxiety disorders and related conditions and consist of difficulty falling asleep, difficulty staying asleep, early morning awakening, and nightmares. Patients with GAD can have increased sleep latency and significant reductions in total sleep time and sleep efficiency. Clinical experience suggests these disturbances are tightly linked to pathologic worry, and improvement in insomnia closely parallels the successful amelioration of core anxiety symptoms. PTSD and, to a lesser extent, panic disorder are often characterized by recurrent frightening arousals from sleep. Sleep complaints are so prevalent in PTSD that it could be argued that every assessment of sleep disturbance should include taking a thorough history of traumatic events. Conversely, every assessment of a patient with PTSD should include a through assessment of sleep symptoms and additional investigation and follow-up as warranted. Available evidence suggests that the predominant sleep pathologic process in PTSD relates to REM sleep, whereas that in nocturnal panic relates to NREM sleep. However, reports of high rates of obstructive sleep apnea and parasomnias in PTSD suggest that a more complex etiology for sleep disturbance might exist in some individual patients with this disorder. Although these sleep disturbances were long believed to be symptoms of the anxiety disorders, the available evidence increasingly suggests that the relationship of sleep problems and anxiety disorders may be more complex and in some cases bidirectional. This evidence strongly speaks for the need to target treatment specifically to the sleep problems in patients with anxiety-related conditions. Some, but clearly not all, evidence-based pharmacologic and psychosocial treatments for anxiety disorders seem to improve sleep as part of their spectrum of therapeutic effects. However, sleep outcomes are not often reported in clinical trials, leaving some uncertainty as to a given treatment's efficacy for sleep disturbances. In some instances, preferred medications with good evidence of efficacy for anxiety disorders (e.g., SSRIs) can actually worsen sleep while they ameliorate daytime and phobic anxiety. As a result, a specific sleep-targeted therapy is needed in many patients with anxiety difficulties. There is a growing set of evidence-based pharmacologic and behavioral interventions available for treating the sleep problems of anxiety patients. Addressing the sleep problems of patients with anxiety and trauma-related disorders using these interventions promises to improve outcomes.

Selected Readings

American Psychiatric Association. Practice guideline for the treatment of patients with acute stress disorder and posttraumatic stress disorder. *Am J Psychiatry* 2004;**161**(Suppl. 11):3–31.

Aurora RN, Zak RS, Auerbach SH, et al. AASM Standards of Practice Committee. Best practice guidelines for the treatment of nightmare disorders in adults. *J Clin Sleep Med* 2010;**6**:389–401.

Breslau N, Roth T, Burduvali E, et al. Sleep in lifetime posttraumatic stress disorder: a community-based polysomnographic study. *Arch Gen Psychiatry* 2004;**61**(5):508–16.

Canadian Agency for Drugs and Technologies in Health. *Treatment for posttraumatic stress disorder, operational stress injury, or critical incident stress: a review of guidelines [Internet]*. Ottawa (ON): 2015 Apr 27. Available at http://www.ncbi.nlm.nih.gov/books/NBK304757/.

Gehrman PR, Harb GC, Cook JM, et al. Sleep diaries of Vietnam War veterans with chronic PTSD: the relationships among insomnia

symptoms, psychosocial stress, and nightmares. *Behav Sleep Med* 2015; **13**(3):255–64.
Jansson-Frojmark M, Lindblom K. A bidirectional relationship between anxiety and depression, and insomnia? A prospective study in the general population. *J Psychosom Res* 2008;**64**(4):443–9.
Krakow B, Hollifield M, Johnston L, et al. Imagery rehearsal therapy for chronic nightmares in sexual assault survivors with posttraumatic stress disorder: a randomized controlled trial. *JAMA* 2001;**286**(5):537–45.
Monti JM, Monti D. Sleep disturbance in generalized anxiety disorder and its treatment. *Sleep Med Rev* 2000;**4**(3):263–76.
Nardo D, Högberg G, Jonsson C, et al. Neurobiology of sleep disturbances in PTSD patients and traumatized controls: MRI and SPECT findings. *Front Psychiatry* 2015;**6**:134.
Pollack M, Kinrys G, Krystal A, et al. Eszopiclone coadministered with escitalopram in patients with insomnia and comorbid generalized anxiety disorder. *Arch Gen Psychiatry* 2008;**65**(5):551–62.
Raskind MA, Peterson K, Williams T, et al. A trial of prazosin for combat trauma PTSD with nightmares in active-duty soldiers returned from Iraq and Afghanistan. *Am J Psychiatry* 2013;**170**(9):1003–10.
Stein MB. Attending to anxiety disorders in primary care. *J Clin Psychiatry* 2003;**64**:35–9.

A complete reference list can be found online at ExpertConsult.com.

Chapter
137

Unipolar Major Depression

Jared D. Minkel; Andrew D. Krystal; Ruth M. Benca

Chapter Highlights

- Major depression is extremely common and is among the leading causes of disability worldwide.
- Self-reports of insomnia and hypersomnia are nearly universal among patients with major depression.
- Polysomnographic indices document objective alterations in sleep continuity, slow wave activity, and rapid eye movement sleep in patients with major depression and control subjects. Their sensitivity and specificity, however, are not such that they can be used as a stand-alone test for major depression.
- Although sleep disturbances have long been thought of as symptoms of major depression, the available evidence indicates a more complex, bidirectional relationship.
- The nature of the relationship of disturbances in sleep and major depression remains to be fully elucidated. It is clear, however, that optimal clinical management of patients with major depression must include the identification and treatment of the sleep problems present.

Mood disorders are the second most common category of psychiatric disorders after anxiety disorders,[1] and major depression alone affects at least 350 million persons worldwide. Moreover, the associated disability of mood disorders is among the highest reported for any disease: In 2000, depression was the leading cause of disability and the fourth leading contributor to the global burden of disease. The World Health Organization has reported that depression is now the leading cause of disability worldwide.[2] The impact of mood disorders includes eventual suicide in 15% of those affected, as well as increased morbidity and mortality from other illnesses.

Subjective sleep complaints are some of the most consistent symptoms associated with major depression. Disruption of typical sleep patterns (insomnia or hypersomnia) is a diagnostic criterion for depressive episodes in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM5)[3] (Box 137-1). Sleep has been studied more extensively in patients with depression than in those with any other psychiatric disorder, and in addition to the subjective reports, objective, robust, and relatively specific changes in sleep architecture have been identified as correlates with the underlying neurobiology of depression. Sleep disturbance is now thought to play an important causal role in the onset and maintenance of many cases of major depression, and efforts are under way to identify novel targets for intervention based on these discoveries.

CLASSIFICATION AND DIAGNOSIS

Mood disorders are subclassified into depressive and bipolar disorders, based on the pattern of depressive and manic episodes.[3] Diagnostic criteria for major depressive episodes are listed in Box 137-1 and include the characteristic finding(s) of depressed mood or anhedonia lasting at least 2 weeks. *Recurrent major depression* includes multiple, distinct major depressive episodes separated by at least 2 months of remission. By contrast, *dysthymic disorder* includes persistent but less severe symptoms lasting for at least 2 years. *Manic episodes* are characterized by elevated or irritable mood lasting at least a week or requiring hospitalization. *Hypomania* is associated with similar but less severe symptoms and may be less persistent. Patients who experience one or more manic or hypomanic episodes are diagnosed with a *bipolar disorder*. This chapter focuses only on *unipolar depression*; see Chapter 138 for discussion of bipolar disorder.

Major depressive episodes may be further categorized with one of several specifiers, including melancholic, atypical, and seasonal, each of which is at least partly related to sleep patterns. Melancholic depression includes a loss of pleasure in all, or almost all, activities; a lack of improvement in mood in response to normally pleasurable stimuli; and early-morning awakening and diurnal variation in mood, with depression worse in the morning. Of interest, patients with melancholic depression are more likely to experience symptomatic improvement with total sleep deprivation. Depression with atypical features includes significant mood reactivity to positive events and weight gain and hypersomnia during periods of depression. Depressive episodes can have a seasonal pattern, with a typical onset in the fall or winter and remission in the spring or summer, suggesting a correlation with diurnal patterns of light exposure (Box 137-2).

EPIDEMIOLOGY AND RISK FACTORS

Major depression is a common disorder and is reported to have a lifetime prevalence of 29.9% and a prevalence

Box 137-1 *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS* (DSM5) CRITERIA FOR MAJOR DEPRESSIVE EPISODE

At least five of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either depressed mood or loss of interest or pleasure. Symptoms that are clearly the result of a general medical condition or that are mood-incongruent delusions or hallucinations should not be included.

- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful); in children and adolescents, can be irritable mood
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day as indicated by either subjective account or observation made by others
- Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month) or decrease or increase in appetite nearly every day; in children, consider failure to make expected weight gains
- Insomnia or hypersomnia nearly every day
- Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
- Fatigue or loss of energy nearly every day
- Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
- Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
- Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

The symptoms are not due to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).

The symptoms are not better explained by a psychotic or delusional disorder such as schizophrenia or schizoaffective disorder.

Responses to a significant loss (e.g., bereavement, financial ruin) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in criterion A, which may resemble a depressive episode. Although such symptoms may be understandable or considered appropriate to the loss, the presence of a major depressive episode in addition to the normal response to a significant loss should also be carefully considered.

There has never been a manic episode or a hypomanic episode (unless the manic-like or hypomanic-like episodes are substance-induced or are attributable to the physiologic effects of another medical condition).

Modified from American Psychiatric Association: *Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5.* Washington, DC: American Psychiatric Publishing; 2013.

Box 137-2 *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS* (DSM5) CRITERIA FOR SEASONAL PATTERN SPECIFIER

This specifier applies to recurrent major depressive disorder if the following four criteria are met.

1. There has been a regular temporal relationship between the onset of major depressive episodes in bipolar I or bipolar II disorder or recurrent major depressive disorder and a particular time of the year (e.g., regular appearance of the major depressive episode in the fall or winter). Do not include cases in which there is an obvious effect of seasonal factor–related psychosocial stressors (e.g., regularly being unemployed every winter).
2. Full remissions (or a change from depression to mania or hypomania) also occur at a characteristic time of the year (e.g., depression disappears in the spring).
3. In the past 2 years, two major depressive episodes have occurred that demonstrate the temporal seasonal relationships defined in the first two criteria, and no nonseasonal major depressive episodes have occurred during that same period.
4. Seasonal major depressive episodes (as described above) substantially outnumber the nonseasonal major depressive episodes that might have occurred over the person's lifetime.

Modified from American Psychiatric Association: *Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5.* Washington, DC: American Psychiatric Publishing; 2013.

of 8.6% for the past 12 months in U.S. adults,[4] with an increased risk (odds ratio [OR], 1.7) in women.[5] The reason for increased rates of major depressive episodes in women is unclear, but hormonal factors as well as differences in psychosocial stressors may be important contributors, and it is likely that the increased rates of both insomnia and depression in women are related. Lifetime and 12-month prevalence rates for insomnia have not typically been reported in large-scale studies, but a 2009 study of 464 good sleepers in Quebec, Canada, followed over 12 months reported insomnia prevalence of 7.4%, with no difference in risk based on gender.[6] Major depression has its onset in early adulthood at a median age of 32 years; a pattern of recurrent episodes with recovery between episodes is typical. However, increased number and severity of episodes and poorer interepisode recovery can lead to an overall worse prognosis, suggesting that chronic treatment may be beneficial for patients with recurrent illness.

PATHOGENESIS

Despite the devastating impact of mood disorders, relatively little is known about their etiology or pathophysiology. It has long been hypothesized that deficiencies in central nervous system monoaminergic systems, including noradrenergic, serotoninergic, and dopaminergic neurotransmission, are responsible for depression, and most effective antidepressant treatments, including medications and

electroconvulsive therapy, increase intrasynaptic concentrations of monoamines.[7,8]

Neural plasticity—modulated by amino acid neurotransmitters, such as glutamate, and brain-derived neurotrophic factor (BDNF), discussed further on—has been shown to play a role in depression.[7,8] In addition to recognizing the role of neurotransmitter systems in mood pathophysiology, researchers have begun to elucidate the neuroanatomic, endocrine, and genetic factors involved.[7,8] Studies of normal sleep and disrupted sleep also demonstrate clearly that most of the systems involved in mood regulation also appear to be involved in the regulation of sleep and wakefulness, suggesting that dysfunction in particular brain systems could lead to both mood and sleep abnormalities.

Emerging evidence also suggests that stress may play an important role in the pathogenesis of depression, especially for people with marked stress sensitivity. Both human and animal studies have demonstrated that stress interferes with normal reward function, which may be one of the primary mechanisms explaining the high rates of depression after stressful experiences.[9] Stress also interferes with sleep, suggesting that it may be a common neurobiologic factor linking poor sleep and depression. The neurogenetics of stress sensitivity currently is being investigated by a number of researchers, and some candidate genes have been identified related to corticotropin-releasing hormone (CRH) and other polymorphisms at different nodes within the hypothalamic-pituitary-adrenal (HPA) axis.[9] Future studies may clarify how stress, sleep, and depression are related, but at present, these relationships are not well understood.

CLINICAL FEATURES

Subjective Sleep Complaints

Problems with sleep are some of the earliest and most commonly reported symptoms of depression. Specific sleep complaints include difficulty falling asleep, frequent nocturnal awakenings, early-morning awakening, nonrestorative sleep, decreased or increased total sleep, and disturbing dreams. Insomnia, hypersomnia, or both are reported by approximately 75% of adults, children, and adolescents with major depression.[10,11]

Association of Sleep Disturbance and Unipolar Major Depression

Sleep problems are extremely common in the general population and often are associated with psychiatric comorbidity. More than 35% of adults report some sleep problem.[12] In the general adult population, 14% to 20% of persons with significant complaints of insomnia and approximately 10% of those with hypersomnia showed evidence of major depression, whereas rates of depression were less than 1% in those without sleep complaints.[12,13] Additionally, the degree and duration of insomnia were positively correlated with more severe or recurrent major depression, or both. An assessment of the lifetime prevalence of sleep disturbance and psychiatric disorders in young adults also found greatly increased rates of major depression in persons with insomnia (OR, 3.8) in comparison with those with no sleep complaints (2.7%).[14]

The association between insomnia and depression may be even greater in clinical samples. More than one half of patients with insomnia and medical or psychiatric patients evaluated by clinical interview in sleep disorders centers had a sleep disorder associated with mood disorder according to the *International Classification of Sleep Disorders* (ICSD) diagnosis.[15] In children seen in general pediatrics clinics, insomnia and daytime fatigue were strongly correlated with elevated scores on the Child Behavior Checklist, and insomnia was particularly associated with symptoms of depression, anxiety, and attentional problems.[16] In one series of 265 consecutive children seen in a hospital-based sleep clinic, 19.6% were diagnosed with behavioral insomnia of childhood, second in frequency only to sleep apnea (67.9%). Of all children seen, 5.3% were diagnosed with unipolar depression.[17]

Predictive Value of Sleep Complaints

It has historically been assumed that major depression causes changes in sleep patterns. Sleep disturbances, however, also can affect major depression, and epidemiologic data support this latter association. Insomnia often precedes the onset of a first episode of major depression.[14,15] Self-reported insomnia symptoms as well as objective abnormalities in sleep latency, continuity, and duration (assessed by polysomnography) have been shown to predict depression longitudinally.[18] By contrast, insomnia was more likely to occur subsequent to the onset of anxiety disorders,[14] although persistent insomnia in childhood has been shown to predict anxiety in early adulthood.[19] Adolescents have been found to show a similar temporal relationship between onset of insomnia relative to depression or anxiety.[14]

In prospective two-wave longitudinal studies, subjects who reported sleep disturbance at both the initial and 1- to 3-year follow-up interviews were much more likely to have experience new-onset major depression.[20,21] The presence of insomnia or difficulty sleeping at an initial assessment also is associated with a long-term (beyond 30 years) increased relative risk for development of major depression.[22,23]

In a meta-analysis of prospective studies on risk factors for depression among community-dwelling elderly persons, 57% of the risk for depression was attributable to insomnia, and insomnia was second only to recent bereavement in predicting depression.[24] Women with more disrupted sleep also had more depression both before and after giving birth, and the presence of initial insomnia seemed to be the most relevant screening question for identifying women at risk for postpartum depression.[25] Children who reported decreased amounts of sleep[26] or insomnia[14] were more likely to acquire symptoms of depression and reduced self-esteem.

In patients who have had a previous episode of depression or mania, sleep changes remain a robust predictor of recurrent mood episodes, and their persistence is associated with a more severe clinical course. Insomnia and fatigue were the most commonly reported symptoms preceding a recurrent major depression.[27]

Interepisodic Persistence of Subjective Sleep Disruption

Although insomnia can resolve with remission from major depression, it also often persists. In fact, insomnia is the most commonly reported residual symptom during remission of major depression.[28] Persistence of sleep disturbances has been shown to predict increased severity and recurrence of major

depression.[28] A three-year prospective study of 267 depressed primary care patients found that sleep problems were present for 85% to 94% of the time during a depressive episode but also persisted for 39% to 44% of the time when depression was in remission.[29] This observation is consistent with findings from the STAR*D trial that insomnia persisted in 55% of those whose depression remitted with citalopram.[30] In a population-based study of more than 24,000 participants, persistent insomnia was associated with a 6.2-fold increased risk of developing major depression.[31] Persistent major depression also increased the likelihood of developing insomnia (6.7-fold increased risk), indicating bidirectional influences of sleep disturbance and depression. These findings demonstrate the importance of addressing sleep disturbance in both clinical practice and scientific investigations of depression.

Polysomnographic Findings

Major depression has been studied polysomnographically more than any other psychiatric disorder, and a majority of patients have been found to have objective sleep disturbances.[32] Since at least the late 1960s, polysomnographic indices have been extensively evaluated for their potential as biologic markers of major depression. Objective sleep abnormalities in depression have been grouped into three general categories: disturbance of sleep continuity, alterations of non–rapid eye movement (NREM) sleep (slow wave activity [SWA] and amount and distribution of stage N3 sleep), and abnormalities of rapid eye movement (REM) sleep.[33] Although these biomarkers have been reliably identified, sensitivity and specificity levels that would allow polysomnography to be used as a stand-alone test for major depression have not yet been established.[34]

Depressed patients showed prolonged sleep latency, increased wakefulness after sleep onset, and early-morning awakening, which results in sleep fragmentation and decreased sleep efficiency. Patients with depression have decreased amount and percentage of N3 sleep. Slow wave sleep loss is most significant during the first NREM period, but depressed patients appear to have reduced delta (1 to 4.5 Hz) power on the sleep electroencephalogram (EEG) and decreased slow wave counts throughout the night.[35] The distribution of N3 during the night is abnormal, with a decreased ratio of SWA in the first relative to that in the second NREM period.[36] The most robust finding in depression is a decreased REM sleep latency (time from onset of sleep to onset of REM sleep).[32] Other REM sleep abnormalities include a prolonged first REM sleep period and increased REM density. Increased percentage of REM sleep also has been observed. Common sleep complaints and polysomnographic abnormalities are listed in Box 137-3. Results from the few studies investigating dysthymia have been variable, but they suggest that some of the characteristic sleep findings of major depression are present but not to the same extent.[37,38]

Sleep disturbances in patients with depression are not limited to periods of acute depressive episodes: Several studies have reported abnormal sleep parameters in patients during remission and during acute illness. Some sleep abnormalities may be more severe in acute than in remission phases, including increased REM density and reduced sleep efficiency.[39] However, reduced REM sleep latency and decreased slow wave sleep can persist for prolonged periods in otherwise asymptomatic persons. Thus sleep disturbances—particularly reduced REM sleep latency and slow wave sleep abnormalities—may be trait markers for some patients with major depression, rather than simply indications of an acute state of illness. The persistence of sleep abnormalities suggests that sleep disturbance could indicate a biologic susceptibility to depression and predate the illness, or that sleep changes may be caused by depression and persist much longer than other affective symptoms.[33] Additionally, first-degree relatives of subjects with major depression also show reduced REM sleep latency and slow wave sleep deficits, whether or not they have a personal history of a mood disorder,[40] suggesting that sleep changes in depression include a hereditary component.

Box 137-3 SLEEP ABNORMALITIES IN DEPRESSION

Subjective Complaints

Insomnia
- Difficulty falling asleep
- Increased awakening at night/restless sleep
- Early morning awakening
- Decreased amounts of sleep

Sleep described as less deep or less refreshing
Disturbing dreams

Polygraphic Findings

Sleep continuity disturbances
- Prolonged sleep latency
- Increased wake time during sleep
- Increased early morning wake time
- Decreased total sleep time

Stage N3 sleep deficits
- Decreased N3 amount
- Decreased N3 percentage of total sleep

REM sleep abnormalities
- Reduced REM sleep latency
- Prolongation of the first REM sleep period
- Increased REM activity (total number of eye movements during the night)
- Increased REM density (REM activity/total REM sleep time)
- Increased REM sleep percentage of total sleep

REM, Rapid eye movement; SWS, slow wave sleep.

Advanced Sleep Electroencephalographic Analysis

Traditional polysomnography studies are recorded from only a few electrodes and have relied primarily on sleep architecture—an indirect assessment of brain activity during sleep—as the primary analysis variable. However, technologic improvements in EEG systems and computerized analysis have provided new tools to more directly assess normal and disrupted brain function in major depression by analyzing the fundamental oscillations of the sleep EEG. Techniques include quantitative analysis of EEG activity across a broad range of frequencies (power spectral analysis), the automated detection and investigation of specific EEG waveforms such as slow waves, spindles, and REMs, as well as measures of the similarity of EEG oscillations across brain regions over time (synchrony and coherence). One advantage of these approaches is the ability to more completely assess specific EEG frequencies or waveforms, independent of sleep stage

(e.g., slow waves can be evaluated in all NREM sleep, not just N3).

Power Spectral Analysis

Sleep EEG power spectra for individual subjects are highly consistent between nights.[41] Power spectral analysis has been used to characterize SWA (1 to 4.5 Hz EEG power) in control and depressed subjects, which confirmed that N3 reductions in major depression were correlated with decrements in SWA, validating the use of EEG power spectra to quantify sleep slow waves.[42] Increases in SWA and N3 on the first night of antidepressant medication treatment have been found to predict subsequent treatment response to antidepressant medications[42]; however this has not yet been replicated. Although these findings might point to SWA as a state marker of depression, the bulk of the evidence indicates that decreased SWA in depressed patients compared with control subjects persists after remission, suggesting that it is more likely to be a trait marker of major depression.[42]

High-Density Electroencephalogram

Traditional polysomnography uses a relatively small number of electrodes (eight or fewer) to collect information about brain activity during sleep. Newer technology allows for a much greater amount of data to be collected, thus providing a more complete view of EEG brain activity. High-density EEG refers to a technique whereby much of the scalp is covered by electrodes (256 electrodes total). This technique offers a versatile, yet less expensive approach to studying more changes in activity in specific brain regions in patients with depression when compared with other brain imaging techniques, such as functional magnetic resonance imaging or positron emission tomography. A series of studies investigating sleep and depression using this technique have identified abnormalities in sleep spindles and SWA that would not have been possible to characterize using traditional EEG.[43,44]

Automated Analysis of Slow Waves

Manual identification of slow waves is complemented by using automated or semiautomated algorithms, which provides a quantitative index with improved sensitivity.[42] The *delta sleep ratio* has been defined as the ratio of the average slow (delta) wave counts per minute in the first NREM sleep period to the average counts per minute in the second NREM sleep period.[42] In control subjects, this ratio (typically greater than 1.6) reflects the higher density of slow waves in the first NREM sleep period. Depressed subjects tend to have a lower ratio (1.10 or less), reflecting this abnormal distribution.[42] The delta sleep ratio has been thought to reflect an abnormal sleep homeostatic process in depression (deficient process S^{38}). In major depression, REM sleep onset is earlier, and SWA is less during the first NREM sleep period, which is the opposite of the normal homeostatic pattern.

The delta sleep ratio has been shown to predict treatment response and likelihood of recurrence: Patients with higher delta sleep ratios were more likely to respond to treatments, and those with lower ratios were more likely to experience recurrent depression.[45,46] Although reduced SWA in the first NREM sleep period also has been found in subjects with schizophrenia, only subjects with depression showed a reduced delta sleep ratio, suggesting that this measure may be more specific for major depression.[47]

Coherence of Sleep Electroencephalogram Rhythms

Coherence is a measure of the similarity of EEG rhythms at different cortical locations. EEG coherence is thought to reflect the functional relationships among different cortical regions. Decreased coherence may indicate impaired brain connectivity associated with psychiatric disorders including major depression.[48] Both intra- and interhemispheric coherence in the delta (0.5 to 4 Hz) and beta (16 to 32 Hz) frequency ranges have been shown to be lower in adolescents and adults with major depression than in control subjects, even in the absence of differences in sleep stage amounts.[48,49] Reduced coherence also appears to be a factor in offspring at risk for developing major depression,[50] and it correlates with risk of recurrence in children and adolescents with major depression,[49] suggesting that it may be a biologic marker of depression.

Mechanisms Associated with Sleep and Unipolar Major Depression

The high coincidence and overlapping symptoms of major depression and insomnia point to a common neurobiology. Attempts have been made to explain the mechanisms for sleep changes in depressive disorders as well as to correlate them with other biologic abnormalities in depression. A number of hypotheses have been advanced and include neurotransmitter imbalance, abnormalities in brain activation, dysregulation of the HPA axis, inflammation, and genetic polymorphisms.

Neurotransmitter Imbalance

The monoamine hypothesis suggests that decreased levels of serotonin, norepinephrine, and dopamine result in the symptoms of depression. This model fits with the proposed action of most antidepressant medications and with the characteristic findings of increased REM and decreased slow wave sleep during depression. Considerable evidence suggests that REM sleep is promoted by cholinergic activation within the medial pontine reticular formation and is inhibited by aminergic (serotoninergic and noradrenergic) activation (see Chapter 8).[51] Thus short REM sleep latency and elevated REM density in depression could be related to enhanced cholinergic neurotransmission or diminished aminergic neurotransmission, or both. Each of these systems is under strong genetic control and could define an endophenotype of depression.[52]

Antidepressant medications produce immediate increases in monoamine levels yet have a therapeutic lag of weeks.[8] This delay is likely to be related to secondary neuroplastic changes that mediate the longer-term effect of antidepressants. The glutamate system is a key neuromodulator related to plasticity, and it also is intimately tied to both sleep and mood regulation.[53] Glutamate interacts with cholinergic neurons to increase activity of the reticular system associated with REM sleep onset. During NREM sleep, excitatory glutamate neurotransmission has a prominent role in the thalamocortical generation of sleep EEG oscillations.[51] Additionally, sleep increasingly has been shown to be necessary for plastic processes such as learning and memory, and it affects the expression of plasticity-related genes.[54] The intertwined processes of mood, sleep, and plasticity, and their modulation through factors such as glutamate and BDNF, make them appealing targets for future therapies for major depression.[53]

Abnormalities in Brain Activation

Neuroimaging has been an invaluable tool for identifying specific brain structures with altered activity or function in major depression.[55] The same modalities can be applied in healthy and depressed subjects during waking and sleep states to identify brain regions involved in regulating sleep and how their activity is altered in depressed patients with sleep disturbances. During normal NREM sleep, metabolic activity is broadly decreased over waking levels in the frontal, temporal, and parietal cortices. Subjects with current major depression exhibited a smaller decrease in these cortical regions from waking to sleep, and they showed a relative hypoactivity during wakefulness.[56] Other brain areas involved in emotional regulation (anterior cingulate cortex, amygdala, parahippocampal cortex, thalamus) also displayed a smaller difference in metabolic rates between waking and NREM sleep and elevated metabolic rates during sleep in depressed subjects relative to control subjects.[55] Altered function in these regions could relate to a failure of arousal mechanisms to decline from waking to sleep, as well as changes in cognition, attention, and emotional regulation in depression.[57] Similarly, increased metabolic activity in diffuse cortical and subcortical structures during REM sleep was seen in depressed subjects, over that in control subjects.[57] These alterations also could reflect the imbalance of monoaminergic and cholinergic systems in altered mood states, and they could explain the increased REM sleep, decreased REM sleep latency, and decreased slow wave sleep in depression.

Other imaging studies have focused on changes in brain activity after total or partial sleep deprivation of depressed patients. Approximately 50% of patients with major depression show rapid symptomatic improvement after sleep deprivation, as discussed later on (reviewed by Berg and associates[58]), further supporting the hypothesis that sleep and mood regulation are controlled by overlapping brain regions. Patients who responded to sleep deprivation initially exhibited increased metabolic activity in the amygdala, orbital prefrontal cortex, inferior temporal lobe, and anterior cingulate that normalized after sleep deprivation.[59,60] Predeprivation perfusion levels correlating with the reduction of depressive symptoms and with single photon emission computed tomography studies suggest that sleep deprivation responders may have a particular deficit in dopaminergic and serotoninergic systems involved with attention, arousal, and mood.[59]

Dysregulation of the Hypothalamic- Pituitary-Adrenal Axis

Some sleep abnormalities observed in patients with major depression may be related to dysregulation of the HPA axis. Supporting this hypothesis, depressed patients demonstrate excessive CRH and cortisol secretion and also show a lack of cortisol suppression in response to challenge with dexamethasone.[61] Chronic stress also increases secretion of cortisol, linking life events to mood episodes. CRH is known to increase arousal, disrupt sleep, and possibly contribute to reduced REM sleep latency.[62,63] Growth hormone–releasing hormone (GHRH) has a reciprocal relationship with CRH and promotes sleep. GHRH and growth hormone also may be decreased in some patients with depression, further contributing to slow wave sleep decrements.[63] Administration of a CRH receptor antagonist was reported to improve sleep EEG patterns in depressed patients.[64]

Additional evidence suggests that poor sleep may contribute to HPA axis dysregulation independently from major depression. Elevated cortisol under stress has been associated with increased awakenings and more time spent in stage 1 and stage 2 sleep.[65] In children, poor sleepers have been shown to have increased morning cortisol over that in good sleepers. Experimental studies have shown that sleep deprivation elevates cortisol release and blood pressure in response to psychosocial stressors.[66,67] Taken together, these findings suggest bidirectional influences of stress and poor sleep mediated by HPA axis dysregulation, which speak for addressing sleep problems in patients with major depression.

Inflammation

Ample evidence shows that depression is associated with a chronic, low-grade inflammatory and immune response.[68] Both chronic and acute sleep deprivation are associated with altered cellular and immune functioning. Although findings have not been consistently reported, some evidence suggests that insufficient sleep may impair immune system function by increasing proinflammatory cytokines (as reviewed by Berk and colleagues[68]). Experimental studies in rats[69] and nonhuman primates[70] suggest that proinflammatory cytokines may cause symptoms of depression. Correlational findings in human studies also are consistent with this hypothesis.[71] It is therefore possible that insufficient sleep contributes to unipolar depression, at least in part, by increasing inflammation.

Genetic Polymorphisms

At least 33% of the risk for major depression can result from genetic factors, and numerous candidate genes have been implicated.[72] Additionally, a growing number of genes have a role in major depression and in sleep regulation. Both the gene for monoamine oxidase A (MAO-A) and the serotonin transporter gene promoter (linked polymorphic region [5-HTTLPR]) have been implicated in depression and can correlate with insomnia scores (MAO-A) or treatment response to sleep deprivation (5-HTTLPR).[73] Similarly, some early work pointed to associations of clock genes, which regulate circadian rhythms, and disrupted sleep in major depression. Multiple reports have now linked a *CLOCK* gene polymorphism (3111 T/C) to the presence of insomnia or decreased sleep time in depressed patients.[74] Of note, however, a recent large study using summary data from the Psychiatric Genomics Consortia found no evidence that genes related to sleep and circadian rhythms increased risk for major depression.[75]

Primary Sleep Disorders and Unipolar Major Depression

An increased association between primary sleep disorders and major depression also has been postulated. Persons with sleep apnea have a 1.8-fold increased risk of developing major depressive disorder, and those with depression, a 1.6-fold increased risk for obstructive sleep apnea (OSA).[76] Similarly, in a longitudinal study, persons with sleep-related breathing disorders showed a severity-related risk of developing depression, with ORs ranging from 1.6 (minimal) to 2.6 (moderate or more severe sleep-related breathing).[76] Depression also may be a modifier of clinical course: In patients with OSA and depression, depression is directly associated with severity of daytime fatigue independent of severity of OSA.[77] Other

studies have shown that treating OSA results in sustained symptomatic improvement in patients with depression, and the degree of improvement in OSA with continuous positive airway pressure also correlated with the amelioration of depression.[78]

The relationship between major depression and restless legs syndrome is less well defined. Several studies have shown an inconsistent correlation between the two disorders, even with implementation of appropriate controls for the potential exacerbation of restless legs syndrome by serotoninergic antidepressants.[79]

TREATMENT

Patients with major depression often present to primary care physicians or sleep clinics with complaints of insomnia or fatigue, or both. However, because of the strong association between major depression and insomnia, any patient presenting with sleep complaints must be screened for depression. Recognition of major depression is important for determining appropriate treatment, and screening also must include a careful assessment of possible suicide risk. In the evaluation for depression, it is important for the clinician to establish rapport and create a permissive and supportive atmosphere for the patient to discuss any psychological symptoms. It is often necessary to ask about depressed mood and anhedonia in several ways (Box 137-4 gives examples) before the patient recognizes or admits to these symptoms. Several self-report measures have been validated for measuring the severity of patients' depression and also may be helpful in the initial detection of these symptoms. These assessment tools include the Patient Reported Outcomes Measurement Information System (PROMIS) depression measures developed by the National Institutes of Health[80] (available free of charge through www.nihpromis.org), the Center for Epidemiological Studies Depression Scale (CES-D),[81] and the Beck Depression Inventory (BDI-II).[82]

Treatment of Sleep Disturbance in Patients with Unipolar Depression

Major depression and the accompanying sleep disturbances may be treated with medication, psychotherapy, or both. Patients with moderate to severe depression usually are treated with medication, either alone or in combination with psychotherapy (Table 137-1).[83]

Box 137-4 EXAMPLES OF QUESTIONS TO ELICIT PRIMARY SYMPTOMS OF DEPRESSION

Depressed Mood

How has your mood been recently?
Have you been feeling down, depressed, sad, or blue lately?
Has your sleep problem affected your mood?

Loss of Interest or Pleasure

Has your ability to enjoy things changed?
Has your interest in things decreased?
Are you less motivated to do things that are normally enjoyable for you?

Treatment with Medication

Depression. Selective serotonin reuptake inhibitors (SSRIs) currently are the most widely prescribed class of drugs for depression in the United States. Along with the serotonin-norepinephrine reuptake inhibitors (SNRIs) (i.e., desvenlafaxine, duloxetine, and venlafaxine) and newer atypical antidepressants (e.g., bupropion, mirtazapine), they have become first-line agents because of their safety and improved side effect profiles in comparison with the tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs). However, SSRIs, SNRIs, and bupropion can cause significant sleep disruption and worsen insomnia in some patients.[83] By contrast, trazodone, nefazodone, and mirtazapine are sedating and can improve sleep initiation and maintenance.[83] It is therefore important for patients to understand the effects of their antidepressant medications on sleep-wake systems, to promote adherence to the appropriate regimen including timing of doses.

Although use of TCAs and MAOIs has diminished because of the associated side effects and greater potential toxicity, these drugs may be effective in patients who fail to improve with the newer antidepressants. Most TCAs are quite sedating and may therefore be helpful to depressed patients with insomnia; they continue to be used for treatment of migraine headaches and other chronic pain conditions. Low doses of TCAs have commonly been prescribed for insomnia, and a few studies have shown good efficacy and tolerability of low-dose TCAs as hypnotics in the absence of significant depression.[84] MAOIs are sometimes used, albeit rarely, in patients who fail to improve with other antidepressants, particularly patients with atypical features (hypersomnia, increased appetite) or significant comorbid anxiety. The most serious potential side effect of MAOIs is the danger of hypertensive crisis if they are used in combination with sympathomimetic drugs or with foods containing tyramine.

For severely depressed patients, including those with psychotic features or strong suicidal intent or whose depression is resistant to antidepressants alone, additional treatment modalities, including electroconvulsive therapy and antipsychotic medications, should be considered. Second-generation antipsychotics (SGAs) (e.g., risperidone, olanzapine, quetiapine, ziprasidone, aripiprazole) are used to enhance the antidepressant response to SSRIs or as mood stabilizers in patients with bipolar disorders (see Chapters 42, 138, and 139 for further discussion of antipsychotic medications). Patients with severe fatigue or hypersomnia also might benefit from adjunctive use of stimulant medications, although the evidence for direct effects on mood or treatment course is limited.[85] Recent large reviews of the effects of antidepressant medications have shown that the magnitude of the effect of medication increases as severity of depression increases.[86,87] For mild depression, medication has not been shown to have advantages over placebo. Therefore medications that include side effects generally should be avoided in this population.

Use of the *N*-methyl-D-aspartate (NMDA) antagonist ketamine is a recent innovation in the treatment of major depression that has provided the first intervention, other than electroconvulsive therapy and sleep deprivation (see further on), that produces a more rapid improvement in depression and does so in treatment-resistant patients. Some work has been completed indicating that changes in sleep may occur in

Table 137-1 Treatment of Major Depression*

Medication	Usual Dosage Range	Pharmacologic Mechanism	Side Effects	Effects on Sleep
Tricyclic Antidepressants (TCAs)				
Amitriptyline (Elavil) Imipramine (Tofranil) Doxepin (Sinequan) Nortriptyline (Pamelor) Clomipramine (Anafranil) Desipramine (Norpramin) Protriptyline (Vivactil)	50–150 mg 50–200 mg 75–300 mg 50–150 mg 150–250 mg 75–200 mg 20–60 mg	Inhibit serotonin and norepinephrine reuptake Anticholinergic Antihistaminergic	*Anticholinergic effects*: blurred vision, dry mouth, urinary retention, orthostatic hypotension, flushing, tachycardia, confusion, others (drugs are listed in decreasing order of severity of these effects) *Other side effects*: liver toxicity, lowering of seizure threshold, sweating, weight gain	Sedation (drugs listed in decreasing order: clomipramine, desipramine, and protriptyline may be nonsedating or activating) REM sleep suppression Increased stage 2 sleep
Monoamine Oxidase Inhibitors (MAOIs)				
Phenelzine (Nardil) Isocarboxazid (Marplan) Tranylcypromine (Parnate) Selegiline transdermal (Emsam)	15–60 mg 40–60 mg 10–30 mg 6–12 mg/24–hr patch	Inhibit monoamine oxidase Increase norepinephrine, serotonin, and dopamine	Hypertensive crisis in combination with tyramine-containing foods or sympathomimetics Anticholinergic effects Dizziness, agitation Liver toxicity Weight gain	Insomnia Potent REM sleep suppression
Selective Serotonin Reuptake Inhibitors (SSRIs)				
Fluoxetine (Prozac) Paroxetine (Paxil) Sertraline (Zoloft) Citalopram (Celexa) Escitalopram (Lexapro) Vortioxetine (Brintellix)	20–60 mg 20–60 mg 50–200 mg 20–40 mg 10–20 mg 5–20 mg	Inhibits reuptake of serotonin	Gastrointestinal disturbances Sexual dysfunction Anxiety, agitation Possible increased risk for suicide	Insomnia REM sleep suppression Increased eye movements in NREM sleep
Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)				
Venlafaxine (Effexor) Desvenlafaxine (Pristiq) Duloxetine (Cymbalta) Levomilnacipram (Fetzima)	150–375 mg 50 mg 30–120 mg 40–120 mg	Inhibit reuptake of serotonin and norepinephrine	Anxiety Anorexia Hypertension Possible increased risk for suicide	Insomnia
Other Antidepressants				
Trazodone (Desyrel)	150–600 mg	Inhibits serotonin reuptake 5-HT_2 receptor antagonist	Few anticholinergic effects Priapism Nausea, dizziness, dry mouth	Decreased REM sleep Increased SWS Sedation
Nefazodone (Serzone)	200–600 mg	Inhibits serotonin reuptake 5-HT_2 receptor antagonist	Few anticholinergic effects Priapism Nausea, dizziness, dry mouth	No change or increased REM sleep Increased SWS Sedation
Bupropion (Wellbutrin)	150–450 mg	Inhibits norepinephrine and dopamine reuptake	Gastrointestinal upset Lowering of seizure threshold Anxiety	Insomnia Increased REM sleep
Mirtazapine (Remeron)	15–45 mg	Alpha$_2$ adrenergic receptor antagonist 5-HT_{1a} agonist 5-HT_2 and 5-HT_3 receptor antagonist	Increased appetite, weight gain Dizziness	Sedation
Vilazodone (Vibryd)	10-40 mg	Inhibits serotonin reuptake Partial agonist of 5HT_{1A} receptors	Gastrointestinal disturbances Sexual dysfunction Anxiety, agitation Possible increased risk for suicide	Insomnia REM sleep suppression Increased eye movements in NREM sleep

*In October 2004, the U.S. Food and Drug Administration (FDA) directed manufacturers to add a black box warning to the health professional labeling of all antidepressant medications, to include warnings about increased risks of suicidal thinking and behavior, known as suicidality, in young adults ages 18 to 24 years during initial treatment (generally the first 1 to 2 months), and to emphasize the need for close monitoring of patients starting on these medications.

5-HT, 5-Hydroxytryptamine; NREM, non–rapid eye movement; REM, rapid eye movement; SWS, slow wave sleep.

conjunction with the antidepressant effects of ketamine. A recent study comprising 30 patients found that a positive response to ketamine was associated with increased SWA during early sleep and elevated BDNF levels.[88] *Delta sleep ratio,* defined as the ratio of SWA between the first two NREM sleep episodes, also has been found to be lower in depressed patients than in healthy control subjects. A positive correlation was found between baseline delta sleep ratio and improvement in depression severity after ketamine administration.[89]

Concurrent Treatment of Depression and Insomnia

Some depression treatment studies have investigated the efficacy of coadministering antidepressants (SSRIs or TCAs) with hypnotic agents such as eszopiclone,[90] zolpidem,[91] or lorazepam.[92] The results from the limited number of studies suggest that addition of a hypnotic generally was well tolerated, improved insomnia associated with depression, and, in one study also was associated with abatement of core symptoms of depression.[90] Hypnotics alone, however, constitute insufficient treatment for depression and are not a substitute for approved antidepressant treatments. The mechanism whereby eszopiclone was found to relieve core depression symptoms and the lack of efficacy for other medications such as zolpidem are unclear. One possibility is that eszopiclone has some direct antidepressant effects owing to its pharmacology, which differs from that of zolpidem. Further studies are needed to resolve this issue.

Insomnia that persists despite treatment with other antidepressants or mood stabilizers may require additional pharmacotherapy or behavioral therapy. Trazodone has been shown to have hypnotic effects in patients with insomnia and depression[93] and often is combined with other antidepressants to improve sleep.[94]

Preliminary evidence from a small randomized, controlled pilot study found that among patients suffering from both insomnia and depression who received cognitive-behavioral therapy (CBT) for insomnia, in addition to antidepressant treatment with citalopram, higher rates of remission from both depression and insomnia were typical.[95] These findings have not yet been supported by a larger study, but the initial data suggest that treatment approaches integrating behavioral and pharmacologic interventions may improve treatment outcome.

Sleep Effects of Psychiatric Medications. Antidepressants have a variety of effects on sleep: more sedating or anticholinergic agents (e.g., trazodone, nefazodone, mirtazapine, TCAs) can cause hangover effects; conversely, activating antidepressants (e.g., bupropion, SSRIs, MAOIs) can cause or exacerbate insomnia, especially if taken near bedtime.[96] SSRIs have been associated with frequent eye movements in NREM sleep, which can persist for prolonged periods after cessation of treatment; the clinical significance of this effect is unknown.[97] TCAs and SSRIs have been implicated in precipitating or exacerbating restless legs syndrome and periodic leg movements, and they can exacerbate symptoms of insomnia or daytime fatigue by this mechanism in some patients.[98] REM sleep behavior disorder also has been reported after administration of various REM sleep–suppressing antidepressants.[97] Despite their increasingly common use, few studies have been conducted to investigate objective sleep changes related to the use of SGAs.[83] In general, SGAs appear to decrease sleep onset latency and to improve sleep continuity and efficiency, but they also commonly cause weight gain, which can increase the risk of OSA.

Nonpharmacologic Treatments

Empirically validated psychotherapies (e.g., CBT and interpersonal therapy) are approximately as efficacious as medications, except in very severe depression. These treatments may be particularly useful for patients who experience strong side effects of medication or who have specific contraindications such as pregnancy. Patients with significant insomnia should be instructed in the basic behavioral interventions for sleep, such as avoiding excessive time in bed and keeping a regular sleep-wake schedule. Behavioral techniques such as stimulus control and relaxation may be useful for some patients as well. For more severe or chronic symptoms, CBT for insomnia is effective in improving sleep disturbances associated with major depression but is not yet widely available.[99] Cognitive therapy also has been shown to be an effective treatment for insomnia that focuses on reducing worry and unhelpful self-monitoring and avoidance behaviors[100,101] and can be easily integrated with psychological treatments for depression. Psychological and behavioral interventions are reviewed in detail in Chapters 85 and 86.

For patients with seasonal depression and hypersomnia, bright light therapy has been shown to be effective, either alone or in combination with antidepressant medication. Light therapy and its use in treating depression are reviewed in detail in Chapter 39.

Treatment of Unipolar Major Depression with Sleep Deprivation

A variety of sleep manipulations, including total and partial sleep deprivation or selective deprivation of REM or slow wave sleep, have been shown to have rapid antidepressant effects, although they have not come into widespread clinical use.[102-104] A single night of total sleep deprivation has been documented to have antidepressant effects (50% reduction of Hamilton Depression Rating Scale scores); response rates are 30% to 60%, with peak antidepressant effects by the afternoon following a night of total sleep loss.[102,103] Patients with a more pronounced diurnal pattern of illness exhibit better responses to sleep deprivation. The major drawback to total sleep deprivation as a therapy for depression, however, has been the immediate reversibility of the antidepressant effects by recovery sleep, including short naps, in at least 50% to 80% of responders, in addition to the problems associated with sleep deprivation itself (e.g., increased sleepiness, limited ability to sustain treatment).

Sleep deprivation has been combined with other treatment modalities—including medications (lithium, antidepressants), sleep phase advance, light therapy, and transcranial magnetic stimulation—with mixed results. The aim is to obtain a rapid response in concert with sustained improvements from other modalities.[103]

Like total sleep deprivation, partial deprivation can be immediately effective and has yielded comparable response rates in some studies.[102,103] The timing of partial sleep deprivation (e.g., sleep restricted to the first or second half of the night) does not appear to be as critical as the degree of

deprivation.[103] Chronic REM sleep deprivation also improved depressive symptomatology,[102,103] although the response was not as robust, and it was not clear that NREM sleep was not affected. Unlike with total sleep deprivation, however, the antidepressant effects of chronic REM sleep deprivation took several weeks to appear and were not immediately reversed after recovery sleep. Selective deprivation of slow wave sleep also has been shown to relieve depression symptoms,[104] with the magnitude of the response correlating with the magnitude of the reduction in SWA relative to baseline.

Although the mechanisms underlying the antidepressant effects of sleep deprivation are not entirely known, some promising correlates have been reported. Functional neuroimaging studies suggest an association with effects on limbic structures. Increased rates of glucose metabolism[59] or blood flow[60] in the amygdala and cingulate before sleep deprivation have been reported in depressed subjects who respond to sleep deprivation. Normalization of activity after sleep deprivation in these limbic structures was associated with a decrease in depression.[59,60] Antidepressant responses to partial sleep deprivation have been shown to be related to BDNF, a protein involved in the pathophysiology of depression.[105] Responders showed significant diurnal BDNF serum variation both before and after partial sleep deprivation that was not present in non-responders, suggesting that BDNF may be involved in positive responses, but is not the primary mechanism. Preliminary data from a sample of 12 healthy young male subjects suggest that sleep deprivation may elevate tryptophan, serotonin, and taurine, possibly contributing to the antidepressant effects of sleep deprivation.[106] The antidepressant response to sleep deprivation also has been correlated with other biologic markers in sleep EEG[46,104] and genetic[73] studies. Abnormalities in slow wave sleep are one of the most consistent biologic markers of depression. Additionally, response to sleep deprivation can serve as a biologic marker of a major depression subtype, as well as the basis for designing novel antidepressants.

These findings should not lead clinicians to conclude that insufficient sleep is generally helpful for patients with depression. On the contrary, patients who obtain suboptimal levels of sleep have been shown to have poorer outcomes. In addition, experimental studies have shown that sleep deprivation tends to have deleterious effects on positive emotion in healthy adults.[107,108] Unless sleep restriction is done therapeutically, it should be avoided in patients with depression.

Clinical Application of Sleep Studies

Sleep studies have potential utility in psychiatry for diagnosis and for evaluating treatment. Although a specific psychiatric diagnosis cannot be made solely on the basis of standard polysomnographic data, sleep studies can sometimes answer specific questions. Sleep complaints in most patients with major depression usually are related to the underlying psychiatric illness; however, diagnostic polysomnography can sometimes be helpful given the risk for primary sleep disorders in psychiatric patients, and vice versa. Furthermore, in patients with major depression and significant sleep complaints that do not respond to therapy, it is important to consider the possibility of a concomitant primary sleep disorder, such as OSA, particularly in view of the overlap of common symptoms. For example, sleep apnea and periodic limb movements can disrupt nocturnal sleep and lead to insomnia, fatigue, and poor concentration, which also are symptoms of depression. It is also important to consider side effects of psychiatric medications that may be responsible for precipitating or exacerbating sleep disorders.

The potential prognostic utility of sleep studies in major depression has not yet been fully realized. It is possible that specific sleep abnormalities, such as REM and slow wave sleep changes, may be diagnostic or predictive of treatment response. However, these findings have not shown consistency that is adequate to justify widespread implementation.[32,42] Alternatively, medication-induced changes in sleep patterns may be correlated with treatment response to antidepressants. The amount of REM sleep suppression during the first night of treatment with TCAs was correlated with eventual antidepressant response in several studies,[109,110] although prolongation of REM sleep latency alone was less consistent in predicting treatment response. Sleep variables also may be helpful in identifying persons susceptible to developing affective illnesses or relapses. For instance, a decreased delta sleep ratio has been shown to correlate with an increased risk of relapse and predict treatment outcome.[39,45]

In interpreting sleep studies in clinical settings, it is important to bear in mind that reduced REM sleep latency can be present in a variety of conditions other than mood disorders. Short REM sleep latency and decreased slow wave sleep have been reported in other conditions, including schizophrenia, borderline personality disorder, eating disorders, and alcoholism.[32] Early appearance of REM sleep also is characteristic of narcolepsy. Short REM sleep latency and REM sleep rebound commonly occur in patients who recently have undergone withdrawal from antidepressant medications, benzodiazepines, or alcohol and in patients with OSA during initial continuous positive airway pressure titration.

CLINICAL COURSE AND PREVENTION

Major depression tends to be a remitting and relapsing disorder with a long course, although some patients experience only a single episode. Increasingly, it is recognized that repeated episodes of depression contribute to a poor prognosis, which underscores the importance of early and effective treatment. Patients with multiple episodes of depression probably require chronic treatment to prevent or minimize recurrence.

After treatment with pharmacotherapy or psychotherapy (or both), approximately one third of depressed patients recover fully, one third experience partial remission, and one third do not respond significantly to treatment. Treatment of depressive episodes reduces episode duration, but it should be continued for at least 6 months to 1 year after remission because of the higher risk of relapse during this period. All patients, including those considered to be in remission, can continue to experience residual symptoms of depression. Sleep disturbance is one of the more common residual symptoms and has been estimated to occur in up to 44% of patients in remission from major depression.[111] In addition, insomnia is one of the most predictive symptoms heralding the onset of an episode of depression,[14,112] as discussed earlier. Thus it is not surprising that patients with major depression often suffer from chronic sleep problems.

PITFALLS AND CONTROVERSIES

Considerable evidence indicates that sleep is biologically linked to major depression, although the mechanisms underlying this link are incompletely understood. Perhaps the most basic question is whether a causal link exists between sleep and depression, regardless of the specific mechanism involved. For example, although sleep abnormalities are extremely common in patients with depression, it is unclear if this association indicates that these abnormalities are caused by major depression or whether they occur independently and predispose affected persons to major depression. Indirect evidence suggests that changes in sleep herald the onset of mood episodes. On the other hand, some changes in sleep that occur in patients with depression, such as short REM latency, are unlikely to be causal. For example, although narcoleptic patients have extremely short REM sleep latencies, they are not invariably depressed, and inducing "depressive" sleep patterns in normal persons with cholinergic agents does not cause depression. Although some studies have begun to address the causal relationship of sleep changes and major depression and have evaluated the impact of insomnia treatment on depression outcomes, the effect of concurrent treatment for sleep problems on the clinical course of major depression is not yet fully elucidated.

Continued study of sleep in major depression and in psychiatric disorders in general can be expected to elucidate neurobiologic and genetic mechanisms involved in sleep regulation and psychiatric illnesses and to provide novel directions for future neuropharmacologic treatments.[113,114] Longitudinal studies may be particularly helpful in clarifying the relationship between depression and sleep.[115]

CLINICAL PEARL

Because of the strong association between sleep disturbance and major depression, all patients with sleep complaints should be screened for mood disorders, and vice versa. Sleep may not normalize with resolution of major depression, and recurrence or worsening of sleep can predict the relapse of depressive symptoms. Concurrent and aggressive treatment of sleep problems, in addition to the treatment for mood, is important for decreasing morbidity and improving the patient's quality of life.

SUMMARY

Major depression is commonly associated with sleep disturbance, and sleep problems are part of the diagnostic criteria for this disorder. Although subjective complaints of insomnia are most common, hypersomnia and fatigue are also sometimes reported during periods of depression. Insomnia and hypersomnia are associated with an increased risk for the development or recurrence of depression and increased severity of symptoms. Polysomnographic studies of depressed patients have consistently revealed abnormalities in sleep architecture in comparison to controls, including decreased time in slow wave sleep, reduced latency to REM sleep onset, and disrupted sleep continuity. These associations provide insight into the neurobiologic relationships between mood and sleep. Both subjective and objective sleep abnormalities commonly persist even during periods of clinical remission, making sleep disturbance a chronic problem for many patients with a history of a mood disorder. Accordingly, assessment for depression is essential for any patient with a sleep complaint, and sleep problems often require specific and potentially ongoing treatment in persons with major depression.

Selected Readings

Fava M, McCall WV, Krystal A, et al. Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. *Biol Psychiatry* 2006;**59**(11):1052–60.

Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;**289**(23):3095–105.

Krystal AD. Psychiatric disorders and sleep. *Neurol Clin* 2012;**30**(4):1389–413.

Le Grande MR, Jackson AC, Murphy BM, Thomason N. Relationship between sleep disturbance, depression and anxiety in the 12 months following a cardiac event. *Psychol Health Med* 2016;**21**(1):52–9.

Murphy MJ, Peterson MJ. Sleep disturbances in depression. *Sleep Med Clin* 2015;**10**:17–23.

Nierenberg AA, Husain MM, Trivedi MH, et al. Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: a STAR*D report. *Psychol Med* 2010;**40**(1):41–50.

Yang H, Sinicropi-Yao L, Chuzi S, et al. Residual sleep disturbance and risk of relapse during the continuation/maintenance phase treatment of major depressive disorder with the selective serotonin reuptake inhibitor fluoxetine. *Ann Gen Psychiatry* 2010;**9**:10.

A complete reference list can be found online at ExpertConsult.com.

Chapter 138

Bipolar Disorder

Allison G. Harvey; Adriane M. Soehner; Daniel J. Buysse

Chapter Highlights

- Different types of sleep and circadian disturbances characterize *mood episodes* (depression, mania, mixed) and *interepisode periods* of bipolar disorder. Sleep disturbances may include insomnia, hypersomnia, reduced sleep need, delayed sleep phase, and irregular sleep patterns. The heterogeneity of these problems in bipolar disorder raises significant challenges for clinical management and research.
- There are several reasons that sleep disturbance may be one important (albeit understudied) mechanism contributing to the mood disorders. Specifically, sleep disturbance is a risk factor for episodes, can contribute to relapse, has an adverse impact on mood regulation, is critical for cognitive functioning, compromises health, and may contribute to substance use comorbidity and suicidality.
- Sleep and circadian rhythms may become perturbed in bipolar disorder because of inadequate, irregular, or inappropriate timing of exposure to light and dark, irregularities in social rhythms, and specific variants in clock gene function.
- Medication treatments for bipolar disorder can produce sedation, and some can also cause insomnia. They also have effects on circadian function.
- Preliminary data suggest that benzodiazepines are associated with worse overall outcome in bipolar disorder. Sedating tricyclic antidepressant drugs should be avoided in the treatment of bipolar depression because of an increased risk of mania.
- Promising psychosocial treatments for sleep and circadian disturbance in bipolar disorder include a combination of cognitive-behavioral therapy for insomnia, interpersonal and social rhythm therapy, chronotherapy, and motivational interviewing.

Bipolar disorder is a common, severe, and chronic psychiatric condition.[1] The lifetime prevalence of the spectrum of bipolar disorders is 2.6% to 6.5%.[2,3] Bipolar I disorder is diagnosed when a patient has experienced at least one manic episode lasting at least 1 week and involving a distinct and abnormal elevated or irritable mood, whereas bipolar II disorder is diagnosed when a patient has experienced at least one episode of major depression and at least one hypomanic episode.[1] The impact of episodes of mania, hypomania, and depression is enormous. People who have been hospitalized as a result of a bipolar episode typically spend 20% of their life in episodes[4] and 50% of their time unwell.[5] Notwithstanding important advances in pharmacologic and nonpharmacologic treatments for bipolar disorder, many patients are seriously symptomatic in the interepisode period despite good adherence and adjunctive interventions, and the risk of relapse is high.

SUBJECTIVE SLEEP COMPLAINTS AND POLYSOMNOGRAPHIC FINDINGS

Clinical Features of Sleep

A range of sleep and circadian disturbances are prominent features of bipolar disorder. Within the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, several types of sleep disturbance are listed as diagnostic criteria for mania, depression, and mixed episodes.[1] Sleep disturbance persists during the interepisode period; up to 70% of adults with bipolar disorder report clinically significant sleep disturbance between episodes.[6] There is also evidence that sleep disturbance can precede the onset of the disorder by several years.[7] Sleep disturbances during mood episodes and interepisode periods of bipolar disorder may include insomnia, hypersomnia, reduced sleep need, delayed sleep phase, and irregular sleep patterns. More specifically, insomnia is highly prevalent in bipolar disorder. As many as 100% of people with bipolar disorder report experiencing insomnia while depressed,[8] and 55% report having insomnia during the interepisode period.[6] Episodes of depression can also be characterized by hypersomnia, or prolonged time in bed with associated excessive sleepiness, although one study carrying out multiple sleep latency testing in this context did not find evidence for sleepiness,[8] and 25% of people endorse symptoms of hypersomnia during the interepisode period.[9,10] Reduced need for sleep is often exhibited during mania and mixed episodes and is characterized by a decrease in total sleep time.[11] Furthermore, people with bipolar disorder are more likely to report circadian dysfunction, such as a delayed sleep phase preference, compared with healthy individuals.[12] Irregular sleep patterns also

characterize bipolar disorder; the mean variability in total sleep time across a week averaged 2.78 hours in a large sample of individuals with bipolar disorder.[13] In interepisode bipolar disorder, lower and more variable sleep efficiency and variability in falling asleep time were related to worse illness course and outcome.[14] Notably, these different types of sleep disturbance can coexist, particularly insomnia and hypersomnia.[9,10] Moreover, in one study, individuals diagnosed with bipolar disorder and individuals at high risk for bipolar disorder both exhibited recurring insomnia, hypersomnia, sensitivity to circadian shifts, and difficulties in waking up relative to a nonpatient group.[15] The heterogeneity of sleep and circadian disturbances in bipolar disorder raises significant clinical management and research challenges.

Polysomnographic Findings

Polysomnographic findings in bipolar depression generally resemble those in unipolar depression but are not consistently observed; the lack of consistent findings is likely related to small sample sizes and considerable heterogeneity among patients. Akin to unipolar depression, patients with bipolar depression tend to experience shortened rapid eye movement (REM) latency relative to healthy adults,[16,17] although this is not a consistent finding.[18] REM density in a bipolar depressed sample was elevated compared with healthy adults and not significantly different from a unipolar depressed group.[17] Reduced non-REM (NREM) stage 3 (N3) has been observed in small samples of bipolar depressed patients relative to healthy adults[19] and at a level equivalent to that of unipolar depressed patients.[20] However, one study found no differences in N3 relative to healthy adults.[18] Finally, in bipolar depressed patients, one investigation reported reduced stage N1 sleep,[18] but most reports have not.[17,19] Within bipolar subtypes, one study observed that depressed bipolar I patients exhibited a trend toward greater fragmentation of REM sleep relative to depressed bipolar II patients.[20] Another study observed no differences between bipolar I and bipolar II patients in NREM and REM sleep parameters, although bipolar II patients demonstrated longer REM latency relative to patients with unipolar depression.[21] Few studies have examined quantitative sleep electroencephalography in bipolar disorder. These studies have reported a trend toward greater sleep spindles in bipolar depressed patients relative to unipolar depressed patients[22] and no difference in delta power among individuals with bipolar depression compared with matched controls ($n = 8$ per group).[23]

Sleep architecture findings during mania are also mixed. Two small studies have shown that adults with mania tend to exhibit shortened REM sleep latency and increased REM density[24,25] relative to healthy nonpatients, although these differences were not observed in another report.[26] One study observed that N1 and N3 percentage was increased in manic patients compared with healthy adults,[27] but another did not report altered NREM sleep in mania.[26]

During the interepisode phase, two studies have found evidence for increased REM density in both unmedicated[28] and medicated bipolar samples.[29] Increased REM density was also observed in a sample of interepisode bipolar disorder patients relative to controls and particularly following a sad mood induction conducted just before bedtime.[30] The latter study also reported a trend toward a shorter first REM episode following the sad mood induction compared with a baseline night in the bipolar disorder group, whereas a trend in the opposite direction was observed for the control group. Talbot et al[30] suggested that these results raise the possibility that the bipolar group may have exhibited an adaptive increase in REM density following a sad mood (as did the control group) but that this attempt to regulate mood during sleep was offset by the shorter length of REM period in the bipolar group relative to the control group. However, one study found no differences in sleep architecture between healthy adults and interepisode bipolar participants.[31] In all three studies, there were no observed differences in NREM sleep in the bipolar samples relative to healthy adults.[28,29,31]

In summary, polysomnographic studies in bipolar disorder are difficult to conduct because of the heterogeneity of clinical features and difficult to interpret because of small sample sizes and medication effects. Most studies indicate that bipolar disorder, like unipolar depression, is characterized by reduced sleep continuity, short REM latency, and increased REM density, with inconsistent findings regarding quantitative electroencephalography.

SLEEP DISTURBANCES AND CORE DOMAINS OF DYSFUNCTION

Figure 138-1 graphically summarizes three pathways, which are discussed in this section, by which sleep disturbance in bipolar disorder may contribute to significant impairment among people with bipolar disorder.

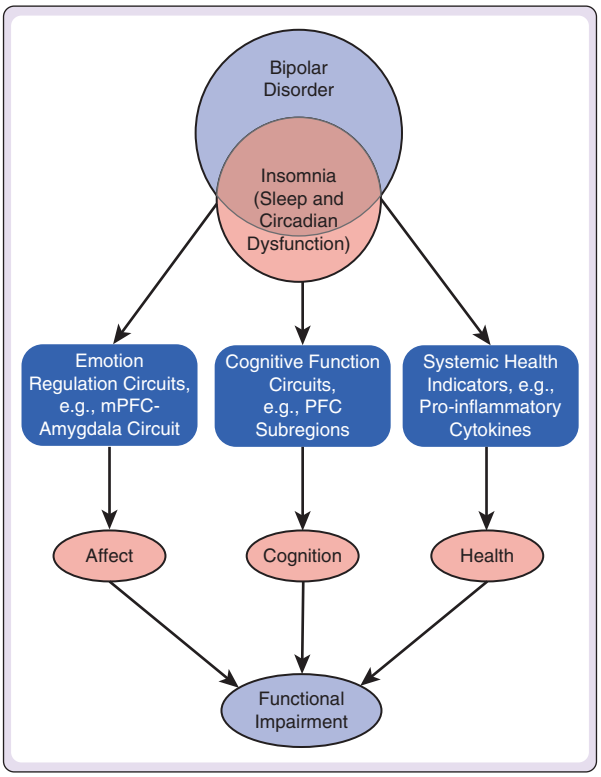


Figure 138-1 Conceptual model depicting sleep disturbance in bipolar disorder as a pathway contributing to poor outcomes (*ovals*) through specific neural and systemic mechanisms (*rectangles*). PFC, Prefrontal cortex; mPFC, medial prefrontal cortex.

Affect

Affective dysregulation in the form of alterations in the processes and neural circuits underlying mood is a hallmark feature of bipolar disorder. Mounting evidence suggests that sleep disturbance contributes to affective dysregulation. Sleep disturbance is a common prodrome, or precursor, of relapse.[32] Experimentally induced sleep deprivation is associated with the onset of hypomania or mania in a subset of patients.[33–35] Patients with short sleep exhibit greater mood symptom severity, poorer daytime functioning, and lower life satisfaction compared with those with longer sleep times.[13] Moreover, shorter total sleep time is associated with increased mania and depression severity during 12 months.[36] In a 7-day diary study, greater total wake time was associated with next-day morning negative mood in individuals with bipolar disorder, and evening negative mood was associated with subsequent total wake time in both bipolar disorder and insomnia.[37]

Few investigations have examined the relationship between sleep architecture and mood in bipolar disorder. In one study of interepisode bipolar disorder, sleep architecture was not correlated with concurrent mood symptoms.[29] However, greater REM density was correlated with more severe depressive symptoms and greater functional impairment at 3 months, whereas prolonged duration of the first REM period and a greater amount of slow wave sleep were positively correlated with manic symptoms and impairment at 3 months. In addition, the amount of N2 sleep was negatively correlated with manic symptoms and impairment at 3 months. In manic patients, Hudson et al[25] found that time spent asleep and REM percentage inversely correlated with mania severity.

The observed associations between sleep and affect in bipolar disorder are not surprising given that in studies of healthy adults without psychiatric illness, sleep has a critical mood-regulatory function. Neuroimaging investigations indicate that sleep loss impairs top-down inhibitory control exerted by medial prefrontal cortex on amygdala and reward circuitry.[38,39] Circuits involved in emotion regulation and sleep regulation also interact in bidirectional ways.[40] Moreover, abnormalities in REM sleep across phases of bipolar illness may be another pathway by which sleep contributes to mood and emotion regulation difficulty in bipolar disorder.[41] Indeed, there is evidence that REM sleep is important for the encoding, regulation, and consolidation of emotional memories.[42]

Health

Medical morbidity and premature mortality are characteristic of bipolar disorder.[43] Indeed, bipolar disorder confers a 1.5- to 2.5-fold increase in risk for mortality from cardiovascular disease (CVD), making CVD the leading cause of death of individuals with bipolar disorder.[44] CVD risk factors, such as hypertension, obesity, and diabetes mellitus, occur with greater frequency in bipolar disorder relative to the general population.[45,46] The accumulation of CVD risk in bipolar disorder might be attributed to long-term psychopharmacologic treatment or unhealthy lifestyles (inadequate exercise, poor diet). However, insomnia may also contribute to poor health[47,48] both indirectly and directly through increased mechanisms, such as increased proinflammatory cytokine activity.[49,50]

Cognition

Cognitive dysfunction in the form of alterations in the processes and neural circuits underlying a broad range of cognitive processes, such as attention, memory, and problem solving, is characteristic of bipolar disorder[51] during acute and interepisode periods and is associated with poorer functional and psychosocial outcome.[52] Given that there are dramatic adverse effects of sleep deprivation in healthy adults on a range of cognitive tasks[53,54] associated with various subregions of the prefrontal cortex[55,56] and that insomnia is also characterized by subtle cognitive deficits,[57] it seems likely that sleep disturbance contributes to cognitive dysfunction in bipolar disorder, although this hypothesis remains to be formally tested.

Additional Domains

Sleep disturbances are risk factors for substance use problems[58] and suicide[59] among individuals *without* bipolar disorder. Because substance use and suicide are particularly prevalent among individuals with bipolar disorder,[60,61] a hypothesis to test in future research is that sleep disturbance contributes to these problems specifically among individuals with bipolar disorder. However, given the relatively low prevalence of bipolar disorder and the actual rate of suicide, it may be most useful to examine sleep-suicide relationships through intermediate mechanisms, such as alterations in executive and reward function.[62,63]

PATHOPHYSIOLOGY OF SLEEP DISTURBANCE IN BIPOLAR DISORDER

Inadequate, Irregular, or Inappropriate Timing of Exposure to Light and Dark

The mood and activity symptoms of bipolar disorder are often cyclic, suggesting that even at the symptom level, a relationship with endogenous rhythmic processes, including circadian rhythms, is core to the disorder. Behaviors that influence the amount of exposure to light and dark are also likely to be important to both sleep and circadian functioning as well as to mood. Circadian processes, including circadian sleep-wake propensity (process C), are controlled by the suprachiasmatic nuclei (SCN) in the hypothalamus. The SCN receive environmental light and dark information. The amount and timing of light influence the increased release or suppression of melatonin. This is because one of the important downstream projections of the SCN is the pineal gland, which secretes melatonin. Secretion of melatonin peaks at night, is suppressed by light, and is practically nonexistent during the day. As melatonin release contributes to sleepiness, exposure to light during this period may suppress melatonin onset. Accordingly, inadequate, irregular, or inappropriate timing of exposure to light and dark, related to the selection and timing of activities, may have an adverse impact on sleep (including the light from computer screens and televisions). There is accruing evidence for the impact of and possible sensitivity to light and dark in patients with mood disorders.[64] This is perhaps not surprising given that the behavioral changes that can be associated with mood disorders may alter exposure to light and dark (inadequate dark during mania, inadequate light during depression).

The endogenous period generated in the SCN is close to but generally not equal to 24 hours. The process by which the

circadian clock is kept in appropriate phase with seasonally shifting astronomical day length is called *entrainment*. In other words, an important feature of the circadian system is its fundamentally open nature. Entrainment occurs through *zeitgebers*, which are environmental events that can affect the phase and period of the clock. The primary zeitgeber in most species is the daily alteration of light and dark. Hence, the selection and timing of daily activities will have an impact on exposure to light and dark, which in turn will influence sleep and circadian rhythms. Indeed, the selection and timing of engagement with activities may be important contributors to mood disorders, a topic to which we return later in the discussion of interpersonal and social rhythm therapy (IPSRT).

Unstable Social Rhythms

Other powerful *nonphotic* cues that influence the SCN include social cues, such as the timing of social interactions and meals,[65] which can be described as social rhythms. For example, higher daily regularity in social rhythms (e.g., social interactions or mealtimes occurring at about the same time each day) is associated with a stronger endogenous temperature rhythm, suggesting stronger overall circadian functioning.[66] Moreover, more regular social rhythms are associated with better subjective sleep quality.[66] Given that depression is associated with lower social support[67] and more social strain,[68] it is possible that disruption to social rhythms is a risk factor for mood disorders. Again, this issue is expanded on later.

Circadian Genes

The molecular mechanisms of the circadian clock in mammals are well understood. Specifically, cells in the SCN generate self-sustained rhythmicity through an autoregulatory transcription-translation feedback loop regulating expression of the *CLOCK*, *Period* (*Per1, Per2, Per3*), *cryptochrome* (*Cry1, Cry2*), *TIM*, *BMal1*, *NPAS*, *DEC1*, and *DEC2* genes.[69] Several studies have reported links between the circadian genes and the mood disorders. For example, bipolar disorder and treatment response have been associated with mutations or polymorphisms in *TIM*, *Clock* (311 T to C) and *BMal1*,[70] and *CLOCK* (homozygous for the C allele). Clock genes have also been implicated in other mood disorders: *Per2*, *NPAS2*, and *Bmal1* are associated with seasonal affective disorder,[71] and *Per1*, *CLOCK*, and *Baml1* mRNA levels are associated with major depressive disorder.[72]

Although this line of research is in its infancy and is often limited by small samples, the findings are consistent with the growing overall picture of sleep and circadian vulnerability as a feature of the mood disorders. Importantly, and consistent with the proposal that psychiatric disorders are likely to be associated with multiple genes of small effect,[73] most of the associations reported are modest, and failures to replicate are common.[74]

TREATMENT OF SLEEP DISTURBANCE IN BIPOLAR DISORDER

Sleep disturbances in bipolar disorder can be treated with pharmacologic, behavioral, and somatic (e.g., bright light) interventions. Pharmacologic treatments can be divided into those aimed at treating the core features of bipolar disorder and those aimed at treating sleep disturbances associated with bipolar disorder. However, this distinction is imperfect because medications that target mood cycling can also have important effects on sleep and circadian function, and medications that target sleep and circadian disturbances can affect mood.

Current treatment guidelines (summarized in[75]) recommend the following medications for first-line treatment of *depression* in bipolar I disorder: quetiapine, olanzapine, olanzapine-fluoxetine combination, lamotrigine, lithium, and valproate. Selective serotonin reuptake inhibitor (SSRI) antidepressants or bupropion may also be used in combination with an antimanic agent (e.g., lithium, valproate). Second-line treatments include combinations of these various agents as well as consideration of treatments such as antidepressants, electroconvulsive therapy, and psychotherapy. Notably, each of the first-line treatments can produce sedation, and some of them (e.g., lamotrigine) can also cause insomnia. Thus, the time of administration may be an important treatment consideration. For instance, medications such as quetiapine and olanzapine may be given at bedtime to help treat insomnia, and lithium may be given in divided doses but with the largest dose at night. In clinical trials, atypical antipsychotic drugs such as quetiapine have been associated with not only relief of bipolar depression but also improved sleep.[76]

The sedative and sleep-promoting effects of medications used for bipolar depression result from a variety of effects on neurotransmitter and receptor function. Antipsychotic drugs antagonize dopamine DA_2 and DA_3 receptors and histamine receptors. Atypical antipsychotic drugs also have significant serotonin receptor *antagonist* effects, particularly at $5\text{-}HT_{2A/2C}$ and $5\text{-}HT_7$ receptors, and serotonin $5\text{-}HT_{1A}$ *agonist* effects.[77] Lithium, valproate, and lamotrigine reduce dopaminergic and glutamatergic neurotransmission and increase GABAergic neurotransmission, likely through second messenger pathways including adenylate cyclase and phosphoinositide pathways.[78] Lamotrigine and valproate also block voltage-gated sodium channels and calcium channels.[79]

Current guidelines (summarized in[80]) recommend the following medications for first-line treatment of mania in bipolar disorder: lithium, valproate, atypical antipsychotic drugs, haloperidol, carbamazepine, and combinations of lithium or valproate with antipsychotic drugs. Second-line treatments involve various combinations of these drugs, electroconvulsive therapy, and other antipsychotic and anticonvulsant drugs. As is the case with treatment of bipolar depression, the timing of sedating medications may be used to help manage the sleep disturbance associated with acute mania.

In addition to their effects on sleep and wakefulness, medications used to treat bipolar disorder also have effects on circadian function (see[81–83] for reviews). For instance, lithium delays circadian rhythm phase and lengthens circadian rhythm period in animals and humans and can increase PER2 rhythm amplitude. This effect may relate to lithium's inhibition of glycogen synthase kinase 3β, which interacts with multiple proteins involved in the transcription-translation feedback loop of the molecular clock. Likewise, valproate has been observed to increase the amplitude of clock gene protein expression, to shift circadian phase, and to lengthen circadian period. SSRI antidepressants, on the other hand, can advance circadian phase, shorten circadian period, and alter phase shifts in response to light or other stimuli. These effects may be modulated by innervation of the SCN by serotonergic neurons from the raphe nuclei. Some of the circadian effects described for SSRIs and mood

stabilizers, like lithium, appear to be in opposite directions. This may have a clinical correlate in the ability of SSRIs to precipitate mania among patients with bipolar disorder. Finally, the antidepressant agomelatine, a combined melatonin MT_1 and MT_2 receptor agonist and serotonin $5\text{-}HT_{2C}$ receptor antagonist, has sleep and circadian effects. These include increased slow wave sleep, synchronization of free-running circadian rhythms, and phase-shifting effects similar to those observed with melatonin.

Medications used to treat insomnia outside of bipolar disorder, such as benzodiazepine receptor agonists and sedating antidepressants, may have mood effects among patients with bipolar disorder. Zolpidem was the most commonly prescribed medication for insomnia in bipolar disorder in one chart review,[84] but case reports on benzodiazepines for insomnia in bipolar disorder have not been encouraging.[85,86] Indeed, drawing from the large sample who participated in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), Perlis et al[87] reported that benzodiazepines were associated with worse overall outcome. Likewise, systematic reviews of the treatment of nonbipolar depression suggest that concomitant treatment with benzodiazepine receptor agonists and antidepressants may improve initial treatment adherence and response but not longer term treatment response.[88] Sedating tricyclic antidepressant drugs, often used as second-line agents for the treatment of primary insomnia, should be avoided in the treatment of bipolar depression because of an increased risk of mania.[89] A case report on gabapentin[90] and open trials of medications targeting circadian rhythms[91,92] are more promising but preliminary.

There may be some advantages to treating sleep disturbance in bipolar disorder nonpharmacologically because there are fewer side effects and interactions with concurrent mood-stabilizing treatments and some insomnia medications pose a risk of abuse, a concern given the prevalence of comorbid substance use in bipolar disorder.[93] Existing evidence-based nonpharmacologic interventions for bipolar disorder include psychoeducation,[94] prodrome monitoring,[95] IPSRT,[96] family therapy,[97] and cognitive-behavioral therapy (CBT) administered individually[95] or in groups[98] as well as combination approaches.[99] These treatments, several of which include some attention to sleep disturbances, reduce relapse among adults with bipolar disorder. However, most of the trials of these treatments have not included sleep as a primary emphasis nor have they included sleep outcome measures or incorporated specific insomnia treatment techniques.

In light of the ongoing need for novel, effective, well-tolerated treatments for bipolar disorder, we recently developed and tested a bipolar disorder–specific modification of CBT for insomnia (CBTI-BP). We present the details of this novel treatment to illustrate how sleep and circadian principles can be incorporated into the treatment of bipolar disorder. The modifications were focused on improving safety and targeting the unique features of sleep in bipolar disorder by integrating elements from IPSRT, chronotherapy, and motivational interviewing. CBT for insomnia (CBT-I) was selected as the basis for the treatment because (1) substantial evidence demonstrates the efficacy of CBT-I,[100] even for insomnia comorbid with depression,[101] post-traumatic stress disorder,[102] and schizophrenia[103]; (2) when insomnia is comorbid with another psychiatric disorder, the symptoms associated with the other psychiatric disorder can also improve after CBT-I[101,103]; (3) previous treatments for bipolar disorder have not been informed by the principles underlying CBT-I[14]; and (4) CBT-I improves sleep efficiency, reduces night-to-night variability, and improves sleep-onset latency, all of which are problematic for patients with bipolar disorder.[14] Elements were added from IPSRT[96] to regularize daily social rhythms (i.e., mealtimes, social contact, exercise) and to build on regular bed and wake-up times traditionally emphasized in CBT-I. Promising data on chronotherapy, such as dark therapy and light therapy,[64] were integrated by adding a 30- to 60-minute wind-down period in dim light and exposure to light on waking within an individualized wake-up routine. However, the optimal duration and intensity of morning light have not been identified empirically. Motivational interviewing[104] was incorporated in every session, given the challenges inherent to behavior change and recognizing the rewarding nature of many sleep-incompatible behaviors. The targets of CBTI-BP along with the treatment elements are summarized in Table 138-1.

The unique features of bipolar disorder necessitated modifications to CBT-I. Specifically, the sleep restriction component of CBT-I was limited to mitigate the risk of sleep loss triggering mania,[33–35] and the nighttime awakening procedures of stimulus control were limited to reduce potential engagement in goal or reward pursuit[105] during prolonged awakenings. The highly variable sleep-wake schedules, tendency toward delayed bedtimes, sensitivity to light,[106–108] sleep inertia, and daytime sleepiness present in many bipolar patients were also areas of emphasis in treatment. Trials of CBT-I in primary insomnia have shown improvements in subclinical mood symptoms[100] but have not determined whether CBT-I can stabilize mood in bipolar disorder. The bipolar disorder treatments have not examined efficacy *on sleep–specific symptoms*.

In a pilot investigation of this approach alongside routine psychiatric care, interepisode bipolar disorder I participants with insomnia were randomly allocated to CBTI-BP ($n = 30$) or to psychoeducation (PE; $n = 28$) as a comparison condition.[109] Outcomes were assessed at baseline, at the end of eight sessions of treatment, and 6 months later. The results indicated that during the 6-month follow-up, the CBTI-BP group had fewer days in a bipolar episode relative to the PE group (3.3 days vs. 25.5 days). The CBTI-BP group also experienced a significantly lower hypomania or mania relapse rate (4.6% vs. 31.6%) and exhibited a trend toward lower overall mood episode relapse rate (13.6% vs. 42.1%) compared with the PE group. Relative to PE, CBTI-BP also reduced insomnia severity and led to higher rates of insomnia remission after treatment and at 6 months. Notably, this approach appears to be safe and tolerable.[110] These results underscore the value of treatments that aim to improve sleep and to reduce circadian disturbances in bipolar disorder and support the proposal that disturbed sleep constitutes a mechanism that substantively contributes to bipolar disorder.

CONCLUSIONS

Thus far, research evidence consistently supports a link among sleep, mood, impairment, and quality of life in bipolar disorder. However, several important gaps in knowledge remain. At a foundational level, more research is needed to

Table 138-1 Summary of a Bipolar Disorder–Specific Modification of Cognitive-Behavioral Therapy for Insomnia (CBTI-BP)

Targets of CBTI-BP	CBTI-BP Treatment Approaches
Building motivation to change and/or not valuing sleep	Sleep and circadian education and motivational interviewing
Irregular bed and wake-up times	Regularize night-to-night bed and wake times
Irregular daytime rhythms	Interpersonal and social rhythms therapy to regularize times for meals, socializing, exercise, and other daytime rhythms
Difficulty getting to sleep, wakings in the night, or waking too early in the morning	Stimulus control, including establishing a wind-down routine, and sleep restriction Dim light exposure at bedtime and during awakenings
Difficulty waking up	Devise a wake-up routine and daytime activity scheduling to establish activities to get up for Bright light exposure on waking
Daytime impairment	Education and behavioral experiments to test contributors to daytime impairment, including the identification of "energy-sapping" and "energy-generating" activities
Unhelpful beliefs about sleep	Education, cognitive therapy, and behavioral experiments to test unhelpful beliefs
Poor sleep efficiency	Stimulus control and sleep restriction
Too much time in bed	Sleep restriction
Delayed phase	Advancing bedtimes by 30–60 min/wk
Sleep-related worry	Cognitive therapy to teach methods to manage thoughts that interfere with sleep
Maintenance of behavior change	Relapse prevention

understand the prevalence of sleep disturbances (insomnia, hypersomnia, delayed sleep phase, reduced sleep need) across the life course of bipolar disorder, in the bipolar disorder subtypes, and during mixed mood episodes (i.e., when diagnostic criteria are met for *both* a manic episode and an episode of depression). In addition, the nature and role of sleep architecture in the course of bipolar disorder remain understudied, with the majority of polysomnographic investigations in bipolar disorder having been conducted approximately two to three decades ago. It will also be crucial to identify causal and potentially bidirectional pathways between sleep disturbance and mood in bipolar disorder. There is robust behavioral and neuroimaging evidence among healthy nonpatient samples that sleep deprivation undermines emotion regulation the following day. Further neuroimaging studies of bipolar patients may elucidate the mechanisms by which sleep and circadian symptoms and treatments influence mood, cognition, and activity in bipolar patients.

Given the high rates of sleep disturbance in bipolar disorder, it is surprising that few controlled trials of medication or psychological interventions for sleep have been conducted in this population. There is exciting preliminary evidence that treating sleep disturbance with adaptations of CBT-I can reduce relapse rates, supporting the notion that treating sleep can improve the course of illness. However, certain treatment complexities in bipolar disorder need to be better understood. For example, phase of bipolar illness may affect selection of treatment approaches. During manic or mixed episodes, a pharmacologic approach for sleep may be most appropriate to aid in fully stabilizing a patient. Yet, during bipolar depression or the interepisode period, psychological approaches may be preferable because of the reduced risk of negative side effects. Larger controlled trials testing psychological and pharmacologic therapies in bipolar disorder hold promise not only for improving sleep but also for increasing mood stability and enhancing quality of life.

CLINICAL PEARLS

- Disturbed sleep and circadian rhythms constitute a mechanism that substantively contributes to symptoms and functional impairments in bipolar disorder.
- Treatments that aim to improve sleep and circadian disturbances in bipolar disorder should be provided as part of front-line treatment.
- Behavioral approaches emphasizing the stabilization of bed and wake times, along with dim light conditions at night and bright light conditions on waking, appear to be helpful for reducing impairment and vulnerability to relapse.

SUMMARY

Sleep and circadian dysfunctions are important mechanisms contributing to the onset and maintenance of symptoms and dysfunction in bipolar disorder. Sleep disturbance is a risk factor for episodes, can contribute to relapse, has an adverse impact on emotion regulation, compromises health, is critical for cognitive functioning, and may contribute to substance use comorbidity and suicidality. Factors driving sleep and

circadian disturbance in bipolar disorder may include inadequate, irregular, or inappropriate timing of exposure to light and dark, irregularities in social rhythms, and specific variants in clock gene function. The accumulated evidence has informed a range of novel, powerful, simple, and inexpensive treatments with potential for significant improvements in quality of life, decreasing length and severity of episodes and reducing the risk of subsequent episodes.

Selected Readings

Abreu T, Bragança M. The bipolarity of light and dark: a review on bipolar disorder and circadian cycles. *J Affect Disord* 2015;**185**:219–29.

Cretu JB, Culver JL, Goffin KC, et al. Sleep, residual mood symptoms, and time to relapse in recovered patients with bipolar disorder. *J Affect Disord* 2015;**190**:162–6.

Frank E. *Treating bipolar disorder: a clinician's guide to interpersonal and social rhythm therapy*. New York: Guilford Press; 2005.

Harvey AG, Soehner AM, Kaplan KA, et al. Treating insomnia improves sleep, mood and functioning in bipolar disorder: a pilot randomized controlled trial. *J Consult Clin Psychol* 2015;**83**(3):564–77.

Kaplan KA, Harvey AG. Behavioral treatment of insomnia in bipolar disorder. *Am J Psychiatry* 2013;**170**:716–20.

Malhi GS, Tanious M, Das P, et al. Potential mechanisms of action of lithium in bipolar disorder. *CNS Drugs* 2013;**27**:135–53.

McClung CA. Circadian rhythms and mood regulation: insights from preclinical models. *Eur Neuropsychopharmacol* 2011;**21**:S683–93.

Wirz-Justice A, Benedetti F, Terman M. *Chronotherapeutics for affective disorders: a clinician's manual for light & wake therapy*. Basel: Karger; 2009.

A complete reference list can be found online at ExpertConsult.com.

Chapter
139

Schizophrenia

Kathleen L. Benson; Irwin Feinberg

Chapter Highlights

- Schizophrenia, perhaps the most devastating neuropsychiatric illness, is associated with immense human suffering and economic cost. Currently schizophrenia can neither be prevented nor cured. No specific or diagnostic laboratory abnormality has yet been identified; consequently, schizophrenia remains a clinical diagnosis. This chapter discusses the complex etiology and pathology associated with schizophrenia. Included in the range of pathologic findings are abnormal sleep patterns.
- Subjective complaints of poor sleep quality have been validated by objective polysomnographic (PSG) assessments. This chapter reviews these PSG assessments. These studies include overnight measures of sleep maintenance and sleep staging using conventional measures of sleep scoring as well as overnight measures of sleep-related brain wave activity. Circadian rhythm abnormalities are also reviewed.
- Most patients with schizophrenia require long-term treatment with antipsychotic medications. These agents are classified as first-generation or second-generation antipsychotics. They differ in their receptor binding profiles, clinical efficacy, and range of side effects. This chapter addresses the effects of antipsychotic agents on measures of sleep maintenance and sleep stages.
- Schizophrenic patients also suffer from a range of dyssomnias including poor sleep hygiene, irregular sleep-wake patterns, parasomnias, obstructive sleep apnea, restless legs syndrome, and periodic limb movements during sleep. This chapter describes how antipsychotics agents may induce or exacerbate dyssomnias.

In many ways, the normal dream experience is similar to psychosis. Hallucinations, perceptual distortions, bizarre thinking, and temporary delusions can intermingle with more normal thought and perceptual processes. Consequently, the discovery of rapid eye movement (REM) sleep and its associated dream reports[1] not only ushered in the modern era of sleep research but also stimulated many of the seminal studies of the sleep of schizophrenics. These studies[2-5] explored the hypothesis that the pathogenesis of schizophrenia might rest with REM sleep abnormalities or, more directly, with the intrusion of the dream state into waking. These early polysomnographic (PSG) studies of sleep in schizophrenic patients found no gross abnormalities of REM sleep or any evidence of an intrusion of REM sleep into wakefulness. However, in subsequent years, a large body of research revealed other abnormalities more frequently present in the sleep patterns in schizophrenia.

Schizophrenia is associated with immense human suffering and economic cost and is often accompanied by disturbances of sleep. This chapter provides an overview of the clinical features of schizophrenia and associated risk factors and neuropathology. It also describes the sleep abnormalities frequently encountered in schizophrenia and their neurobiologic correlates. Finally, we address the issues surrounding antipsychotic medications and their side effects, particularly those associated with the development of comorbid dyssomnias.

EPIDEMIOLOGY AND RISK FACTORS

Epidemiologic studies of schizophrenia suggest a median lifetime prevalence of 0.7% with a male-to-female incidence ratio of 1.4:1 worldwide.[6] The age of onset is usually in the second decade of life and the disorder seems to begin earlier in men than in women.

Although the etiology of schizophrenia is enigmatic and its elucidation a continuing challenge, several factors have been associated with an increased likelihood for disease development. First, a genetic predisposition greatly elevates the risk for developing schizophrenia. Meta-analysis of twin studies indicates that the heritability of schizophrenia is high, roughly 81%.[7] Broadly speaking, the prevailing genetic models encompass two scenarios: first, an additive polygenic model consisting of multiple combinations of common mutations that cumulatively increase the risk for schizophrenia; and second, a heterogeneity model postulating highly penetrant, individually rare mutations that are associated with high risk for disease development.[8] A recent genome-wide association study identified 108 schizophrenia-associated loci.[9] The additive polygenic model would suggest that multiple forms or phenotypes of schizophrenia might exist. The twin study meta-analysis[7] also demonstrated significant shared or common environmental effects influencing liability to schizophrenia. The environments of twins tend to increasingly

diverge during childhood, adolescence, and adulthood but are similar or shared in utero and in the postnatal period. Several environmental factors (from the prenatal period through adolescence) may increase the liability to schizophrenia. Epigenetic mechanisms, largely through modulation of gene expression, could be the vehicle that mediates these gene-environment interactions.[10] In conclusion, multiple genetic variants in combination with a range of environmental factors suggest that schizophrenia is best viewed less as a unitary disorder but more as a family of disorders.[11]

DIAGNOSIS

The term *schizophrenia* derives from Bleuler's description of the splitting or disintegration of normal thought processes.[12] His belief that cognitive impairment or thought disorder is the defining symptom of schizophrenia shaped the course of diagnostic criteria developed during the 20th century. Current criteria for the diagnosis of schizophrenia are defined in the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5; Box 139-1).[13] Black and Andreasen[14] describe schizophrenia as a clinical diagnosis that has no specific laboratory abnormalities that are diagnostic. Schizophrenia is also, in large part, a diagnosis of exclusion, which requires eliminating psychotic disturbances that might be produced by a wide range of medical, psychiatric, and substance-abuse disorders. The number and diversity of symptoms that can arise in individual patients can challenge the clinical diagnosis.

The characteristic symptoms of schizophrenia[13] (see Box 139-1) are organized into two main categories: positive and negative symptoms. According to DSM-5, "positive symptoms appear to reflect an excess or distortion of normal functions, whereas the negative symptoms appear to reflect a diminution or loss of normal functions." Positive symptoms can be further subdivided into a psychotic dimension that includes hallucinations and delusions and a disorganization dimension that includes disorganized speech and behavior. Negative symptoms include flattening of affect, avolition, and poverty of speech. The second diagnostic criterion[13] (see Box 139-1) reflects a marked deterioration in occupational and social functioning.

The sleep abnormalities found in schizophrenia lack diagnostic specificity. Consequently, they do not reliably differentiate schizophrenia from other psychiatric disorders. It is unlikely, therefore, that a sleep clinic would be asked to diagnose schizophrenia. However, the sleep clinic could help differentiate schizophrenia from narcolepsy, which can manifest with a strong hallucinatory component.[15] Given the high frequency with which sleep problems occur in patients with schizophrenia, sleep clinicians may also be faced with managing sleep problems in these patients.

Box 139-1 DSM-5 DIAGNOSTIC CRITERIA FOR SCHIZOPHRENIA

Diagnostic Criteria

A. Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated). At least one of these must be (1), (2), or (3):
   1. Delusions
   2. Hallucinations
   3. Disorganized speech (e.g., frequent derailment or incoherence)
   4. Grossly disorganized or catatonic behavior
   5. Negative symptoms (i.e., diminished emotional expression or avolition)

B. For a significant portion of the time since the onset of the disturbance, level of functioning in one or more major areas, such as work, interpersonal relations, or self-care, is markedly below the level achieved before the onset (or when the onset is in childhood or adolescence, there is failure to achieve expected level of interpersonal, academic, or occupational functioning).

C. Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or by two or more symptoms listed in criterion A presented in an attenuated form (e.g., odd beliefs, unusual perceptual experiences).

D. Schizoaffective disorder and depressive or bipolar disorder with psychotic features have been ruled out because either (1) no major depressive or manic episodes have occurred concurrently with the active-phase symptoms, or (2) if mood episodes have occurred during active-phase symptoms, they have been present for a minority of the total duration of the active and residual periods of the illness.

E. The disturbance is not attributable to the physiologic effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.

F. If there is a history of autistic disorder or another pervasive developmental disorder, the additional diagnosis of schizophrenia is made only if prominent delusions or hallucinations are also present for at least 1 month (or less if successfully treated).

DSM-5, *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition. Modified from the American Psychiatric Association: *Diagnostic and statistical manual of mental disorders*. 5th ed. Arlington, VA: American Psychiatric Press; 2013. Copyright 2013 American Psychiatric Association.

PATHOLOGY

Although the etiology of schizophrenia is poorly understood, accumulating evidence reveals a wide range of brain abnormalities.[14] Brain structural abnormalities have been found in postmortem studies and in living subjects by computed tomography and magnetic resonance imaging.[13,14] Regarding the latter, structural dysmorphologies have included enlarged lateral and third ventricles; loss of total gray matter, frontal, and temporal lobe volume; and a reduction in total brain size. These findings seem to be present at the onset of illness and do not seem to be the result of progressive degeneration; however, most findings are nonspecific and are observed in other psychiatric disorders. Functional imaging studies using positron emission tomography or regional cerebral blood flow have reported decreased metabolism in the frontal cortex (hypofrontality) and left hemisphere dysfunction.

Abnormalities of neurotransmitter systems have also been extensively investigated. For many years, the prevailing theory of schizophrenia has centered on the dopamine (DA) system. The DA hypothesis of schizophrenia derived from two

observations. First, the potency of standard antipsychotic medication correlates with the amount of D_2 receptor blockade. Second, drugs such as amphetamines, which enhance DA activity, can cause a psychosis that mimics paranoid schizophrenia and can exacerbate schizophrenic symptoms. The hypothesis holds that psychotic symptoms such as hallucinations and delusions are associated with hyperactivity of the mesolimbic DA system. Coupled with or independent of DA dysfunction, an abnormality of the excitatory neurotransmitter glutamate may also play a role in the pathophysiology of schizophrenia.[16,17] Agonists of metabotropic glutamate receptors are currently being investigated as treatments for schizophrenia.[18] Metabotropic glutamate receptor agonists actually decrease brain excitability. They also have profound effects on the sleep electroencephalogram (EEG), diminishing REM sleep and non–rapid eye movement (NREM) fast frequencies in the rat.[19] Serotonin (5-hydroxytrypamine [5-HT]) and norepinephrine have also been associated with the pathophysiology of schizophrenia because the potency of second-generation antipsychotics has been linked to 5-HT and α-adrenergic receptor blockade. Finally, cortical and subcortical abnormalities of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) may have a potential role in the pathophysiology of schizophrenia.[20,21]

Additionally, immune system abnormalities with associated neuroinflammation may play a role not only in elevating the risk for schizophrenia but also in schizophrenia-related brain dysfunction after the onset of the disease. Most research has focused on early life exposure to infection, the development and effects of proinflammatory cytokines, and immune-related genes.[22,23] Some investigators believe that schizophrenia is an autoimmune disease triggered by maternal antibodies to viral infection.[24]

Because no discrete pathologic abnormality has emerged as an etiologic factor, it has been proposed that schizophrenia might be an abnormality of neuronal connectivity[25] or of integrative neuronal circuits.[26-29] Failures of communication within and between neuronal substrates are consistent with the current prevailing view that schizophrenia is a neurodevelopmental disorder.[30,31] Although abnormal events can occur early in development (prenatal or perinatal), maturational abnormalities can appear during the second decade of life[32] or even into middle age.[33]

CLINICAL COURSE AND PREVENTION

In the prevailing pathophysiologic model, schizophrenia is viewed as a neurodevelopmental disorder with onset typically (but not always) in late adolescence. The onset may be abrupt or insidious—that is, beginning with a prodromal phase characterized by subtle changes in behavior, mild thought disorder, and social withdrawal. The prodromal phase is followed by an active phase marked by positive psychotic symptoms such as hallucinations and delusions. This phase can wax and wane, but some degree of psychoticism usually persists during the waning or residual phase. Positive symptoms (relapse or acute exacerbation) can recur episodically. Over the course of the illness, positive psychotic symptoms can gradually decline; in contrast, negative symptoms such as affective flattening and avolition tend to increase with the progression of the disease.

For about 50% of patients, the onset of the illness is both progressive and insidious; for the remainder, the onset is acute, with little or no prodromal syndrome. Also, for about 50% of patients, the course of the illness is continuous; for others the course is marked by episodic flare-ups. The likelihood of relapse has been associated with stressful life events, a critical and hostile family environment, and drug abuse.

If a positive outcome is defined as the absence of psychotic symptoms and normal levels of social functioning, long-term follow-up studies suggest that 20% to 30% of patients have a full recovery. Another 20% to 30% of patients recover to levels where they can function occupationally and socially, but they have residual symptoms. Approximately 50% continue to show moderate to severe dysfunction requiring numerous outpatient interventions or rehospitalization with each relapse. Approximately one fifth require long-term institutionalization. Although antipsychotics can lower the number of hospitalizations, antipsychotic treatment does not fundamentally alter these outcomes. Better outcomes are associated with acute onset, episodic course, female sex, and lack of family history of schizophrenia.

The course of illness in schizophrenia can include dysphoria or depressive episodes, particularly early in the course. Relative to other major psychiatric disorders, schizophrenic patients tend to marry less, are more likely found in an institutional setting, and have poorer occupational functioning. Schizophrenic patients also have a twofold to threefold increased risk for dying; this higher mortality rate encompasses both suicide as well as physical illness.[6] One intriguing question, still unresolved, is why the incidence of schizophrenia has not declined over the past century despite both greater mortality and decreased fecundity in this illness. One possible explanation is that the illness occurs as a stochastic error in the genetic development of the enormously complex human nervous system.[34]

Schizophrenia is certainly among the most devastating psychiatric illness in its human and economic costs. At present, we can neither cure this disorder nor prevent its occurrence. Most schizophrenic patients, after an early onset, will endure lifelong mental disability as well as social and economic marginalization.

SLEEP-RELATED FEATURES

Subjective Sleep Complaints

With the onset of psychotic symptoms, and with each subsequent relapse, sleep is usually markedly impaired. The sleep of schizophrenic patients who are in a state of psychotic agitation usually, but not invariably, is manifested by prolonged periods of sleeplessness. In times of less severe psychotic agitation, sleep is often characterized by pronounced insomnia—long sleep-onset latencies, reduced total sleep time, and sleep fragmented by bouts of waking. Recurrence or exacerbation of symptoms is often heralded by increasing insomnia. Even among clinically stable, medicated patients with schizophrenia, ongoing subjective sleep disturbance is common, particularly early and middle insomnia.[35,36] There may also be a reversal of sleep and wake so that the patient sleeps during the day and remains awake at night. Subjective complaints of poor sleep quality are correlated with sleep-wake reversals.[37]

Schizophrenic patients also complain of poor sleep quality, including restlessness and agitation, disturbing hypnagogic hallucinations, and nightmares. Subjectively assessed poor sleep quality is predictive of self-assessed poor quality of life

and impaired coping skills.[38,39] Although there are no systematic studies, anecdotal clinical reports suggest that alcohol and substance abuse can disturb sleep and cause the patient to relapse. Alternatively, use of these drugs may be an attempt to self-medicate and attenuate the psychic misery of this illness.

Polysomnographic Features

PSG studies have provided a comprehensive and objective description of the range of dyssomnias found in schizophrenia, and they have been broadly consistent with subjective complaints. However, these PSG studies have, on occasion, produced discrepant findings, owing perhaps to differences in protocol design, composition of control groups, sample size, inclusion criteria (e.g., age, medication status and history, clinical features, and clinical history), and algorithms used to quantify sleep parameters. In this section, we rely on meta-analyses[40,41] and a review[42] to summarize the diversity of these findings.

The reader should note that sleep stages reported in this chapter are based on the sleep scoring rules and terminology that prevailed before 2007 and that are broadly consistent with those developed by Rechtschaffen and Kales.[43] They are designated "classic terminology" in Table 139-1. Table 139-1 provides a simplified translation of classic terminology to terminology as revised in 2007 by the American Academy of Sleep Medicine.[44]

Total Sleep, Sleep Maintenance, and Sleep Continuity

Overnight PSG studies have shown that the sleep of schizophrenic patients is characterized by poor sleep efficiency. Often this takes the form of a reduction in total sleep time as well as early, middle, and late insomnia. The most consistently reported abnormality is early insomnia or difficulty reaching a state of persistent sleep. Even a daytime measure of sleep propensity reported that mean sleep latency in the Multiple Sleep Latency Test was 36% longer in unmedicated patients with schizophrenia than in nonpsychiatric controls.[45] Furthermore, many of the overnight PSG studies that have evaluated schizophrenic patients treated with antipsychotics have observed a reduced sleep propensity suggesting that some degree of residual insomnia is not an uncommon outcome of standard antipsychotic treatment. Note also that severe insomnia is one of the prodromal signs of impending psychotic decompensation or relapse subsequent to discontinuing antipsychotic medication.[46-48]

Table 139-1 Comparison of Sleep Stage Designations

Classic Terminology	Revised Terminology
Wakefulness	Stage W
Stage 1 sleep	Stage N1 (NREM 1 sleep)
Stage 2 sleep	Stage N2 (NREM 2 sleep)
Stage 3 sleep	Stage N3 (NREM 3 sleep)
Stage 4 sleep	Stage N3
Slow wave sleep (S3 + S4)	Stage N3
Stage REM sleep	Stage R (REM sleep)

Abnormalities of REM Time and REM Eye Movements

Despite early speculation regarding potential REM sleep abnormalities in schizophrenia, studies comparing schizophrenic patients with healthy control subjects have consistently shown that REM sleep time is not systematically augmented or reduced.[40,41] Eye movements during REM sleep have also been studied. Visual scoring of REM sleep eye movements report no difference in density of eye movements between schizophrenic patients and control subjects.[49,50] An automated eye movement detection system[51] yielded the same conclusion but extended the observation in finding no differences in eye movement density in schizophrenic patients, nonpsychiatric control subjects, and patients with major depressive disorder.

REM Sleep Latency

Many PSG studies have quantified the latency to the onset of the first REM sleep period.[40-42] Several have compared the REM latency of unmedicated schizophrenic patients with that of nonpsychiatric control subjects. Approximately half have reported significant between-groups differences, with the schizophrenic patients demonstrating abnormally short REM latency. Even in studies finding no between-groups differences, a bimodal distribution of REM latency values in schizophrenic patients has been observed, suggesting that there are subgroups of schizophrenic patients with sleep-onset REM periods.[49,52] Short REM latency might represent an active or primary alteration of REM sleep mechanisms. Alternatively, as suggested by Feinberg and colleagues,[53] a slow wave sleep (SWS) deficit in the first NREM period might permit the passive advance or early onset of the first REM period.

Abnormalities of NREM Sleep

SWS deficits are often, but not consistently, observed in PSG recordings of schizophrenic patients. In visually scored PSG, SWS is reported as the summation of sleep stages 3 and 4, with stage 4 sleep having the greater incidence of underlying slow wave activity. Documentation of SWS or stage 4 deficits has been reported in many studies of schizophrenic patients.[38-40] The possibility that SWS deficits might serve as a trait marker in patients with schizophrenia was suggested by a study reporting no SWS differences between schizophrenics and first-degree relatives.[54] Although some research has suggested that prior exposure to, or withdrawal from, antipsychotics might explain these inconsistencies,[50] SWS deficits have been observed in antipsychotic-naive patients in their first episode of schizophrenia.[55] In addition to clinical heterogeneity, another factor might contribute to the inconsistency, notably the insensitivity of visual scoring to variations in the incidence and amplitude of the slow or delta wave (0 to 3 Hz) EEG oscillations that make up SWS.

Sleep-Related Brain Wave Activity

Studies[56-59] using computer quantification of the frequency content of the sleep EEG (e.g., Fourier analysis, period and amplitude analysis) have confirmed the degradation of sleep-related delta activity in schizophrenic patients relative to age-matched nonpsychiatric control subjects. Significant decrements in delta amplitude between schizophrenic patients and nonpsychiatric controls have been observed despite

comparable amounts of visually scored SWS.[60] Laterality of frontal cortex delta counts is reduced in schizophrenic patients relative to healthy controls, suggesting right frontal pathophysiology among the schizophrenic patients.[59]

In contrast to degradation of delta (0 to 3 Hz) activity, power in the high-frequency beta (20 to 35 Hz) and gamma (35 to 45 Hz) ranges of the sleep EEG may be greater in unmedicated schizophrenic patients relative to healthy controls.[61] Also, the number, duration, and amplitude of NREM sleep spindles (12 to 16 Hz) may be reduced in medicated schizophrenic patients, a finding that suggests some abnormality in thalamic-reticular and thalamocortical function in schizophrenia.[62] In addition, reductions in the number and density of sleep spindles in schizophrenic patients may be associated with impaired memory consolidation.[63]

Slow Wave Sleep Homeostasis

Deficits of SWS, whether scored visually or quantified by computer algorithm, raise concern about the integrity of homeostatic regulatory mechanisms in schizophrenia. In 1974, Feinberg advanced the homeostatic model of SWS; according to this model, the homeostatic drive builds up during waking and dissipates in SWS across successive NREM cycles.[64] In healthy subjects, SWS increases in proportion to the amount and intensity of prior waking, suggesting that the homeostatic or dynamic response might serve a restorative role in the central nervous system. This homeostatic drive is demonstrated by a rebound in SWS or EEG delta activity following the naturalistic probe of total sleep deprivation.

Two studies have looked at potential homeostatic dysregulation in schizophrenia. In the first, no homeostatic recovery of visually scored SWS was found after 85 hours of total sleep deprivation.[65] More recently, SWS visual scoring and computer analysis of delta activity were examined after 1 night of total sleep deprivation.[66] Stage 4 sleep was absent on both baseline and recovery nights, but period and amplitude analysis of EEG delta activity revealed a significant (although modest) recovery night increase in delta incidence and amplitude relative to baseline. Therefore it would appear that the homeostatic drive in schizophrenia is operative but diminished. In support of this conclusion, it has been shown that on an intranight basis, the time course of delta activity dissipation over all NREM sleep cycles is normal in schizophrenic patients.[60]

Circadian Rhythm Abnormalities

Although circadian rhythm disturbances have also been reported in patients with schizophrenia, the findings are largely inconsistent.[67-69] Studies of the circadian organization of sleep-wake and rest-activity cycles have reported both phase-delayed and phase-advanced sleep patterns, irregular sleep-wake patterns, and free-running rest-activity patterns.[70-73] A phase-delayed pattern would be congruent with the well-documented difficulty of sleep initiation as well as the observation that schizophrenics tend to favor "eveningness" on the morningness-eveningness scale.[74]

Because disturbed patterns of melatonin secretion have been associated with insomnia, patterns of melatonin secretion have been examined in schizophrenic patients. Both phase-advanced[75] and phase-delayed[76] melatonin secretion has been reported. There are also reports of blunted nocturnal secretion of melatonin, including a report that normalization of melatonin levels may not occur following clinical improvement with antipsychotic treatment.[77,78]

The inconsistencies associated with studies of circadian rhythm abnormalities in patients with schizophrenia may, in part, be attributed to a wide range of methodologic shortcomings inherent in these circadian rhythm studies,[69] to a diminished exposure to sunlight, activity, and other effective zeitgebers, or to lifestyle choices often involving preferences for staying awake at night and sleeping during the day. Circadian rhythm abnormalities are not diagnostic of schizophrenia but occur with sufficient frequency to merit further research.

Correlation with Clinical and Neurobiologic Measures

Many of the sleep abnormalities associated with schizophrenia have been studied in relationship to global symptom severity, the severity of positive and negative symptoms, clinical prognosis, and neurocognitive impairment.[42]

Brain dysmorphologies have also been studied in relationship to sleep abnormalities in patients with schizophrenia. On a theoretical level, SWS deficits have been associated with microstructure brain abnormalities. Feinberg[32] has proposed that schizophrenia is a neurodevelopmental disorder arising from a malfunction in the normal maturational process of synaptic elimination during the second decade of life; excess synaptic pruning would result in less capability for synchronous EEG slow wave activity and a deficit of SWS. Empirically, SWS deficits or a reduction in its stage 4 component have been associated with enlarged ventricular system volume.[79,80] A related study using antipsychotic-naive schizophrenic patients did not concur.[50] PSG studies have also noted a relationship between brain dysmorphologies in schizophrenia and sleep maintenance abnormalities.[50,79,81]

These reports of brain dysmorphologies and their relationship to sleep abnormalities in patients with schizophrenia must be viewed with some interpretive restraint. Terms such as "enlargement" and "atrophy" need longitudinal assessment, but all the reported studies employed a cross-sectional design. Volumetric status of the reported brain structures before the point of cross-sectional assessment is only speculative. We must also view with some restraint any contention that sleep abnormalities with a neuroanatomic correlate must have a stable or traitlike presentation over time. First, many of the dyssomnias of schizophrenia do change over time in response to antipsychotic treatment. Second, consistent with the observation of adult neurogenesis[82] in the human brain are demonstrations of longitudinal changes in the brain structure of schizophrenic patients undergoing antipsychotic treatment.[83-85]

Our understanding of how neurochemical mechanisms might correlate with, or possibly mediate, the sleep abnormalities of schizophrenia is limited to a small number of studies whose findings may be confounded by long-term exposure to antipsychotics. Furthermore, although dopamine and glutamate dysfunction might occupy prominent roles in the pathophysiology of schizophrenia, the relationship of sleep abnormalities in schizophrenia to either of these neurochemical systems has not been directly demonstrated.

Limited reports involving four other neurotransmitter systems have been published. Cholinergic supersensitivity has been associated with short REM latencies in patients with

schizophrenia.[86] Among unmedicated schizophrenic patients, SWS deficits have been linked to a 5-HT dysfunction; this study reported a positive correlation between SWS time and cerebrospinal fluid (CSF) levels of the 5-HT metabolite 5-hydroxyindole acetic acid.[87] A separate study reported that increased CSF levels of norepinephrine and its principal metabolite, 3-methoxy-4-hydroxyphenylglycol, accompanied psychotic decompensation and relapse-related insomnia.[46] CSF hypocretin, a wake-promoting neurotransmitter that excites midbrain dopamine neurons, has been correlated with sleep latency, suggesting a relationship between hypocretin and hyperarousal in schizophrenia.[88]

TREATMENT

Treating Schizophrenia with Antipsychotic Agents

Consistent with standards of practice, most patients with schizophrenia are treated with one or more of the antipsychotic agents listed in Table 139-2. Antipsychotics differ in their chemical design and have differential effects on DA, 5-HT, α-adrenergic, cholinergic, and histaminic receptors and their various subtypes.[89,90] The receptor binding profiles of these agents are associated with different side effects and clinical outcome.

Types of Antipsychotic Agents

First-Generation Antipsychotics

The therapeutic efficacy of the first generation of antipsychotics is attributed to their ability to block the dopamine D_2 postsynaptic receptor, particularly the D_2 receptors in the striatum, but the clinical benefits associated with these first-generation antipsychotics have occurred with some cost. Because of their signature-binding profile, these antipsychotics often result in extrapyramidal side effects (EPS) such as akathisia, dystonia, and parkinsonism. A more damaging side effect associated with D_2 receptor blockade is tardive dyskinesia. The traditional antipsychotics are also associated to varying degrees with antihistaminergic and anticholinergic side effects, including sedation, changes in blood pressure and myocardial conduction, sexual dysfunction, and weight gain. A rare but potentially fatal side effect associated with hyperthermia is the neuroleptic malignant syndrome.

Table 139-2 Commonly Used Antipsychotic Medications

Antipsychotic Medication	Usual Adult Daily Maintenance Dose (mg)
First-Generation Antipsychotics	
Chlorpromazine (Thorazine)	50–400
Fluphenazine (Prolixin)	1–15
Haloperidol (Haldol)	1–15
Perphenazine (Trilafon)	8–24
Thioridazine (Mellaril)	50–400
Thiothixene (Narvane)	6–30
Trifluoperazine (Stelazine)	4–30
Second-Generation Antipsychotics	
Aripiprazole (Abilify)	10–30
Clozapine (Clozaril)	200–600
Olanzapine (Zyprexa)	5–20
Quetiapine (Seroquel)	150–750
Paliperidone (Invega)	6–12
Risperidone (Risperdal)	2–8
Ziprasidone (Geodon)	80–160

Second-Generation Antipsychotics

Several factors played a role in motivating the development of the second-generation antipsychotics. First, up to 30% of patients with chronic schizophrenia have a poor or inadequate response to first-generation antipsychotics. Second, although traditional antipsychotics have demonstrated success in ameliorating positive symptoms, they have been less successful in treating negative symptoms. Third, the side effects associated with traditional antipsychotics, particularly EPS and tardive dyskinesia, were a source of noncompliance and raised difficult management issues. Clozapine was the first of the second-generation antipsychotics to demonstrate good clinical efficacy with little or no EPS. The atypical antipsychotics now include six other agents, listed in Table 139-2. Although they demonstrate unique receptor binding profiles, as a group they are characterized by weaker affinity for the D_2 receptor (relative to traditional antipsychotics) and strong affinity for serotonin receptors.[90] Second-generation antipsychotics are now considered first-line treatment for schizophrenia. Relative to the traditional antipsychotics, they may offer better compliance.

However, second-generation antipsychotics are not the panacea once envisioned.[91] Only clozapine has clearly demonstrated superior effectiveness in treatment-refractory schizophrenia and in childhood schizophrenia. None of the atypical antipsychotics has been strongly effective against negative symptoms. Like the traditional antipsychotics, the atypical antipsychotics can have serious side effects. Although the incidence of EPS and tardive dyskinesia remains low with atypical antipsychotics, both EPS and tardive dyskinesia can occur with atypical antipsychotics, particularly at higher doses. Clozapine is associated with increased risk for agranulocytosis and seizures at higher doses. The atypical antipsychotics are associated with increased morbidity owing to weight gain, dyslipidemia, and impaired glucose regulation including type 2 diabetes mellitus. Actigraphic monitoring of schizophrenic patients treated with olanzapine or risperidone documented reduced 24-hour activity (owing to a longer time in bed) and, notably, reduced 10-hour daytime activity; low amounts of physical activity may also contribute to the metabolic side effects of these antipsychotic agents.[92] Finally, antipsychotic-related weight gain may be particularly striking and puts the patient at increased risk for developing sleep-related breathing disorders and other obesity-related morbidities.

The National Institute of Mental Health–sponsored Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study[93] compared a single first-generation antipsychotic to several second-generation antipsychotics in terms of effectiveness and side effects. This study found no significant difference in efficacy between perphenazine and three atypical antipsychotics (risperidone, quetiapine, and ziprasidone). Olanzapine, which was found to be more effective than the other drugs in terms of discontinuation rate, was also

associated with greater weight gain and metabolic morbidities. The CATIE study underscores the need for a more systematic comparison of first- and second-generation antipsychotics with respect to their efficacy, side effects, and expense, all of which have a major impact on treatment compliance. Discontinuation rates of antipsychotics are alarmingly high for a chronic illness. Within 1 year of initiating treatment with atypicals (aripiprazole, olanzapine, risperidone, ziprasidone, or quetiapine), approximately 90% of schizophrenic patients had discontinued their antipsychotic medication.[94]

Sleep-Promoting Effects of Antipsychotics

Broadly speaking, first-generation antipsychotics alter sleep architecture by improving sleep maintenance with augmentations of total sleep time and sleep efficiency and reductions in sleep latency and waking.[95-99] With less consistency, the traditional antipsychotics have been associated with increases in REM latency,[95-98] REM sleep time,[96,97] and REM eye movement density.[95,97,99] In one study, chlorpromazine was associated with increased SWS time.[95]

With regard to the effect of second-generation antipsychotics on sleep patterns in schizophrenic patients, clozapine has been the most widely studied. Clozapine has been shown to have strong consolidating effects on sleep, including increased total sleep time, sleep efficiency, and stage 2 sleep and decreased sleep latency and wake time after sleep onset.[99-101] Relative to a medication-free baseline, a significant decline in SWS and stage 4 was observed in schizophrenic patients treated with clozapine.[100] Olanzapine has also been shown to be a sleep-promoting agent with increases in both sleep efficiency and SWS[102,103]; a reduction in sleep spindle density has also been reported.[104] Likewise, a significant enhancement of SWS has been noted in schizophrenic patients treated with risperidone.[105] Paliperidone's effects on sleep patterns in patients with schizophrenia revealed increased sleep efficiency, total sleep time, and stage 2 and stage REM minutes as well as decreased sleep latency, waking, and stage 1 minutes.[106] PSG evaluations of schizophrenic patients treated with quetiapine, ziprasidone, and aripiprazole have not been undertaken; however, effects of quetiapine and ziprasidone on the sleep of healthy controls include improvements in sleep induction and consolidation.[107,108]

Treating Comorbid Psychiatric Illness

Although monotherapy with second-generation antipsychotics is standard practice, augmentation with additional antipsychotics is not uncommon and can optimize therapeutic response. In addition to antipsychotics, schizophrenic patients may be prescribed other psychoactive agents to treat comorbid psychiatric illness. For example, schizophrenic patients with problems of impulse control and patients with schizoaffective disorder may be prescribed adjunctive mood stabilizers such as valproate. It is not uncommon for comorbid depression to be diagnosed in schizophrenic patients, and adjunct antidepressants may be prescribed. Mood stabilizers often demonstrate beneficial effects on sleep, whereas antidepressants may either ameliorate or exacerbate sleep problems, depending on the agent and individual. Schizophrenic patients may be enrolled in dual-diagnosis programs to treat comorbid problems of alcohol and substance abuse; by themselves, alcohol and substance abuse can contribute to significant sleep disruption.

Treating Sleep Disorders in Patients with Schizophrenia

After the symptoms of schizophrenia and other comorbid medical and psychiatric illnesses have been optimally treated, clinicians might need to address a range of residual sleep disorders. These residual sleep disorders can encompass insomnia, hypersomnia, and parasomnias.

Insomnia

The CATIE study[93] reported that rates of residual insomnia in antipsychotic-treated schizophrenic patients ranged from 16% to 30%. It is not uncommon for insomnia to be attributed to poor sleep hygiene, sleep-related movement disorders, and untreated hyperarousal associated with the schizophrenia illness.

Inadequate Sleep Hygiene. Many schizophrenic patients lack daytime structure and prefer to avoid social interactions. Consequently, they can develop bad habits in relation to sleep initiation and sleep maintenance, including sleep reversals and polyphasic sleep patterns with extended periods of daytime napping. Daytime naps affect nocturnal sleep more when the naps are taken later in the day. This situation may be further complicated by excess caffeine intake and the self-prescribed alcohol or psychoactive drugs such as cocaine. Patients with schizophrenia are always good candidates for sleep hygiene counseling.

Sleep-Related Movement Disorders. Because prevalence studies of sleep disorders in antipsychotic-naive schizophrenic patients are lacking, baseline rates for comorbid sleep disorders in this population have not been established. Although in some cases insomnia may be associated with lifestyle habits, in others sleep disturbance may be a side effect of treatment with antipsychotic medication.

Because restless legs syndrome (RLS) and periodic limb movements of sleep (PLMS) respond to DA agonists, DA deficiency is thought to be central to their pathophysiology.[109] Consequently, the D_2 receptor blockade of antipsychotic agents might induce RLS and PLMS in antipsychotic-treated schizophrenic patients. Consistent with this hypothesis, the prevalence of RLS in antipsychotic-treated schizophrenic patients has been reported to be more than twice that of healthy controls.[110] Although the diagnosis of RLS is made by clinical interview, the diagnosis of PLMS requires overnight PSG; consequently, the prevalence of PLMS in patients with schizophrenia has been less systematically examined. PLMS prevalence rates for schizophrenic patients with a history of treatment with first-generation antipsychotics have been reported to be in the 13% to 14% range.[111,112] Certain atypical antipsychotics, such as olanzapine[113-114] and risperidone,[115] have been associated with the development of RLS and PLMS. Although DA agonists such as ropinirole and pramipexole are first-line treatments for RLS and PLMS, treatment options for schizophrenic patients with RLS or PLMS should be carefully considered. One might first consider a reduction in antipsychotic dose or a change to a different antipsychotic agent rather than introduce another agent with an opposing mechanism of action. Patients might also be evaluated for possible iron deficiency, a known risk factor for the development of RLS, or use of substances, such

as caffeine, that might worsen sleep-related movements. Although agents such as ropinirole and pramipexole are not first-line choices for patients with schizophrenia, such agents might be necessary when their benefits outweigh the risk. Chapter 95 addresses evaluation and treatment options for patients with RLS or PLMS.

Clinicians should note that akathisia, which is characterized by intense motor restlessness and pacing, is a common side effect of antipsychotic treatment. By itself, akathisia can cause significant sleep disruption. Although RLS might be confused or overlap with the restlessness of akathisia, the clinical presentation of RLS is more likely to be associated with a circadian worsening of symptoms in the evening or at night and symptomatic relief by movement.

Schizophrenia-Related Hyperarousal. Antipsychotic-treated schizophrenic patients may also experience a residual and persistent insomnia directly associated with hyperarousal or an inadequately treated component of their clinical illness. A range of options for treating residual insomnia in patients with schizophrenia has been recently reviewed.[116,117] Given sufficient severity, the clinician may opt to treat residual insomnia by changing to a different, more sedating antipsychotic agent, changing the dose of the prescribed antipsychotic agent, or adding an adjunctive, more sedating antipsychotic medication. Alternatively, supplemental use of anxiolytics or sedative-hypnotics might be considered[118]; however, these agents should be prescribed cautiously, particularly for patients with a sleep-related breathing disorder or a history of alcohol or drug abuse. Because disturbed patterns of melatonin secretion have been associated with insomnia, the use of melatonin as an alternative adjunctive treatment for residual insomnia in schizophrenia has been the subject of investigational study. Two studies examined the effect of exogenous melatonin on residual insomnia in antipsychotic-treated schizophrenic patients. Melatonin replacement was associated with improved sleep maintenance measured by actigraphy[119] and by self-report.[120]

Hypersomnia

The CATIE study[93] also reported rates of antipsychotic-related somnolence. Rates of somnolence in antipsychotic-treated schizophrenic patients ranged from 24% to 31%. Medication compliance is associated with patients' attitudes toward antipsychotic treatment and related side effects, and it is important to note that sedation has a significant negative effect on these attitudes.[121] Somnolence in antipsychotic-treated schizophrenic patients may be a side effect of antipsychotic treatment, or it may be symptomatic of a sleep disorder such as a sleep-related breathing disorder enhanced by, or induced by, antipsychotic treatment.

Somnolence as a Side Effect of Antipsychotic Medication. Among first-generation antipsychotics, sedation is a side effect associated with high-milligram, low-potency agents such as chlorpromazine. In contrast, the low-milligram, high-potency agents such as haloperidol have lower sedation rates. Among second-generation antipsychotics, clozapine and olanzapine are the most sedating. The remaining atypical agents have lower rates of sedation. Sedation rates also reflect the half-life of the agent as well as dosing levels. Clinicians typically address excessive antipsychotic-related sedation by changing antipsychotic agents or reducing dosage.

As an alternative to changing antipsychotic agents or reducing their dosage, several studies have examined the effect of modafinil as an adjunct to antipsychotic treatment. Modafinil is a wake-promoting agent whose mechanism of action remains unclear. Its primary U.S. Food and Drug Administration–approved target is the excessive daytime sleepiness associated with narcolepsy. However, it has been suggested that modafinil may improve excessive daytime sleepiness in a variety of off-label conditions such as antipsychotic-induced daytime somnolence. Case studies[122] and an open-label pilot study[123] have shown that modafinil can increase wake time, reduce total sleep time, and ameliorate fatigue in antipsychotic-treated schizophrenic patients. However, several randomized, double-blind, placebo-controlled trials of modafinil as an adjunct to antipsychotic treatment reported that modafinil, relative to placebo, did *not* significantly improve antipsychotic-associated sedation or fatigue.[124-127] Taken together, these studies suggest that the use of modafinil to counter somnolence in antipsychotic-treated schizophrenic patients is not justified at this time. Off-label use of modafinil as an adjunct to antipsychotic treatment in schizophrenia needs further study with larger patient samples and with particular scrutiny given to the possibility of exacerbating psychosis in patients with schizophrenia.[128]

Somnolence Associated with Sleep-Disordered Breathing. The prevalence of sleep-disordered breathing (SDB), such as obstructive sleep apnea (OSA), in antipsychotic-naive schizophrenic patients has not been established. Antipsychotic-treated schizophrenic patients who have volunteered for research protocols have demonstrated a high prevalence rate for SDB, with reported estimates of 17%,[111] 19%,[129] 25%,[130] and 48%.[112] These studies varied in terms of sample size, age, sex, type of antipsychotic, and measurement instrument (PSG, oximetry, sleep questionnaire, and ambulatory apnea monitor, respectively).

In contrast, among schizophrenic patients referred to a sleep clinic for a suspected sleep disorder, more than 46% had a respiratory disturbance index greater than 10 events per hour; the mean respiratory disturbance index was 64.8 events per hour.[131] In this study, the most powerful predictor of OSA syndrome was obesity. Weight gain has long been a reported side effect of antipsychotic treatment, and as discussed previously, weight gain secondary to the use of the atypical antipsychotics has been associated with morbid obesity, diabetes, and the development of moderate to severe SDB.[132]

Daytime somnolence is a relatively common side effect of antipsychotic treatment, but clinicians must consider the possibility of comorbid SDB for schizophrenic patients who have a history of obesity or who have gained weight secondary to antipsychotic treatment. Schizophrenic patients with comorbid SDB can be treated effectively with nasal continuous positive airway pressure and can demonstrate relatively good compliance and significant clinical improvement.[133-135]

Parasomnias

First-generation antipsychotics, particularly in combination with lithium,[136,137] can induce somnambulism. Sleepwalking also has been documented in patients treated with second-generation agents such as olanzapine.[138,139] Because

sleepwalking is associated with impaired arousal from SWS, patients treated with lithium and olanzapine may develop somnambulism because these agents have been credited with SWS enhancement. It has been reported that clonazepam may be a treatment option for antipsychotic-induced somnambulism[140] but with the same caveats as those noted for hypnotic use. Another parasomnia, sleep-related eating disorder, may be induced by antipsychotics such as haloperidol,[141] olanzapine,[142] and risperidone.[143] Treatment options for sleep-related eating disorders might include topiramate, a selective serotonin reuptake inhibitor, or a sedative agent, although this latter class has itself been associated with parasomnias.

FUTURE DIRECTIONS

One of the ongoing controversies is the diagnostic specificity of sleep abnormalities. Although a range of abnormalities is observed in schizophrenia, they lack diagnostic specificity. Long sleep-onset latencies and other sleep maintenance abnormalities found in schizophrenia are observed with insomnia occurring in many settings, including in patients with other psychiatric conditions such as major depression. Sleep-onset REM periods and SWS deficits are also observed in depressive illness.

A goal less ambitious than establishing diagnostic specificity is to find a brain abnormality that would reliably characterize a subgroup of patients in a meaningful way. SWS deficits (measured as sleep stages 3 and 4 with visual scoring or as delta power or integrated amplitude with computer analysis) could furnish a biologic basis for separating subgroups of schizophrenic patients. Although low SWS is not specific to schizophrenia, it is a highly reliable, enduring trait that is consistently found in 40% to 50% of samples and might provide a meaningful phenotype. Despite its reliability, SWS deficits have received attention from only a few schizophrenia researchers and remain a neglected research area in a critically important psychiatric illness that has few consistently documented brain abnormalities. Given the paucity of clues to brain dysfunction in schizophrenia, SWS investigations could prove a fruitful field for further research. Such studies might explore dynamic features of SWS, such as its trends across NREM periods[144] and the more recently discovered inverse oscillation of sigma and delta frequencies.[145]

The synaptic pruning model of schizophrenia noted that it was possible that too few, too many, or the wrong synapses might be pruned during adolescence.[32] One testable hypothesis might be that patients characterized by an enduring decrease in SWS would show different trajectories or greater amounts of cortical thinning detectable by structural magnetic resonance imaging.[146] There are, of course, many other hypotheses that could be advanced. Given the paucity of biologic abnormalities in schizophrenia, it is puzzling that so robust and consistent a brain abnormality as low SWS should have been so neglected.

Further empirical research might also address the role of SWS deficits in both clinical presentation and outcome. The homeostatic property of SWS suggests that SWS serves some central nervous system restorative function whose loss would have negative neuropsychiatric consequences; if so, are SWS deficits in schizophrenia associated with greater cognitive impairment and thought disorder, and do schizophrenic patients with less SWS impairment face a better clinical outcome? As a corollary, do antipsychotics that augment SWS produce a greater likelihood of positive clinical outcome?

Although insomnia is a defining characteristic of schizophrenia, we do not understand the role that sleep disruption might play in the illness. Is insomnia only a byproduct of schizophrenia, or does ongoing sleep disturbance also affect the course of the illness? Because ongoing sleep disturbance might have a negative effect on clinical outcome, treatment plans for schizophrenic patients should explicitly incorporate strategies to deal aggressively with complaints of poor sleep quality. The restorative effect of antipsychotics on sleep maintenance and structure may enhance their clinical efficacy and increase tolerance. A positive clinical outcome may require the restoration of sleep's recuperative powers.

Comorbid sleep disorders in schizophrenics should be treated. Untreated sleep disorders can augment the fragility of underlying sleep processes in schizophrenia, with potentially adverse effects on clinical outcome. Research into sleep and circadian mechanisms in these patients may yield novel treatments.[147]

CLINICAL PEARL

Schizophrenia is often associated with severe insomnia that can be a harbinger of relapse. Caregivers should routinely query schizophrenic patients about changes in their sleep patterns and should modify treatment in the face of severely increasing insomnia. Caretakers should also be instructed in the specific sleep disorder syndromes that tend to occur in this population and should use this knowledge to routinely evaluate the possibility of emergent sleep disorders stimulated by antipsychotic treatment.

SUMMARY

Subjective and objective assessments of sleep in schizophrenia reveal a wide range of sleep abnormalities with insomnia, principally sleep-onset insomnia, being the most common presentation. Among schizophrenics, worsening insomnia may be a harbinger of impending relapse and psychotic decompensation. Other sleep abnormalities, such as SWS deficits, can be found with sleep staging or measures of sleep EEG frequency content. Broadly speaking, most antipsychotic agents have a positive effect on the insomnia of schizophrenia. However, some schizophrenic patients medicated with antipsychotics experience undertreated or residual insomnia, whereas others may experience significant somnolence. Heightened sedation is an unwanted side effect of some second-generation antipsychotics. Antipsychotic agents may also induce or exacerbate comorbid dyssomnias such as RLS and SDB. All dyssomnias in schizophrenia should be vigorously addressed because a positive clinical outcome may be associated with the normalization of sleep and its restorative processes.

Selected Readings

Annamalai A, Palmese LB, Chwastiak LA, et al. High rates of obstructive sleep apnea symptoms among patients with schizophrenia. *Psychosomatics* 2015;**56**(1):59–66.

Baandrup L, Jennum P, Lublin H, et al. Treatment options for residual insomnia in schizophrenia. *Acta Psychiatr Scand* 2013;**127**:81–2.

Bridler R, Umbricht D. Atypical antipsychotics in the treatment of schizophrenia. *Swiss Med Wkly* 2003;**133**:63–76.

Chemerinski E, Ho B, Flaum M, et al. Insomnia as a predictor for symptom worsening following antipsychotic withdrawal in schizophrenia. *Compr Psychiatry* 2002;**43**:393–6.
Chouinard S, Poulin J, Stip E, et al. Sleep in untreated patients with schizophrenia: a meta-analysis. *Schizophr Bull* 2004;**30**(4):957–67.
Feinberg I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J Psychiatr Res* 1983;**17**:319–34.
Ferrarelli F. Sleep in patients with schizophrenia. *Curr Sleep Med Rep* 2015;**1**(2):150–6.
Kane JM, Sharif ZA. Atypical antipsychotics: sedation versus efficacy. *J Clin Psychiatry* 2008;**69**(Suppl. 1):18–31.
Kang SG, Lee HJ, Jung SW, et al. Characteristics and clinical correlates of restless legs syndrome in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2007;**31**:1078–83.
Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005; **353**(12):1209–23.
Monti JM, BaHammam AS, Pandi-Perumal SR, et al. Sleep and circadian rhythm dysregulation in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2013;**43**:209–16.
Rapoport JL, Addington AM, Frangou S, et al. The neurodevelopmental model of schizophrenia: update 2005. *Mol Psychiatry* 2005;**10**:434–49.
Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait. *Arch Gen Psychiatry* 2003;**60**:1187–92.

A complete reference list can be found online at ExpertConsult.com.

Chapter
140

Medication and Substance Abuse

Timothy Roehrs; Thomas Roth

Chapter Highlights

- Most psychoactive drugs with abuse liability have effects on sleep and wakefulness.
- The mechanisms underlying substance abuse are known, but the role of the drug's sleep-wake state–altering effects in substance abuse is not fully known, although it is likely to be important.
- Some of the drugs of abuse are legal and widely used socially and may be the cause of patients' sleep or alertness complaints.
- Other drugs with abuse liability are drugs indicated in the treatment of sleep disorders.
- This chapter provides guidelines for sleep disorders clinicians to differentiate drug-seeking behavior from therapy-seeking behavior.

Various legal medications and all illegal central nervous system (CNS)–acting drugs have a high abuse liability, that is, the likelihood for development of physiologic or behavioral dependence on these substances is heightened. The various terms often used in discussing substance abuse are confusing, are controversial, and need clarification. Physiologic dependence is a state induced by repeated drug use that results in a withdrawal syndrome when the drug is discontinued or an antagonist is administered. Many legal medications and illegal drugs can produce physiologic dependence, although the syndrome intensity, relation to dose, and necessary duration of use vary among different drugs. The fact that a drug produces physiologic dependence, meaning that a withdrawal syndrome appears when the drug is discontinued, does not necessarily imply substance abuse.

In the sleep field, a phenomenon suggestive of the presence of physiologic dependence on a drug is rapid eye movement (REM) sleep rebound. When drugs that suppress REM sleep are discontinued, a REM rebound (i.e., increased REM pressure) is seen, which is manifested by reduced REM sleep latency, by increased REM sleep time and REM density, and subjectively with reports of nightmares. Most of the antidepressant medications at therapeutic doses suppress REM sleep, and a REM sleep rebound occurs when the drug is discontinued. However, a REM sleep rebound after antidepressant use does not lead to resumption of antidepressant use. On the other hand, a reduced REM latency (e.g., one sign of REM pressure and an underlying REM sleep disturbance) in abstinent alcoholics is predictive of alcoholic relapse.

Physiologic dependence may be a component of but is neither a necessary nor a sufficient condition for behavioral dependence. Behavioral dependence is a pattern of behavior characterized by repetitive and compulsive drug seeking and consumption, despite considerable substance-related problems. The formal diagnostic criteria according to the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5) are discussed later. Whereas the sleep disorders clinician will not make formal diagnoses, awareness of the criteria is important for referral decisions. Substance abuse and dependence are common, as 18% of the U.S. population will experience a substance abuse disorder during their lifetime, and about 20% of patients in general medical practice and 35% of psychiatric patients present with substance abuse disorders.

Virtually all drugs with a high abuse liability have profound effects on sleep and wake. For this reason, sleep disorders clinicians should assess *all* the drug-taking behavior of their patients, including prescribed and over-the-counter drugs, recreational drugs, tobacco and caffeine, health foods, steroids, botanicals, and natural substances. This chapter reviews the sleep-wake alterations produced by administration and discontinuation of various drug classes associated with abuse. Also discussed is the way that the state-altering characteristics (i.e., their disruptive effects on sleep or daytime alertness) of these drugs contribute to their dependence liability. Finally, what is known about the neurobiologic and behavioral mechanisms that underlie these drugs' abuse liability and their state-altering effects is discussed.

Some drugs of abuse have no legal therapeutic indications (i.e., various inhalants, LSD); others have narrowly defined therapeutic indications (i.e., amphetamine, methylphenidate), and some have broader therapeutic indications (i.e., benzodiazepine receptor agonists). Other drugs of abuse have wide use as social drugs (i.e., alcohol and caffeine). Marijuana occupies a unique position as it remains illegal federally but as of 2014 is legal in 20-plus states for specific therapeutic indications and in a few states for recreational and social use. For these various reasons, guidelines are provided for sleep disorders clinicians that will help them differentiate drug-seeking behavior from therapy-seeking behavior. In drug-seeking behavior, the drug and its effects are the focus of the drug use, whereas in therapy-seeking behavior, the medications' ability to reverse the signs and symptoms of the disease is the focus of the drug use. This distinction is important because many of the drugs used by sleep disorders clinicians are used chronically, and hence it is important to differentiate long-term use

from abuse. Many of these medications also have scientifically documented efficacy and are hence indicated for specific sleep disorders; some of these drugs also are abused and, importantly, may be the cause of a sleep disorder.

DIAGNOSIS OF SUBSTANCE-RELATED DISORDERS

The generally accepted diagnostic classification system for substance-related disorders is DSM-5. DSM-5 reflects a major departure from DSM-IV and DSM-IV-TR. Substance-related disorders are divided into two major classes: (1) substance use and (2) substance-induced disorders. Substance use disorders, formerly termed substance abuse and substance dependence (DSM-IV-TR), are characterized by behaviors and consequences (a total of 11 criteria are listed) in groupings that show *impaired control, social impairment, risky use, and pharmacologic consequences* (i.e., tolerance, withdrawal) associated with use of 10 classes of substances. Rather than distinguishing abuse and dependence as in DSM-IV-TR, DSM-5 rates the severity of the disorder by the number of symptoms present: mild, two or three; moderate, four or five; and severe, six or more. Substance-induced disorders are characterized by symptoms reflecting the presence of intoxication, withdrawal, or a mental disorder. As this category name implies, the disorder has to be associated with current or very recent use of the substance.

Whereas most of the drugs of abuse are disruptive of sleep or daytime alertness, such disturbances are not major criteria for substance abuse in DSM-5. They are mentioned as possible symptoms in a withdrawal syndrome, which is one of the 11 criteria for a substance use disorder. DSM-5 emphasizes that symptoms of tolerance development and withdrawal, *only* occurring in the context of medical treatment with prescribed medications, should not receive a diagnosis of substance-related disorder (distinguishing drug seeking from therapy seeking in clinical practice is discussed later). Also as discussed in this chapter later, the role of disruptions of sleep and daytime alertness in substance use disorders is not elaborated in DSM-5.

The attempt to understand substance abuse scientifically has proceeded at two levels of analyses, behavioral and neurobiologic. Interestingly, the scientific evidence at these two levels of analysis has converged. For sleep scientists and clinicians, an obvious question is, "How does a drug's effect on sleep and daytime alertness relate to the behavioral and neurobiologic changes that occur in a person abusing the substance?"

Behavioral Mechanisms

Substance use, whether in a therapeutic and socially accepted recreational form or in an excessive, socially unacceptable, and physically hazardous form, is a behavior that can be analyzed to determine factors important to the initiation and maintenance of substance abuse. Drugs are viewed as reinforcers when they promote and maintain drug seeking and drug self-administration as a function of the effects the drug produces. Those effects may include the drug's pharmacologic effects as well as various nonpharmacologic consequences. Two mechanisms of drug reinforcement are hypothesized. First, the drug produces an inferred "mood-elevating" or "euphorigenic" effect and thus acts as a positive reinforcer. In the second, the drug acts as a negative reinforcer by reversing an inferred "aversive" state, such as a withdrawal syndrome, insomnia, or excessive sleepiness-fatigue. The two processes, positive and negative drug reinforcement, are not necessarily mutually exclusive and may operate concurrently or at the different stages in a drug abuse cycle (i.e., its initiation, maintenance, or relapse). These two reinforcement processes lead to the initiation and maintenance of excessive and hazardous drug use. How the sleep-wake state–altering consequences of the drugs addressed in this chapter may function as positive or negative reinforcers can be illustrated.

The alerting effects of stimulants are reinforcing for individuals who are sleepy, fatigued, and having difficulty in functioning to their desired level. Healthy normals will self-administer a stimulant when they are sleepy but not when alert.[1] That self-administration does not necessarily imply abuse. In substance abuse, however, the sleepiness may be present as part of a withdrawal syndrome due to abstinence following chronic nonmedical use of a stimulant. It has been hypothesized that continued substance use, difficulty in reducing use, or relapse may reflect "self-medication" to reverse the excessive sleepiness of the abstinence. For example, in chronic caffeine or nicotine dependence, the 8-hour sleep period is functionally an enforced abstinence, and given the pharmacokinetics of these drugs, the 8-hour abstinence during the sleep period is followed by enhanced sleepiness in the morning and, in extreme cases, smoking during the night. Caffeine or nicotine taken immediately on arising reverses the sleepy state. "When do you have your first morning cigarette?" is a question clinicians can use to gauge the severity of nicotine addiction. In amphetamine or cocaine abuse, excessive sleepiness during the initial drug abstinence has been consistently reported. Again, use of these stimulants will reverse the sleepiness.

During a period of chronic drug use, daytime sleepiness may also result from a drug-induced disturbance of nocturnal sleep. All the stimulants as reviewed later disrupt nocturnal sleep to some degree, depending on dose and proximity of their use to the sleep period. Disrupted and fragmented sleep produces daytime sleepiness. One could hypothesize that a drug-induced sleep disturbance at night leads to daytime sleepiness, which then enhances the likelihood of the self-administration of a stimulant. This is the common vicious circle seen in heavy coffee drinkers.

The state of sleepiness may not necessarily be drug induced. It may also be due to chronic insufficient sleep in healthy normals or to disturbed sleep efficiency and circadian dysrhythmia seen in altered work schedules. We have already noted that healthy normals will self-administer a stimulant when experiencing sleepiness. Night workers and rotating shift workers have shortened and disturbed sleep when sleeping during the day as well as increased sleepiness when awake at night. Rotating shift workers and night workers report a disproportionate use of sedating drugs, especially alcohol, to improve sleep and stimulants, especially caffeine, to improve alertness.[2,3] This substance use may increase risks of substance abuse.

The sedative-hypnotics, tetrahydrocannabinol (THC), and alcohol may become reinforcers and lead to substance dependence or abuse through their capacity to induce sleep in persons with insomnia or to reverse a waking "hyperaroused" state. People with insomnia and no history of alcoholism or drug abuse, given an opportunity to choose between previously experienced color-coded ethanol and placebo beverages

at night before sleep, chose ethanol, whereas healthy normals with a similar level of self-reported social drinking chose placebo.[4] Interestingly, people with insomnia not only showed nighttime self-administration of benzodiazepine receptor agonists but also self-administered them during the daytime. However, only those who showed evidence of daytime physiologic hyperarousal self-administered during the day.[5] Whether these drug self-administration patterns reflect drug seeking or therapy seeking needs to be considered, and the issue is further discussed later.

Neurobiologic Mechanisms

The positive-reinforcement neurobiology of drug self-administration is generally accepted to involve mesocorticolimbic projections originating in the ventral tegmental area of the rostral reticular activation system and terminating in the nucleus accumbens. It is a dopaminergic system.[6] Most drugs of abuse interact in some way with this system. This reinforcement system is part of a broader dopaminergic system that projects into several forebrain regions and as a whole is considered to have executive and integrative functions.[6] The dopaminergic neurons of the ventral tegmental area are modulated by a number of other neurotransmitter systems. It is through this modulation that sleep-wake state and level of sleepiness-alertness during wake could have an impact on a drug's reinforcing properties.

In chronic use, it is hypothesized by Koob that the neurobiologic systems involved in acute positive reinforcement adapt by establishing "opponent" processes.[6] These opponent processes neutralize the acute reinforcing effects of the drug and during abstinence are left unopposed. Consequently, they produce the abstinence syndrome, an inverse state to the drug state, which becomes a basis of negative reinforcement (i.e., reversal of abstinence symptoms). For example, stimulant abstinence produces sleepiness, which in turn leads to self-administration of stimulant drugs. Koob hypothesized that the opponent processes do not necessarily develop through the same neurobiologic system that produces positive reinforcement. Other neurobiologic systems and, for the purposes of this chapter, possibly sleep-wake systems could be one basis of the opponent processes. Two important sleep-wake findings are consistent with this model. They are the REM suppression and REM rebound associated with administration and discontinuation of most all of the drugs with an abuse liability. Second, the persisting disturbance of sleep that is found after weeks and months of abstinence suggests that some type of neurobiologic adaptation within sleep-wake systems to the chronic drug exposure has occurred.

SLEEP-WAKE ALTERATIONS AND SPECIFIC SUBSTANCES

Alcohol and Alcoholism

Epidemiology and Risk Factors

The U.S. prevalence of alcohol dependence by age, sex, and race-ethnicity is available at www.niaaa.nih.gov. For 2012, the past-month prevalence of alcohol dependence among those 21 years and older was estimated to be 6.9%, with the rate being 10.8% for men and 3.4% for women. The rates for those aged 12 to 20 years were 4.3% for heavy drinking and 15% for binge drinking. These rates can be much higher in general medical or psychiatric clinics, with 20% to 50% of men and 6% to 10% of women suffering from alcoholism.

Alcohol in Healthy Adults

Alcohol in large doses is mildly stimulating on the rising phase of the plasma concentration curve and sedating on the declining side of the curve.[7] In lower doses, it has little stimulatory effects on the rising phase and is mildly sedating on the declining phase. Studies of alcohol effects on sleep typically administer alcohol 30 to 60 minutes before sleep, which results in peak concentrations occurring before or at bedtime. The studies have used doses ranging from 0.16 to 1.0 g/kg, the rough equivalent of one to six standard drinks, producing breath ethanol concentration (BrEC) up to 0.105% at bedtime.[8] Sleep latency is reduced over this dose range, and one study reported increased sleep time at a low 0.16 g/kg dose but not at the higher 0.32 and 0.64 g/kg doses of that study. Improved sleep *only* at low doses is likely due to a second half of the night sleep disruptive rebound wakefulness that occurs with the higher doses. The typical BrEC at lights out for higher dose studies is between 0.05% and 0.09%, and given that ethanol is metabolized at a rate of 0.01% to 0.02% BrEC per hour, within the first 4 to 5 hours of the sleep period, ethanol has been completely metabolized. This leads to rebound wakefulness during the last hours of the sleep period.[9] Thus, for the whole night, sleep time is not increased at high doses and often is actually decreased.

In addition to these effects on sleep induction and maintenance, ethanol affects the normal progression of sleep stages.[10] A dose-dependent suppression of REM sleep, at the least in the first half of the night (i.e., when ethanol blood levels are present), and in *some* studies increased slow wave sleep in the first half of the night are reported. Slow wave sleep enhancement occurs in subjects with less than age-corrected amounts of slow wave sleep at baseline. When first-half REM sleep suppression is observed, a second-half REM sleep rebound is reported as well. As with rebound wakefulness, the second-half REM sleep rebound likely relates to the timing of complete ethanol elimination from the body. Repeated nightly administration of ethanol leads to tolerance development to both the sleep induction and sleep stage effects but interestingly not to the rebound effects. Finally, discontinuation of ethanol is followed by a REM sleep rebound, although the appearance of a REM sleep rebound is likely related to dose, duration of use, basal level of REM sleep, and extent of prior REM sleep suppression and tolerance development.

The aftereffects of heavy alcohol consumption, commonly referred to as hangover, are typically experienced after peak BrECs of 0.100% and greater.[11] Some laboratory studies of heavy drinking, hangover, and next-day cognitive and psychomotor performance have demonstrated impairment in the morning with BrEC zero 14 hours after ethanol ingestion the previous night and approximately 4 hours before going to bed.[12] Some studies have related the next-day impairment to the degree of ethanol-related sleep loss and fragmentation during the previous night.[13] Even with low alcohol doses and in the absence of hangover symptoms, the sleep-disruptive and performance-impairing effects may continue after alcohol is completely metabolized and BrEC is zero. Late afternoon drinking with BrEC zero at bedtime disrupted sleep in the second half of the night, and morning or midday drinking

continues to impair performance for 2 to 3 hours after BrEC is zero.[14,15]

Several clinical implications of these alcohol effects in healthy people might be discussed. Regular heavy drinking and chronic insomnia complaints, particularly sleep maintenance insomnia, are easily identified; but occasional insomnia can be related to a patient's occasional heavy drinking. Sleep diaries that query about social drug use as well as the sleep complaints can help identify for the clinician and patient this relation of the alcohol use and sleep problems. Inquiry as to the quantity and timing of the patient's alcohol consumption and attempt to sleep may reveal the sleep-disruptive potential of the alcohol. Just as important are the potential next-day impairing effects of evening alcohol consumption as discussed earlier. Recognition that sleep-disruptive alcohol consumption combined with sleepiness due to other causes can have additive daytime impairing effects is also important.

Alcohol Effects in Primary Sleep Disorders

Approximately 30% of individuals with insomnia in the general population use alcohol to help them sleep, and 67% of those people have reported that the alcohol was effective in inducing sleep.[16] The laboratory studies of healthy normals showing sleep-disruptive ethanol effects used higher doses (e.g., BrEC >0.05% at bedtime, about five or six drinks) than the doses (e.g., one or two drinks) reportedly used by insomniacs. In laboratory studies of people with primary insomnia, ethanol administered 30 minutes before sleep, which raised BrEC to only 0.04% at bedtime, improved sleep without producing a second-half wakefulness rebound.[17] Compared with age-matched people without insomnia and with a similar social drinking history, those with insomnia chose to self-administer alcohol before sleep more frequently. The risk associated with using alcohol as a sleep aid is that tolerance to its initial beneficial effect develops within three to five nights, and people with insomnia increase their self-administered dose to compensate.[18] Later in the chapter, guidelines are offered for determining when what could be initially considered therapy seeking (e.g., use of alcohol as a sleep aid) has shifted to drug seeking.

Alcohol is a mild respiratory depressant in the wake state, and during sleep it exacerbates obstructive sleep apnea syndrome and may precipitate sleep disordered breathing in at-risk persons. In one study, patients with moderate obstructive sleep apnea, defined as an average respiratory disturbance index (RDI; number of apneas and hypopneas per hour of sleep) of 22, received 300 mL of bourbon 2 hours before bedtime.[19] That dose of ethanol increased the RDI to 28. In another study, patients with a range of sleep-related breathing disorders were studied.[20] An unspecified dose and BrEC increased the RDIs of every patient with baseline RDI between 14 and 54. The effects of ethanol are also found in other sleep-related respiratory disorders. For example, in several patients with chronic obstructive pulmonary disease, ethanol worsened the degree of hypoxemia, and in several patients with only a snoring history, it induced apnea.[20] That an asymptomatic snorer with no apnea will develop apnea after ethanol was convincingly shown in a later study.[21] In one study, ethanol (BrEC = 0.08%) in snoring men increased their RDIs to a pathologic level (i.e., RDI >10).[22] On the other hand, the findings in asymptomatic individuals without risk factors have been inconsistent.

The risk of periodic limb movements (PLMs) during sleep was increased in association with self-reported alcohol consumption in a sample of sleep disorders clinic patients.[23] However, in a sample of patients abstinent of alcohol, the rate of PLMs was not different from the rate of a general sleep disorders clinic sample.[24] Alcoholism is associated with deficiencies in iron, ferritin, magnesium, and vitamin B_{12} and polyneuropathy, all of which are also associated with PLMs and restless legs syndrome (RLS). To the extent that the patient with alcoholism has any of these deficiencies, RLS and PLMs may be precipitated or exacerbated.

Sleep and Alcohol Effects in Alcoholism

Patients with alcoholism commonly complain of sleep problems, daytime sleepiness, and parasomnias. During drinking binges, people with alcoholism report polyphasic sleep patterns, with short sleep periods followed by short wake periods that are distributed across the 24-hour day. This type of sleep pattern is seen in organisms without a circadian pacemaker (i.e., lesioned suprachiasmatic nuclei). It is possible that during these binges, the sleep-wake cycles of people with alcoholism are so chaotic that they are arrhythmic. It would be of interest to see the circadian pattern of temperature, dim light melatonin onset, and other phase markers in this population. Laboratory studies of patients with alcoholism show that sleep latency and total sleep time are disturbed on both drinking and discontinuation nights.[25] The prolonged sleep latency in alcoholism when drinking contrasts with the reduced sleep latency that alcohol produces in people without alcoholism and suggests tolerance development and possible neurobiologic changes as discussed before. In addition, as in insomnia and some healthy normals with basally low amounts of slow wave sleep, drinking in people with alcoholism is associated with increased slow wave sleep and REM sleep suppression, with rapid tolerance development to these effects.[25]

The acute alcohol abstinence phase lasts 1 to 2 weeks, although some of the withdrawal symptoms, such as mood instability, disturbed sleep, and craving, remain beyond this period. During the acute abstinence phase, slow wave sleep is reduced, sometimes to minimal levels, and REM sleep continuity is disrupted.[26] There are frequent REM episodes and shortened non-REM (NREM)–REM cycles during the acute abstinence.

Recovery and Long-term Abstinence

Some of these abnormal sleep patterns can persist for up to 3 years in some patients with alcoholism. Sleep remains shortened and REM sleep pressure elevated, as evident in elevated levels with shortened latencies to REM sleep.[27] Figure 140-1 illustrates the abnormally short sleep and elevated REM sleep percentage in abstinent primary alcoholism. Whereas it is tempting to attribute these sleep abnormalities to the excessive alcohol drinking of the patients, the sleep problems could have preceded the development of the alcoholism or they could be secondary to the development of other medical and psychiatric disorders during the alcoholic drinking of the patient.

Regardless of the cause of the sleep disturbance evident in alcoholism, both objective and subjective measures of sleep taken after the immediate acute abstinence predict the likelihood of relapse during the long-term abstinence. Early laboratory studies suggested that low levels of slow wave sleep are

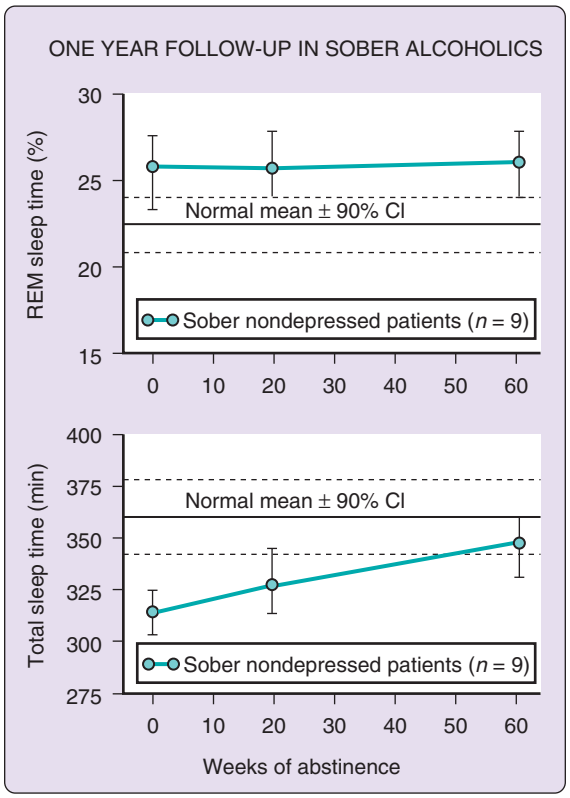


Figure 140-1 Polygraphic sleep recordings for rapid eye movement (REM) sleep percentage and total sleep time in nine male, sober, recovering, primary nondepressed, alcoholic patients for more than 1 year. Note that both measures differ from those of age-matched healthy control subjects for most of the recovery period. Values for the patients are mean ± standard error of the mean; values for healthy control subjects are mean + 90% confidence interval (CI). (Modified from Drummond SP, Gillin JC, Smith TL, Demodena A. The sleep of abstinent pure primary alcoholic patients: natural course and relationship to relapse. *Alcohol Clin Exp Res* 1998;22:1796–802.)

predictive of alcoholism relapse.[28] More recent studies have identified REM sleep disturbances, either elevated REM sleep percentage or shortened REM sleep latency, as predictive of relapse.[29] Interestingly, that risk was greater than that associated with other variables, such as age, marital status, employment, duration and severity of alcoholism, hepatic enzymes, and depression ratings.

Treatment of the Sleep Disturbance of Alcoholism

Treatment of sleep disturbance in patients with alcoholism is challenging as there are few placebo-controlled studies to guide the clinician. The benzodiazepine receptor agonists are the drug class of choice for insomnia treatment in patients without alcoholism. However, whereas these drugs have a relatively low abuse liability in those without alcoholism, their risk in outpatients with alcoholism after acute inpatient withdrawal is unknown.[30] These drugs are effective for the immediate inpatient withdrawal syndrome because they share the same mechanism of action as that of alcohol itself, promotion of γ-aminobutyric acid inhibition. A further caution to their outpatient use is that they have a high potential for toxicity and overdose when combined with alcohol, thereby being dangerous for the people with alcoholism who relapse.

Sedating antidepressant medications, such as trazodone and doxepin, are often used in the United States to treat primary insomnia and insomnia comorbid with depression. Trazodone has been used to treat sleep disturbance in alcoholism patients in open-label, noncontrolled studies.[30] In doses of 50 to 300 mg, it improved self-reported measures of insomnia. Currently, the three Food and Drug Administration–approved alcoholism treatments are disulfiram, naltrexone, and acamprosate. Unfortunately, sleep problems have not been major outcome measures in those studies testing the efficacy of these agents used as alcoholism treatments. One exception is a recently completed large trial of the anticonvulsant gabapentin, in which both sleep and drinking outcomes were improved.[31]

Cognitive-behavioral therapy for insomnia (CBT-I) is an alternative to medications, and several trials in alcoholism with sleep disturbance have been conducted. A randomized controlled trial of brief CBT-I associated with alcoholism was done in 60 patients without comorbid depression.[32] The CBT-I treatment, compared with wait-list controls, improved sleep diary measures of sleep quality, sleep efficiency, awakenings, and time to fall asleep. However, CBT-I had no impact on drinking relapse rates during the 6-month follow-up period. An open trial of CBT-I in patients with alcoholism similarly found improved sleep but not improved drinking outcomes.[33]

It is tempting to assume that the excessive alcohol intake is the sole or primary cause of insomnia in alcoholism. Earlier it was noted that other primary sleep disorders and possibly circadian disorders are present in patients with alcoholism at a higher rate than in those without alcoholism. Furthermore, depression is often present in alcoholism, and the sleep disturbances of depression have been well documented. Appropriate treatment of the comorbid sleep or psychiatric disorder should be pursued. Insomnia treatment in patients with alcoholism must begin with referral to addiction treatment specialists. Treating the insomnia alone will not initiate abstinence. Patients with alcoholism often deny their drinking problems and may focus on their sleep problems to avoid confronting their drinking problems.

Epidemiology and Risk Factors for Drugs of Abuse

Prevalence data on abuse of specific drugs other than alcohol are highly variable across years and regions of the country. U.S. prevalence estimates on use and dependence are collected annually by the U.S. government Office of Applied Studies in the Substance Abuse and Mental Health Services Administration. Prevalence tables by drug, age, sex, and region of the country for lifetime, past year, and past month use and for dependence or abuse are available at www.oas.samhsa.gov. In the 2013 National Survey on Drug Use and Health, 9.4% of Americans 12 years and older were current illicit drug users, which includes marijuana, cocaine, heroin, hallucinogens, inhalants, and the nonmedical use of prescription medications.

Stimulants

Caffeine

Caffeine often is not considered a drug of abuse by society, and even within the medical community, its potential for abuse is not fully appreciated. Laboratory data indicate there are conditions under which caffeine is persistently

self-administered and that it does have abuse liability, albeit a relatively low liability compared with other recognized drugs of abuse.[34] As might be expected, caffeine in doses of 150 to 400 mg administered immediately before sleep prolongs the onset of sleep and reduces total sleep time in healthy normals.[34-39] To compare the disruptive effects of caffeine with the psychomotor stimulants, in one study 300 mg of caffeine reduced sleep efficiency from 89% to 74%, whereas 40 mg of pemoline reduced it to 80%, and 20 mg of methylphenidate reduced it to 61%.[36] As to sleep stage effects, some studies report reductions of stage 3–4 sleep,[35,36] but unlike with the psychomotor stimulants, stage REM sleep is not affected.

Discontinuation of the chronic use of caffeine is associated with mood and performance disturbance. A withdrawal syndrome was observed after a double-blind, placebo-controlled cessation of chronic but moderate (235 mg daily on average) caffeine consumption.[37] Putting the dose in context, an 8-ounce cup of coffee contains 100 mg of caffeine. On the second day of caffeine cessation (20 hours after caffeine use), in addition to the ubiquitous headache, reduced vigor and increases in sleepiness, fatigue, and drowsiness were experienced. For moderate to heavy caffeine users, the morning cup of coffee immediately after arising probably restores caffeine levels and alertness as the 8-hour sleep period is essentially a caffeine discontinuation.

Nicotine

That nicotine is a drug of abuse is undisputed, and some have argued that it rivals the abuse liability of cocaine. The study of nicotine's effects on sleep has been facilitated with the development of nicotine delivery systems (i.e., nicotine gum and patches) for use in clinical smoking cessation programs. In healthy normals, transdermal nicotine (7 to 14 mg) produced a dose-dependent increase in wakefulness and a reduction in percentage REM sleep relative to a placebo patch.[40] A second study of nonsmoking normals also found that 17.5 mg transdermal nicotine increased wake time and decreased REM sleep percentage.[41] In obese, nonsmoking patients with sleep disordered breathing, 15 mg transdermal nicotine reduced both total sleep time and percentage REM sleep; it did not improve the sleep disordered breathing of these patients, which was the study's primary purpose.[42]

The discontinuation of nicotine in nicotine-dependent individuals is associated with a disturbance of sleep and alertness. In an uncontrolled study that compared the sleep of smokers during the week before and after cessation of chronic smoking, the number of arousals, awakenings, and sleep stage changes were all increased during the cessation week.[43] In a double-blind study using a placebo versus nicotine patch during the discontinuation of the average daily use of 30 cigarettes, the number of arousals was increased relative to the smoking baseline in the placebo group, whereas in the nicotine (22 mg) patch group, arousals were reduced and stage 3–4 sleep was increased relative to the smoking baseline.[44] However, given the pharmacokinetics of nicotine, any smoking baseline is a partial discontinuation during the usual 8-hour sleep period of nonsmoking. Because the nicotine patch was worn continuously, it is probable that sleep was improved relative to a partial discontinuation (i.e., a smoking baseline). Such is consistent with the several questionnaire studies that find smokers are more likely than nonsmokers to report problems falling asleep and staying asleep.[45]

Psychomotor Stimulants

Cocaine. Cocaine is the CNS stimulant with the greatest abuse liability. Some laboratory studies have examined the effects of cocaine and its discontinuation on sleep and daytime alertness. Clinical assessments have described continued and prolonged wakefulness during "cocaine runs," days of protracted cocaine use. The cocaine runs are then followed by "crashes," which are characterized by excessive sleep and sleepiness.[46] Several polysomnography (PSG) studies of cocaine discontinuation in cocaine-dependent persons have reported elevated sleep time and a REM sleep rebound during the initial abstinence. That disturbance was then followed by a persisting insomnia and REM sleep disturbance for the 3-week study duration.[47] A laboratory study of cocaine administration and discontinuation found that 600 mg/day intranasal cocaine (insufflated from 1900 to 2100 hours) severely disrupted sleep, delaying its onset at the 2300 hours bedtime for up to 3 to 4 hours and suppressing REM sleep.[48] During the first two discontinuation days, average daily sleep latency on the Multiple Sleep Latency Test (MSLT) was less than 5 minutes (i.e., a pathologic level of sleepiness), most probably because of the severe sleep disturbance of the prior 5 days of cocaine self-administration. The MSLT also showed multiple sleep-onset REM periods, probably due to the prior REM sleep suppression during the cocaine administration nights. Then, even after 14 days of abstinence, a nocturnal sleep and REM sleep disturbance remained, although the MSLTs were free of sleep-onset REM periods and the sleep latencies returned to normal levels. Later, we discuss how these state-altering effects may serve to maintain cocaine dependence. Some PSG studies have now shown that slow wave sleep and REM sleep disturbances are predictive of relapse during short-term abstinence.[49]

Amphetamine. Amphetamine is also a drug of abuse. During the 1950s, before its manufacture and distribution became tightly controlled, an epidemic of amphetamine abuse occurred in the United States. Although it is not the drug of choice, it has therapeutic indications for the treatment of narcolepsy and attention deficit/hyperactivity disorder (ADHD). Few PSG studies of amphetamine administration and discontinuation have been done. In one of the earliest PSG drug studies, 10 or 15 mg of *d*-amphetamine doubled sleep latency and suppressed REM sleep in healthy young adults.[50] Another study of healthy adults and patients with narcolepsy reported that a 7 to 8 AM administration of methamphetamine (10 mg) reduced sleep efficiency that night (11 PM bedtime) in control subjects and at 40- to 60-mg doses in the narcoleptics.[51] No REM sleep effects were observed in the normals. However, the patients received higher doses, and their REM sleep latency was increased and REM sleep time was reduced. A study of amphetamine-dependent subjects assessed sleep during the drug's discontinuation.[52] On the second night after the last amphetamine dose, REM sleep time rebounded and remained elevated for three to five nights. This REM sleep rebound was delayed relative to that associated with cocaine discontinuation, which may be due to the pharmacokinetic differences between the drugs (i.e., the longer half-life of amphetamine). The rebound also was longer lasting, which may be due to differences in the duration or amount of prior use or in the level of REM sleep suppression caused by the

drug. Total sleep times were elevated during the same 3- to 5-day period, probably reflecting recovery sleep due to the prior drug-induced sleep loss, but subsequently sleep times became shorter than normal, suggesting continuing insomnia. As in abstinence from cocaine dependence, the continued insomnia in amphetamine abstinence raises questions about a stimulant-induced persisting alteration of CNS sleep-wake mechanisms.

No studies of the effects of amphetamine discontinuation on daytime levels of sleepiness-alertness have been done in amphetamine-dependent persons. One would predict that amphetamine discontinuation is associated with increased daytime sleepiness and multiple sleep-onset REM periods on the MSLT similar to what is reported for the discontinuation of cocaine.

Methylphenidate. Methylphenidate is generally considered a safer drug as regards its abuse liability, although case reports of its abuse have appeared over the years. Its indications are for ADHD and narcolepsy. It is the drug of choice for ADHD and a second-choice drug for narcolepsy. In healthy normals, 20 mg reduced total sleep time, increased the latency to REM sleep, and reduced the minutes of REM sleep, whereas 10 mg increased only REM sleep latency.[53] In an earlier study, 5 mg was reported to increase the latency to REM sleep and to reduce the percentage of REM sleep without affecting other sleep measures.[54] In children with ADHD, studies of the effects of methylphenidate taken during the day on subsequent sleep are inconclusive. Sleep time was shortened in one study[55] and increased in another,[56] and in one study REM sleep was fragmented.[56] However, these studies can be questioned for a variety of methodologic and control issues.

We are unaware of studies that have documented the discontinuation effects of methylphenidate in dependent individuals. Increased daytime sleepiness with multiple sleep-onset REM periods would be predicted. The role that the excessive sleepiness following discontinuation of any of these stimulant drugs may have in their continued use and abuse is illustrated by self-administration studies done in healthy normals. When given an opportunity to self-administer methylphenidate (20 mg), healthy normals, with no current or previous substance abuse history, choose active drug on only 20% of the opportunities after 8 hours of time in bed the previous night but on 80% of the opportunities (i.e., 20% placebo choice) after 4 hours of time in bed.[1] These data are consistent with and extend the cocaine data, showing that stimulants have performance-enhancing effects in sleep-deprived healthy normals and profoundly reinforcing effects for sleepy individuals or normal individuals experiencing sleep loss.

Opioids

The opioids are drugs with a high abuse liability. They have disruptive effects on sleep continuity and sleep staging. Heroin (3, 6, and 12 mg/70 kg) administered intramuscularly in a double-blind placebo-controlled design to abstinent opioid-dependent people produced dose-related decreases in total sleep time, stage 3–4 sleep, and REM sleep.[57] Heroin is metabolized to morphine, and as would be expected, morphine (7.5, 15, and 30 mg/70 kg) also produced a dose-related decrease in sleep time, stage 3–4 sleep, and REM sleep in the abstinent opioid-dependent people.[58] Finally, methadone (7.5, 15, and 30 mg/70 kg) also decreased total sleep time, stage 3–4 sleep, and REM sleep in abstinent opioid-dependent people.[59] All these drugs also produced increased brief arousals and frequent sleep stage changes (i.e., sleep fragmentation). Several studies have indicated that tolerance to the sleep-disruptive effects develops within weeks.[60] With tolerance development, the sleep fragmentation is lessened, and the REM sleep–suppressive effects tend to diminish.

The discontinuation of chronic opioid use is associated with sleep disturbance. Heroin-dependent people being maintained on methadone in an outpatient research treatment program were discontinued from their treatment and sleep was recorded in the laboratory.[61] Sleep latency and latency to REM sleep were prolonged and percentage stage 3–4 sleep was reduced compared with normals. Interestingly, after a week of buprenorphine 4 mg, a partial μ opioid agonist, the sleep patterns normalized.

Opioids also have an impact on other sleep disorders. Although not the first line of treatment, opioids improve RLS and are often used in patients who do not respond to dopamine agonists. In contrast, given negative effects on respiratory drive, they are contraindicated in sleep apnea.

Sedative-Hypnotics

Within society and the medical community, sedative-hypnotics are considered to have a relatively high abuse liability. The scientific evidence from epidemiologic and laboratory studies indicates that the abuse liability of modern sedative-hypnotics is relatively low. The distinction "modern" is critical in that the early sedative-hypnotics, barbiturates and ethanol-based drugs (i.e., ethchlorvynol, choral hydrate), clearly produce both physical and behavioral dependence, but the modern benzodiazepine receptor agonists are not as likely to do so.

In the 1980s, daytime self-administration studies were done in normals, persons with substance abuse histories, and patients with anxiety disorders; the results showed a generally low behavioral dependence liability. The benzodiazepine receptor agonists were self-administered by people with a substance abuse history at low and declining rates over time[62] and were not differentially self-administered relative to placebo by the normals or patients with anxiety disorders.[63,64] Research has found that some patients with anxiety disorders self-administered alprazolam relative to placebo.[65] Our short-term studies of pre-sleep zolpidem or placebo self-administration in patients with insomnia have found that triazolam and placebo are self-administered at similar rates (67% to 88%) in a single-choice paradigm (i.e., choice of taking the available capsule or not).[66-68] However, when subjects were forced to choose on a given night between triazolam and placebo, triazolam was preferred 80% of the time.[68] Most important, when subjects were given an opportunity, in a single-choice paradigm, to self-administer multiple capsules on the same night, a 0.27-mg average nightly triazolam dose was self-administered (i.e., one capsule), whereas the placebo dose was escalated to the three-capsule maximum. Long-term studies of nightly use of zolpidem or placebo for 1 year showed that (1) zolpidem was preferred to placebo by people with insomnia, but its overall rate of self-administration did not increase during 12 months of nightly use (i.e., a stable average nightly dose of 9.5 mg was self-administered), whereas placebo self-administration did increase during the 12 months, again to the three maximum capsules[69]; (2) 12 months of

nightly zolpidem use did not produce rebound insomnia or withdrawal signs and symptoms after chronic use[70]; and (3) relative to placebo, zolpidem increased total sleep time, and this hypnotic efficacy did not diminish during 8 months of nightly use.[71]

Hallucinogens

THC

THC is one of the principal active ingredients in marijuana. In the United States, marijuana still remains the most frequently abused illicit drug, although its popularity appears to cycle, trending up and down over decades. THC is classified as a mild sedative at low doses and a hallucinogen at high doses.

Study of THC's effects on sleep and wakefulness occurred predominantly during the 1970s. Low doses of THC (4 to 20 mg), in either experienced marijuana users or nonusers, had mildly REM sleep–suppressive effects.[72-74] In several of the studies, total sleep time or stage 3–4 sleep was increased,[72,74] which then decreased after a week of repeated nightly use.[74] At high doses (50 to 210 mg) in naive or experienced marijuana users, THC again suppressed REM sleep, but effects on total sleep time were not observed,[75] and stage 3–4 sleep was reduced in one report.[76] Some of the studies reported a REM sleep rebound on the THC discontinuation nights,[72,75] and some reported a reduction in sleep time or an increase in sleep latency.[73,75] All of these studies have involved pre-sleep laboratory administration of THC. Several studies in situ or in semicontrolled laboratory situations with moderate and heavy marijuana users smoking their usual marijuana cigarettes during the day or early evening have also been done. Self-reported or rater-observed sleep time was increased in two such studies.[77,78] In a PSG study, little or no effects on sleep measures were observed.[79] These mild sedative effects have also been observed during daytime studies using performance assessments. In situ assessment of daily activity levels also showed reduced activity during moderate or heavy marijuana use.[78]

MDMA

An amphetamine derivative that has become increasingly popular as a recreational drug and drug of abuse is (±)3,4-met hylenedioxymethamphetamine (MDMA) or "ecstasy." This drug has hallucinogenic properties and acts indirectly by stimulating the release of brain monoamines.[80] Exposure to MDMA decreases total sleep time with a decrease in NREM sleep but no significant effects on REM sleep. Among NREM sleep stages, individuals who use MDMA have less stage 2 sleep, but there are no apparent differences in stage 1 or slow wave sleep (stages 3 and 4).[81] In a placebo-controlled study, acute MDMA shortened sleep primarily by increasing sleep latency and reducing stage 3–4 sleep and suppressing REM sleep.[82] The MDMA-reduced sleep time was not associated with increased daytime sleepiness the following day as seen in a 4-hour sleep restriction condition (Figure 140-2). Average daily sleep latency on the MSLT the day after nighttime placebo was increased in MDMA users compared with age-matched controls, and MDMA users had an elevated number of sleep-onset REM periods compared with controls.

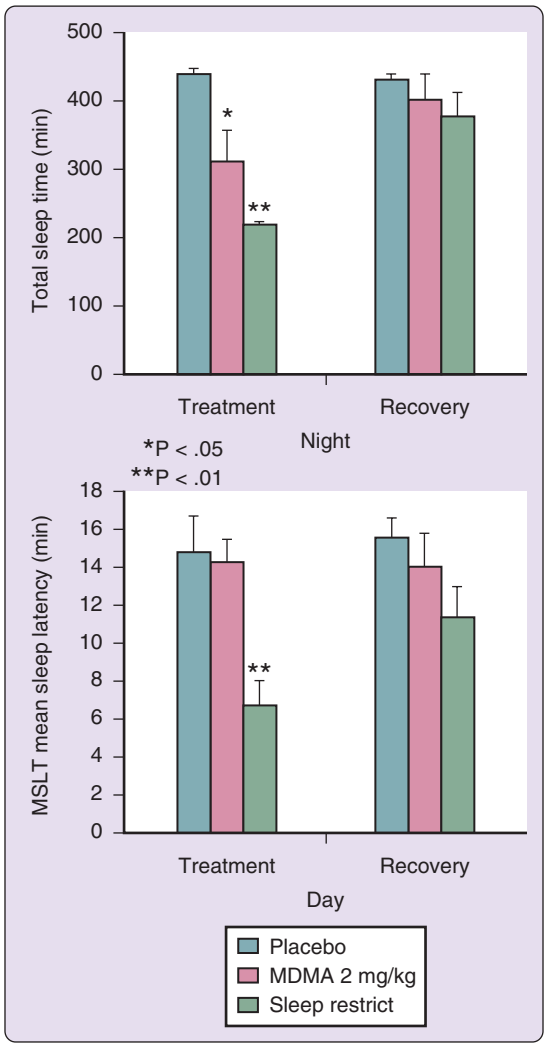


Figure 140-2 Polygraphic sleep recordings for total sleep time and mean sleep latency on the Multiple Sleep Latency Test (MSLT) were less than 5 minutes (i.e., a pathologic level of sleepiness), most probably because of the severe sleep disturbance the following day in moderate MDMA users (n = 7) after placebo, MDMA 2 mg/kg, or a 4-hour restricted bedtime. Drug or placebo was administered at 1800 hours, and bedtime was 2300 to 0700 hours on drug or placebo nights and 0300 to 0700 hours on sleep-restricted nights. Note the unusually high MSLT times after placebo and again after MDMA despite the MDMA-shortened sleep time the previous night. After sleep restriction, the MSLT time did decline. (Modified from Randall S, Johanson CE, Tancer M, Roehrs T. Effects of acute 3,4-met hylenedioxymethamphetamine on sleep and daytime sleepiness in MDMA users: a preliminary study. *Sleep* 2009;32:1513–9.)

Treatment of the Sleep Disturbance in Drug Abuse

The literature on treating sleep disturbance in drug abuse is even more limited than that for alcoholism, but the few studies are provocative. A trial of CBT-I and daytime sleepiness in adolescents improved the sleep of those completing more than four sessions.[83] The improved sleep showed a trend toward reducing substance abuse problems at the 1-year follow-up. A treatment trial of nicotine-dependent adults compared a 16-hour versus 24-hour nicotine patch during smoking abstinence.[84] The patches reduced smoking urges, with the 24-hour patch having a greater effect than the 16-hour patch. The 24-hour patch improved sleep, specifically the amount of slow wave sleep. This result in nicotine-dependent adults is in contrast to the sleep-disruptive effects of a nicotine patch on the sleep of nonsmokers. In the drug abuse treatment literature, sleep is rarely included as an outcome measure. The need for

Table 140-1 Characteristics of Drug-Seeking versus Therapy-Seeking Behavior

Drug-Seeking Behavior

- Drug chosen over other commodities or activities
- Drug taken in excessive, nontherapeutic amounts
- Drug taken on chronic basis, leading to tolerance and physical dependence
- Drug taken in nontherapeutic context

Therapy-Seeking Behavior

- Drug has demonstrated efficacy
- Duration and dose of self-administration are limited to therapeutic effects
- Patient has signs and symptoms for which drug is indicated
- Drug is believed to be and is experienced as being efficacious

clinical trials that focus on treatment of sleep complaints in substance abuse is clearly evident.

CONTROVERSIES AND PITFALLS

Drug Seeking versus Therapy Seeking

It is often difficult to differentiate drug seeking from therapy seeking in clinical practice. In drug seeking, the drug and its effects, typically its euphorigenic effect, are the focus of the drug use, whereas in therapy seeking, the alleviation of disease-related symptoms is the focus of the drug use. However, in the clinic, drug seeking and therapy seeking can become closely intertwined, and what was once therapy seeking can shift to drug seeking. The clinical challenge to sleep disorders clinicians is to differentiate the two phenomena in making diagnoses and appropriately treating patients. Some of the potentially differentiating and defining characteristics of drug seeking versus therapy seeking are presented in Table 140-1. The defining characteristic of drug seeking is evidence that the drug is taken in excessive amounts and in nontherapeutic contexts and is preferred to other commodities (e.g., money) and various social and occupational activities. The degree to which the drug is chosen over other commodities or social activities is evidence for the extent of its risk for abuse. Supportive of its reinforcing capacity is evidence in the scientific literature that the drug is readily discriminated from placebo by behavioral and subjective assessment. These assessments typically rate the drug for it euphorigenic and drug-liking effects, that is, the drug's subjective effects are the focus of its use. Then, to the degree that the dose is escalated over time, one has evidence of the development of tolerance and possible physical dependence.

In contrast, therapy seeking is evident if the drug has demonstrated efficacy for the disorder or condition being treated. As well, the patient has the signs and symptoms and the appropriate diagnosis for the indicated use of the drug. The pattern of drug taking, its dose and duration of use, is consistent with its therapeutic effects. Finally, the patient believes that the drug is effective and readily experiences its therapeutic effects. However, the distinction between drug seeking and therapy seeking becomes difficult in situations in which therapy-seeking behavior shifts to drug-seeking behavior. For example, one is concerned about the use of ethanol as a hypnotic by an insomniac. Whereas pre-sleep ethanol use may initially be effective in improving sleep, rapid tolerance development is likely, which may lead to dose escalation. Further, other reinforcing effects of ethanol (i.e., its euphorigenic effects) may be discovered by the person, especially as dose is escalated, and its use may then extend beyond the therapeutic context (i.e., before sleep as a hypnotic). A similar shifting pattern can be described for stimulant or opiate use. On the other hand, drug seeking may be maintained because the drug, in addition to its mood-altering and euphorigenic effects, also has therapeutic effects (i.e., the stimulant effects of cocaine or amphetamine do reverse the excessive sleepiness that is experienced during drug discontinuation). Thus, the dependence is maintained by both its mood-altering effects and its therapeutic effects, what others have termed self-medication (i.e., both positive and negative reinforcing effects).

Sleep-Wake Disturbance in Alcoholism and Drug Abuse

The role of sleep-wake disturbance in alcoholism and substance abuse is not well understood. It is known that insomnia is a risk factor for substance abuse. Yet, earlier in discussing diagnoses, the point was made that sleep-wake disturbances are not major criteria in making DSM-5 abuse diagnoses. The extent to which insomnia or daytime sleepiness leads to new cases of alcoholism or drug abuse is not known. Furthermore, the degree to which treatment of insomnia or daytime sleepiness in abstinent alcoholism and drug abuse reduces risk of relapse is not known. To date, the few alcoholism treatment trials have failed to clearly demonstrate that improved sleep reduces relapse, and the only drug abuse treatment trial, although encouraging, is not conclusive. There is the inherent assumption in this discussion that sleep disturbance is causally related to alcoholism or drug abuse, as either the precipitant or consequent, but it may be comorbid and independent or related to a third common factor.

CLINICAL PEARL

The sleep clinician, in assessing patients with sleep complaints, must query about substance use, including prescribed and over-the-counter drugs, legal recreational drugs, and illegal recreational drugs. Attention to the quantity and pattern of use is important. When substance abuse is suspected, referral to a substance abuse specialist is desirable. When drugs with a known abuse liability are prescribed, patients should be observed closely to monitor their appropriate therapeutic use. Use of alcohol as a sleep aid should be discouraged as tolerance develops rapidly to its low-dose beneficial effects.

SUMMARY

Almost all of the psychoactive drugs with a known abuse liability have profound effects on sleep and wakefulness. Substance abuse is characterized by physiologic (e.g., withdrawal syndrome on drug discontinuation) and behavioral (e.g., repetitive, compulsive drug seeking and consumption and

relapse) dependence, in which the relative contribution of each can vary according to the drug and its duration of use. The neurobiologic mechanisms underlying both physiologic and behavioral dependence are known, but the role of the abused drug's sleep-wake state–altering effects in substance abuse is not fully known, although it is likely to be important. Some of the drugs of abuse are legal and widely used socially and may be the cause of patients' sleep or alertness complaints. Other drugs with abuse liability are drugs indicated in the treatment of sleep disorders. For that reason, guidelines have been provided for sleep disorders clinicians to help in differentiating drug-seeking behavior from therapy-seeking behavior.

Selected Readings

Bonnet MH, Arand DL. Caffeine use as a model of acute and chronic insomnia. *Sleep* 1992;**15**:526–36.

Brower KJ. Insomnia, alcoholism and relapse. *Sleep Med Rev* 2003;**7**:523–39.

Garcia AN, Salloum IM. Polysomnographic sleep disturbances in nicotine, caffeine, alcohol, cocaine, opioid, and cannabis use: a focused review. *Am J Addict* 2015;**24**(7):590–8.

Griffiths RR, Weerts EM. Benzodiazepine self-administration in humans and laboratory animals. Implications for problems of long-term use and abuse. *Psychopharmacology (Berl)* 1997;**134**:1–37.

Johanson CE, Roehrs T, Schuh K, Warbasse L. The effects of cocaine on mood and sleep in cocaine-dependent males. *Exp Clin Psychopharmacol* 1999;**4**:338–46.

Johnson EO, Roehrs T, Roth T, Breslau N. Epidemiology of alcohol and medication as aids to sleep in early adulthood. *Sleep* 1998;**21**:178–86.

Koob GF, Stinus L, LeMoal M, Bloom FE. Opponent theory of motivation: neurobiological evidence from studies of opiate dependence. *Neurosci Biobehav Rev* 1989;**13**:135–50.

Ogeil RP, Phillips JG, Rajaratnam SM, Broadbear JH. Risky drug use and effects on sleep quality and daytime sleepiness. *Hum Psychopharmacol* 2015;**30**(5):356–63.

Randall S, Roehrs TA, Roth T. Efficacy of eight months of nightly zolpidem: a placebo controlled study. *Sleep* 2012;**35**:1551–7.

Roehrs T, Papineau K, Rosenthal L, Roth T. Ethanol as a hypnotic in insomniacs: self administration and effects on sleep and mood. *Neuropsychopharmacology* 1999;**20**:279–86.

Troxel WM, Ewing B, D'Amico EJ. Examining racial/ethnic disparities in the association between adolescent sleep and alcohol or marijuana use. *Sleep Health* 2015;**1**(2):104–8.

Wetter DW, Fiore MC, Baker TB, Young TB. Tobacco withdrawal and nicotine replacement influence objective measures of sleep. *J Consult Clin Psychol* 1995;**63**:658–67.

Willoughby A, de Zambotti M, Baker F, Colrain I. Partial K-conplex recovery following short-term abstinence in individuals with alcohol use disorder. *Alcohol Clin Exp Res* 2015;**39**:1417–24.

A complete reference list can be found online at ExpertConsult.com.

Chapter
141

Sleep Disturbances in Attention-Deficit/Hyperactivity Disorder

Samuele Cortese; Michel Lecendreux

Chapter Highlights

- Although sleep disturbance is not mentioned in the criteria for attention-deficit/hyperactivity disorder (ADHD) in the most recent version of the *Diagnostic and Statistical Manual of Mental Disorders* (fifth edition), in the past decade there has been an increasing interest in the relationship between ADHD and sleep.
- In this chapter, we first review brain correlates possibly underpinning both ADHD and conditions characterized by sleep disruption. These include both cortical (e.g., dorsolateral prefrontal) and subcortical (e.g., locus coeruleus) structures.
- The second part of the chapter focuses on the empiric evidence supporting the management of sleep disorders in individuals with ADHD. Because, with the possible exception of melatonin for sleep-onset delay, the number of randomized controlled trials is still limited, most of the recommendations are based on a recent expert consensus workshop.

ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

Attention-deficit/hyperactivity disorder (ADHD) is the most common neurodevelopmental disorder in children. According to the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5),[1] ADHD is defined by a persistent and impairing pattern of inattention or hyperactivity/impulsivity. The disorder is often comorbid with other neuropsychiatric conditions, such as oppositional defiant disorder/conduct disorder and mood and anxiety disorders. Besides the core symptoms of ADHD, deficits in executive functions (a set of cognitive skills that are necessary to plan, monitor, and execute a sequence of goal-directed complex actions) are commonly although not consistently associated with ADHD.[2]

ADHD is a major public health issue. Its estimated worldwide pooled prevalence is around 5% in school-age children, as found in a 2007 meta-analysis[3] and confirmed by its recent update.[4] Impairing symptoms of ADHD have been shown to persist in adulthood in up to 65% of individuals,[5] with an estimated pooled prevalence of ADHD in adulthood of approximately 2.5%.[6] In adults, ADHD may be comorbid with personality disorders, such as borderline personality disorder.[7] The core symptoms of ADHD and its comorbid conditions impose an enormous burden in terms of psychological dysfunction, adverse vocational outcomes, stress on families, and societal financial costs. Average annual incremental costs of ADHD have been estimated at $143 to $266 billion in the United States.[8] In Europe, annual national costs range between €1041 and €1529 million.[9]

Available therapeutic strategies for ADHD include pharmacologic and nonpharmacologic strategies. Pharmacologic treatments, which are generally well tolerated,[10] are an important element of therapeutic strategies for ADHD and are recommended as the first-choice option in some countries[11]; in other countries, they are recommended only for severe cases and those who fail to respond to nonpharmacologic interventions.[12,13] ADHD medications include psychostimulant (e.g., methylphenidate or amphetamine derivatives) and nonpsychostimulant (e.g., atomoxetine or guanfacine) drugs. Strong evidence from randomized controlled trials (RCTs) shows that different classes of psychostimulants (e.g., methylphenidate[14-17] or mixed amphetamine salts[18] and other amphetamines[19]) and some nonpsychostimulant treatments, such as atomoxetine,[20-23] are significantly more efficacious than placebo in considering short-term improvement in ADHD core symptoms. The efficacy of nonpharmacologic strategies, such as behavioral therapies, cognitive training, neurofeedback, and dietary treatment, for ADHD core symptoms is still unclear,[24] although these interventions may be effective for ADHD-related conditions and consequences of ADHD (e.g., oppositional behaviors in the case of behavioral therapies and executive impairment in relation to cognitive training) and may be preferred to medication by some patients and families.

ADHD and Sleep Disturbances: Historical Perspective

Whereas the comorbidity between ADHD and psychiatric disorders has been extensively addressed in previous research, the association with sleep disorder/disturbances has received less attention until recently. Indeed, sleep problems were mentioned in early descriptions of the disorder. For instance, in 1957, Laufer and Denhoff[25] wrote, "Generally the parents of hyperkinetic children are so desperate over the night problems that the daytime ones pale in significance." Later

on, Douglas[26] mentioned alterations in the "regulation of arousal and alertness to meet task demands" associated with ADHD.

Alterations in sleep, and more precisely "moves about excessively during sleep," were one of the possible diagnostic criteria for ADHD in the DSM-III.[27] On the grounds of the lack of specificity for ADHD, sleep disturbance was removed as a criterion for the disorder in the following editions of the DSM, likely contributing to a decrease in the attention paid to the relationship between ADHD and sleep in both research and clinical settings. However, in the past decade, there has been a renewed interest in the relationship between ADHD and sleep, as shown by the results of a simple PubMed search with the search terms "(ADHD OR Hyperkinetic Disorder OR Attention Deficit) AND sleep" (Figure 141-1). Although the DSM-5[1] ADHD criteria do not mention the word "sleep" (neither in the diagnostic criteria nor in the differential diagnosis), several recent guidelines and clinical recommendations on ADHD do refer to sleep disturbances as possible comorbidities or as disorders to consider in the differential diagnosis. For instance, the clinical practice guideline for the diagnosis, evaluation, and treatment of ADHD by the American Academy of Pediatrics[28] indicated that sleep disorders are among the conditions that can coexist with ADHD and recommended that primary care clinicians include sleep apnea as one of the conditions that are likely to be comorbid with ADHD. In addition, the European consensus statement on diagnosis and treatment of adult ADHD[29] pointed out that ADHD patients are at increased risk of having a sleep phase delay as a result of having chronic sleep-onset difficulty and recommended that a sleep assessment be included in the diagnostic process of ADHD.

ADHD AND SLEEP: GENERALITIES

Up to 70% of children with ADHD have been reported to have mild to severe "sleep problems."[30] In terms of adult ADHD, there are no epidemiologic studies that investigate subjective sleep complaints, nor are there large polysomnographic investigations in this population, but sleep disturbances are also considered to be common in adults with ADHD.[31]

Prevalence rates of sleep disturbances/disorders differ as a function of ADHD subtypes, with highest prevalence in the combined subtype,[32] although sleepiness may be more frequent in the inattentive subtype.[33]

The most recent meta-analysis on sleep disturbances in ADHD, which focused on children and adolescents,[34] found significantly more sleep problems in children with ADHD than in healthy controls without ADHD based on subjectively rated sleep items, including bedtime resistance, sleep-onset difficulties, night awakenings, difficulties with morning awakenings, sleep-disordered breathing, and daytime sleepiness. The meta-analysis[34] also found that compared with controls, children with ADHD had (1) significantly greater sleep-onset latency and less total sleep time based on actigraphy, (2) significantly greater number of stage shifts/hour of sleep and apnea-hypopnea index (AHI) and significantly lower sleep efficiency based on polysomnography, and (3) shorter average sleep-onset latency on the Multiple Sleep Latency Test (indicating that children with ADHD have higher levels of daytime sleepiness than controls do). Of note, studies assessing children treated with medications or with comorbid anxiety/mood disorders were excluded from the meta-analysis; therefore, the authors concluded that the significant differences in sleep parameters (both subjective and objective) are not accounted for exclusively by ADHD drugs or psychiatric comorbidities. However, other conditions occurring with ADHD were not addressed in this meta-analysis and could be responsible for the observed sleep disturbance. Indeed, it is likely that several mutually exacerbating factors ranging from behavioral/environmental conditions to primarily physiologic alterations may underlie sleep complaints in ADHD.[35] For a list of such factors that might contribute to sleep-onset difficulties, see Table 141-1. Although a previous meta-analysis[36] limited to polysomnographic studies showed only significantly higher values of the periodic limb movement index in ADHD compared with controls, the lack of significant findings on the other objective parameters may be accounted for by the smaller number of studies included in this older meta-analysis.

The association between ADHD and alterations of sleep processes has prompted investigation on possible

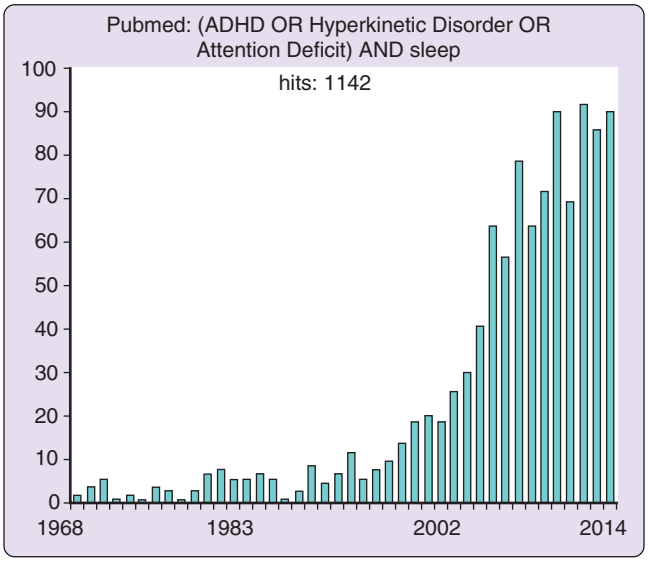


Figure 141-1 Scientific publications on ADHD and sleep, per a PubMed search (May 5, 2014).

Table 141-1 Conditions Underlying Sleep Disturbances (in particular, sleep-onset delay) in Individuals with ADHD

- Sleep disorders such as circadian rhythm sleep disorder or restless legs syndrome
- Poor sleep hygiene (e.g., use of electronics before bedtime)
- Psychiatric comorbidities (e.g., mood/anxiety disorders)
- Associated medical conditions and their treatments (e.g., obesity)
- Medication used to treat ADHD or comorbid conditions
- ADHD per se (individuals with ADHD may find it difficult to slow down their thoughts to settle for sleep, with consequent sleep-onset delay[33])

neurobiologic overlap at the brain level. In the following section, we provide an overview of the possible common brain mechanisms underlying both ADHD and sleep disorders. We then present the current evidence and, in the absence of empiric studies, provide recommendations for the management of sleep disturbances in individuals with ADHD based on recent expert consensus.[37] Because most of the studies have been conducted in pediatric samples, we focus on children, but we also present evidence from adult studies where it is available. The vast literature on the impact of sleep deprivation/disorders on neurocognitive functions, in particular, executive functions, that may be abnormal in individuals with ADHD goes beyond the scope of this chapter; the interested reader is referred to specific reviews of the literature on this topic.[38]

NEUROBIOLOGIC OVERLAP OF ADHD AND SLEEP DISORDERS

In a consensus conference paper on future perspectives of research on ADHD and sleep, Owens et al[39] pointed out that the cortical and brainstem regions involved in the regulation of sleep-wake function are also key areas implicated in the pathophysiologic mechanism of ADHD.[40] These include the frontal, dorsolateral prefrontal, ventrolateral prefrontal, and dorsal anterior cingulate cortices as well as the striatum (caudate and putamen) and lateral temporal and parietal regions. In addition, altered connectivity in key circuits pertaining to sleep-dependent learning, including disruption of limbic prefrontal circuits and alterations in connectivity between prefrontal regions (especially the ventral medial prefrontal cortex) and the hippocampus and amygdala, have been demonstrated in ADHD.[41]

Kirov and Brand[42] have noted that dopaminergic, noradrenergic, and serotoninergic abnormalities may underlie both ADHD and alterations in the pattern of sleep stage distribution across the night (changes in sleep architecture). They proposed a neurochemical model predicting possible changes in sleep architecture and other sleep problems in ADHD (Figure 141-2). More specifically, these authors suggested that (1) a nigrostriatal deficit may underlie periodic limb movements in sleep and restless legs syndrome (RLS) in ADHD, (2) a serotoninergic deficit may lead to deficient melatonin synthesis that in turn would contribute to delayed sleep phase syndrome, and (3) a mesolimbic (ventral tegmental area/nucleus accumbens) dopamine deficit may result in an altered monoamine-acetylcholine ratio, which would contribute to rapid eye movement (REM) sleep overdrive.

Another brain structure that may be functionally abnormal in both ADHD and sleep/circadian rhythm disorders is the locus coeruleus. Imeraj et al[43] postulated that a dysfunction of this structure could underlie both a disruption in circadian rhythm and disturbances in arousal-related processes that are considered to be abnormal in ADHD according to the cognitive energetic model originally developed by Sergeant (Figure 141-3).

Finally, evidence from genetic studies points to the involvement of the catecholaminergic system in both ADHD and sleep regulation. Specific genetic alterations, such as those of *CLOCK* genes, have been suggested to be common to both ADHD and circadian rhythm disorders.[39]

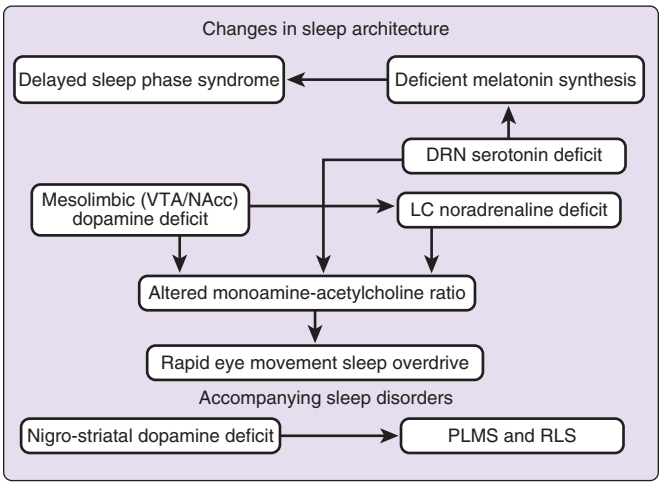


Figure 141-2 Neurochemical model of sleep disturbances in ADHD proposed by Kirov and Brand. DRN, Dorsal raphe nucleus; LC, locus coeruleus; NAcc, nucleus accumbens; PLMS, periodic limb movements in sleep; RLS, restless legs syndrome; VTA, ventral tegmental area. (From Kirov R, Brand S. Sleep problems and their effect in ADHD. *Expert Rev Neurother* 2014;14:287–99, with permission.)

SPECIFIC SLEEP DISORDERS IN ADHD

This section focuses on the management of sleep disturbances in ADHD. The levels of evidence are graded on the basis of the Scottish Intercollegiate Guidelines Network (SIGN) grading system (http://www.sign.ac.uk/) (Table 141-2). As previously mentioned, the etiology of sleep disturbances in ADHD is likely multifactorial; therefore, we present the management of the most common conditions (Table 141-1) underlying sleep complaints in ADHD. In particular, we focus on the most common sleep disorders and on the effects of ADHD medication on sleep; we refer the reader to other chapters of this book for information on the management of comorbid psychiatric disorders and somatic conditions associated with ADHD. When no empiric evidence is available, suggestions on the clinical management of specific disorders are based on the recommendations of the participants in a recent consensus conference.[37]

Behavioral Insomnia of Childhood

Precise estimates regarding the prevalence of behavioral insomnia of childhood as defined in the *International Classification of Sleep Disorders* (third edition) in children with ADHD are not available. However, clinical experience and expert consensus[37] suggest that this disorder is fairly common in children with ADHD referred to ADHD clinics.

Whereas the efficacy of behavioral treatments for behavioral insomnia in children in general is supported by empiric evidence,[44] at the time of this writing there is very little research evaluating behavioral sleep interventions specifically for children with ADHD. One case series found that a structured five-session therapy provided to three children with ADHD was successful at improving symptoms of insomnia.[45] This structured five-session therapy, which consisted of psychoeducation about sleep in children (e.g., basic sleep structure and function, sleep problems, and impact on daytime

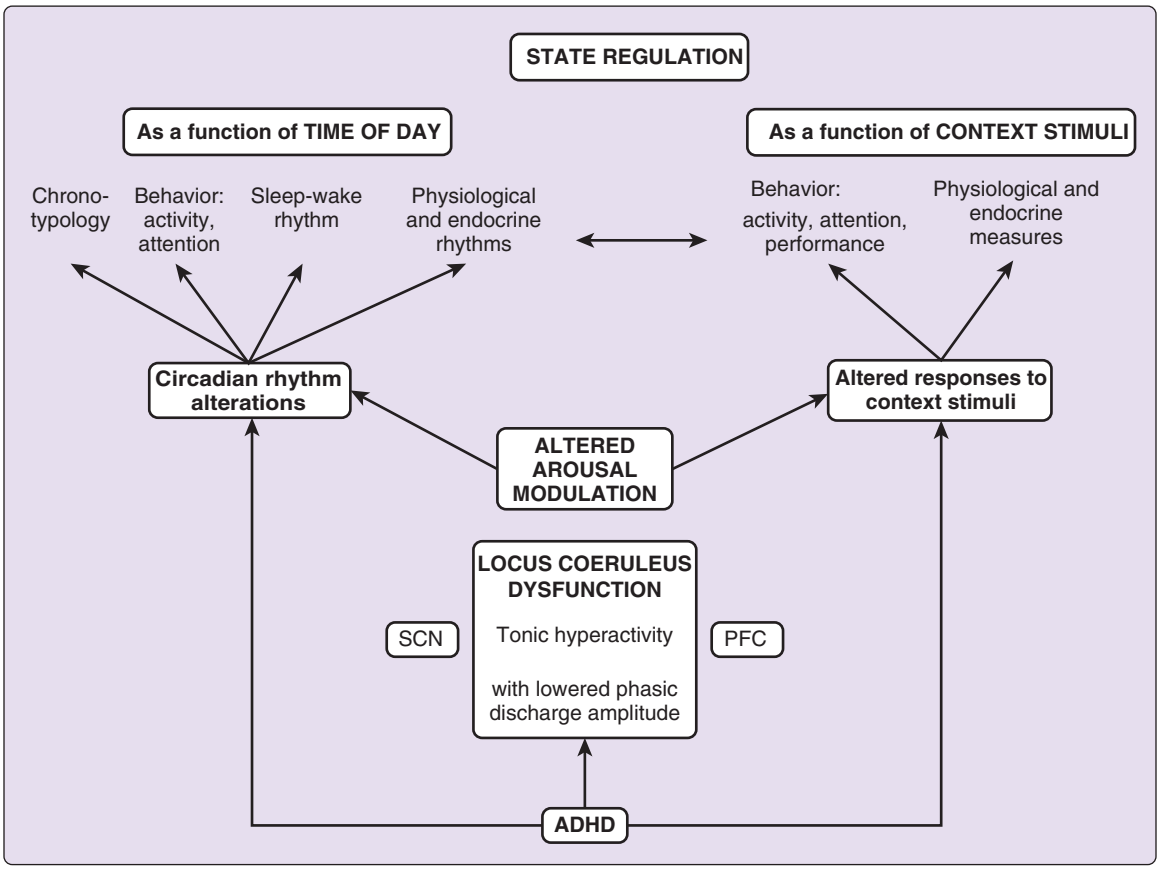


Figure 141-3 Model proposed by Imeraj et al on the involvement of the locus coeruleus in arousal dysfunctions and circadian rhythm disorders in ADHD. SCN, Suprachiasmatic nucleus; PFC, prefrontal cortex. (From Imeraj L, Sonuga-Barke E, Antrop I, et al. Altered circadian profiles in attention-deficit/hyperactivity disorder: an integrative review and theoretical framework for future studies. *Neurosci Biobehav Rev* 2012;36:1897–919, with permission.)

Table 141-2 Scottish Intercollegiate Guidelines Network (SIGN) Grading System

Levels of Evidence

1++ High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+ Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1– Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias
2++ High-quality systematic reviews of case-control or cohort studies
High-quality case-control or cohort studies with a very low risk of confounding, bias, or chance and a high probability that the relationship is causal
2+ Well conducted case-control or cohort studies with a low risk of confounding, bias, or chance and a moderate probability that the relationship is causal
2– Case-control or cohort studies with a high risk of confounding, bias, or chance and a significant probability that the relationship is not causal
3 Nonanalytic studies (e.g., case reports, case series)
4 Expert opinion

Grades of Recommendation

A At least one meta-analysis, systematic review, or RCT rated as 1++ and directly applicable to the target population; *or*
A systematic review of RCTs or a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; *or*
Extrapolated evidence from studies rated as 1++ or 1+
C A body of evidence including studies rated as 2+, directly applicable to the target population, and demonstrating overall consistency of results; *or*
Extrapolated evidence from studies rated as 2++
D Evidence level 3 or 4; *or*
Extrapolated evidence from studies rated as 2–

RCTs, Randomized controlled trials.
For critical appraisal of studies, see http://www.sign.ac.uk/guidelines/fulltext/50/annexoldb.html.

functioning), sleep hygiene, bedtime routines, and a faded bedtime with response cost procedure, was also included in an RCT of children aged 6 to 12 years (two thirds of the sample were healthy controls without ADHD and one third of the sample had ADHD). Unpublished results of this RCT indicate that children with ADHD responded as favorably as the control subjects did.[46] There is also another ongoing RCT[47] studying treatment of sleep problems in children with ADHD using two face-to-face consultations and one follow-up phone call. Preliminary results of this study[48] indicate that intervention was feasible and resulted in improved sleep, at least as rated by parents.

Because of the lack of research in this area, it is necessary to rely on expert consensus recommendations regarding treatment,[37] which include using behavioral therapies demonstrated to be effective in children without ADHD,[49] and to modify these for children with ADHD. For example, in promoting positive bedtime routines, children with ADHD may require more warnings before being called to bed so that they know that bedtime is coming. In addition, the routine should be structured to allow sufficient time to complete each activity (e.g., getting ready with school bag, brushing teeth) but not so much time that the child has opportunities to become distracted. It is critical that the instructions are given by parents one step at a time. Further, the bedtime routine should be posted where it can be easily seen by the child.[33]

Recommendation by the expert consensus group[37]

Behavioral interventions, adapted for ADHD children, should be the first line of treatment.

Level of evidence: Pending the full publication of two ongoing RCTs.[46,48]

Circadian Rhythm Sleep-Wake Disorder

Although estimates of prevalence of this disorder in individuals with ADHD based on formal criteria are lacking, evidence from studies on melatonin patterns, both in children[50] and in adults,[51] suggests that individuals with ADHD may present with delayed sleep phase. (For a detailed discussion of delayed sleep phase syndrome, see Chapter 40.)

One case report in a child[52] and one open-label trial[53] in adults with ADHD showed the effectiveness of phototherapy. Melatonin represents an alternative treatment approach. The expert consensus[37] recommended administration of melatonin (3 to 5 mg) 2 hours before the estimated dim light melatonin onset or 4 hours before the average sleep-onset time. Two RCTs,[54,55] a follow-up study of one of them,[54] and an open-label study[56] have shown that evening melatonin is an effective treatment of sleep-onset insomnia, possibly related to delayed sleep phase syndrome, among children with ADHD.

Recommendation by the expert consensus group[37]

Options for sleep-onset insomnia within the framework of delayed sleep phase syndrome include bright light therapy, chronotherapy, and melatonin.

Level of evidence for bright light therapy and chronotherapy: E (evidence level: E, case reports[52,57]); *for melatonin:* B (extrapolated evidence from two RCTs rated as 1++ or 1+[54,55]).

Restless Legs Syndrome/Periodic Limb Movement Disorder

According to a review conducted in 2005,[58] about 44% of subjects with ADHD have been found to have RLS or RLS symptoms, and up to 26% of subjects with RLS have been reported to have ADHD or ADHD symptoms. However, these figures need to be considered with caution because of methodologic issues in the included studies, such as the lack of ADHD diagnosis or possible recruitment bias of subjects in specialized clinics for sleep. More recent and likely less biased studies suggest that the prevalence of RLS in individuals with ADHD is approximately 10% to 20%.[59,60]

As detailed in Chapter 95, the management of RLS/periodic limb movement disorder (PLMD) includes nonpharmacologic as well as pharmacologic strategies for moderate to severe cases. Nonpharmacologic strategies include establishing healthy sleep habits, physical exercise, and avoiding exacerbating factors, such as irregular sleep schedule, low body iron stores, pain, caffeine, nicotine, alcohol, and certain medications (e.g., selective serotonin reuptake inhibitors).[61] A case series and a case report support the effectiveness of dopaminergic agents (L-dopa, pergolide, or ropinirole) for RLS in children with ADHD.[62,63] Finally, another recent case series suggests that levetiracetam may be an option for ADHD children with RLS and interictal epileptic discharges. No data on the long-term effectiveness and tolerability of these agent in children are available.

Although there is no formal consensus on the treatment of iron deficiency, thought to be involved in the pathophysiologic mechanism of RLS, some experts have recommended oral iron supplementation with low iron stores, 50 to 65 mg of elemental iron once or twice a day, and suggested rechecking of serum ferritin levels in 2 to 3 months.[61]

Recommendation by the expert consensus group[37]

Establish healthy sleep habits, avoid factors exacerbating RLS (e.g., pain, caffeine, nicotine, alcohol); oral iron supplementation if iron deficiency is found. For severe cases: consider use of L-dopa or other dopaminergic agents (off label on children).

Level of recommendation for the treatment of RLS/PLMD in children with ADHD (for dopaminergic agents): C/D (one RCT with risk of bias, plus case reports).

Sleep-Related Breathing Disorders

Some studies suggest that the prevalence of sleep-related breathing disorders (SRBDs) in children with a diagnosis of ADHD is at least 18% to 32%[64,65] and may be as high as 50% to 65%.[66-69] However, some authors do not agree with these high figures.[70-72] This discrepancy is mainly related to definitions of SRBD in children (for a detailed discussion, see[73]). The majority of children with ADHD appear to have mild rather than severe forms of SRBD (i.e., AHI >1 and <5).[73] Thus, it has been suggested that severe SRBD and its associated daytime manifestation, such as sleepiness, may mask the hyperactivity.[73] In addition, symptoms of SRBD, such as snoring, are also common in children with ADHD or ADHD-like behaviors. Between 20% and 66% of children with a diagnosis of ADHD or symptoms of ADHD have been reported to habitually snore,[59,74,75] with the

ADHD-hyperactive/impulsive being the most common presentation.[76]

We located three studies that investigated SRBD treatment in a group of children diagnosed with ADHD according to formal criteria (DSM-IV).[69,77,78] Weber et al[78] assessed children with a clinical history of obstructive ventilator disorders (because many children did not complete polysomnography); they found that after adenotonsillectomy, there were reductions in reported attention deficit (87.5% to 33.3%), hyperactivity (75% to 50%), and impulsivity (50% to 33%) for 8- to 10-year-olds. However, because of small sample sizes, these reductions were not statistically significant. In a prospective controlled trial that aimed to assess hyperactivity and inattention before and 1 year after adenotonsillectomy,[77] 28% of children had ADHD at baseline, but half of the children no longer met criteria for an ADHD diagnosis at 1 year after surgery. In another prospective, nonrandomized, open trial,[69] in children with ADHD (DSM-IV criteria) and mild SRBD (i.e., AHI >1 and <5), treatment groups included (1) methylphenidate, (2) adenotonsillectomy, and (3) no treatment. At 6-month follow-up, ADHD symptoms had improved in both intervention groups compared with the no-treatment controls. In addition, the adenotonsillectomy group demonstrated significantly greater improvement than the medication group in the hyperactivity and inattention scales as well as in the total score of the ADHD rating scale. In summary, data from these studies suggest that treatment of SRBD improves ADHD symptoms.

Recommendation by the expert consensus group[37]

All children undergoing evaluation for ADHD should be screened for symptoms of SRBD, particularly snoring; if findings are positive, polysomnography should be performed. An AHI > 1 on polysomnography and the presence of large tonsils/small airway on examination should alert the practitioner to consider adenotonsillectomy as a treatment intervention.

Level of evidence for the treatment of sleep disordered breathing according to the SIGN system: B/C (one RCT rated as 2++[77] and open-label study rated as 2+[69]).

EFFECTS OF ADHD MEDICATIONS ON SLEEP

Subjective and objective studies of the effect of stimulants on sleep have produced mixed results.[79] Whereas polysomnographically determined lengthened total sleep time, increased sleep stage shifts, increased number of REM periods, elevated indexes of REM activity and REM period fragmentation, and parent-reported longer latencies to sleep onset or higher rates of insomnia in ADHD children treated with stimulants versus controls have been reported in some studies, others did not confirm these findings. It is challenging to combine the results of these studies because of different stimulant formulations, dose, and dose scheduling. Moreover, some of the studies compared children with ADHD receiving medication with nonmedicated healthy controls. Therefore, it is unclear whether the sleep disturbances resulted from the children's having ADHD or from their being treated with methylphenidate. However, clinical experience suggests that stimulants may negatively affect sleep. The vulnerability to these negative effects is likely to be related to individual factors. As correctly pointed out by Brown and McMullen,[80] whereas some patients with ADHD are able to get to sleep easily within just a few hours of taking a dose of stimulant, others need an interval of 6 to 8 hours.

As for atomoxetine, in a randomized, double-blind, crossover study comparing the effect of methylphenidate (given thrice daily) and atomoxetine (given twice daily) on the sleep of children with ADHD, Sangal et al[79] found that methylphenidate increased both polysomnographic and actigraphic sleep-onset latency significantly more than did atomoxetine. Moreover, both child diaries and parent reports indicated a better quality of sleep with atomoxetine compared with methylphenidate. Both medications decreased nighttime awakenings, but the decrease was greater for methylphenidate.

Implementation of healthy sleep habits is recommended as a first option even if the sleep complaints are thought to be medication related; in one RCT,[55] sleep hygiene was effective in a subsample of children with ADHD treated with stimulants. If, after several weeks (during which the negative effect of medication may also spontaneously decrease), sleep hygiene is not effective, possible further options are reducing the total dose, changing the dose regimen or formulation so that less medication is administered later in the day, changing to a different stimulant (e.g., switching to methylphenidate from amphetamine or vice versa), changing to a nonstimulant (e.g., atomoxetine), adding an alpha$_2$-agonist, and adding another medication (e.g., melatonin or sedative-hypnotics as "last resort" option). An RCT[81] showed that adding melatonin (3 to 6 mg) can significantly decrease sleep-onset delay in ADHD children treated with methylphenidate (1 mg/kg).

Recommendation by the expert consensus group[37]

- Promoting healthy sleep habits
- If sleep difficulties persist: try alternative dosages, formulations, or ADHD medications or add a sleep-promoting medication (e.g., melatonin).

Level of evidence for melatonin: B (one RCT rated as1+[81]); *for other approaches:* D (expert consensus).

CLINICAL PEARLS

- The key to the management of sleep problems in ADHD is an appropriate differential diagnosis. Healthy sleep practices are the foundation of management strategies.
- Behavioral interventions should be the first-line treatment of insomnia. If behavioral strategies are not effective, melatonin should be considered for sleep-onset difficulties. Bright light therapy, chronotherapy, and melatonin are options for circadian rhythm disorders.
- Adenotonsillectomy should be considered if the AHI is > 1 and there is evidence of enlarged tonsils/small airway on examination.
- Off-label dopaminergic agents should be considered only for severe cases of RLS/PLMD.
- Options for sleep disturbances induced by ADHD medications include alternative dosages, dose regimen, formulations, or ADHD medications and adding a sleep-promoting medication (e.g., melatonin).

SUMMARY

Although long overlooked, sleep disturbances are highly relevant in the clinical management of individuals with ADHD.

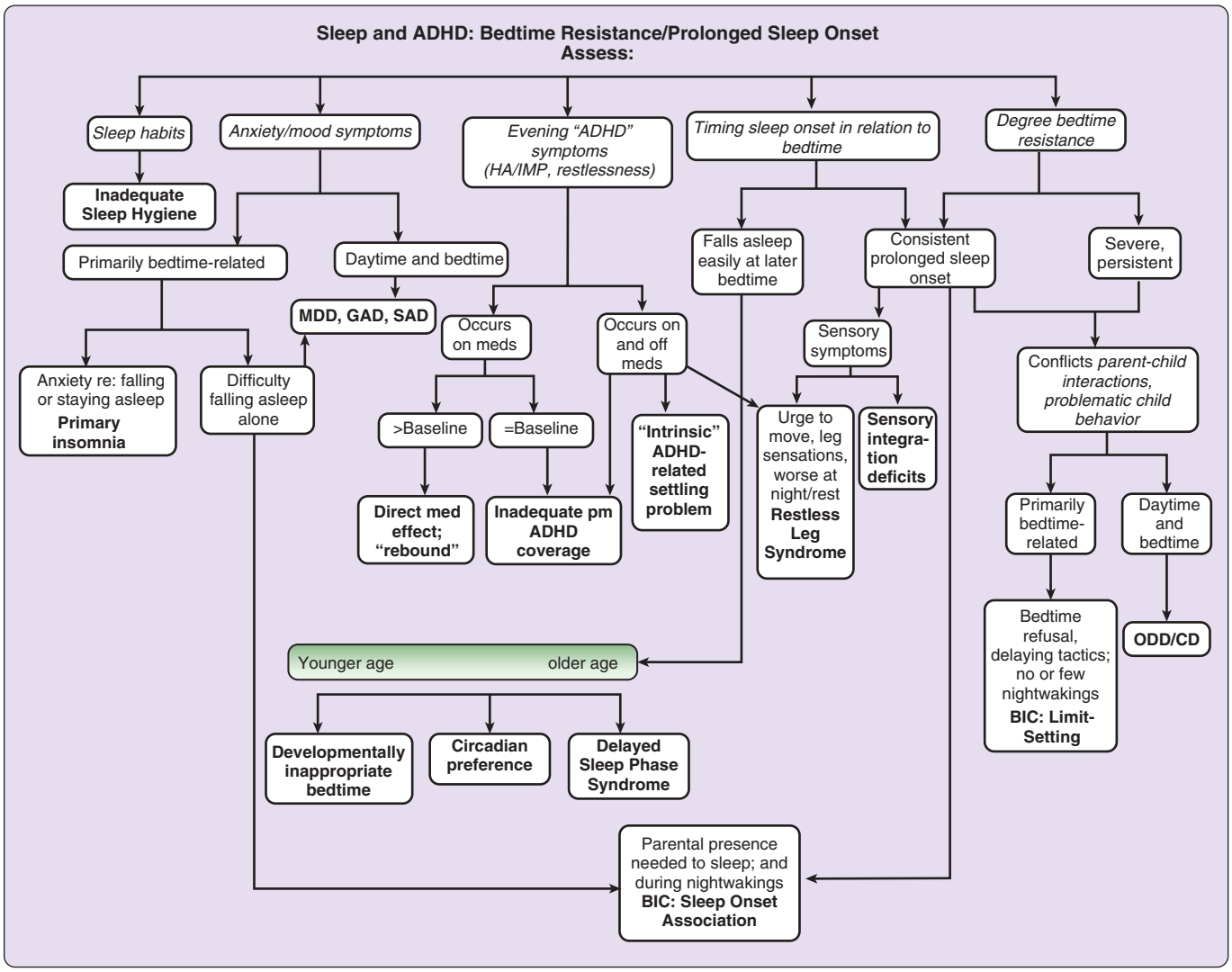


Figure 141-4 Diagnostic algorithm for the management of sleep-onset difficulties/bedtime resistance in children with ADHD. HA/IMP, Hyperactivity/impulsivity; MDD, major depressive disorder; GAD, generalized anxiety disorder; SAD, social anxiety disorder; ODD/CD, oppositional defiant disorder/conduct disorder; BIC, behavioral insomnia of childhood. (Reproduced with permission from Owens JA. A clinical overview of sleep and attention-deficit/hyperactivity disorder in children and adolescents. *J Can Acad Child Adolesc Psychiatry* 2009;18:92–102.)

The relationship between ADHD and altered sleep/arousal mechanisms is also relevant from a research standpoint because it might lead to further insights into the pathophysiologic process of both sleep and attentional processes. Brain regions possibly involved in mediating both conditions include cortical (e.g., dorsolateral prefrontal) and subcortical (e.g., locus coeruleus) structures. The most common conditions affecting sleep quality or quantity in ADHD include behavior insomnia, circadian rhythm sleep-wake disorder, delayed sleep phase type, sleep-related disordered breathing, RLS/PLMD, and sleep disturbances due to ADHD medications. Sleep hygiene therapy is recommended as the first step in management. Behavioral therapies should be considered as first-line treatment of behavioral insomnia, although further evidence from RCTs is needed to prove their efficacy in individuals with ADHD. Among pharmacologic options, RCTs support the use of melatonin to reduce sleep-onset delay, whereas there is more limited evidence for other medications. Figure 141-4 summarizes the algorithm for the management of bedtime resistance/prolonged sleep onset in children with ADHD.

Selected Readings

Corkum P, Davidson F, MacPherson M. A framework for the assessment and treatment of sleep problems in children with attention-deficit/hyperactivity disorder. *Pediatr Clin North Am* 2011;**58**:667–83.

Cortese S, Brown TE, Corkum P, et al. Assessment and management of sleep problems in youths with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2013;**52**:784–96.

Cortese S, Faraone SV, Konofal E, Lecendreux M. Sleep in children with attention-deficit/hyperactivity disorder: meta-analysis of subjective and objective studies. *J Am Acad Child Adolesc Psychiatry* 2009;**48**:894–908.

Gruber R. Sleep characteristics of children and adolescents with attention deficit-hyperactivity disorder. *Child Adolesc Psychiatr Clin N Am* 2009;**18**:863–76.

Huang YS, Tsai MH, Guilleminault C. Pharmacological treatment of ADHD and the short and long term effects on sleep. *Curr Pharm Des* 2011;**17**:1450–8.

Konofal E, Lecendreux M, Cortese S. Sleep and ADHD. *Sleep Med* 2010;**11**:652–8.

Lecendreux M, Cortese S. Sleep problems associated with ADHD: a review of current therapeutic options and recommendations for the future. *Expert Rev Neurother* 2007;**7**:1799–806.

Lecendreux M, Lavault S, Lopez R, et al. Attention-deficit/hyperactivity disorder (ADHD) symptoms in pediatric narcolepsy: a cross-sectional study. *Sleep* 2015;**38**:1285–95.

Owens J, Gruber R, Brown T, et al. Future research directions in sleep and ADHD: report of a consensus working group. *J Atten Disord* 2013;**17**:550–64.

Philip P, Micoulaud-Franchi JA, Lagarde E, et al. Attention deficit hyperactivity disorder symptoms, sleepiness and accidental risk in 36140 regularly registered highway drivers. *PLoS One* 2015;**10**(9):e0138004.

Sadeh A, Pergamin L, Bar-Haim Y. Sleep in children with attention-deficit hyperactivity disorder: a meta-analysis of polysomnographic studies. *Sleep Med Rev* 2006;**10**:381–98.

Stein MA, Weiss M, Hlavaty L. ADHD treatments, sleep, and sleep problems: complex associations. *Neurother* 2012;**9**:509–17.

A complete reference list can be found online at ExpertConsult.com.

Section
18 Dentistry and Otolaryngology in Sleep Medicine

Chapter
142 Role of Dentistry and Otolaryngology in Sleep Medicine

Gilles Lavigne; Michael Simmons; Nelly Huynh; Fernanda R. Almeida; Olivier M. Vanderveken

Chapter Highlights

- This new section of this volume links medicine and dentistry experts' knowledge on dealing with sleep disorders.
- Reviewed in separate chapters of this section are methods to improve upper airway function during sleep by addressing underlying skeletal imbalances by orthodontic treatment in children, by surgery in adults, and by oral appliance therapy with the treatment goal of reducing upper airway collapsibility.
- Additional chapters address tooth grinding/sleep bruxism, alone or in presence of orofacial pain, with indications for and limitations of oral appliances and surgical interventions in managing sleep-disordered breathing.
- Finally, this section addresses reducing poor outcomes related to use of medications for oral or intravenous sedation or general anesthesia in patients undiagnosed but at potential risk for sleep-disordered breathing.
- Interdisciplinary collaboration is an important goal in the efforts to improve sleep behavior, with an overall impact on patient health and well-being. Correction of facial growth disturbances may be an adjunctive treatment option that may be used depending on the unique presentation of each patient.
- A better understanding of patient phenotypes requires a collaboration of multiple disciplines to improved treatment outcomes in the future.

The addition of a section on ear-nose-throat (ENT) and oral-dental aspects of sleep medicine—a major component of the evolving field of "integrated sleep medicine"—to *Principles and Practice of Sleep Medicine* has been a long-time goal of one of the editors, Dr. Meir Kryger, which is realized in this sixth edition of the book. Dr. Kryger has been a strong advocate of interdisciplinary collaborations in managing patients with sleep disorders and their associated comorbidities.

In the past several decades, sleep medicine has been dominated by the presence of respiratory medicine, neurology, and psychology. With recognition of the high prevalence of sleep disorders and the technological innovations currently available for the management of sleep-disordered breathing (SDB), however, rapid and strong growth of sleep medicine as a subspecialty of both medicine and dental science is under way.

SLEEP SOCIETIES: HISTORICAL PERSPECTIVE ON INTEGRATED SLEEP MEDICINE

All over the world, sleep medicine has become organized and integrated into several health disciplines, merging knowledge and nurturing exchanges between basic science, medicine, dentistry, pharmacology, nutrition, psychology, nursing, social health, health economics, and others.

International Societies

For more than 25 years, the World Sleep Federation (WSF) and later the World Association of Sleep Medicine (WASM) have been dedicated to improving dialogue within the sleep community, promoting innovation in research and technologies, elevating the standard of care, and optimizing access to evidence-based diagnostic and management paradigms. Interdisciplinary meetings are held by these associations and their allied societies. The WSF and the WASM have organized more than six world meetings since they were founded (in 1988 and 2003, respectively).

Continental and National Societies

In North America, the American Academy of Sleep Medicine was created in 1975 (already 40 years ago!), in its precursor state as the Association of Sleep Disorders Centers. The American Academy of Sleep Medicine is the main body for the following journals: *Sleep, Journal of Clinical Sleep Medicine,* and the most recent addition from the American Academy of Dental Sleep Medicine, *Journal of Dental Sleep Medicine.* By 2013 the American Academy of Sleep Medicine had approximately 10,000 members, of whom approximately 5% were dentists. Many special-interest groups developed along the way in neurology, ENT, and respiratory medicine and in dental sleep medicine. Reflecting the importance of sleep medicine, many physicians also hold subspecialty certification in sleep medicine from the American Board of Otolaryngology and the American Board of Psychiatry and Neurology. Creation of the WASM was driven mainly by neurologists and neuroscientists, but this body is now enlarging its scope, facing the needs of integrated expertise in sleep medicine; this is reflected in the content of its journal, *Sleep Medicine.*

The American Academy of Dental Sleep Medicine evolved from the Sleep Disorders Dental Society, originating in 1991 with its first meeting of 25 members in 1992, expanding to more than 3000 members in 2013. This body provides a bridge for sleep clinicians and researchers in the fields of dentistry, otolaryngology, surgery, and respiratory and behavioral sleep medicine. In Canada, the Canadian Sleep Society has a dental medicine chapter, with members also attending the Sleep Disorders Dental Society meetings.

The European Sleep Research Society (ESRS) was created from a symposium held in Würzburg, 1971, on "Die Natur des Schlafes"("The Nature of Sleep"). In 2015, the ESRS held a joint meeting with the European Respiratory Society in Barcelona, Spain, and the ESRS and the WSF will co-host the Seventh World Congress of the World Sleep Federation in Istanbul, Turkey. The ESRS is the main body editing the *Journal of Sleep Medicine*. The European Academy of Dental Sleep Medicine (EADSM), formerly known as the European Dental Sleep Medicine Academy and officially founded in 2006, aims at being a European platform for dental sleep medicine. In 2015, the EADSM organized its sixth conference in Antwerp, Belgium. In addition, many European countries such as Germany, United Kingdom, France, The Netherlands, Belgium, Spain, Portugal, Italy, and Sweden, among others, have their own national dental sleep medicine societies.

The Federal Latin American Sleep Societies, founded in 1984, also has grown, and delegates from Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Mexico, Peru, and Uruguay have been actively involved to organize and disseminate sleep medicine in their countries. In Brazil, the Brazilian Sleep Association, founded in 1985, has held biannual meetings since 1993, and today has approximately 1000 members. It recently has helped with and endorsed the creation of the Brazilian Dental Sleep Association, which sponsored approximately 40 certified dentists in dental sleep medicine.

Other sleep societies such as the Asian Sleep Society and the Australian Sleep Society include regular activities in dental sleep medicine in their annual programs. The position of dentistry in sleep medicine is well illustrated by the fact the journal *Sleep and Breathing,* an interdisciplinary sleep journal (as most are these days), is the official journal of the Australasian Academy of Dental Sleep Medicine (AustADSM), the British Society of Dental Sleep Medicine (BSDSM), the European Academy of Dental Sleep Medicine (EADSM), the Japanese Academy of Dental Sleep Medicine (JADSM), and the Korean Academy of Dental Sleep Medicine (KADSM).

Recently, the Indian Association of Surgeons for Sleep Apnoea was created to bring together specialists from various fields such as ENT surgeons, pulmonologists, oral-maxillofacial surgeons, anesthesiologists, and radiologists. Such trends seem now irreversible and illustrate the need of integrated sleep medicine in the world.

INTEGRATION OF HEALTH DISCIPLINES IN SLEEP MEDICINE AND RESEARCH

Nowadays, it is not surprising to observe that more and more otorhinolaryngologists (also called ENT specialists), anesthesiologists and surgeons, and dentists and orthodontists are collaborating with pulmonologists, neurologists, and psychologists to manage various sleep disorders and concomitant health conditions. Improvement in oropharyngeal airway patency during sleep in children and adults can be achieved by orthodontic and surgical jaw corrections (see Chapters 143 and 149)[1]. Use of general anesthesia done in patients presenting with several concomitant medical conditions is a

challenge, and risks need to be assessed if obesity and sleep apnea are present (see Chapter 148). In addition, with recognition of breathing dysfunction in children with a small lower jaw or mandible (retrognathic cases), deep palate, or large tonsils or adenoids, careful evaluation is warranted, with use ofspecific precautions for anesthesia.

Sleep bruxism, characterized by tooth grinding and repetitive jaw muscle activity, and comorbid temporomandibular dysfunction have been in the domain of dentistry for many years but became integrated into dental sleep medicine in the past two to three decades (see Chapters 144 and 145). The putative interrelationship and/or the role of risk factors of "wake-time" breathing disorders, neurologic conditions, insomnia, and pain cannot be overlooked. A single health care discipline cannot manage complex complaints or persistent sleep disorders in the presence of these comorbid conditions. Sleep bruxism can be concomitant with sleep apnea (association possible in a subgroup of patients/phenotypes to be identified; low evidence for cause-and-effect link), headache, and temporomandibular dysfunction, REM behavior disorder, epilepsy, and other disorders.

It is obvious that the informed physician and dentist have to collaborate in increasing numbers for several reasons: (1) to identify causal and contributory disorders more efficiently in a wider range of the population of patients with undiagnosed problems, (2) to understand the role of comorbid conditions that can contribute to the development of chronicity of poor sleep, (3) predict and explain why a given subject is likely to be a poor responder to specific therapeutic interventions, and (4) to select the most suitable management approaches in accordance with individual patient characteristics including economic status. Furthermore, recent advances in complementary and alternative medicine have to be integrated into sleep medicine because patients are frequent users, as illustrated in Chapter 150.

ENTs, oral-maxillofacial surgeons, and orthodontists are key specialists in achieving functional upper airway maximal patency to achieve the best results in patients with pathologic conditions or atypical growth of the nose, oropharyngeal, and jaw bone structures. It also is obvious that dental specialists in sleep medicine and orofacial pain are key health practitioners in managing simple and idiopathic sleep bruxism, alone or with comorbid temporomandibular dysfunction and/or headaches. Dentistry's role is important in the detection of concomitant sleep disorders and referral for diagnosis when insomnia, SDB, or concomitant neurologic conditions are suspected. Dentists, orthodontists, and oral-maxillofacial surgeons with sufficient training and experience in the subspecialty of sleep disorders are key players in SDB management with oral appliances, orthodontic correction, or telegnathic maxillofacial surgery.

FUTURE DIRECTIONS: EDUCATION AND COLLABORATION

The interface between sleep medicine and dentistry is currently promoted by the international Oral Appliance Network on Global Effectiveness research group, created to evaluate the long-term effectiveness of oral appliance therapy in patients with obstructive sleep apnea and to assess long-term health outcomes of oral appliance therapy related to cardiovascular disease.[2] This is a truly multidisciplinary team, with respiratory and sleep medicine physicians working alongside dentists and researchers.

To achieve more efficient early detection of SDB or sleep-wake disorders, training of the current and next generations of family physicians, family dentists, nurses, psychologists, physical therapists, health educators, and school teachers is critical.[3] The health and well-being of the patient population can be markedly improved by increasing awareness at an earlier age, early prevention, and overall public education in the importance of sleep. In fact, teachers spend more time with children during their wake periods than their parents; dentists are the health care providers who see more than 70% of persons of all ages at least once a year; and primary care providers often have the family history on file to address issues such as childhood obesity, lack of exercise, and other factors potentially affecting patient health in adulthood. To meet the major public health challenge of "better sleep = better health," the role of first-line health professionals and educators should be encouraged in early detection of sleep disorders and access to simple prevention measures (e.g., diet recommendations, sleep hygiene, oropharyngeal exercise) at an early age to prevent growth and development abnormalities triggering sleep disorders later in life.

References

1. Huynh NT, Desplats E, Almeida FR. Orthodontics treatments for managing obstructive sleep apnea syndrome in children: a systematic review and meta-analysis. *Sleep Med Rev* 2015. [Epub ahead of print] Review.
2. Almeida FR, Vanderveken OM, Cistulli P, et al. Oral appliance network on global effectiveness (orange): start-up and design description. *JDSM* 2014;**1**:17–20.
3. Phillips B, Gozal D, Malhotra A. What is the future of sleep medicine in the US? *Am J Respir Crit Care Med* 2015. [Epub ahead of print].

Oropharyngeal Growth and Skeletal Malformations

Chapter 143

Stacey Dagmar Quo; Benjamin T. Pliska; Nelly Huynh

Chapter Highlights

- Sleep-disordered breathing (SDB) is marked by varying degrees of collapsibility of the pharyngeal airway. The hard tissue boundaries of the airway dictate the size and therefore the responsiveness of the muscles that form this part of the upper airway. Thus, the airway is shaped not only by the performance of the pharyngeal muscles to stimulation but also by the surrounding skeletal framework.
- The upper and lower jaws are key components of the craniofacial skeleton and the determinants of the anterior wall of the upper airway. The morphology of the jaws can be negatively altered by dysfunction of the upper airway during growth and development. In turn, the altered morphology of the jaws can be positively influenced by orthodontic treatment.
- The association between altered dentofacial morphology and SDB has been well documented in children, adults, and patients with craniofacial syndromes. Whether this disease of childhood has the same origins as adult obstructive sleep apnea but more subtle manifestations has not been determined. The length and volume of the airway increase until the age of 20 years, at which time there is a variable period of stability, followed by a slow decrease in airway size after the fifth decade of life. The possibility of addressing the early forms of this disease with the notions of intervention and prevention can change the landscape of care.
- Correction of specific skeletal anatomic deficiencies can improve or eliminate SDB symptoms in both children and adults. It is possible that the clinician may adapt or modify the growth expression, although the extent of this impact is uncertain. These strategies seek to alter an abnormal facial growth pattern wherein SDB worsens over time. Future research should focus on determining in which individuals dentofacial morphology makes a significant contribution to the pathogenesis of SDB. This may bring clinicians one step closer to targeting specific treatments that more effectively treat the disorder.

This chapter is directed to provide an integrated vision to both the scientist and clinician dealing with growth and development of the oropharyngeal structures. It is divided into three sections. The first reviews the fundamentals of growth of the craniofacial complex and the development of the upper airway. The second section outlines those characteristics based on current definitions of the sleep-disordered breathing (SDB) syndrome, that is, there is a broad range of anthropomorphic characteristics and OSA severity that needs to be better delineated. The last section reviews management and treatment strategies based on what is known about craniofacial development and anatomy. This chapter is a primer to other chapters in this section and to the ones in Section 14.

CRANIOFACIAL GROWTH AND DEVELOPMENT

Early theories of craniofacial development were based on the belief that growth of the face and jaws was essentially immutable because of intrinsic regulation by inherited genetic traits. Research centered on finding the location or sites where these traits were expressed that drove normal form and function of the other surrounding structures. During the early twentieth century, it was believed that differential deposition and resorption on the surface of bones were largely responsible for growth of the craniofacial skeleton. This remodeling theory then gave way to one heralding the role of the sutures, which held that similar to the epiphyses of the long bones, it was the connective tissue and cartilaginous joints of the craniofacial skeleton that produced expansive proliferative growth forcing the bones and soft tissues away from each other.[1] Exemplifying this theory was the concept that the mandible was much like a horseshoe-shaped long bone, with the condylar cartilage acting like open-ended epiphyseal plates, pushing the mandible down and away from the rest of the head.

Several inconsistencies in the hypothesis that sutures alone could be the determinants of craniofacial growth surfaced. Sutural growth was more similar to periosteal apposition of bone than previously understood, and sutures acted as reactive sites of bone growth rather than growth centers. Scott, in 1953, proposed the nasal septum theory of growth, which held as its main tenet that the anterior and inferior growth of the nasal septal cartilage was the determining, driving force of facial growth.[2] Whereas Scott's theory contributed to the understanding that growth at the sutures and surface

remodeling were essentially reactive sites of bone growth, it was still founded in the paradigm that craniofacial development and morphogenesis were genetically predetermined and unalterable.[1] In modern perspective, the nasal septal cartilage is considered an important growth center; however, the mechanism behind this growth has been appropriately updated as described here.

A fundamental shift in the field of craniofacial biology emerged with Moss' introduction of the functional matrix hypothesis in 1960.[3] Whereas all previous theories deemed craniofacial growth to be predetermined by genetic traits, the functional paradigm introduced the idea of plasticity of development and growth of the craniofacial skeleton.[1] According to this theory, the role of our genes was to initiate the process by setting the initial context under which development could occur; beyond that, it was the extrinsic environmental and functional demands of the various craniofacial components that determined all future aspects of growth. This theory revolutionized the field by introducing two important concepts. First, it brought the possibility of growth modification as a treatment option for malocclusions or facial malformations by changing the direction of facial development to a more desired outcome. Second and perhaps more important, the plasticity of development opened a new area of research, focusing on the critical time and specific factors that lead to the maladaptive plasticity of the form and function of the craniofacial complex.[4]

With the emergence of developmental molecular biology, the genetic and external or epigenetic regulation of craniofacial growth is now better understood, and the modern synthesis is that both genomic and epigenetic factors are necessary contributors to craniofacial development.[5,6] Several genes and gene products regulate the morphogenesis and intrauterine development of the craniofacial complex. There is an extremely complex interaction between these genetic factors and epigenetic influences. The plasticity of early development depends not only on the effects of certain environmental conditions on the underlying genetic code but also on previous environmental conditions that may have directly upregulated or downregulated specific genetic regulatory factors and indirectly further influenced the response.

Prenatal Craniofacial Growth

The earliest form of the face appears in the fourth week of life with the enlargement and movement of the frontonasal prominence as well as the paired maxillary and mandibular prominences stemming from the first branchial arch. These five prominences emerge to encircle the stomodeum, or primitive mouth. A critical aspect of this process is the migration of cranial neural crest cells into the developing facial prominences. Unlike in the rest of the body, these neural crest cells develop into the majority of the craniofacial hard tissues, including the bone, cartilage, and teeth of the craniofacial complex.[7] The specific end tissue into which these cranial neural crest cells develop is determined largely by the family of Homeobox or HOX genes. The variable expression of HOX transcription factors causes the groups of cranial neural crest cells in the distinct prominences to respond differently to the same growth factors.[8]

After rapid growth of the two mandibular prominences, there is midline fusion in the fifth week of development.[9] Initial mandibular development is dependent on tightly controlled molecular signaling between the oral ectoderm and the underlying core of cranial neural crest.[7] Further development is contingent on the formation and growth of Meckel cartilage, the first skeletal element of the mandible. The subsequent elongation of this rod of cartilage leads to promotion of outgrowth of the mandible, and the primary ossification of the mandible occurs as intramembranous bone formation along this cartilaginous core.[10] As the bony mandible further develops, Meckel cartilage largely disappears, eventually to persist only as small portions of the incus and malleus of the middle ear. In addition to the mandible, the mandibular processes also form the lower lip and the lower areas of the cheeks.

The lateral and medial nasal processes originate as ectodermal thickenings on the surface of the frontonasal process early in the fifth week of embryonic life. Subsequent broadening of the head and medial growth of the maxillary processes result in medial displacement of the early nasal processes. As the two medial nasal processes merge at the midline, the philtrum and columella of the nose are formed. Deeper aspects of the medial nasal processes will differentiate to form the nasal septum, which is a key growth center of the midface postnatally. Fusion of the medial nasal process with the maxillary process leads to formation of the majority of the upper lip, the zygomas, and the maxilla bilaterally. The lateral nasal processes go on to form the sides and alae of the nose as the subsequent 2 weeks sees the formation of the future nostrils and nasal cavity with the development of the primary and secondary palates.[11] Secondary palate formation relies on the coordinated growth and movement of the primordia of the tongue and both lower and upper jaws. During the sixth week of development, paired lateral palatal shelves arise as medial projections from each maxillary process. Critically, at this time, there is rapid anterior growth of the early mandible by proliferation of Meckel cartilage, which displaces the tongue forward, lowering it relative to the palatal shelves.[12] Once the tongue has descended, during the seventh week, the palatal shelves rotate from a vertical to a horizontal position directed toward the midline. Further growth of the shelves sees them fuse at the midline as well as with the primary palate anteriorly and the nasal septum superiorly.[13]

By the ninth week of fetal development, the initial cartilaginous facial skeleton is well established, composed of the chondrocranium forming the skull base, the nasal capsule in the upper face, and the Meckel cartilage in the lower face. Within the twelfth week of fetal growth, areas of ossification begin to appear and bone begins to rapidly replace this cartilaginous template to form the early cranial base. At this same time, the bones of the cranial vault as well as the mandible and maxilla develop through intramembranous ossification.

Postnatal Craniofacial Growth

The general pattern of postnatal development is the cephalocaudal gradient of growth, in which there is an axis of increased growth that extends away from the head. Structures away from the brain tend to grow more and later than those structures closer to the brain. Mandibular growth begins later and continues longer than does the growth of the midface.[14] This pattern of growth continues until maturity and is exemplified in the proportionality of head size to total body length. At birth, the head makes up almost a quarter of the total body length, which decreases to around

12% in the adult. Facial growth can be summarized as being driven forward initially by growth of the cranial base, then the maxilla and mandible both grow back and up to fill in the space created as they are being pulled down and forward away from the cranial base by the soft tissues in which they are embedded.

During infancy and early childhood, the cranial base increases in length through endochondral ossification that occurs at important growth sites called synchondroses. The synchondroses push the growth of the face forward until around the age of 7 years, when they begin to become less active and later ossify and fuse. Much of the forward movement of the maxilla is due to the growth of the cranial base pushing it downward and forward from behind. Further forward displacement of the maxilla results from bone apposition at the sutures located posteriorly and superiorly that connect it to the cranial base. Unlike the forward displacement generated by the synchondroses, the bone formation at these sutures is instead responsive in nature to the downward and forward pull of the maxilla from the growth of the associated soft tissues and nasal septum. Much of the increase in size of both the nasal and oral cavities occurs from surface remodeling of the maxilla and not from sutural growth. As the maxilla is translated downward and forward, the periosteum acts to remove bone at the floor of the nose, while at the same time bone is formed on the roof of the mouth. Over time, this results in a hollowing out and widening of the nasal cavity.[15] In addition, the palatal vault deepens with age despite the bone apposition on its surface because of the increased growth of the alveolar process that accompanies tooth eruption. Transverse growth in maxillary width results from growth at the midpalatal suture and appositional remodeling along the lateral aspects of the posterior region of the maxilla and the maxillary tuberosity.[16] This bone deposition at the tuberosity allows sagittal lengthening of the maxilla. The midpalatal suture begins to fuse in early adolescence but stays amenable to orthopedic force required for maxillary expansion treatment until around the age of 14 years in most individuals.

Lacking open sutures necessary for suture apposition of bone, the mandible instead grows by endochondral ossification at the condyle as well as by a combination of extensive surface bone remodeling. Unlike an epiphyseal plate or synchondroses, the growth of the condyle is responsive to translation of the mandible rather than driving it.[17] Transverse increases in the body of the mandible occur through surface apposition and remodeling of bone. The dentoalveolar structures develop with the eruption of the teeth, which continue to erupt throughout life to maintain occlusal contact, matching the vertical growth of the ramus.

First described nearly a century ago,[18] the hard and soft tissues undergo different rates of growth throughout childhood development. This is evident in the upper airway of children and has important implications for obstructive sleep disorders. Because of the significance of suckling to the newborn infant, the epiglottis is located close to the soft palate, which facilitates separation of the pathways for respiration and deglutition.[19] Neonates are born as obligate or preferential nasal breathers, but this changes as the upper airway matures. Between the ages of 1 and 2 years, vertical growth allows the larynx to descend to the level of the fifth cervical vertebra, and the epiglottis descends, which accommodates the newly acquired function of speech for the child.[20] The hyoid bone descends to a lower position in the neck, and the posterior third of the tongue descends to form the anterior wall of the oropharynx. Due mainly to hypertrophy of the adenoids and tonsils that frequently can exceed the growth of surrounding skeletal structures, adenoid and tonsillar tissue is found to be largest relative to the surrounding anatomy between the ages of 4 and 6 years.[21-23] This not coincidentally is the same age range at which OSA is most frequently seen in children. The upper airway volume then increases in adolescence because of both the concurrent increase in vertical skeletal growth and involution of the lymphoid tissue, which decreases in size after 12 years of age.[22,24] During the adolescent years, the upper airway also becomes larger in the transverse dimension and more elliptical.[24] The length and volume of the airway increase until the age of 20 years, at which time there is a variable period of stability, followed by a slow decrease in airway size after the fifth decade of life.[25]

DENTOFACIAL MORPHOLOGY ASSOCIATED WITH SLEEP DISORDERED BREATHING

Dentofacial morphology in children and in adults has been assessed by lateral and anterior-posterior cephalography, dental casts of the upper and lower teeth, digital photography, and three-dimensional magnetic resonance imaging (Table 143-1).[26] Cephalography is limited by landmark identification, measurement variability, and two-dimensional assessment of a three-dimensional anatomy. In children, only a few studies assess three-dimensional dentofacial measures.[27,28] However, all these imaging methods are performed while the patient is either awake or sedated, which does not reflect upper airway volume and soft tissue sleep-related changes. Whereas dentofacial morphology is an important component to the multidisciplinary assessment and management of SDB, there is no single cephalometric measurement that can effectively predict OSA severity,[29] as outlined in Table 143-1.

Children

In children, although adenotonsillar hypertrophy and obesity often contribute to SDB, dentofacial morphology can further contribute to the narrowing of the upper airway. Behavioral or functional oral breathing in SDB is associated with altered craniofacial growth,[30] altered muscle recruitment in the nasal and oral cavities,[31] and change in posture.[32] These ideas will be further explored in this chapter, in the section on treatment strategies. For children between 6 and 8 years, dentofacial morphology is a stronger risk factor for SDB than obesity is.[33] Cephalometric studies suggest that a long and narrow face, transverse deficiency, and retrognathia are craniofacial morphologic factors associated with a narrow upper airway and SDB in children,[34-37] which are also particular to oral breathing.[38]

Studies measuring differences in position between the maxilla and mandible (ANB: A point, nasion, B point) show an increased ANB angle in children with OSA or with primary snoring compared with controls.[36,39-43] In the primary snorers, this was associated with a decreased SNB angle (sella, nasion, B point),[36,42,43] which is a measure of retrognathia. In children with OSA, a decreased SNB angle and lower hyoid bone position and mandibular volume[28] were observed with three-dimensional imaging.

Table 143-1 Craniofacial Morphology Associated with Obstructive Sleep Apnea

	Adults	Children
Long and narrow face		Marino 2009,[34] Pirila-Parkkinen 2009,[35] Pirila-Parkkinen 2010,[36] Tsuda 2011[37]
Transverse deficiency	Johal 2004,[78] Poirrier 2012[80]	Marino 2009,[34] Pirila-Parkkinen 2009,[35] Pirila-Parkkinen 2010,[36] Tsuda 2011,[37] Cozza 2004,[40] Lofstrand-Tidestrom 1999,[44] Katyal 2013[43]
Retrognathia	Paoli 2001,[59] Guilleminault 1984,[60] Lowe 1995,[61] Lowe 1986,[62] Lyberg 1989,[64] Miles 1996,[65] Johal 2007,[69] Ishiguro 2009,[70] Chi 2011,[26] Iked 2001,[66] Sforza 2000,[67] Tangugsorn 2000,[68] Banhiran 2013,[71] Gungor 2013,[72] Okubo 2006,[75] Riha 2005[76]	Marino 2009,[34] Pirila-Parkkinen 2009,[35] Pirila-Parkkinen 2010,[36] Tsuda 2011,[37] Cappabianca 2013[28]
Increased ANB angle		Pirila-Parkkinen 2010,[36] Deng 2012,[39] Cozza 2004,[40] Zucconi 1999,[41] Katyal 2013[43]
Decreased SNB angle		Cappabianca 2013[28]
Lower hyoid bone position	Paoli 2001,[59] Guilleminault 1984,[60] Lowe 1995,[61] Lowe 1986,[62] Lyberg 1989,[64] Miles 1996,[65] Johal 2007,[69] Ishiguro 2009,[70] Chi 2011,[26] Iked 2001,[66] Sforza 2000,[67] Tangugsorn 2000,[68] Banhiran 2013,[71] Gungor 2013,[72] Riha 2005,[76] Lowe 1997[74]	Cappabianca 2013[28]
Increased mandibular plane angle		Linder-Aronson 1970,[30] Pirila-Parkkinen 2009,[35] Pirila-Parkkinen 2010,[36] Deng 2012,[39] Cozza 2004,[40] Zucconi 1999,[41] Lofstrand-Tidestrom 1999,[44] Zettergren-Wijk 2006,[45] Juliano 2009,[46] Ozdemir 2004,[47] Guilleminault 1989[48]
Increased lower anterior facial height		Linder-Aronson 1970,[30] Pirila-Parkkinen 2009,[35] Pirila-Parkkinen 2010,[36] Deng 2012,[39] Cozza 2004,[40] Zucconi 1999,[41] Lofstrand-Tidestrom 1999,[44] Zettergren-Wijk 2006,[45] Juliano 2009,[46] Ozdemir 2004,[47] Guilleminault 1989[48]
Retrusive maxilla	Paoli 2001,[59] Guilleminault 1984,[60] Lowe 1995,[61] Lowe 1986,[62] Lyberg 1989,[64] Miles 1996,[65] Johal 2007,[69] Ishiguro 2009,[70] Chi 2011,[26] Iked 2001,[66] Sforza 2000,[67] Tangugsorn 2000,[68] Banhiran 2013,[71] Gungor 2013[72]	
Palatal morphology, increased length and thickness of soft palate	Lyberg 1989,[64] Johal 2007,[69] Kurt 2011[77]	

An increased mandibular plane angle and increased lower anterior facial height are associated with OSA.* However, a meta-analysis showed significant heterogeneity across selected studies,† suggesting insufficient evidence of a strong association. Retrognathia creates a posterior displacement of the tongue base, which further narrows the upper airway and is associated with a high-arched (ogival) palate due to tongue position.[41,49] Although a narrow maxilla is associated with OSA and snoring, only a few studies have reported this from dental impressions of hard and soft tissues[35,40,43,44] as this cannot be measured from lateral cephalograms. Moreover, orthodontic correction of a narrow maxilla (rapid maxillary expansion) has been reported to reduce respiratory indexes.[50-52] These earlier studies suggesting an association between morphology and sleep apnea do not fully explain causality of dentofacial morphology in the pathophysiologic process of SDB.

In children with OSA, it was suggested that 50% also have sleep bruxism, a concomitant sleep movement disorder[53] that can affect their dentofacial health, although no causal relationship has been established. Parents report tooth grinding twice more often in children who are habitual snorers than in non-snorers[54] and more often in younger children than in older ones.[54] Up to 60% of children with sleep bruxism but without sleep apnea have a retrusive mandible; 28% have short faces (brachyfacial type).[55] Adenotonsillectomy reduced sleep bruxism muscle activity in 75% of OSA children as reported by questionnaires.[56] A temporary maxillary occlusal splint of 3 mm in thickness was worn 3 months by children (aged 6 to 8 years) with a history of sleep bruxism, oral breathing, and snoring.[57] As reported by questionnaires, tooth-grinding noises were decreased in 89% of patients, whereas snoring was reduced in 55.5% of patients.[57] This could be explained by potentially restoring nasal breathing during sleep as all

*References 27, 30, 35, 36, 39-41, 44-48.
†References 27, 35, 36, 39-41, 43-45.

participants adapted from oral breathing to nasal breathing after treatment.[57] The associations between SDB and sleep bruxism in children need further investigation and objective data. A full description of sleep bruxism is found in Chapters 147 and 150.

Adults

In adults, obesity is the main anatomic risk factor for SDB. Like children, dentofacial morphology can also contribute to a compromised upper airway, and this is more often observed in nonobese patients with OSA.[58,59]

Overall, studies have reported a retrusive mandible, macroglossia, lowered hyoid bone position, and retrusive maxilla to be associated with OSA[26,59-72] and snoring.[73] A lower hyoid bone position is suggested to be a proxy of tongue shape, posture, and tone, which could increase upper airway collapsibility.[26,74] Magnetic resonance imaging studies in Asian and white populations observed a shorter and smaller mandible in patients with sleep apnea than in controls.[26,69,75] This was significant in men with sleep apnea[26,75] and not reported in women.[26] Mandibular morphology seems to be a stronger risk factor than maxillary morphology for OSA in adults.[26] Taking into account the genetic influence on dentofacial morphology, comparison of a cohort of siblings with and without sleep apnea found a short mandible and a lower hyoid bone position to be the most important risk factors for sleep apnea.[76]

Palatal morphology and increased length and thickness of the soft palate are also risk factors for OSA[63,64,69,77] and for snoring.[77] In addition, patients with sleep apnea are reported to have longer soft palates than snorers.[73] In comparing lateral cephalograms and dental casts of OSA adults with those of controls, an increased palatal depth was seen as measured at the first and second premolars and molars,[78] although this was not a consistent finding.[79] By anteroposterior cephalometric measurements, patients with OSA had narrower maxillas.[80]

Craniofacial Syndromes

Multiple craniofacial syndromes can have a higher incidence of SDB because of modified craniofacial morphology or hypertrophy of soft tissues. Outlined in Table 143-2, among the mandibular hypoplasia syndromes are Pierre Robin, Prader-Willi, Treacher Collins, and Marfan. Moreover, some syndromes can be associated with neuromuscular disorders, which can further negatively affect breathing during sleep. In children with craniofacial syndromes and SDB, the causes of obstruction or restriction of the upper airway can be at multiple levels, requiring multidisciplinary management and multiple treatments.[81-85]

SDB is often associated in children with midface hypoplasia; 50% of nonsyndromic and syndromic craniosynostosis children,[86] such as children with Apert, Crouzon, and Pfeiffer syndromes, develop SDB.[82,87,88] SDB improves during the first 3 years of life in the absence of midface hypoplasia.[89] However, this is not observed in children with syndromic craniosynostosis and midface hypoplasia (Apert or Crouzon/Pfeiffer).[89] Midface advancement surgery was successful in the short term with improved respiratory outcomes in 55% but was ineffective in 45% of children.[82] In this latter group, endoscopy and volume measurements showed obstruction of the hypopharynx.[82] After adenotonsillectomy, 60% of children with syndromic craniosynostosis had less desaturation events (Sao_2 4%).[81] Achondroplasia, an autosomal dominant congenital disorder, is also characterized by midface hypoplasia, leading to increased risks for development of SDB. Approximately 35% of patients with achondroplasia have SDB,[90] associated with increased lower facial height and retrognathia.[91]

Table 143-2 Syndromic and Craniofacial Morphology Associated with Obstructive Sleep Apnea

- **Maxillary hypoplasia or midface hypoplasia**
  such as syndromic craniosynostosis (Apert, Crouzon, Pfeiffer), achondroplasia, trisomy 21, and cleft palate
- **Mandibular hypoplasia**
  such as Pierre Robin, Prader-Willi, Treacher Collins, and Marfan
- **Mandibular hypoplasia or micrognathia**
  such as Pierre-Robin, Smith-Lemli-Opitz syndrome, and trisomy 21
- **Orofacial hypotonia**
  such as Smith-Lemli-Opitz syndrome and trisomy 21
- **Cleft lip and palate**
  such as Pierre Robin
- **Maxillary and mandibular hypoplasia**
  such as Turner syndrome
- **Neuromuscular disorders**
  such as Duchenne muscular dystrophy, myopathies, Guillain-Barré syndrome, and myasthenia gravis

In addition to midface hypoplasia, patients with trisomy 21 also have micrognathia and orofacial hypotonia, which predisposes them to SDB[92] in 50% of pediatric and adult cases.[83,93-95] Although suggested as the first-line treatment, adenotonsillectomy resolves SDB in only 27% to 34% of cases.[96,97] Other alternative treatments are positive airway pressure therapy, mandibular distraction osteogenesis, midface advancement, and oral appliances. Oral appliances with tongue-stimulating knobs have been shown to improve orofacial muscle function.[83,84,98-103]

Cleft lip and palate children showing midface hypoplasia have a significantly higher incidence of SDB (22% to 37.5%)[85,104] than do healthy children (5%).[105] Furthermore, 34% of syndromic children with cleft lip and palate reported symptoms, whereas 17% of nonsyndromic children with cleft lip and palate did.[85] Following various surgical interventions (e.g., adenotonsillectomy, flap takedown, tonsillectomy, and partial adenoidectomy), only 38.5% of patients had improved SDB.[85]

Syndromes characterized by micrognathia and orofacial hypotonia, such as Smith-Lemli-Opitz syndrome, or by smaller maxilla and mandible, such as Turner syndrome, can also have increased incidence of SDB.[106,107]

Pierre Robin sequence is characterized by a triad of craniofacial anomalies consisting of micrognathia, cleft palate, and glossoptosis leading to respiratory and feeding issues. Most infants with Pierre Robin sequence (85%) also have sleep breathing disorders.[108] Nonsurgical management of Pierre Robin sequence in infants includes positional therapy, placement of nasopharyngeal airway, and oral appliances with velar extension. Positional therapy is successful in 49% to 52% of cases.[109,110] Oral appliances with velar extension were reported in one study to effectively reduce OSA at hospital

discharge and 3 months later, in the absence of any adverse events.[111] Surgical management includes tongue-lip adhesion or glossopexy, subperiosteal release of the floor of the mouth, mandibular distraction osteogenesis, and tracheotomy. Mandibular distraction osteogenesis is reported to have better outcome measures than tongue-lip adhesion in regard to oxygen saturation, apnea-hypopnea index (AHI), and number of postprocedure tracheostomies.[112] However, both surgeries had comparable complications.[112]

TREATMENT STRATEGIES

Management and treatment acknowledge that SDB has a cumulative positive feedback loop in which repetitive respiratory-related arousals cause changes to the properties of the upper airway, and this causes end-stage sequelae that exacerbate the initial stimulus.[113] This section focuses on the treatment strategies for the pediatric patient. For more information on adult SDB treatment, refer to Section 14 of this volume.

The indication for treatment underlies what is pathologic and what is normal. What constitutes disease that necessitates treatment? Recent evidence shows that simple snoring carries consequences of cognitive impairments, suggesting that benign primary snoring should be treated as a disease. The incidence of snoring is much higher than the incidence of OSA,[114] and snoring is now recognized as an abnormality in children.[115] If snoring is the start of the cascade of abnormal breathing, early treatment may be correlated to improved outcomes. Even with early intervention, the recurrence of SDB has been documented in studies of pediatric patients observed through adolescence.[116,117] So, whereas the timing of treatment is relevant, early detection and treatment of SDB children are only part of the solution because there can be a familial inheritance of both symptoms and anatomic risk factors.[118-120] Consequently, asymptomatic children with at-risk morphologic characteristics and a familial history of SDB should be monitored for further evaluation.

There is not a linear relationship of symptoms to severity of the disease, and similarly, if there is a disproportion in anatomic structures, it may not necessarily correlate with the symptoms or the severity of the disease.[121] A study that observed SDB children 4 years after treatment showed that despite the improvement in respiratory parameters, complaints of daytime sleepiness persisted in the treated and untreated SDB group compared with asymptomatic, non-SDB controls. For the health care practitioner treating children, there are screening measures to decide if a child needs further care. Again, see Section 14 of this volume for more information on SDB and OSA diagnosis and management. These factors can be distilled down to the presence of daytime or nighttime symptoms, the orofacial anatomy, and the familial history of SDB (Box 143-1). The treatments for pediatric SDB are multiple and listed in Box 143-2.[122,123] The intent of early treatment is to halt the continued cycle of worsening OSA and also to prevent early systemic complications. The consequences of neurocognitive deficits and cardiovascular changes are evident in children,[124,125] similar to the systemic changes seen in adults. It is possible that OSA in an adult could have begun in childhood or adolescence.[126] However, there are no long-term outcome studies that demonstrate the progression of these changes from childhood into adulthood, nor is there any evidence that these effects seen in children are severity dependent or evolve into the adult presentation with end-organ morbidity.

Box 143-1 SCREENING MEASURES FOR PEDIATRIC TREATMENT

Presence of daytime and nighttime symptoms
Orofacial anatomy
Familial history

Box 143-2 TREATMENT OPTIONS FOR PEDIATRIC OBSTRUCTIVE SLEEP APNEA SYNDROME

Soft tissue reduction or removal (tonsils, adenoids, turbinates)
Allergy management
Jaw expansion
Myofunctional therapy
Nasal continuous positive airway pressure
Skeletal surgery

Box 143-3 CURATIVE STRATEGIES FOR PEDIATRIC OBSTRUCTIVE SLEEP APNEA SYNDROME

Increase in airway size (nasopharynx, oropharynx, hypopharynx)
Improvement in muscle response
Craniofacial growth modification

There are three general treatment strategies for SDB. The problem of collapsibility affecting airflow exchange in SDB can be primarily related to inadequate airway size creating airflow resistance, and the first line of treatment is directed at enlarging the airway. Magnetic resonance imaging studies of OSA children show a smaller upper airway cross-sectional area.[127] However, children with normal oropharyngeal anatomy may suffer from OSA,[128] and the AHI has not been shown to directly correlate with airway volume.[121] First-line treatment approaches to increasing space in the nasopharynx, oropharynx, and hypopharynx are reviewed. SDB may also encompass a component of muscle alteration that may be a primary (etiologic) or secondary effect.[113] The second component of therapy addresses the muscle remodeling and myopathy that may be associated with SDB. The third strategy incorporates the challenge to complete care, and that raises the question of whether a cure or complete resolution of the disease is possible by changing the underlying facial growth pattern and modified anatomy to eliminate this as an etiologic factor in the disease. This paradigm for care is summarized in Box 143-3.

Strategy 1: Increase in Airway Size

Location–Nasopharynx

Although the area of greatest collapsibility is the soft tissue of the oropharynx, the properties along the entire upper airway will affect this collapsibility. Each part of the pharynx (i.e., nasopharynx, oropharynx, hypopharynx) serves different

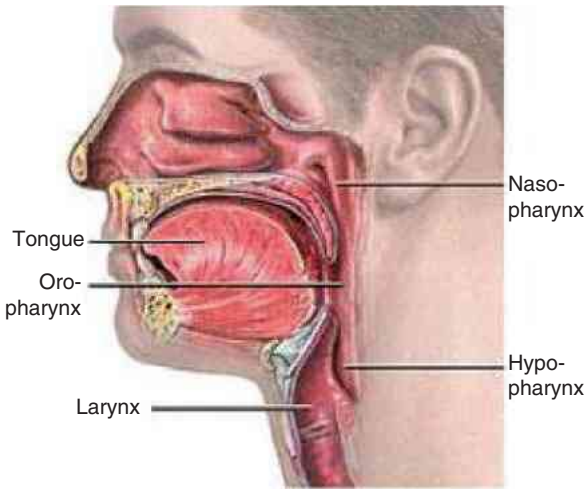


Figure 143-1 The pharynx, where airway collapsibility occurs, is divided into three sections: nasopharynx, oropharynx, and hypopharynx.

Box 143-4 INCREASING THE NASOPHARYNX SIZE

Reduction of inflammation through medication
Orthodontic expansion
Surgical removal of soft tissue

functional roles, and so treatments to increase airway size are reviewed according to site in the pharynx (Figure 143-1).

The entrance to the airway at the initial site of airflow is the nasopharynx. The treatments to increase the size of the nasopharynx (Box 143-4) address either removal of obstructions inside the nose or enlargement of the space itself. The nasal influences on snoring and SDB are widely known because nasal obstruction can cause sleep disturbances that affect daytime performance.[129] The degree of nasal obstruction does not correlate with the severity of OSA,[130] probably because the extent of nasal resistance does not correlate with the amount of nasal airflow.[131] This is evident in patients with choanal atresia who have nasal obstruction as a clinical feature. In this population, it was reported that 65% are diagnosed with OSA[132] versus the entire patient group. Systematic review of the relationship between nasal obstruction and OSA shows that nasal obstruction plays a modulating role in OSA but is not a direct causative factor.[133] Whereas the relationship between nasal obstruction and SDB is not linear, it is thought either to be linked to an increase in nasal resistance initiating unstable oral breathing or to result from impaired nasal reflexes that hinder continued ventilation.[134] Increased nasal resistance depresses the critical closing pressure of the pharyngeal muscle walls, rendering the airway more collapsible. The critical closing pressure is correlated with the severity of SDB.[121] Pharyngeal compliance is impaired in a cycle in which SDB can both result from and be worsened by nasal obstruction.[135]

Reduction of Inflammation and Medication Management. SDB is more prevalent in patients with allergic diseases.[136] Allergic rhinitis hinders nasal respiration by increasing nasal resistance and is considered a risk factor for SDB.[137] Allergic rhinitis is one of the major causes of impaired nasal function, and it affects up to 40% of the general population in developed countries, with an increasing prevalence. Although the evidence generally supports a connection between SDB and allergic rhinitis, this connection is not definitive, and the mechanism linking these two diseases and the mechanism of how nasal inflammation causes SDB are unclear.[138] However, nasal resistance measured by anterior rhinometry is increased in children with OSA compared with control subjects.[139] Larger population studies[140] and several smaller studies demonstrated an improvement in SDB by treatment of allergic rhinitis through medication management.[141] Nasal corticosteroids in children with moderate sleep OSA was effective in improving symptoms but not in eliminating the disorder,[140] thus suggesting the role of inflammation in the disease.[142] The cause and effect link is not yet fully understood or explained between allergic rhinitis and SDB. This topic is explored in more detail in Chapter 119.

Orthodontic Expansion. As early as the 1860s, the midpalatal suture was separated within 2 weeks,[145] using an orthodontic screw-type expander device to widen the transverse dimension of the upper jaw either to correct a transverse deficiency or to create space for the permanent teeth. Early studies of maxillary expansion show that both dentoalveolar and craniofacial structural changes were created.[146,147] The amount, location, and rate of force application to the facial skeleton from expansion appliances create localized changes in the bony housing surrounding the teeth, with a potential effect at the sutural level of the maxilla.[148,149] Rapid maxillary expansion (RME) which refers to a rate of expansion of at least 0.25 mm/day, was described in the medical literature dating back to 1975[150] as a therapy for medical ailments and referenced in the dental literature for medical problems since 1974.[151] This early work described maxillary expansion to treat problems such as enuresis, nasal congestion, and asthma. These symptoms also describe the OSA syndrome, although the syndrome was not coined until 1976.[152]

In 1980, surgical widening of the maxilla to increase the lateral dimension of the nasopharynx was first described as a treatment for adult OSA. Nonsurgical rapid palatal expansion was first suggested as a therapy for OSA in 1998.[143] Pirelli et al[144] published seminal work in 2004 using rapid maxillary expansion (RME) to successfully treat OSA syndrome in children with narrow maxillas. Several other groups have corroborated this work, and now other published studies also describe the effectiveness of maxillary expansion in treating children with or without narrow palates, with or without retrognathic mandibles. These studies are outlined in Table 143-3, and the general appliance design is depicted in Figure 143-2. In the maxilla, the applied force from the expansion appliance creates midpalatal (also called median palatal or palatal) suture separation. This results in distraction osteogenesis across the palate, resulting in an increase in width of the maxilla (Figure 143-3).[147] Both oral and nasal volume is increased as the triangular nasal fossa is enlarged by widening of the nasal floor and outward movement of the lateral nasal walls.[153] The volumetric increase evident in the nasal fossa but not posteriorly in the nasopharynx[154] has an effect on nasal airflow, but this is not a proportional relation.

Several studies show that maxillary expansion reduces nasal resistance.[153,155-157] In a systematic review of all expansion studies across several different databases, there were

Table 143-3 Expansion Outcome Studies for Obstructive Sleep Apnea

Author	Year	Mean Age (years)	N	Responders
Pirelli et al[144]	2004	8.7	31	31
Villa et al[51]	2007	6.6	14	12
Miano et al[249]	2009	4–8	9	9
Villa et al[52]	2011	6.6	10	8 (3-year follow-up)
Guilleminault et al[167]	2011	6.5	31	31
Marino et al[250]	2012	5.9	15	8
Guilleminault et al[117]	2013	7.5	24	24
Villa et al[245]	2014	3.7	22	18

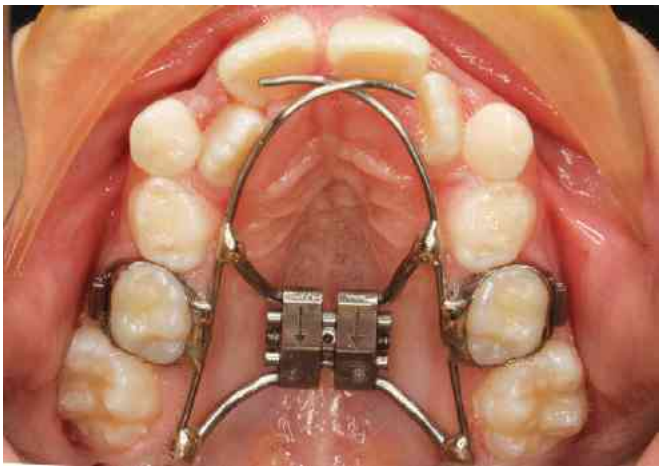


Figure 143-2 Example of maxillary expansion appliance in the mixed dentition.

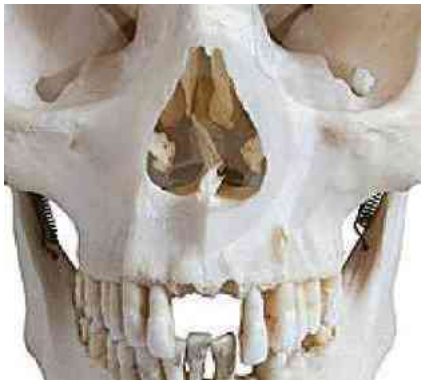


Figure 143-3 Skeletal boundaries of the nasal cavity. The maxilla forms the floor of the nose (F); the lateral walls are the sides of the maxilla (LW). This intranasal space is increased volumetrically with maxillary expansion.

eight controlled studies with 6-month follow-up after therapy that measured changes in airway dimensions and functions after RME. These studies support a moderate level of evidence that RME therapy in a growing child causes increases in nasal cavity width and in the posterior nasal airway, associated with reduced nasal resistance and increased total nasal flow.[158,159] The stability of the results can be expected for at least 11 months after the orthopedic therapy.[160] In one study, nasal airflow measured by rhinomanometry improved in the supine position in 65% of the patients.[161] Changes in nasal geometry of increases in intranasal width[153,162,163] and increases in both nasal cross-sectional area and nasal volume[153,164] were noted. Airway properties were examined, showing decreases in nasal resistance measured by rhinomanometry[153] and acoustic rhinometry,[164] along with changes in head position and decreases in the craniocervical angle.[161,165,166] RME therapy through sutural opening creates an increase in the nasal cavity width, area, and volume in children, which allows a reduction in nasal resistance.

The studies in Table 143-3 support maxillary expansion as a treatment modality for pediatric OSA syndrome. The majority of children responded to expansion therapy, and OSA was eliminated in a few children.[144,167] Most of the expansion studies looked at children with narrow upper jaws (selection criteria for treatment) and malocclusions, including crossbites, dental crowding, and mandibular retrusion. Expansion as a first-line therapy was initiated in a few of the studies. Mandibular retrusion was not specifically a factor in selection of patients, although it was noted in more than half of the patients studied. Two of the studies used bimaxillary expansion.[117,167] Bimaxillary expansion was employed because of the dental compensation in both the maxilla and mandible from a narrowed maxilla. The dentition is tipped inward toward the tongue, which creates a narrowed intraoral space. Dental expansion of the lower dentition aids in achieving maximum skeletal expansion of the upper jaw. The effectiveness of bimaxillary expansion as a treatment option for pediatric SDB was first described a decade ago.[168] The overall expansion data of all studies are varied, but mostly with a general improvement in both sleep parameters and subjective symptoms of SDB.

The therapeutic effects of maxillary expansion include increasing space for dental crowding,[169] increasing airway dimensions, and decreasing nasal resistance[153,155,156] in addition to treating SDB (Table 143-3). These studies also demonstrate a greater gain when RME treatment is rendered before the pubertal growth spurt peak, demonstrating an age-related phenomenon to long-term success. In the studies outlined in Table 143-3, expansion was started as early as age 3 years up to age 15 years. Whereas nonsurgical expansion of the palate is still possible beyond the age of 15 years, because of increased interdigitation of the suture in an older child, the widening that occurs is more often dental tipping that increases the intraoral volume and not true skeletal expansion

that increases the intranasal space. Greater changes in nasal width were evidenced when expansion was done early in maturation versus late in maturation.[162] By the age of 4 years, the craniofacial skeleton has reached 60% of its adult size; by the age of 7 years, 75% total craniofacial growth is complete, and by the age of 12 years, 90% total craniofacial growth is reached.[170] This suggests that RME therapy should be considered as an early-stage treatment in pediatric SDB, as the intervention timing seems critical for predicting RME orthopedics outcomes.

Maxillary expansion is a common noninvasive orthodontic treatment that is well tolerated by children. Advantages as therapy for SDB include little or no risk of morbidity or discomfort, with treatment performed in an outpatient setting during a 4- to 6-month period. There is a high level of acceptance for initiating this type of therapy, especially because many children with SDB also have concomitant dental crowding. However, some type of holding or retaining appliance is needed to maintain the expanded arch form, and so there is long-term ongoing care. If early expansion is advocated, a holding appliance may be in place for many years, and the timing is dependent on the eruption status of the dentition and the ability to sustain nasal respiration.

The reported risks of RME therapy include bite opening, relapse, microtrauma of the temporomandibular joint and the midpalatal suture, gingival recession, and root resorption.[158,160] Whereas these effects are not usually encountered, the incidence is also age dependent, seen more with maturation. There are no current guidelines for selection of patients, other than the studies that treated children with narrow palates and dental crossbites, in which teeth of the upper arch do not horizontally overlap teeth of the lower arch. Future work will help identity which types of patients will benefit most, how much expansion can be gained, and at what age to start expansion.[171] An example of the shape changes and the amount of space that can be gained is shown in Figure 143-4.

Whereas the expansion results for SDB are promising, these studies were not controlled or randomized and were limited to only a few groups that reported data. Few new therapies of pediatric SDB have been validated with randomized controlled trials.[251] Current pediatric guidelines recommend referring children with maxillary transverse narrowing for orthodontic therapy and possible RME treatment for persistent OSA syndrome after adenotonsillectomy.[122] Orthodontic expansion therapy has three potential effects for the SDB child. It can widen the intranasal volume to reduce nasal resistance, which improves airway collapsibility and SDB; it can facilitate other SDB treatment, such as positive pressure therapy, by allowing improved nasal airflow; and it can facilitate the transition from oral to nasal respiration, which can have a secondary effect on oropharyngeal growth. Because maxillary expansion is so well tolerated by children, it is expected that in the coming years, larger scale trials will examine the efficacy of jaw expansion as one component of improving the properties of the developing airway in a child with SDB. The effect of RME on orofacial growth and facilitating nasal respiration is discussed in the next part of this chapter.

Figure 143-4 Example of the amount of expansion that can be created. Preexpansion model shows a width of 42 mm, measured from the central groove of the upper primary second molar, compared with postexpansion width of 52 mm.

Surgical Expansion: Removal of Soft Tissue. An alternative mode of increasing airway size is through the surgical removal of structures or obstructions. This is further described in Chapter 149, and for children, these procedures may include reduction of the inferior nasal turbinates, sinus surgery, or adenoidectomy. There is a nonlinear relationship between the level of nasal resistance and the severity of SDB. This may explain why the reported cure rate of adenotonsillectomy for OSA in meta-analyses of children with a mean age of 6.5 years was only 59.8%, showing that adenotonsillectomy as the first-line and most common therapy for pediatric OSA may be insufficient.[172]

Location–Oropharynx

The retropalatal area is represented here, which is often the site of greatest airway narrowing in children. SDB children have more fluctuation in airway size, with narrowing during inspiration that is more prominent in higher oropharyngeal levels.[127] Surgical adenotonsillectomy can enlarge this space, as shown in Box 143-5. Whereas maxillary expansion increases the intraoral space, imaging studies using cone beam computed tomography and lateral cephalography do not depict any changes at the oropharyngeal level with RME. OSA children have narrowed space at this level compared with controls, but after RME therapy, there was no evidence of increased oropharyngeal airway volume.[173,174] However, it has been postulated that the increased oral volume resulting from maxillary expansion induces a postural change in the tongue position.[175]

Surgical Expansion: Removal of Soft Tissue. Hypertrophy of the tonsils (pharyngeal and palatine) is the second major cause of respiratory obstruction in childhood, followed by allergic rhinitis, and is found to be associated with allergic rhinitis in many children, exacerbating the respiratory symptoms.[176] As described earlier in this chapter, tonsils generally initiate hypertrophy within the first 3 years of life, the period of highest immunologic activity during childhood. Because tonsil growth outpaces craniofacial growth from 3 to 7 years

Box 143-5 INCREASING THE OROPHARYNX SIZE

Orthodontic expansion
Surgical removal of soft tissue

Box 143-6 INCREASING THE HYPOPHARYNX SIZE

Oral appliances (Chapter 147)
Surgical expansion (Chapter 149)
Nasal continuous positive airway pressure (Chapter 115)

of age, most symptoms are observed during this period, coinciding with the peak incidence of childhood OSA syndrome.[177] Tonsil atrophy starts after 10 years of age and is completed in adulthood.[178] The first-line therapy for pediatric SDB is adenotonsillectomy, which has been associated with a decline in the critical closing pressure of the muscles along the pharynx, rendering the upper airway less collapsible.[121] Surgical therapies are further described in Chapter 149.

Location–Hypopharynx

Hypopharyngeal airway obstruction can be caused by the prominence or relaxation of the base of the tongue, the lateral pharyngeal wall, and, occasionally, the aryepiglottic folds of the epiglottis. The treatments to increase the size of the hypopharynx are outlined in Box 143-6.

Oral Appliances. Several case report studies show the effectiveness of oral appliances that hold the lower jaw forward in treating SDB and OSA in children. The use of oral appliances in the treatment of adults is well established, and the specific mechanics of how these appliances work is further discussed in Chapter 147. Oral appliances that advance the jaw forward are similar to the functional appliances used in orthodontics that address problems of mandibular deficiency. In children, these appliances create changes in the dentition and the growth of the maxillomandibular complex, and these alterations are thoroughly described in the orthodontic literature. A 2007 Cochrane review[179] examined studies in children aged 15 years and younger using mandibular advancement appliances. It concluded that there was not sufficient evidence to support treatment of pediatric OSA syndrome using oral appliances. In the more than 200 studies, only one study was recognized for inclusion,[180] which examined children aged 4 to 10 years against a control group of no treatment, showing a reduction in respiratory parameters in the majority of the treatment group. It must be highlighted that oral appliances can affect the forward growth of the upper and lower jaw, which likely makes their use inappropriate unless the child presents with retrognathia. The long-term side effects of this growth on the pediatric developing airway warrant further examination.

Surgical Expansion: Skeletal Surgery. Orthognathic advancement surgery is not considered a treatment option for the pediatric patient until after jaw growth cessation. Because of the late-stage timing, other therapies would be enacted before orthognathic surgery is planned. Mandibular advancement surgery enlarges the hypopharyngeal space, whereas maxillary advancement surgery enlarges the oropharyngeal cavity. Bimaxillary or maxillomandibular advancement surgery creates expansion across the entire pharynx, primarily at the oropharynx and hypopharynx, with a secondary impact on the nasopharynx. Expansion/advancement surgery is described in Chapter 149.

Continuous Positive Airway Pressure. Nasal continuous positive airway pressure (CPAP) stents the pharyngeal airway open and prevents the muscular walls from collapsing. It is not a curative strategy as it does not increase the airway size or change the neuromotor properties of the surrounding musculature. Studies in children[181] demonstrate the efficacy of CPAP therapy in reducing or eliminating symptoms and improving respiration but also acknowledge the challenges of compliance with and adherence to routine use. Consequently, CPAP is used as a secondary measure when adenotonsillectomy, bimaxillary expansion, or pharmacologic management has not improved SDB or as a primary option for obese children or children with craniofacial syndromes. This form of therapy is further outlined in Chapter 115.

Strategy 2: Improvement in Muscle Response

If OSA is modeled as a disease of progressive muscle degeneration, one strategy for treatment would counter these resultant muscular changes that stem from pharyngeal muscle damage. The collapsible pharynx is a tube composed of paired muscles with no bony perimeter, mediating airflow from the nose in the upper airway to the lungs of the lower airway. With repetitive collapse in OSA, the sustained pressure in the collapsible pharynx can result in repetitive microtrauma to the pharyngeal muscles.[182,183] Over time, these insults can lessen muscular neural control of the upper airway muscles that regulate airway opening and collapse, the interplay of the airway dilating opening muscles against the airway contracting or closing muscles. The pharyngeal muscle activation can become altered such that the response to neurochemical (hypoxemia or hypercapnia), neuromechanical (respiratory effort), or sensory (afferent input) stimuli becomes diminished or absent.

The tongue muscle is one of the largest structures defining the oropharyngeal airway and bounds its anterior aspect with motor innervation by the hypoglossal nerve and sensory innervations by the lingual nerve and glossopharyngeal nerve. As the largest and most studied pharyngeal dilator muscle, it almost directly controls airway patency by its forward movement, enlarging the airway. It is composed of extrinsic muscles (genioglossus, hyoglossus, and styloglossus) that alter its position and intrinsic muscles that alter its shape, both of which can affect airway size and shape.

Although SDB is often an anatomic problem of small size, the tongue volume in children with OSA syndrome does not differ from that of controls.[184] Similarly, the other muscles that border the pharyngeal airway, such as the pterygoid muscles, pharyngeal constrictor muscles, and parapharyngeal fat pads, were similar in size in both normal and OSA syndrome–affected children. In children with OSA syndrome, there was no muscle thickening of the lateral pharyngeal wall as reported in adults, suggesting that the surrounding craniofacial structures create the deficient anatomic volume differences and not the muscular soft tissues. In contrast to OSA children, OSA adults show evidence of muscle remodeling in phenotype of muscle size or fiber type,[185] which affect the sensory properties and strength performance of the muscle system. So, whereas the size of the muscle tissue mass is not enlarged in children with OSA syndrome, it implies that repetitive activations to the muscle can create muscle enlargement and changes in muscle properties (i.e.,

length, force) and innervations (i.e., spindle, motoneuron) that then can become problematic in adults.[186] It is not known if the upper airway neural abnormality is related to muscle activation or muscle recruitment; much more work is need to understand such complex interactions and impact on breathing.

At sleep onset, pharyngeal muscle activity is reduced, and it becomes slightly atonic during rapid eye movement sleep (see Chapters 21 to 23 for more information on muscle physiology and breathing). During adolescence, the dilating upper airway reflexes show a gradual reduction of responsiveness, and so the collapsibility of the upper airway increases.[187] Some amount of airway reflex attenuation or blunting occurs with age. Normal children and adolescents show neuromuscular compensations of increasing genioglossus electromyographic (EMG) activity in response to increased resistive loading, indicating that these children and adolescents have active upper airway neuromuscular reflexes during sleep. OSA syndrome children and adults show higher EMG activity during wake than normal controls and lower EMG activity during sleep.[188,189] These daytime increases in EMG activation impart a resistance to collapse,[190] and these mechanisms have been described in obese non-OSA children[191] as a neural compensation. Snoring may or may not be present in children, but over time it creates vibratory stress as mediated in the upper airway and is hypothesized to induce change or injury to the affected pharyngeal muscles.

Because of this eventual blunted airway muscle reflex to motor and sensory stimuli, efforts have been targeted toward muscle rehabilitation, specifically to address deficiencies in muscle activation. It is thought that these changes in the upper airway muscles are reversible. Electrical neurostimulation as a form of muscle training activation on animals and human subjects was experimented to maintain pharyngeal airway patency during sleep. Stimulation was applied directly on the hypoglossal nerve[192] or the dilator muscles.[192-195] These studies show mixed results, but current investigations are ongoing.

Myofunctional Therapy

More recent work in muscle rehabilitation explores pharyngeal muscle exercise therapy to change muscle strength, posture, and responsiveness. These oropharyngeal exercises of myofunctional therapy (MFT, also called orofacial myofunctional therapy and orofacial myology) are intended to improve the neuromechanical performance of the pharyngeal dilators. History dates MFT to the field of speech pathology, with references to the orthodontic literature in 1906, recognizing the role of the facial musculature in the development of malocclusions. The term myofunctional therapy was coined in 1935.[196] MFT by definition is a tongue and lip muscle exercise therapy to treat malocclusions and other dental and speech disorders, usually as a result of an anterior tongue displacement pattern.[197] It has also been implemented as a cotherapy for orthodontic management of craniofacial growth along with the treatment of crossbites and anterior open bite malocclusions.[198,199] With a high level of evidence, a form of oral MFT has been used to treat dysphagia resulting from neuromuscular movement disorders or stroke patients.[200] The potential mechanisms implicated in the growth deceleration of OSA include dysphagia resulting from adenotonsillar hypertrophy.[201]

The tenets of MFT address muscle hypotonia and weakness by strengthening the nasal mode of respiration using resistance and range of motion exercises. Performance of these exercises during the day supposes that airway collapsibility during sleep will be reduced. As the pharyngeal muscles are involved in creating speech, swallowing and deglutition, and breathing, therapy will treat dysfunctions in abnormal drinking, chewing, and swallowing while correcting the resting posture of the lips, tongue, and jaw by repatterning of facial muscles.[202] This is thought to restore the altered sensory input and proprioception to the affected musculature. Swallowing as an upper airway reflex is impaired in OSA patients,[203] suggesting that targeted therapy for swallowing could improve upper airway motor tone, although the degree of swallowing impairment does not correlate with OSA severity.[204] Preliminary studies of muscle rehabilitation exercises combined with oral shields were effective in reducing reports of snoring in normal and overweight (but not obese) middle-aged individuals as measured on an analog scale in non-OSA adults.[205] OSA young adults in a small-scale trial using MFT for 2 months showed overall improvement in AHI, oxygen saturation, and strength of the perioral musculature. The small sample size of six patients warrants further investigation to strengthen these findings.[206]

Randomized and controlled studies using MFT in adults reinforce the efficacy of muscle exercises on decreasing airway collapsibility in the middle-aged population with moderate OSA. One study with almost daily didgeridoo instrument playing for 4 months' duration in combination with breathing training yielded improved results of AHI and SDB symptoms as measured by various scales.[207] Another investigation using a highly specified regimen of oropharyngeal exercises examined an overweight, middle-aged, and predominantly male cohort that underwent 3 months of upper airway muscle rehabilitation. Compared with controls, the treatment group showed improvements in all outcome variables of improved AHI severity, lowest oxygen saturation, and symptoms and thus laid the foundation for further exploration.[208]

The muscle remodeling phenomenon seen in adults has not been evidenced in children, but the tenets of muscle strengthening can be applied to pediatric SDB. A retrospective study of adolescent youth showed resolution of the breathing abnormalities seen during sleep, but only after MFT was supplemented to the treatment regimen. Prior adenotonsillectomy and orthodontic expansion were other treatments used to increase nasopharyngeal and oropharyngeal space; however, flow limitations and daytime symptoms of mild OSA persisted after these therapies.[123] A prospective study of MFT and twice-daily nasal lavage was instituted for a 2-month period for children older than 4 years with residual moderate to severe OSA after adenotonsillectomy. This randomized and controlled work showed considerable improvement in AHI severity in all 14 children using MFT.[209]

These studies provide evidence to identify MFT as a potential adjunctive treatment for SDB in adults and children. MFT yielded improvements in both respiratory parameters and symptoms of SDB. The aspect of inadequate muscle activation or recruitment is not completely managed with mechanical (CPAP), surgical, orthodontic, or medication therapy. MFT may address these problems of diminished pharyngeal muscle motor tone from neural motor and

sensory deficiencies in all parts of the pharynx, extending from the nasopharynx to the hypopharynx. Future work in this area will likely outline screening and treatment parameters and identify the most effective muscle rehabilitation techniques for managing SDB. Most patients who undergo MFT do not have imaging to demonstrate an abnormal swallow reflex or have EMG tests to show abnormal muscle activation, and so identification of selection criteria is ambiguous. Longer term studies can address questions of treatment timing, age, and severity appropriateness and lend more definition to the initiation, end point, and duration of MFT. (See also Chapter 150 for other approaches to improving SDB.)

Strategy 3: Craniofacial Growth Modification

Changing the Pattern of Respiration

Does oral breathing cause abnormal orofacial development, and if so, does this cascade of altered growth exacerbate SDB or predispose a growing child to SDB? A simple cause and effect relationship between upper airway respiratory function and dentofacial and skeletal morphology has not yet been determined, despite the myriad studies in the clinical specialties of otolaryngology, oral surgery, and orthodontics—all areas that are specific to treating the upper airway, the jaws, or the teeth. However, there are associations between OSA syndrome, craniofacial anatomy, facial growth, jaw size and shape, and malocclusions as outlined earlier in this chapter. The genesis of these relationships stems from the hypothesis that oral breathing from impaired nasorespiratory function can affect craniofacial growth. Although controversial, the relationship between nasal obstruction and facial growth has been demonstrated in animal experimental and human clinical trials.[210]

Because the nasal respiratory pattern is critical in the development of a normal upper airway, initial SDB treatment strategies in children are targeted to promote and to sustain nasal respiration. Removal of nasal obstructions can promote nasal respiration, during the day and night, but the relevance of establishing daytime nasal respiration in affecting nighttime upper airway properties is not known. If nasal obstruction is sufficiently severe, a transition to oral breathing may occur. Nasal breathing is the primary route of airflow, responsible for most inhaled ventilation during wake and sleep states. The transition from nasal to oronasal breathing at night increases with age.[211] Transitioning to primarily oral airflow can be detrimental because it can change the upper airway properties. Open mouth breathing during sleep can create lengthening of the pharynx and lowering of the hyoid bone,[212] which increases the upper airway resistance from increased collapsibility of the pharyngeal lumen and posterior retraction of the tongue. Upper airway resistance during sleep is significantly lower during nasal breathing than during oral breathing,[213-215] which may further compromise the airway and increase the effort of breathing if oral breathing is sustained during sleep.[135]

Oral breathing in children can also have a potentially dysmorphic effect on the developing orofacial complex. Based on the earlier described functional matrix theory of Moss, it is thought that nasal respiration provides continuous airflow through the nasal passage during breathing to induce a constant stimulus for the lateral growth of the maxilla and for lowering of the palatal vault. The position and shape of the bone are the result of air pressures through the nasal and oral cavities.[216] The second effect of oral breathing on the facial skeleton is mediated by changes in muscle recruitment patterns, which result in an altered soft tissue and skeletal morphology and posture.[217] Increased nasal resistance can yield a more collapsible airway, and it potentiates mouth breathing that can create the postural and thus muscular changes that lead to unfavorable jaw growth. So the oral breathing cycle is perpetuated once it is initiated. This suggests that one of the tenets of SDB therapy would be to change the respiratory pattern from oral breathing to a predominantly nasal breathing mode.

Before strategies to create oral breathing can be created, it is challenging to understand what oral breathing means. There is not a standard well-accepted test to define an oral breather. The controversy stems from the inability to quantify nasal versus oral respiration or if spontaneous transitions from oral to nasal breathing occur and the lack of long-term data with growth maturation. Reliable tests to assess continuous airflow through the nose and mouth are lacking, and often the assessments are made on clinical presentation and subjective perceptions of the patient. Tests that measure airflow and nasal resistance, such as anterior rhinometry, acoustic rhinometry, nasal peak flow, rhinomanometry, and pneumotachography, are not routinely used because the testing results do not consistently correlate with the patients' subjective complaints.[218,219] Oral breathing is thought of as an oral habit or as an adaptive response to a perturbation. Despite the associations to maxillomandibular growth, few studies explore how to cure oral breathing. The most common respiratory mode is a combination of simultaneous oral and nasal airflow.[220] During sleep, normal subjects without nasal disease or SDB are nasal breathers, with only 4% of total ventilation time spent breathing orally.[221] Neonates are born as obligate or preferential nasal breathers, but as outlined earlier in this chapter, this changes as the upper airway matures. Even patients with severe nasal obstruction from either allergies or soft tissue enlargement display some measure of nasal respiration.[222]

Given the multifactorial nature of the nasorespiratory pattern and function on the expression of SDB and facial growth, relief of nasal obstruction to create nasal respiration is only part of the solution. There are likely other multilevel variables that should also be addressed, such as the development of nasal breathing at the habitual level. Some children with adequate upper airways breathe through the mouth because of habit.[222] Nasal obstruction can be the stimulus for oral breathing, but removal of nasal obstruction to create nasal patency does not always induce a spontaneous change in the respiratory pattern.[223] Although RME therapy can enlarge the intranasal space and reduce nasal resistance, the oral respiratory mode does not automatically revert,[155] suggesting that a patterned mode of breathing develops.[224] Increased nasal airflow is not enough to achieve nasal breathing because other factors, such as nasal concha hyperplasia, nasal polyps, adenoidal hypertrophy, and septal deviation, are responsible for oral breathing.[225] Even after surgical removal of enlarged tonsils in children, changes in airway muscle tone were variable after surgery,[226] demonstrating a spontaneous but partial motor tone and posture improvement.

Oral breathing is associated with tonic changes in the orofacial musculature but also in head posture, where the head is extended and forward of the spine.[227] In this position, the

cervical musculature shows higher muscle activity, measured by electromyography, in at-rest oral breathers compared with nasal breathers, showing that there is cervical musculature recruitment.[228]

This modification of the muscle recruitment pattern when the mode of respiration changes from nasal to oral breathing was demonstrated in experimental animals.[217] Therapy exercises aimed at these cervical muscles can create reduced EMG activity as seen in nasal breathers.[228] So the postural mode can influence the respiratory mode as in habitual oral breathers,[229] and this affects muscle balance. This suggests that muscle exercise therapy treatment be incorporated to reinforce a habitual mouth-closed mode of breathing. MFT could target the orofacial and cervical musculature either to reverse the adapted muscle recruitment patterns or to strengthen tonicity to change the upper airway caliber. Use of MFT as adjunctive therapy to firmly establish the nasal mode of respiration is an area to be explored.

Passive Force Application

Shape Changes in the Upper Jaw. As described earlier in this chapter, the muscle pattern exerts a passive tension on the developing airway through the force mediated on the maxilla. Numerous studies have shown a relationship between nasal airway obstruction and aberrant facial growth. In monkeys, the nasally obstructed animals had longer faces and unusual dental malocclusions. In children, these same morphologic effects of impaired nasal respiration are evident, including narrowing of the maxilla, reduced development of the mandible, increased vertical growth of the lower face, dental crowding and malocclusion,[30,230] and altered head posture. A new posture compensates for the decrease in nasal airflow to allow respiration. However, these are not consistent phenomena in children with oral breathing tendencies.[231]

These changes are typified in Figure 143-5 and occur passively because there is no directly applied force or load. In response to chronic oral breathing, three distinct

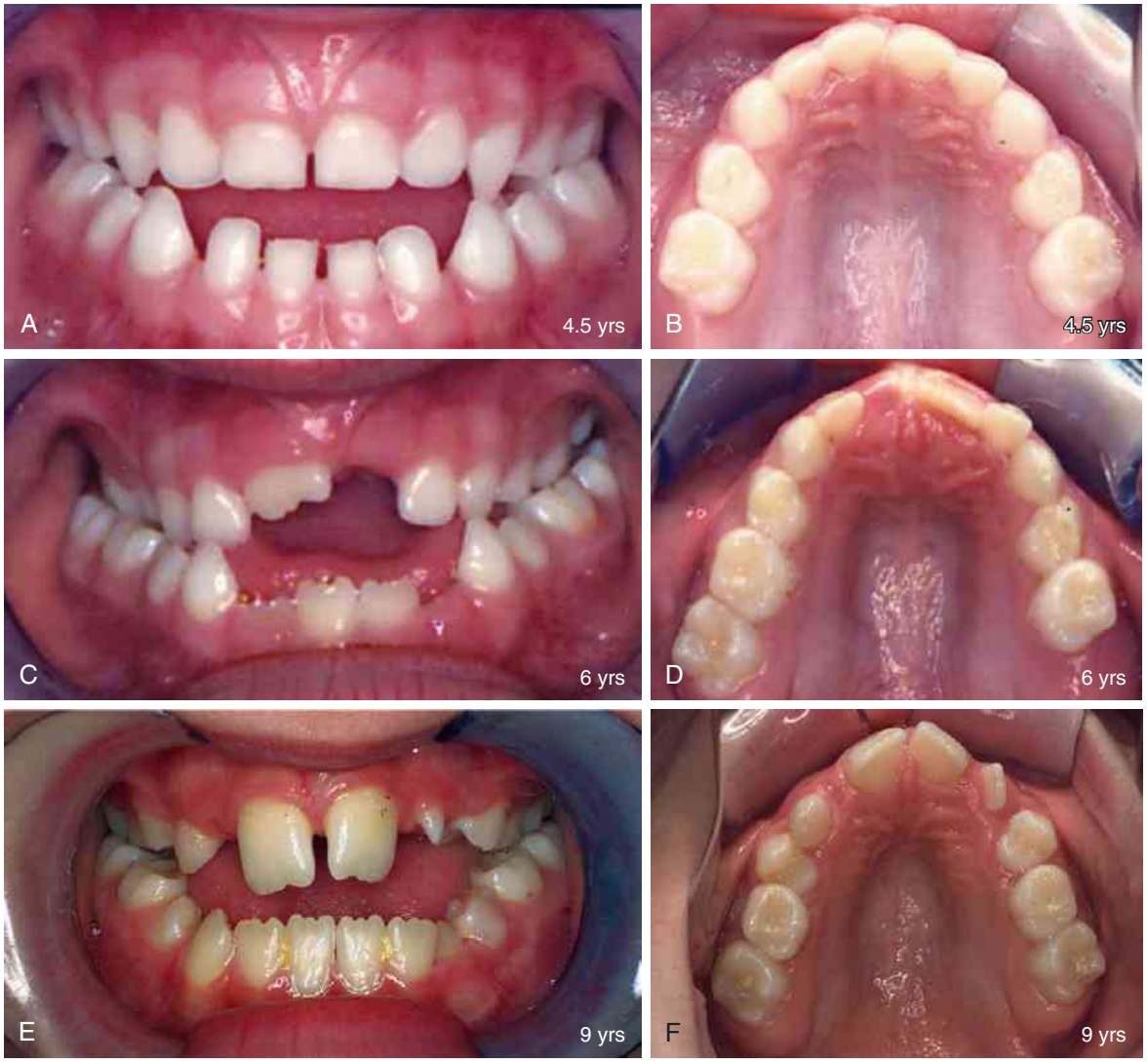


Figure 143-5 A–F, Natural growth of the maxillomandibular complex in response to adenotonsillar hypertrophy and oral breathing. **A** and **B,** Taken at age 4.5 years. **C** and **D,** Taken at age 6 years. **E** and **F,** Taken at age 9 years. Note the gradual constriction of the upper jaw from the age of 4.5 to 9 years, while the lower jaw remains an intact shape.

developmental changes may be manifested. The upper jaw narrows, suggesting that the volume of the nasal cavity becomes smaller as the upper jaw continues to narrow. As nasal volume decreases, the nasal resistance increases, which can exacerbate upper airway collapsibility, causing a cycle of progressively worsening SDB. This induced distortion in morphogenesis could partly explain the pathogenesis of SDB. The second developmental change is seen in the dentition, as the teeth add a second layer of insult to this distortion by compensatory inward tipping in both the upper and lower jaws. This restricts the intraoral volume even more, which potentiates a postural and neuromuscular adaptation of the genioglossus muscle. It can be argued that aggressive efforts are needed to combat the inherent tendency for upper jaw narrowing in oral breathers, even after the obstacles to nasal airflow are removed, because of the muscle alterations. The subsequent effect on the mandible is the third level of passive distortion that intensifies the pharyngeal collapsibility as the hypopharynx narrows with the backward rotation of the mandible. This narrowing is specifically evident at the retroglossal area.

Shape Changes in the Lower Jaw. In a growing child, abnormal nasal resistance may affect the maxilla and the mandible.[232] During puberty, these deleterious changes can be magnified because of the growth velocity seen during teenage years. Deleterious effects of mouth breathing are the sensory stimulus that propagates abnormal tongue position and altered orofacial musculature tone. The muscle recruitment pattern is modified when nasal respiration changes to oral breathing.[217] Some of these growth responses can be exaggerated because of the concurrent effect of molar tooth eruption occurring at the same time.

In oral breathers, it has been reported that mandibular growth is redirected posteriorly. Oral breathing that develops from nasal obstruction has an effect on the largest pharyngeal dilator muscle that keeps the airway open, the genioglossus. In an open mouth posture, the genioglossus adopts a low-lying position in the floor of the mouth. The relationship of breathing to airway problems shows a lowered position of the hyoid bone in OSA syndrome patients. In this dorsocaudal position, EMG activation of the tongue on experimental animals shows a weaker protrusive force,[210,217,233,234] which suggests a possible mechanism for increased collapsibility as decreased genioglossus tension led to an increased muscle length in experimental animals.[233] The increased size of the genioglossus would narrow the pharyngeal airway, and this increases collapsibility. Genioglossal activity is increased with nasal compared with oral breathing, with supine versus sitting body position, and during inspiratory resistive loading,[235] probably because of an altered respiratory drive and reflex activation. This effect coupled with the potential edema from the negative airway pressure generated during sleep[236] exacerbates the disorder, exemplifying the necessity to treat the sequelae of the root cause. A cyclic pattern of distortion develops in which the lower jaw rotates backward because of mode of breathing. This changed positvion narrows the retroglossal airway, which could potentiate increased repetitive activation stimuli to keep the airway open, leading to an increase in pharyngeal muscle mass. Further muscle injury can result from snoring, vibration, or increased airway resistance.

The term adenoid facies has historically been used to describe the long, narrow, flat face of the individual with oral hypotonia and protruding teeth and lips that are widely separated at rest, often accompanied by an open bite malocclusion.[30] The original work on experimental animals by Harvold et al[210] demonstrated that there is a variable response to this insult of forced oral respiration. Oral breathing was not linked to a specific skeletal structure. Similarly, there is not a defined facial type or malocclusion that accompanies adenotonsillar hypertrophy in an oral breather. Several studies corroborate this conclusion because there is not a consistent relationship between nasal resistance and dentofacial morphology, as described earlier in this chapter. These differences are illustrated in Figure 143-6, which shows varied facial phenotypes with OSA. Medical intervention has not been shown to influence the pattern of facial growth in children with allergies. Surgical therapy to relieve nasal airway obstruction in children (whether adenoidectomy or turbinectomy) has not been shown to predictably affect ultimate facial form.[237]

The three cases illustrated in Figures 143-7 to 143-12 demonstrate the variable response of mandibular growth to impaired nasal respiration. All three patients had therapy for their malocclusion, but the problem of allergies and oral respiration persisted. The records shown were taken during a period of adolescent growth, illustrating a lack of forward upper jaw development. Whereas there is an inherited growth pattern that accounts for the familial tendency or inheritance in OSA, there can be different epigenetic expressions of craniofacial malformation overlaid on the inherent preexisting growth presentation.

Comparison of the general superimposition tracings, from three subjects, shows a general descent and slight rotation of the maxilla but variable responses of the mandible (Figures 143-8, 143-10, and 143-12). All three subjects had nasal obstruction that persisted after expansion treatment, and the disproportionate jaw growth was expressed during adolescence. The mandible showed posterior rotation (Figure 143-8) and straight vertical descent with no forward growth (Figure 143-12) or forward growth rotation, versus forward anterior mandibular growth rotation (Figure 143-10). The dentoalveolar development from tooth eruption can affect the forward (Figure 143-10) or backward rotation (Figures 143-8 and 143-12) of the mandible, and this important effect cannot be discounted. These are three different expressions of mandibular growth, from the same stimulus of increased airway resistance from nasal obstruction causing differences in muscle recruitment patterns that changed muscle tone and function. This created variable adaptations of soft tissue, which influence the skeletal morphology. Thus, the oral breather may present with a normal appearance to severe skeletal and dental irregularities. "Nasal obstruction presents the trigger factor, but it is the deviant muscle recruitment which causes maldevelopment."[210]

These cases illustrate the answer to the question about disease propagation if craniofacial malformation is not only a cause of SDB but a consequence of SDB. It is a variable response, depending on how the effect of impaired nasal respiration affects the muscle recruitment and activation. For some patients, this paradigm of propagation would hold true. The challenge is to identify those patients who are more at

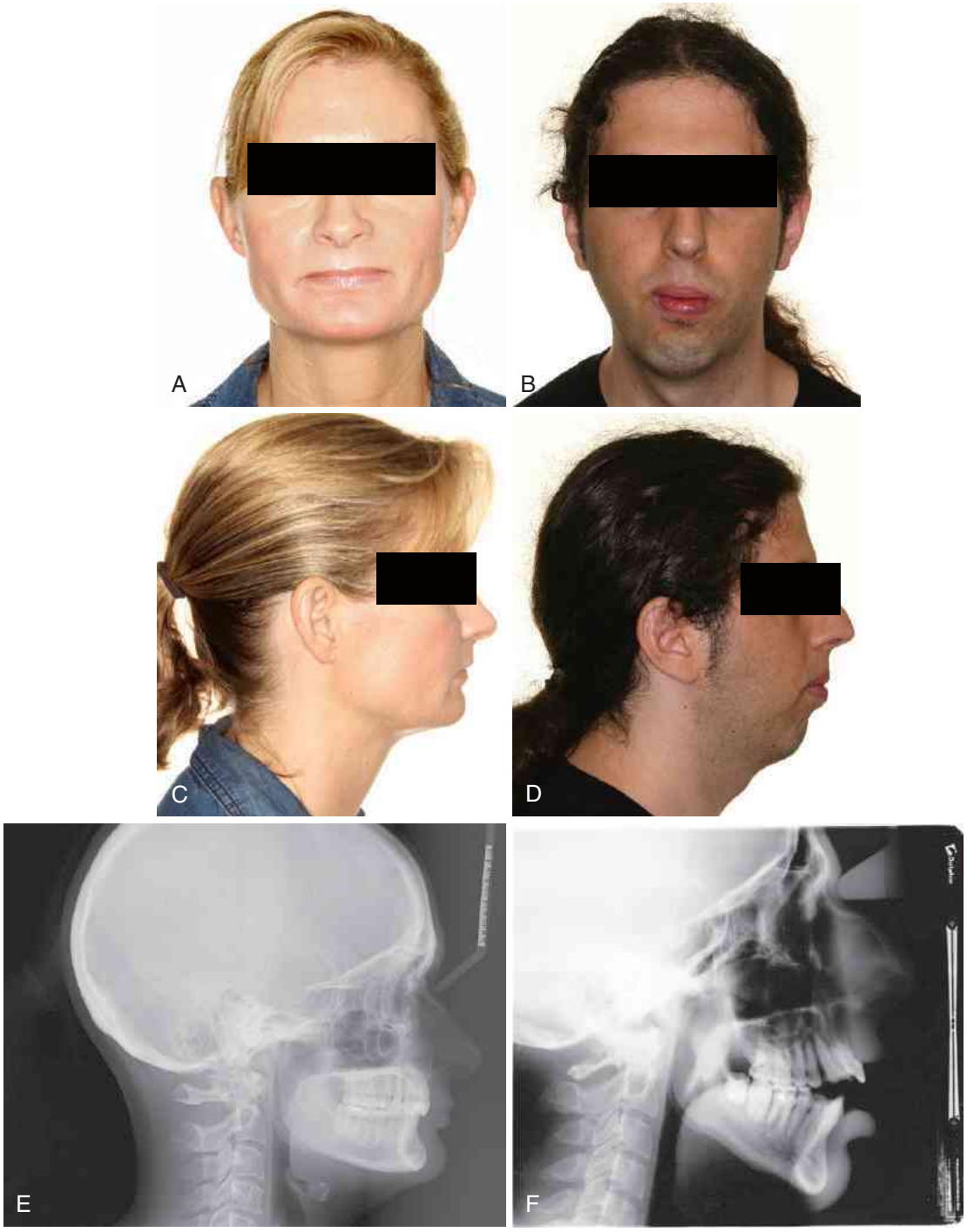


Figure 143-6 Contrasting facial morphologies associated with OSA. **A, C,** and **E** Decreased lower facial height of an individual with a closed gonial angle and bimaxillary retrusion. **B, D,** and **F** Increased lower facial height of an individual with bimaxillary dental protrusion and an open gonial angle.

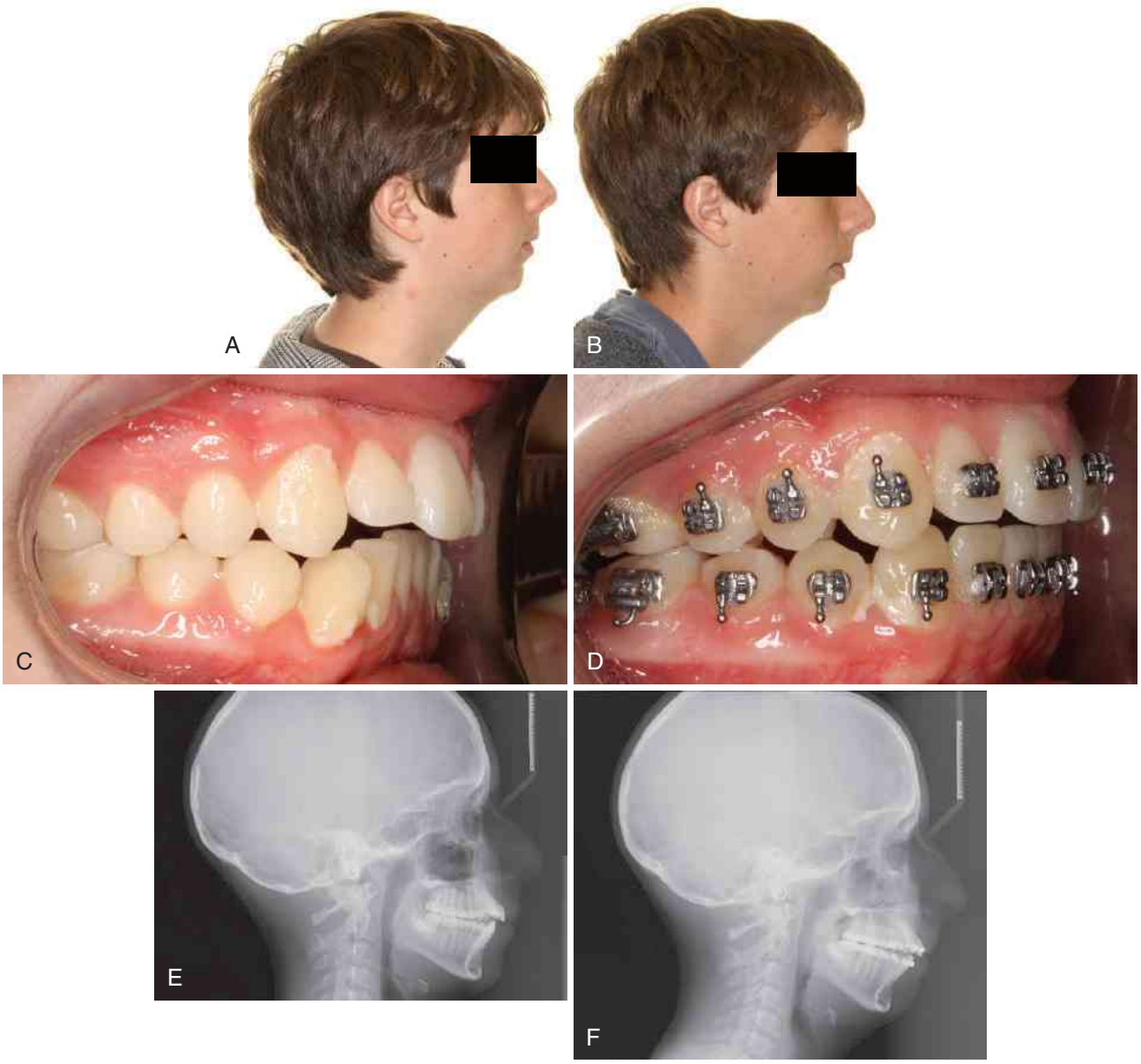


Figure 143-7 Growth changes during adolescence. **A, C,** and **E** Presentation at the age of 14 years before the initiation of orthodontic therapy for malocclusion correction. **B, D,** and **F** Growth and dental changes 12 months later.

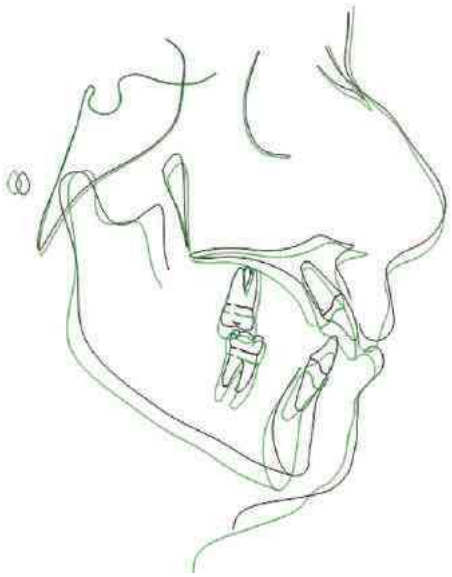


Figure 143-8 General growth superimposition during a 7-month period in adolescence of the patient in Figure 143-7. Method of Bjork, superimposed on cranial base. Vertical lowering of the maxilla is noted, with backward growth rotation of the mandible.

risk for aberrant muscle alterations, and for these patients muscle rehabilitation may be beneficial.

Active Force Application

Induced Bone Remodeling. If bone remodeling is the result of passive force application, active force application as an epigenetic effect to direct bone development could be a strategy to increase airway size or to redirect unfavorable growth patterns, beyond the inherent genetic expression. Not all patients develop a dysmorphic growth pattern as a result of impaired nasal respiration, so interventional therapy to modify existing growth patterns may not be justified when an aberrant growth pattern cannot be predicted. However, if these therapies help improve the critical closing pressure, the pharyngeal collapsibility would ultimately be improved or even normalized.

The consideration for this direction of care lies in the shortcomings of current therapies. The "gold standard" for treatment of OSA in adults is nasal CPAP. Whereas adenotonsillectomy is the first-line therapy for children, nasal CPAP is considered after other therapies have not completely improved respiration. Nasal CPAP is burdened with problems of compliance due to mask fit, comfort, or initiation of CPAP use. As an extraoral appliance, it will exert a molding force on the facial skeleton and the dentoalveolar bone. This force can cause remodeling and redirection of maxillomandibular jaw growth in the pediatric patient.[238-241] This retrusive remodeling force from the positive pressure from the mask or machine can create a skeletal jaw imbalance and negate all earlier efforts to enlarge the airway size, as seen in Figure 143-13, *A* and *B*. As increased nasal resistance is both a cause and consequence of SDB, an often prescribed option for treatment not only treats the problem but can perpetuate the problem too. However, induced bone remodeling through active force application can rectify these iatrogenic craniofacial alterations (Figure 143-13, *C*). There remains the question as to how long this imbalance is to be sustained.

Placing applied pressure (or force or loading) on the craniofacial bone causes shape and size changes. Future directions in care will test these boundaries. RME appliances use the dentition for force delivery, but there can be unwanted tooth movement side effects that accompany the skeletal change. Removable appliances that attach to the teeth to change the muscle tension on the underlying bone have been used with a small degree of success. One future avenue for force delivery uses temporary attachment devices, called TADs, anchored directly into the bone, that bypass the dentition.[242,243] The advantage of this type of direct force application to the bone is the direct and possible increased effect of loading on bone modeling while minimizing the side effects to the adjacent teeth and periodontium.

Through the combination of therapies, it may be possible to modify the inherent growth pattern. Efforts to target therapy at the site of greatest airway narrowing or obstruction may not be adequate because airway narrowing and collapse are thought to occur at multiple sites.[244] There are currently no studies that detail the effectiveness of a comprehensive and combined approach of treatments and growth modification to the nasopharynx, oropharynx, and hypopharynx while also implementing muscle rehabilitation to address the concomitant neuromuscular adaptations of an impaired airway. Although there are studies that document the efficacy of adenotonsillectomy and jaw expansion together as a treatment modality[167] and studies that describe adenotonsillectomy and MFT,[245] there remains to be examined the full complement of creating the largest patency of the airway using the modalities described in Box 143-3. These types of studies would be difficult to design, and longitudinal data across several different populations would be needed. The 11 known collections of longitudinal craniofacial growth records of untreated children in the United States and Canada could serve as the comparison of treated to untreated populations. This monumental effort would necessitate the collaboration of the multiple specialists involved in the treatment of upper airway disorders.

Prophylactically treating the entire pharynx, instead of targeting the site of greatest obstruction or narrowest opening, may seem aggressive and extreme. However, children with SDB have blunted responses to hypercapnia,[246] demonstrating early neural deficits that are thought to be reversed with treatment, but it is not known if there is complete recovery. Reports of a high recurrence of SDB in later teenage years[117] suggest that some neural dysregulation persisted and perhaps worsened after previous treatment. Treatment is usually initiated from daytime or nighttime symptoms. The onset of symptoms may underlie a larger problem, as it is not known how long the primary cause needs to be present before symptoms develop. The most viable treatment may be consistent early screening to identify patterns before the onset of symptoms. By the time treatment is rendered, even if it is done at any early age, there may already be alterations in muscle recruitment and tone that initiate the development of a craniofacial malformation. A recent population-based longitudinal study observing children from 6 months to 7 years of age found that early symptoms of snoring, mouth breathing, and witnessed apneas had statistically significant effects on behavior in later

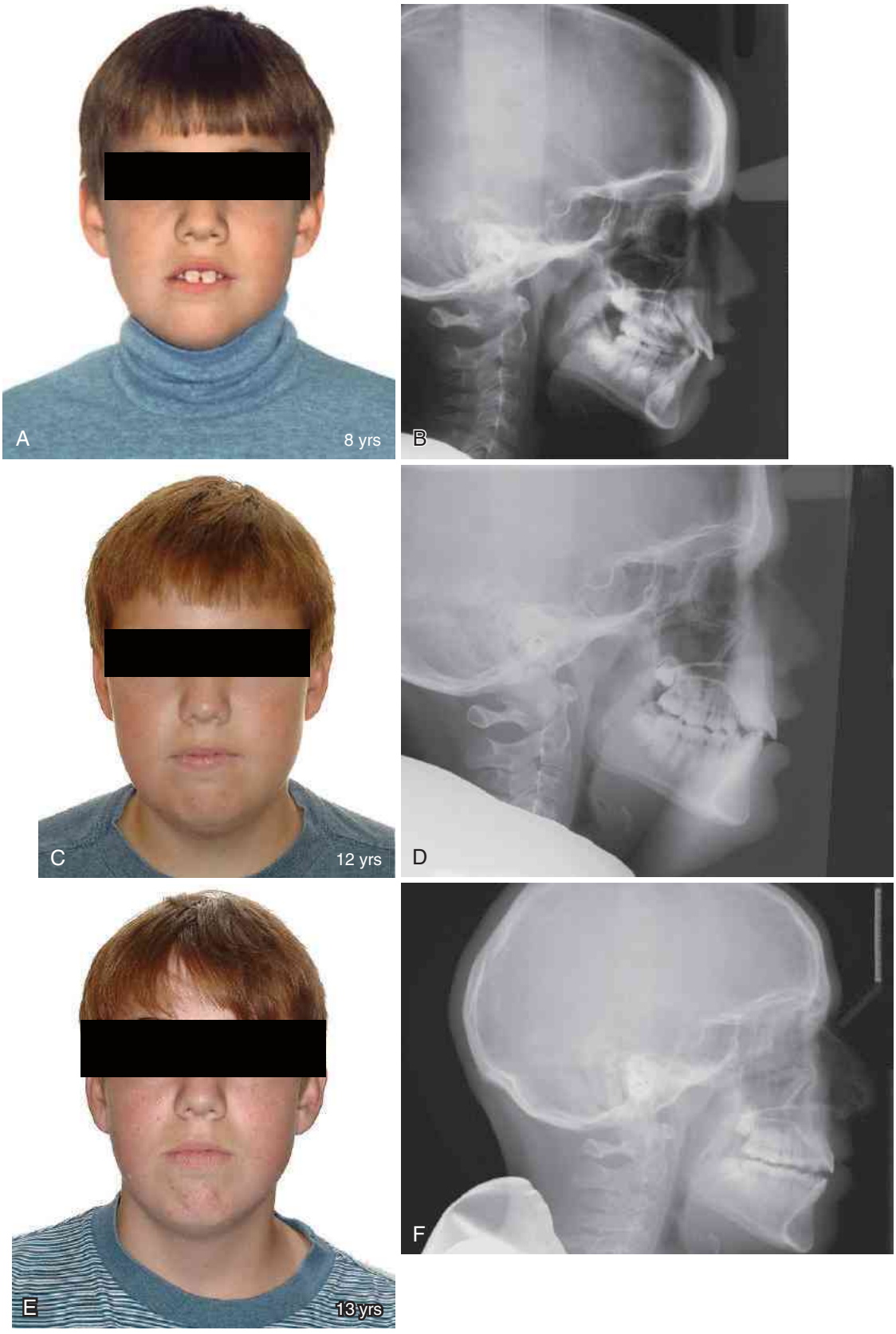


Figure 143-9 This patient had orthodontic intervention of maxillary expansion for the malocclusion. **A** and **B,** Facial proportions and jaw relationship at the age of 8 years, initial presentation. **C** and **D,** Vertical lowering of the mandible after 4 years. **E** and **F,** Narrowed posterior airway space, continued vertical lowering of mandible, and forward jaw rotation after 5 years. No treatment interventions were rendered between 12 and 13 years of age. Note the dramatic change in growth velocity of the lower jaw from the age of 12 to 13 years.

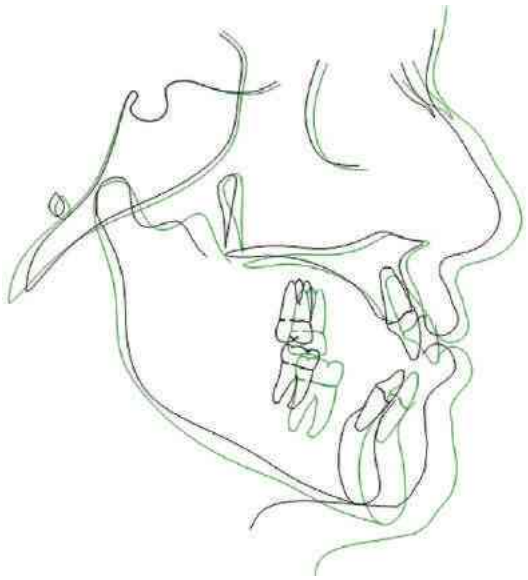


Figure 143-10 General superimposition from 12 to 13 years of age of the patient in Figure 143-9. Method of Bjork, superimposed on cranial base. Forward growth rotation of the maxilla and mandible are evident.

childhood, suggesting that early symptoms warrant examination as early as the first year of life.[247] Herein lies the challenge of defining reliable screening patterns that are predictive of the disease.

A stepwise evidence-based approach for the diagnosis and multitherapeutic management of childhood SDB has been recently presented.[248] This approach, starting with weight control followed in succession with nasal corticosteroids, adenotonsillectomy, dentofacial orthopedics such as mandibular advancement or maxillary expansion, CPAP, and maxillofacial surgery, is illustrated in Figure 143-14. This case demonstrates that despite early best efforts at recognition, diagnosis, and intervention of multiple therapies, including adenotonsillectomy, allergy management, multiple rounds of bimaxillary expansion, and nasal CPAP, the upper airway problems may still persist. For this particular case, problems of CPAP adherence rendered it ineffective as long-term treatment. Ultimately, on growth cessation, maxillomandibular advancement was the final treatment rendered, which normalized respiratory parameters and symptoms. Anatomic insufficiencies were improved to maximize the enlargement of the pharyngeal space.

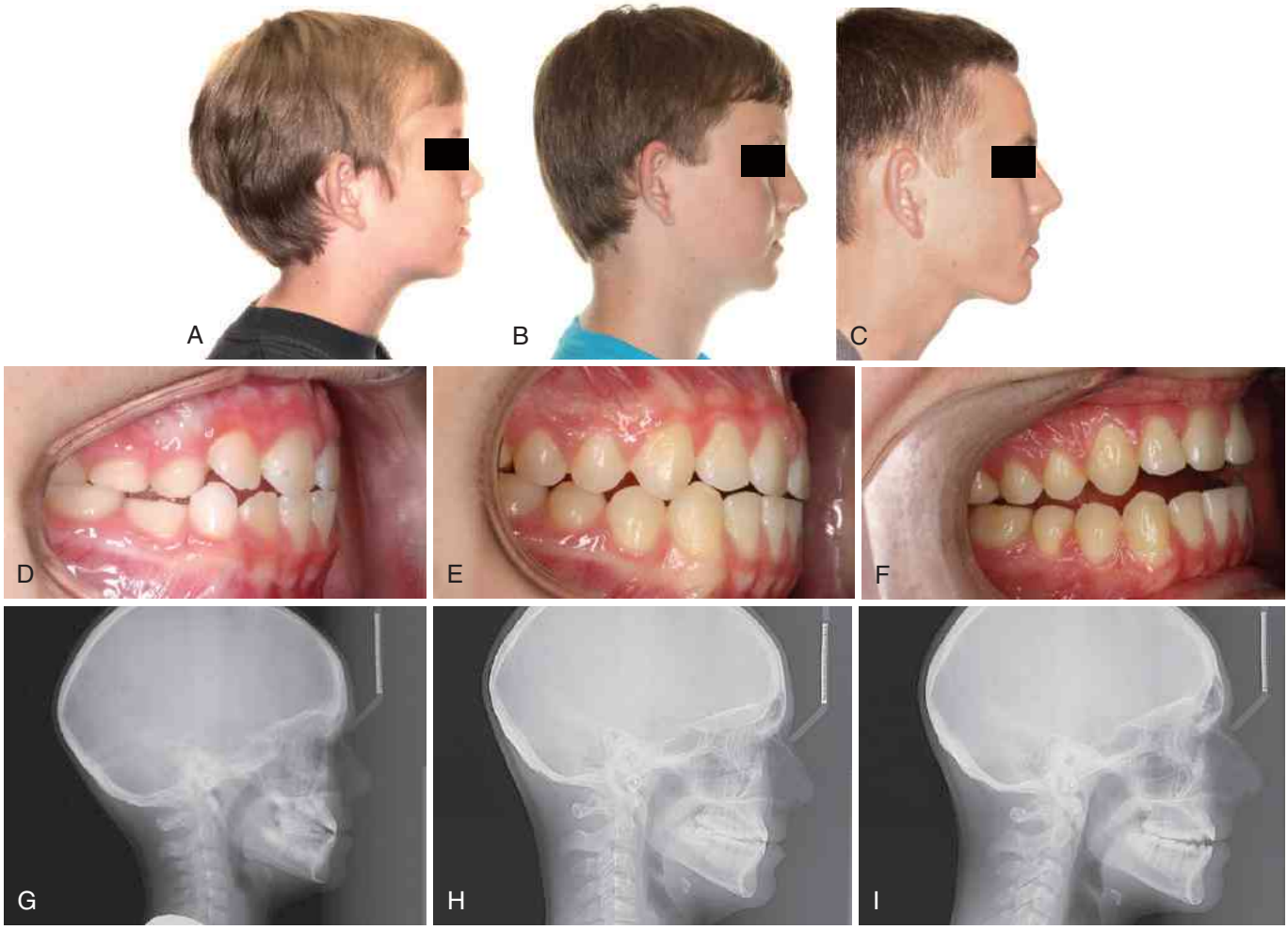


Figure 143-11 **A–C,** Growth changes during adolescence. Initial presentation at the age of 8 years, when upper jaw expansion was done. **D–F,** Growth presentation 5 years later. **G** and **H,** Growth presentation with no treatment interventions between the ages of 13 and 15 years.

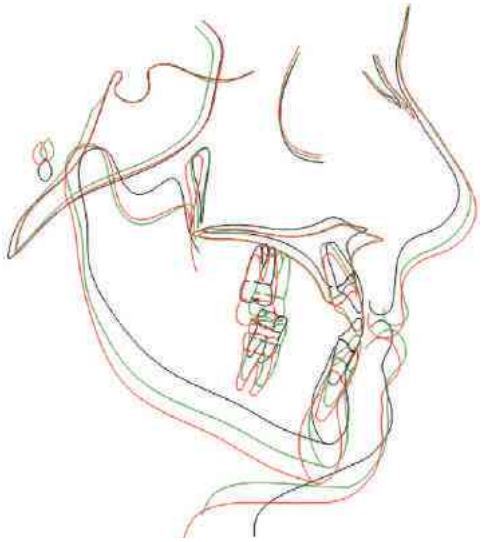


Figure 143-12 General superimposition, method of Bjork, superimposed on cranial base of the patient in Figure 143-11. The lowering of the mandible for oral respiration was followed by a downward displacement of the maxilla with an increased extrusion of the teeth.

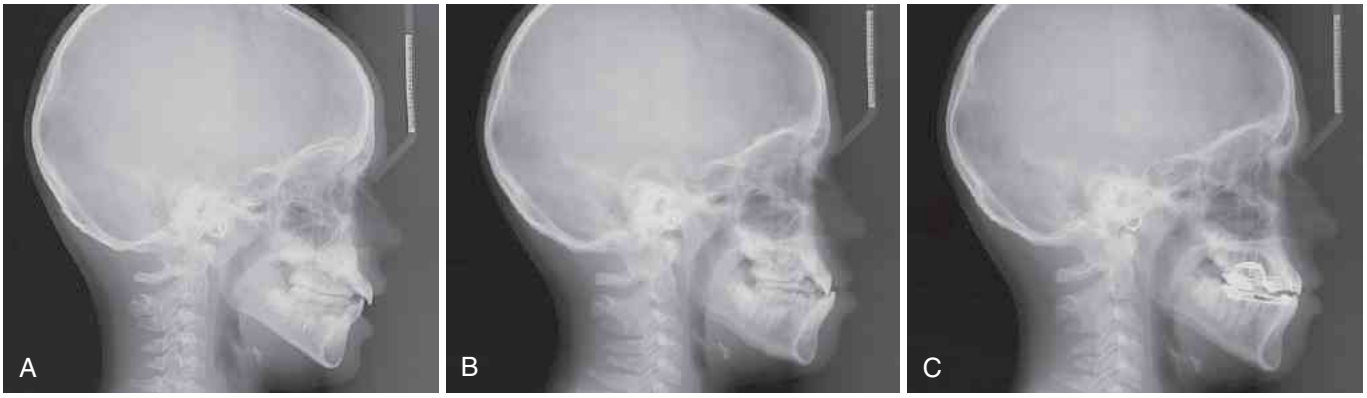


Figure 143-13 Effect of CPAP on facial growth. **A,** Lateral cephalogram of an adolescent boy before initiation of nasal CPAP. **B,** Lateral cephalogram showing retrusion of the upper jaw in response to nasal CPAP (mask type: nasal mask) for 12 months. **C,** Lateral cephalogram showing protraction of upper jaw from orthodontic orthopedic forward traction of 8 months to the maxilla, while still using CPAP (mask type changed to nasal pillows).

CLINICAL PEARLS

- The altered craniofacial morphology associated with SDB has been well classified and trends to growth of longer and narrower facial structures. The exact mechanism by which this maladaptive plasticity of growth occurs is not well understood but is most likely related to impaired function that is both caused and perpetuated by genetic and epigenetic factors.
- Nasal CPAP as an extraoral appliance will exert a molding force on the facial skeleton and the dentoalveolar bone, causing remodeling and redirection of maxillomandibular jaw growth in the pediatric patient. This retrusive remodeling force from the positive pressure from the mask or machine can create a skeletal jaw imbalance and negate all earlier efforts to enlarge the airway. Pediatric mask selection should minimize face contact to reduce the retrusive molding pressure to the maxilla.
- When children with SDB present with the specific abnormality of narrowed palate or posterior dental crossbite, treatment with rapid palatal expansion is an effective, noninvasive therapy to improve disease symptoms. Another common means to orthodontically increase airway volume, using a mandibular advancement device in children, has limited but promising evidence in retrognathic SDB patients.
- Improvement of both subjective and objective measures of SDB by MFT has been demonstrated in children and adolescents through the mechanisms of muscle strengthening and reinforcing nasal respiration. Despite impressive results, the overall number of patients is small, and future research should develop the full extent to which this form of treatment can be effectively applied.

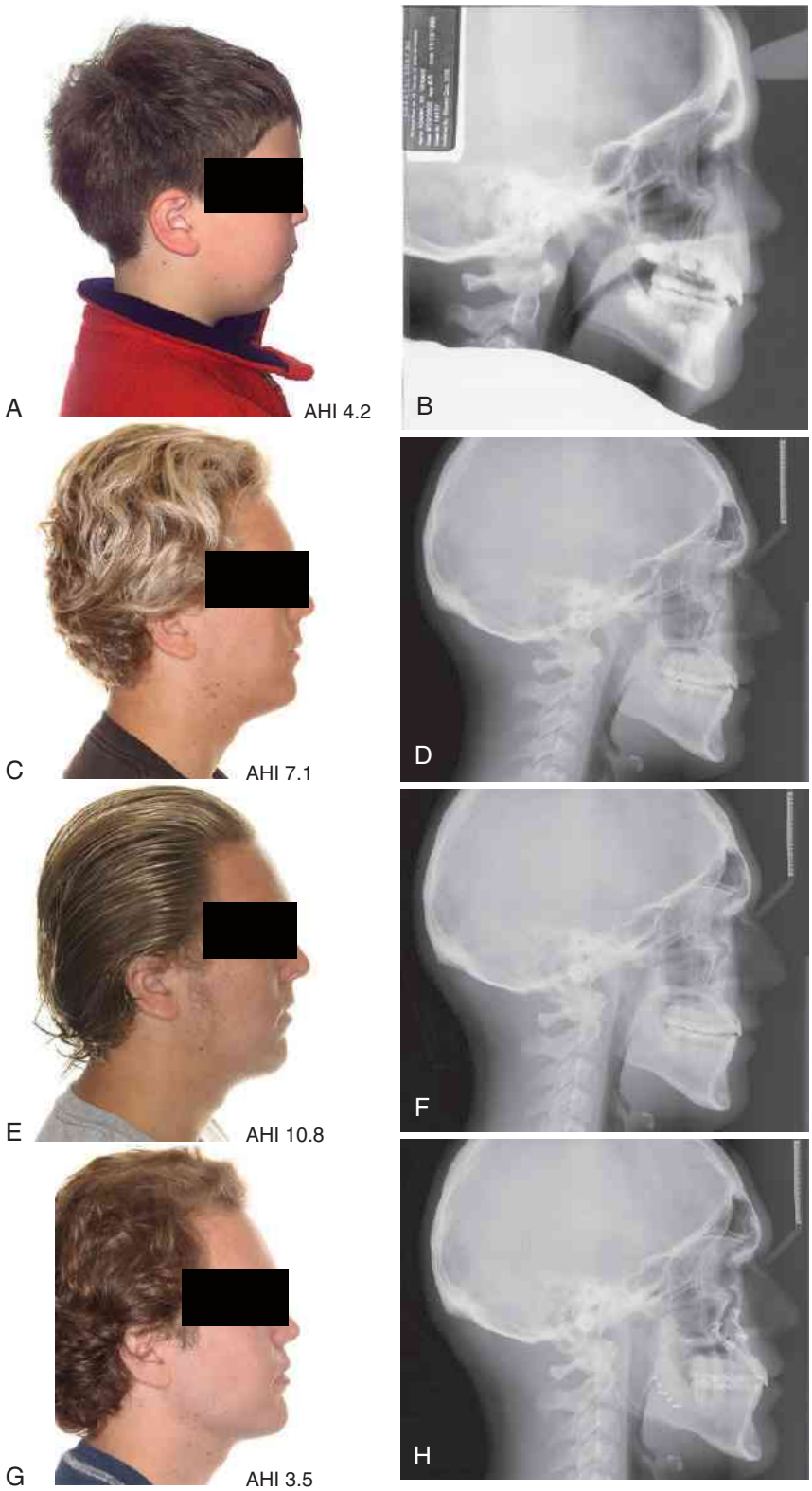


Figure 143-14 Craniofacial development from the ages of 8 to 19 years. **A** and **B,** Records at age 8 years show airway and facial proportions after adenotonsillectomy, indicating residual SDB/OSA (AHI 4.2, residual symptoms). Bimaxillary orthodontic expansion initiated. Initial diagnosis of OSA at age 5 years (AHI 2.9 with daytime symptoms). **C** and **D,** Age 15 years, after bimaxillary expansion was completed at age 8 years and again at age 10 years. AHI increased to 7.1. **E** and **F,** Age 17 years, in preparation for maxillomandibular advancement. AHI increased to 10.8 with disproportionate increase in symptoms. **G** and **H,** Age 19 years after maxillomandibular advancement, with normalization of AHI and resolution of symptoms. Note the pharyngeal airway enlargement and accompanying facial soft tissue profile changes.

SUMMARY

Many of the factors leading to pediatric SDB create secondary morphologic changes that exacerbate and perpetuate the syndrome. The approach to establish normal daytime and nighttime respiration is a multidisciplinary endeavor, involving the treatment of all three components of the pharynx and the attendant surrounding musculature. As multiple treatments may act synergistically, a greater degree of collaboration between specialists in sleep medicine, otolaryngology, allergy, and orthodontics is warranted to establish the contribution made by each discipline to the outcome of pediatric OSA. To achieve successful treatment of upper airway disorders, future efforts aimed at modifying the anatomy can hold the promise of prevention of problems of constricted size and impaired muscle function because oropharyngeal malformation may be not only a cause but also a consequence of SDB.

Selected Readings

Camacho M, Certal V, Abdullatif J, et al. Myofunctional therapy to treat obstructive sleep apnea: a systematic review and meta-analysis. *Sleep* 2015;**38**:669–75.

Guilleminault C, Monteyrol PJ, Huynh NT, et al. Adeno-tonsillectomy and rapid maxillary distraction in pre-pubertal children, a pilot study. *Sleep Breath* 2011;**15**:173–7.

Katyal V, Pamula Y, Martin AJ, et al. Craniofacial and upper airway morphology in pediatric sleep-disordered breathing: systematic review and meta-analysis. *Am J Orthod Dentofacial Orthop* 2013;**143**:20–30.e3.

Marcus CL, Smith RJ, Mankarious LA, et al. Developmental aspects of the upper airway: report from an NHLBI Workshop, March 5–6, 2009. *Proc Am Thorac Soc* 2009;**6**:513–20.

Pirelli P, Saponara M, Guilleminault C. Rapid maxillary expansion (RME) for pediatric obstructive sleep apnea: a 12-year follow-up. *Sleep Med* 2015;**16**(8):933–5.

Tapia IE, Marcus CL. Newer treatment modalities for pediatric obstructive sleep apnea. *Paediatr Respir Rev* 2013;**14**:199–203.

A complete reference list can be found online at ExpertConsult.com.

Chapter 144

Sleep Bruxism: Definition, Prevalence, Classification, Etiology, and Consequences

Peter Svensson; Taro Arima; Gilles Lavigne; Eduardo Castrillon

Chapter Highlights

- Sleep bruxism (SB) is a sleep-related movement disorder characterized by teeth grinding or clenching frequently but not exclusively associated with sleep arousal. SB has traditionally been considered an oral parafunction with involuntary rhythmic or spasmodic nonfunctional gnashing, grinding, or chewing movements of the mandible, triggered by occlusal discrepancies, which then could lead to occlusal trauma and dysfunction of the stomatognathic system. Current lines of research, however, have downplayed the importance of occlusal and anatomical factors for SB and, rather, emphasized central nervous system factors and regulation of the autonomic nervous system. Knowledge on definitions, pathophysiology, and mechanisms of SB has distinct implications for management of SB in the clinic.

DEFINITIONS

Gnashing and clenching of the teeth have been reported since early times in human history. Statements on this particular oral activity can be found even in the old testament of the Holy Bible.[1] *Bruxism* has been the preferred term to characterize these oral behaviors and is originated from the Greek word *brygmos* (βρυγμός) that means gnashing of the teeth. Bruxism has been the focus for many health professions and, more specifically, in dentistry for many years and was described for the first time in the scientific literature by Marie and Pietwiekwicz.[2]

A variety of definitions of bruxism has been proposed and used, but most of them have been criticized for not providing specific or operationalized criteria.[3] A recent publication from an international expert group addressed this issue, discussing all of the different definitions available. They proposed the following new consensus-based definition:

> Bruxism is a repetitive jaw-muscle activity characterized by clenching or grinding of the teeth and/or by bracing or thrusting of the mandible. Bruxism has two distinct circadian manifestations: it can occur during sleep (indicated as SB) or during wakefulness (indicated as awake bruxism: AB)[3]

This chapter will focus on sleep bruxism (SB).

PREVALENCE

The prevalence of SB varies, depending on the specific criteria to define bruxism and the methods used to establish the diagnosis. Most of the literature available on the epidemiology of SB are based on self-reports, that is, simple questions such as: Are you aware of clenching or grinding your teeth during sleep?—or questionnaires (e.g., oral behavior check list).[4] Using this methodology, SB is reported between 8% and 31% of the general population.[5] SB has been reported in the literature as occurring as "frequent as three times a week" with a prevalence of 9.3%, as "frequent" bruxism in 14%, and as "often" bruxism in 15.3%.[5] For comparison, awake bruxism (AB) is reported as "often" in 22.1% and "any AB during the last 6 months" in 31% of the population. Overall, these prevalence figures suggest that SB is, indeed, a very common phenomenon in the population.

There are nevertheless obvious limitations with prevalence data based on self-reports of SB, for example, if the patients sleep alone, they may not be aware of their SB behavior or, if their sleep partner sleeps very deeply, they may not be told. Furthermore, information from health care providers can cause bias. A classical example is that the majority of patients who may report SB have received the information from their dentist.[6]

The gold standard for a more definite SB diagnosis is based on the use of polysomnographic (PSG) analyses, including audio and video recordings as described chapter 145.[7] (Figure 144-1). Due to its practical complications, heavy load in data analysis, and relatively high costs, there is not much epidemiologic information on SB using PSG criteria. One recent large-scale PSG study reported a prevalence of 7.4% SB in adults using exclusively PSG methodology.[8] Unfortunately, this report lacked video and audio recordings that are needed to fulfill the gold standard requirements and remove false-positive events, for example, derived from body movements or scratching the skin/electrodes.[7,9] It has also been suggested that an overestimation of the SB prevalence is possible when only self-reports are used.[8,10] Interestingly, SB prevalence appears not to differ between genders but to decrease with age.[5,11] The reason for the age-dependency, so far, is not well understood but must await confirmation from PSG-based SB

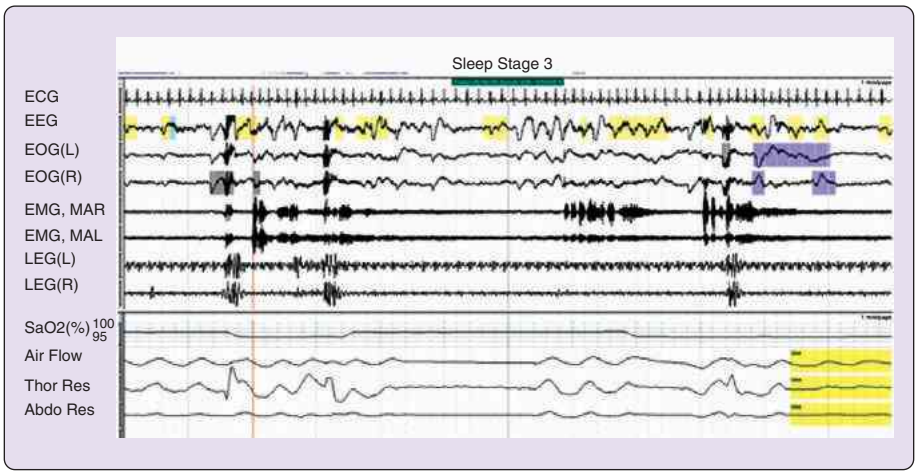


Figure 144-1 A typical polysomnographic (PSG) recording for the assessment of rhythmic masticatory muscle activity (RMMA). This recording includes electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG) of the right masseter (MAR) muscle, EMG of the left masseter (MAL) muscle, movement recording of the right leg (LEG[R]), movement recording of the left leg (LEG[L]), oxygen saturation (SaO_2[%]), airflow, thorax expansion resistance (ThorRes), and abdomen expansion resistance (AbdoRes). Gold standard criteria for sleep bruxism demand that PSG recordings for RMMA also include synchronized audio and video recordings (not included in this figure). (Courtesy Faramarz Jadidi, PhD.)

studies. Despite the small amount of scientific information about genetics in SB, a recent review suggested that SB may, at least in part, be genetically determined and environmentally expressed (see the later section on etiology and pathophysiology for more information).[12-15] However, it will be a complex task to identify specific candidate genes for SB, and more research in this field will be needed.

CLASSIFICATION

One of the most logical ways to classify bruxism is to use the circadian rhythm to divide this oral activity into AB and SB, recognizing it may overlap in some subjects[16] because the physiology/pathophysiology may differ.[5,17] There are, however, many additional possibilities to subclassify SB given the definition of "repetitive jaw-muscle activity," for example, depending on the specifics of the electromyographic (EMG) activity (type, duration, frequency, amplitude, total amount, etc.). One could speculate that concentric, short-lasting, infrequent, and low-intensity jaw muscle contractions would have different clinical consequences (e.g., tooth wear, orofacial pain, and headache) than eccentric, longer-lasting, frequent, and high-intensity jaw muscle contractions; further studies are obviously needed to test such hypotheses. Lavigne and colleagues[7] used operationalized research criteria, derived from the literature, to define phasic, tonic, and mixed EMG bursts and episodes and established cutoff values for SB and non-SB. These three types of EMG burst are scored as an episode, called *rhythmic masticatory muscle activity* (RMMA). These research criteria have subsequently been refined in a follow-up study.[18] Before these criteria are extrapolated for clinic use, further validation is needed with various recording systems, full PSG or limited number, or EMG channels.[19,20] Furthermore, the specifics of such EMG criteria need to be elaborated, for example, to take into account the amplitude, total amount of jaw muscle activity (area under the EMG curve), and so forth.

The International Classification of Sleep Disorders—Second Edition (ICSD-2) has placed SB under the group of sleep-related movement disorders and emphasized the link to arousal responses; it is again included into the same sleep disorder category for the third edition (2014) (www.aasmnet.org/store/product.aspx?pid=849). Sleep-related movement disorders are characterized by simple and stereotypic movements that disturb sleep, while the patients may or may not be aware of such movements. A recent study showed that SB events and periodic limb movement are time-linked, suggesting that, at least some of the underlying mechanisms may be in common. (See Chapter 95 for more information on Restless Legs Syndrome and Periodic Limb Movements during Sleep.)[21] The more mechanism-based ICSD classification of SB is entirely compatible with the consensus-based descriptive definition of SB by Lobbezoo and colleagues.[3]

It should be recognized that the clinical diagnosis of SB is associated with considerable diagnostic uncertainty like most other medical diagnoses and, for example, neuropathic pain.[22] Another approach that has been suggested to classify bruxism is the use of grading system criteria proposed by the group of experts in the consensus report.[3] This diagnostic grading system takes into account the methods and their validity to determine the presence of SB. Three levels were proposed: possible, probable, and definitive. Possible AB or SB is based on self-reports using the history or questionnaires. Probable AB or SB will be the same as for possible AB or SB plus the outcome from the clinical examination, for example, tooth wear or hypertrophic jaw muscles. Definitive will be the same as probable plus evidence from a PSG study following the gold standards for SB,[7] whereas for AB it was proposed to use the criteria for probable plus ecological momentary assessment of EMG.[3] It is suggested to apply this grading system to better characterize and classify SB, which should allow better insights into both risk factors and pathophysiology of SB, but also to better understand the potential pathophysiological consequences of SB.

ETIOLOGY AND PATHOPHYSIOLOGY

Basically, the etiology of SB is still controversial, and it may be more appropriate to talk about risk factors instead of etiological factors. Historically, SB has been associated with three major domains: anatomical, psychological, and central nervous system (CNS) factors. A short review of these topics is provided in the following.

In the dental community, the anatomical factor has for a long time been considered a main factor for SB because the prevalence of SB was much higher in population groups with malocclusion than in comparable groups with normal occlusion.[23] This tendency was also seen in patients with so-called occlusal discrepancies.[24] A classical mechanistic hypothesis was that SB would be caused by dental factors, for example, occlusal disharmony/supracontacts, which would "irritate" the CNS and trigger excessive jaw-muscle activity, that is, bruxism.[25] Logically, the adjustment or correction of occlusal disharmony, therefore, would result in the immediate disappearance of the habitual grinding of the teeth because the disharmony/supracontacts would be "equilibrated" by selective grinding or advanced dental restorative procedures.[24,26] Further support to this hypothesis was derived from the observation that SB is most prevalent in 8- to 15-year-old children, that is, with mixed dentitions and more instable occlusions.[27] However, these views are now downplayed because of a lack of good evidence to support any close relationships or causality between occlusal factors or craniofacial morphology and SB in well-designed studies with operationalized criteria.[17,28-30] Further, experimental studies with insertions of artificial and reversible supracontacts have shown a decrease, rather than an increase, in jaw-muscle activity during sleep, contradicting the occlusal-based hypothesis of SB.[31] In other studies, it was shown that the elimination of interferences in occlusion and articulation did not influence jaw-muscle activity during sleep.[32,33] Today, therefore, SB is considered to be mainly regulated and influenced by the CNS autonomic nervous system (CNS-ANS) with psychological-hyperarousal factors with little, if any, contribution from anatomical factors (occlusion, craniofacial morphology)[29,34-36] (see also Chapter 23).

Psychological factors such as anxiety,[37-43] neuroticism,[44] competitiveness,[45] emotional tension,[46] aggressiveness,[47] stress, and maladaptive/less positive coping strategies[48] have frequently been associated with SB, although some controversy remains on their specific contribution; some phenotype may be present in some individuals and not in others.[49] In particular, questionnaires and self-report data have linked SB to stress,[37,39,41,43,50-56] and some studies have also used excretion of urinary catecholamine or salivary chromogranin A as biomarkers of stress and found significant relationships.[50,57] It appears that a variety of psychological vulnerability factors, indeed, may predispose, maintain, or exacerbate SB.

Finally, the role of comorbidity and CNS should be considered in our search of understanding the pathophysiology of rapid eye movement (REM) sleep motor anomaly SB. Restless legs syndrome (RLS)[7,58] and another sleep-related movement disorder, such as REM-sleep behavior disorder (RBD), are considered to be concomitant to SB and seems to be a significant risk factor for SB in some individuals (see Chapter 145).[58,59] It remains to be demonstrated whether this co-occurrence is due simply to age (intersecting epidemiology, that is, prevalence of SB and RLS or RBD merge with a higher probability of comorbidity across ages: SB in younger ages and RLS or RBD in older ages) or whether there are some common pathophysiological bases (see also Chapter 23). Sleep apnea is considered, in some individuals, as another significant risk factor of SB. The clinical interest in the putative link between SB and sleep breathing disorders is growing; however, the strength of such an association is still debated, and causality is not yet fully proven.[60-65] Use of several types of medications or neuroactive substances, such as selective serotonin reuptake inhibitors,[66,67] dopamine antagonists,[68] or recreational drugs (amphetamine, alcohol, nicotin, caffeine),[11,69] is also considered as risk factors for some vulnerable SB patients; however, the level of evidence is often modest. Genetic susceptibility, as mentioned previously, has been relatively little examined, but there is some evidence of a hereditary predisposition based on twin studies, but accounting only for 20% to 50% of the variance.[14,15,70,71] The lessons learned, for example, from pain genetics would predict that it is highly unlikely to identify a single gene responsible for SB,[12] but rather that several genes and their haplotypes will interact in a complex manner with environmental, psychological, and endogenous factors.[72,73]

In general, SB patients are not bad sleepers; it is their sleep partners who have most sleep disturbed by the grinding sounds. Comparisons between young and healthy patients with SB and control subjects have shown normal sleep organization and macrostructure.[59,74] Several critical parameters of sleep, such as sleep latency, total sleep time, percentage of time spent in the various sleep stages, and number of awakenings, are within normal limits in sleep bruxers.[18,74,75,76] They also report a normal amount of time spent awake during the sleep period, and they demonstrate sleep efficiency that falls within the usual range of good sleepers (greater than 90%). Moreover, sleep bruxers without pain or insomnia rarely complain about poor sleep quality. Several studies have analyzed the sleep in details in sleep bruxers, and there is emerging evidence of a sequence of biological events, which culminate with excessive jaw-muscle activity, that is, teeth grinding or clenching. First, SB may occur in all sleep stages, but it has been established that SB episodes are most frequently (>80% of the time) encountered in sleep stages 1 and 2 (N 1 and 2), in the minutes before REM stage and not as previously suggested during REM sleep.[7,77-84] Furthermore, studies have now linked SB to a cyclic alternating pattern (CAP).[76,85] The CAP consists of a cyclic pattern of electroencephalographic (EEG), electrocardiographic (ECG), and EMG activation every 20 to 60 seconds,[86] and about 80% of SB episodes are observed in association with the CAP.[76] This may serve as a resetting mechanism for physiologic functions in relation to sleep environment or endogenous factors that act as a permissive window to allow the occurrence of otherwise inhibited motor activity such as SB or other body or limb movements.[76,87] The association between SB and CAP is further supported by findings showing that more than 50% of SB episodes occur in clusters (within 100 seconds), and that approximately 15% to 20% occur in the transition from deep sleep (Stage 3/4) to REM sleep.[85] These findings are also consistent with the observation that SB is preceded by alpha EEG activity and is associated with a tachycardia.[88-91] Lavigne and colleagues have made a series of important contributions to the understanding of the SB in the sequence of biological events during sleep.[7,18] The sequence seems to be initiated by a change in the

autonomic-cardiac sympathetic and parasympathetic balance, which subsequently is followed by a rise in EEG activity with more delta activity. This arousal response is followed by tachycardia and increased jaw-opener muscle activity with an increase in respiratory amplitude (1 to 2 breaths), and, finally, EMG activity in the jaw-closing muscles is often described as rhythmic jaw movements or as RMMA, as described previously.[36,92] Also, the blood pressure surge is linked to RMMA episodes, and combined findings support the concept that SB is secondary to exaggerated transient motor and ANS activation in relation to micro-arousals.[93-95] The presence of RMMA and SB-tooth grinding is frequent in young and healthy SB patients, but it is not always the case in the general population; older subjects may have concomitant sleep breathing disorders or insomnia, as described previously, who could also contribute to more or less vulnerability of RMMA-SB.[8,64] The sleep arousal axis allows a better understanding of the integration between the ANS and the CNS; the sympathetic/parasympathetic balance during sleep may influence SB and also explain why cardiovascular/respiratory factors may interact with SB.[59,94,96] Micro-arousals seem to be important, although not the only factor contributing to the genesis of RMMA-SB, and, indeed, experimental studies with application of vibrotactile stimuli during sleep indicate that such stimuli may cause micro-arousal without awakenings but with activation of SB and RMMA.[93]

In summary, anatomical factors like dental occlusion appear to have little, if any, influence on the genesis of SB. There appears to be no single etiological factor that explains SB but rather a set of factors, which may interact and exacerbate normal sleep-related bruxism. To avoid confusion, the term *etiological factors* should be used cautiously and better substituted by *risk factors*. So far, the most evidence supports the hypothesis that SB is centrally mediated and under the influence of the autonomic system function and brain arousal responses.

It is important to note that knowledge on pathophysiology and mechanism of SB is not only an academic exercise because there are important clinical consequences of the phenomenon. First, it should be recognized that the consequences of SB can form a continuum from no effects to significant and deleterious impact on oral structures and quality of life (e.g., severe tooth attrition, jaw muscle soreness and pain, and headache). This means that SB should not be treated in each and every patient but rather should be based on a careful examination and analysis of the problems and needs in the individual patient. Second, oral appliances may reduce tooth wear and grinding sounds, but SB cannot be cured by occlusal rehabilitation procedures; the occurrence of sleep-related RMMA will persist. The use of psychological methods (relaxation and cognitive behavioral therapy) may help, although evidences are yet scarce; in some cases, medication may be justified to alleviate the ANS and CNS arousal permissive influences on SB genesis. Subsequent chapters will address in more details the management of SB (see chapters 147 and 150).

CLINICAL PEARLS

- SB is a frequent condition and based on self-reports affecting between 8% and 15% of the population.
- SB has a counterpart during awake states, but the pathophysiology as well as clinical consequences may differ.
- A diagnostic grading system is suggested to distinguish between possible, probable, and definitive types of SB.
- SB cannot be explained by only dental factors such as occlusion or craniofacial morphology but is influenced by a range of psychological factors and CNS function in combination with ANS.
- Micro-arousal responses seem essential for a sequence of biological events and may contribute to have a permissive effect leading to SB.

SUMMARY

SB is considered a common condition associated with activation of the jaw-closing muscles leading to teeth grinding or jaw clenching. The potential consequences of SB form a continuum from no deleterious effects to severe tooth wear and tooth destruction to headache upon awakening and craniofacial pain complaints. In addition, sleep partners can be disturbed by the noise from tooth gnashing, and SB can have a negative impact on the quality of life. It is a consistent finding that children (15%) more often report SB than adults (8%). There is a general consensus that there is no single cause of SB but that CNS and psychological factors, including stress, may contribute; in severe cases, excessive sleep arousal responses are influenced by autonomic factors. Diagnosis and clinical examination of the orofacial region are essential to be performed. A definite diagnosis of SB, however, can be made with only PSG recordings of jaw muscle activity, together with audio-video signals. Due to the multifactorial etiology and complex pathophysiology of sleep bruxism, management is always symptomatic.

Selected Readings

Carra MC, Huynh N, Fleury B, Lavigne G. Overview on sleep bruxism for sleep medicine clinicians. *Sleep Med Clin* 2015;**10**:375–84.

Lavigne GJ, Khoury S, Abe S, et al. Bruxism physiology and pathology: an overview for clinicians. *J Oral Rehabil* 2008;**35**:476–94.

Lavigne GJ, Rompré PH, Montplaisir JY. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. *J Dent Res* 1996;**75**:546–52.

Lobbezoo F, et al. Bruxism defined and graded: an international consensus. *J Oral Rehabil* 2013;**40**:2–4.

Lobbezoo F, Ahlberg J, Manfredini D, et al. Are bruxism and the bite causally related? *J Oral Rehabil* 2012;**39**:489–501.

Lobbezoo F, Visscher CM, Ahlberg J, et al. Bruxism and genetics: a review of the literature. *J Oral Rehabil* 2014;**41**:709–14.

Mayer P, Heinzer R, Lavigne G. Sleep bruxism in respiratory medicine practice. *Chest* 2015 Jul 30. [Epub ahead of print].

Rompré PH, Daigle-Landry D, Guitard F, et al. Identification of a sleep bruxism subgroup with a higher risk of pain. *J Dent Res* 2007;**86**:837–42.

A complete reference list can be found online at ExpertConsult.com.

Sleep Bruxism: Diagnostic Considerations

Frank Lobbezoo; Kiyoshi Koyano; Daniel A. Paesani; Daniele Manfredini

Chapter 145

Chapter Highlights

- The main challenge in relation to sleep bruxism concerns its diagnosis. So long as consensus on how to diagnose sleep bruxism is lacking, studies on the condition will lack comparability and global acceptance. The recent suggestion for a diagnostic grading system of "possible," "probable," and "definite" sleep bruxism is an important step toward consensus.
- For a "possible" or " probable" diagnosis of sleep bruxism, self-report and clinical approaches are indicated, whereas for a "definite" diagnosis, instrumental assessments such as polysomnography, preferably in combination with audio-video recordings, are required.
- As yet, insufficient evidence has been accumulated to support the use of ambulatory electromyographic devices as stand-alone tools for the diagnosis of sleep bruxism, as tested against full polysomnographic recordings.
- In the absence of full consensus on the diagnosis of sleep bruxism, an accurate differential diagnosis that considers oral-motor disorders, such as orofacial dyskinesia and oromandibular dystonia, will be difficult. Likewise, purported associations of sleep bruxism, such as those suggested for rapid eye movement behavior disorder, obstructive sleep apnea, and gastroesophageal reflux disease, will be hard to interpret unequivocally in terms of cause-and-effect implications.
- It will be a stimulating challenge for the near future to operationalize the foregoing suggestions for diagnostic grading, taking into consideration the important work that has already assessed this topic.

DEFINITION OF BRUXISM

Recently, an international group of experts has defined bruxism as a repetitive jaw muscle activity characterized by clenching or grinding of the teeth and/or by bracing or thrusting of the mandible. The condition has two distinct circadian manifestations: sleep bruxism (SB) and awake bruxism.[1] The new consensus definition, which is suggested for clinical and research purposes in all relevant dental and medical domains, has already been included in leading documents such as the orofacial pain guidelines of the American Academy of Orofacial Pain[2] and the third edition of the *International Classification of Sleep Disorders* (ICSD3),[3] showing its emerging international acceptance.

This chapter reviews recent insights into the methodologic aspects and the differential diagnosis for bruxism restricted to sleep. More details on the definition, epidemiology, consequences, and management of SB can be found in Chapters 144, 146, 147, and 150.

DIAGNOSIS OF SLEEP BRUXISM

For the assessment of SB, multiple diagnostic tools are available. Whether any of them, alone or in combination, is uncontestably valid, however, remains unknown and thus a matter of debate. As briefly described in Chapter 144, in a consensus effort of experts, Lobbezoo and colleagues[1] have suggested a grading system for SB: For a "possible" diagnosis, self-report of the condition or a proxy (such as a bed partner or a parent) report is sufficient. For a diagnosis of "probable" SB, clinical features suggestive of SB should be present as well, whereas for a "definite" diagnosis, positive findings on polysomnography, preferably in combination with positive findings on audio-video recordings and with bruxism outcome measures above a predefined threshold, are required in addition to the patient's self-report and consistent clinical features on physical examination.

Presented next are various diagnostic approaches to SB, elaborated in some detail.

Self-Report

For a "possible" diagnosis of SB, a self-report of the condition can be obtained by questionnaires and/or by oral history taking. Questionnaires can be used to obtain information about the condition itself, and on its possible causes and consequences, as well as possible comorbid conditions and differential diagnostic considerations (see Differential Diagnosis, further on). An example of a questionnaire for the assessment of sleep bruxism that can be used in the research setting as well as in the clinic is the so-called BRUX scale, developed and tested for its psychometric properties by van der Meulen and associates.[4] The BRUX scale consists of four questions

that can be answered on a 5-item verbal scale (ranging from never to always), two of which deal with SB (Box 145-1). The BRUX scale is part of a more extensive tool: the Oral Parafunctions Questionnaire (OPQ).[4] Apart from four bruxism questions (the BRUX scale), the OPQ inquires about chewing on pens or pencils, nail biting, and chewing gum (the BITE scale) as well as about vacuum sucking with the tongue, playing or pushing with the tongue, lip biting, sucking on lips/cheeks, and playing with a denture (the SOFT scale). In addition, the OPQ assesses behaviors that may strain the jaw, such as sleeping on the belly. Not only the latter behavior but also some of the other behaviors ("parafunctions") can occur during sleep, and their effects should thus be distinguished from SB.

Another comprehensive tool for the assessment of SB and other (possibly comorbid) oral behaviors is the Oral Behaviors Checklist (OBC).[5,6] This 21-item tool was composed based on expert opinions and on patient contributions. Like the OPQ, the OBC uses a 5-point scale for its responses (for sleep-related behaviors: none of the time, <1 night/month, 1 to 3 nights/month, 1 to 3 nights/week, 4 to 7 nights/week).

Box 145-1 BRUX SCALE OF THE ORAL PARAFUNCTIONS QUESTIONNAIRE (OPQ)* FOR THE ASSESSMENT OF SELF-REPORTED BRUXISM

Questions

1. How often do you clench your teeth during sleep?
2. How often do you grind your teeth during sleep?
3. How often do you clench your teeth while awake?
4. How often do you grind your teeth while awake?

Response Options

Never (0)
Sometimes (1)
Regularly (2)
Often (3)
Always (4)

*Developed by van der Meulen MJ and colleagues; reported in van der Meulen MJ, Lobbezoo F, Aartman IH, Naeije M. Self-reported oral parafunctions and pain intensity in temporomandibular disorder patients. *J Orofac Pain* 2006;20:31-5.

While the daytime behaviors were tested for reliability by means of electromyographic recordings, the sleep-related behaviors were not subjected to reliability assessments so far. Of importance, whereas the OPQ asks about the presence or absence of single parafunctions, the OBC contains several questions that ask about the presence or absence of multiple behaviors in one and the same question—for example: "Based on the last month, how often do you clench or grind teeth when asleep, based on any information you may have?" Because clenching and grinding are commonly considered distinct conditions, this mode of inquiring yields less information than when single behaviors are being assessed.[7] Nevertheless, the OBC was well received by the international community (see further on; see also Table 145-1).

Multiple self-report tools are available for the assessment of the possible causes and consequences of SB. For one of the most commonly suggested consequences, specifically, temporomandibular disorders (TMD), and for several possible causes of SB (see Chapter 144), the recently published *Diagnostic Criteria for TMD* (DC/TMD)[8] suggests a comprehensive protocol for the assessment of several physical and psychological domains by means of (in most cases) validated questionnaires. The domains and their respective assessment tools are shown in Table 145-1. Although the DC/TMD is a leading and largely evidence-based system, clinicians are free to select their own questionnaires for the assessment of SB as well as of its possible causes and consequences.

When a questionnaire is completed and analyzed before the clinical consultation, the clinician knows better what to expect. In addition, patients usually have gained an improved awareness of their oral behavior, which helps improve the reliability of the oral history—the second way of obtaining self-reported data on SB. An important advantage of an oral history over the use of questionnaires is the way in which the questions are formulated: They can be adapted to the individual patient's knowledge and cognitive abilities. For research purposes, on the other hand, an oral history lacks standardization. Also, oral history taking may be influenced by the clinician's preconceived ideas about SB. A disadvantage of the use of questionnaires for the diagnosis of SB is that approximately 80% of bruxism episodes are free of grinding sounds.[9] Consequently, owing to lack of awareness for many such events, self-reports of the condition may be characterized by

Table 145-1 Questionnaires for Assessment of Physical and Psychological Domains in Temporomandibular Disorders and Sleep Bruxism as Recommended by the Diagnostic Criteria for TMD (DC/TMD)

Domain	Questionnaire	Abbreviation
Pain intensity, physical function	Graded Chronic Pain Scale	GCPS
Pain locations	Pain drawing	N/A
Limitation	Jaw Function Limitation Scale	JFLS
Distress	Patient Health Questionnaire-4	PHQ-4
Depression	Patient Health Questionnaire-9	PHQ-9
Anxiety	Generalized Anxiety Disorder-7	GAD-7
Physical symptoms (somatization)	Patient Health Questionnaire-15	PHQ-15
Parafunction	Oral Behaviors Checklist	OBC

Data from Schiffman E, Ohrbach R, Truelove E, et al: Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. *J Oral Facial Pain Headache* 2014;28:6-27, Table 7.

underscoring, with respect to the standard reference for a definite SB diagnosis (see further on under Instrumental Assessment). At present, the only investigation reporting both polysomnographic and self-reported SB diagnosis in a large sample showed polysomnographic confirmation of SB in less than one half of the patients self-reporting this behavior.[10] Furthermore, the occurrence and severity of SB fluctuate over time,[11] which further hampers the accuracy of self-reports of SB. Hence, questionnaires should be interpreted with caution, and preferably, they should be used in combination with other diagnostic tools.[12]

As a final remark, it must be pointed out that, despite the aforementioned disadvantages, self-reported approaches to SB diagnosis have been used to collect data for several research purposes. The generalization of findings on most SB issues is thus limited.[13,14]

Clinical Examination

A clinical examination that assesses "probable" SB (as described previously) focuses on the extraoral assessment of masticatory muscle hypertrophy (see earlier) as well as on the intraoral assessment of hyperkeratosis (i.e., tooth-induced impressions in cheeks ["linea alba"], tongue ["indentations"], and/or lips),[15,16] tooth wear (Figure 145-1, *A, B*),[17] fracture or failure of restorations and implants,[15,18,19] and signs such as tooth mobility, pulp necrosis, and traumatic ulcers.[16] Of importance, all of those clinical features lack specificity in the assessment of SB. Masseter muscle hypertrophy, for example, should be distinguished from parotid salivary gland pathology, although it also can be associated with awake bruxism activities or other poor oral habits.

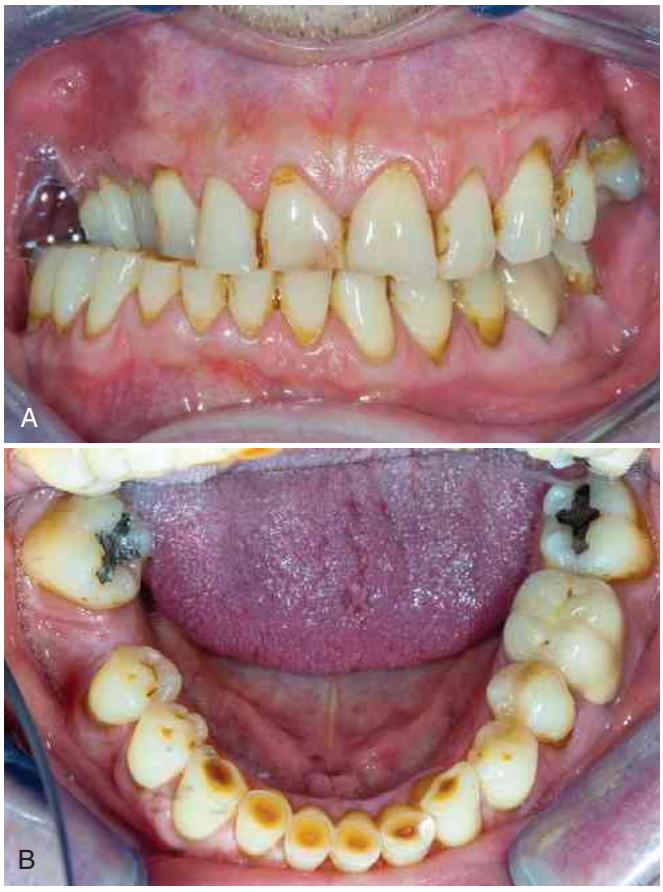


Figure 145-1 Example of sleep-related grinding damage. **A,** Frontal view. **B,** Occlusal view on the lower jaw.

Tooth wear can have multiple causes.[20] Whereas attrition is caused by tooth-tooth contacts as in sleep grinding, abrasion is due to tooth contact with objects (e.g., toothbrushes, pens, or fingernails), and erosion is a chemically caused type of tooth wear (as with reflux of stomach contents or an acidic diet). The interpretation of tooth wear in SB recognition is not simple, because no single cause is present and the wear could have happened years before the examination. As a consequence, this complicates the interpretation of a worn dentition in terms of tooth-grinding activities as well. This is illustrated by the study of Abe and coworkers,[17] who concluded that although the presence of attrition is indeed associated with the presence of SB, the degree of attrition did not show a dose-response relationship with the severity (specifically, the frequency of jaw muscle contractions per hour of sleep) of SB.

From a dental point of view, the direction of tooth grinding is an important factor to consider in restorative dentistry: a restoration should be loaded toward the central axis of the tooth, and not away from the center as to prevent fractures of the fillings. So long as the tooth wear is confined to the enamel, the direction of grinding can be established with the so-called "scratch test," whereby a small scratch is placed on the wear facet by means of a scalpel. At baseline and after a brief time interval of a few nights, dental precision impressions are made of the scratches and subsequently studied under scanning electron microscopy. Such scratches will show most microwear on their leading edge—that is, on the edge opposite to the grinding direction.[21] When extensive dental restorations are planned, this information can be taken into account.

Several reviews have been published on the effects of SB on dental implants[18,19,22] as well as on implant-supported prostheses.[23] In short, SB seems to be unrelated to biologic complications (e.g., failing osseointegration of the implants), although some evidence suggests that bruxism causes mechanical complications (e.g., fractures of implants or suprastructures/prostheses).

Apart from the aforementioned aspects, the clinical examination may include a functional examination of the masticatory system, based on the paradigm that bruxism and TMD pain are potentially causally related. To that end, the clinician may use the highly sensitive muscle and joint palpation tests, a negative outcome of which confirms the absence of TMD pain, and/or the highly specific dynamic/static pain tests, a positive outcome of which confirms the presence of TMD pain.[24] Whether or not SB and TMD are actually causally related, however, is highly debated in the literature. According to Lobbezoo and Lavigne,[25] the nature of this relationship is still unclear. A more recent, systematic review by Manfredini and Lobbezoo[14] yields a slightly different, although not opposing, insight into the purported causal association between SB and TMD: Investigations based on self-report or clinical examination, or both, suggested a positive association between the two conditions, although these studies are characterized by some potential bias and confounders at the diagnostic level. By contrast, studies based on use of more quantitative and specific methods to diagnose SB (e.g., polysomnography; see

later) showed much weaker associations with TMD symptoms. Apparently, the more precisely SB is defined, the less clear is its purported association with TMD. Consequently, the value of a TMD pain diagnosis is limited in the context of the assessment of SB.

A recent study assessed the correlation between questionnaire-based SB and a diagnosis of SB based on an oral history in combination with a clinical examination.[26] Proxy-reported sleep-related grinding showed a high correlation between both diagnostic approaches ($\Phi = 0.93$), whereas a lower (albeit still acceptable) correlation value was found for sleep-related clenching behavior ($\Phi = 0.64$). Interestingly, the non-proxy self-report yielded a lower correlation value ($\Phi = 0.63$) than the proxy-based one, which underlines the importance of including proxy (notably partner) reports in dedicated questionnaires and oral history taking.

Of note, at present, data are lacking on the correlation between clinically diagnosed SB and the definite diagnosis of this condition obtained by instrumental techniques (as described next). Thus further research on this topic is strongly encouraged.

Instrumental Assessment

A "definite" diagnosis of SB can be based on a positive polysomnographic recording, preferably in combination with audio-video recordings and with bruxism outcome measures above a predefined threshold, in concert with self-report of such episodes and expected clinical features on the physical examination. Because, as elaborated further on, polysomnography is a rather complicated, inaccessible, and expensive tool, several alternative instrumental techniques have been developed over the years with the aim of objectifying the presence of SB in individual patients. Following is an overview of the most commonly applied tools for the instrumental assessment of SB.

Intraoral Appliances

A commonly used tool to assess the presence and progression of sleep-related grinding is the hard occlusal stabilization appliance (Figure 145-2, *A, B*). This appliance (or "splint"), that is indicated for TMD pain as well as for SB, is worn over the upper or lower dental arch. A systematic review of the available randomized clinical trials demonstrated that the available evidence is insufficient for a recommendation for or against the use of stabilization splint therapy for the treatment of TMD.[27] Likewise, for the treatment of SB, a systematic review concluded the following:

> There is not sufficient evidence to state that the occlusal splint is effective for treating sleep bruxism. Indication of its use is questionable with regard to sleep outcomes, but it may be that there is some benefit with regard to tooth wear.[28]

In other words, it is likely that in at least some patients with SB, the grinding behavior persists—thereby wearing down the stabilization splint. Hence, the splint may function as a tool for the assessment of the presence and progression of SB. However, quantifying splint wear is fraught with difficulties, both of a technical nature and in terms of interpretation.[16]

Another frequently used intraoral appliance for the assessment of SB is the Bruxcore plate (BBMD; Bruxcore, Boston, Mass.)—a thin (0.51-mm) polyvinyl chloride sheet with four layers of two alternating colors that is fitted to the dental arch. When worn during the night, sleep-related grinding activities will lead to wear of the colored layer(s), thus yielding a measure for the amount of SB (Figure 145-3). Unfortunately, as outlined in detail by Koyano and associates,[12] several drawbacks are attached to this method, the most important one being the finding that the Bruxcore plate scores do not correlate with jaw muscle activities detected on the electromyogram (EMG) during sleep.[29] In all likelihood, this is due to the fact that bruxism is characterized not only by tooth grinding but also by clenching activities—the latter leading to less material wear than from grinding. Clearly, other methods are needed for the establishment of a definite SB diagnosis.

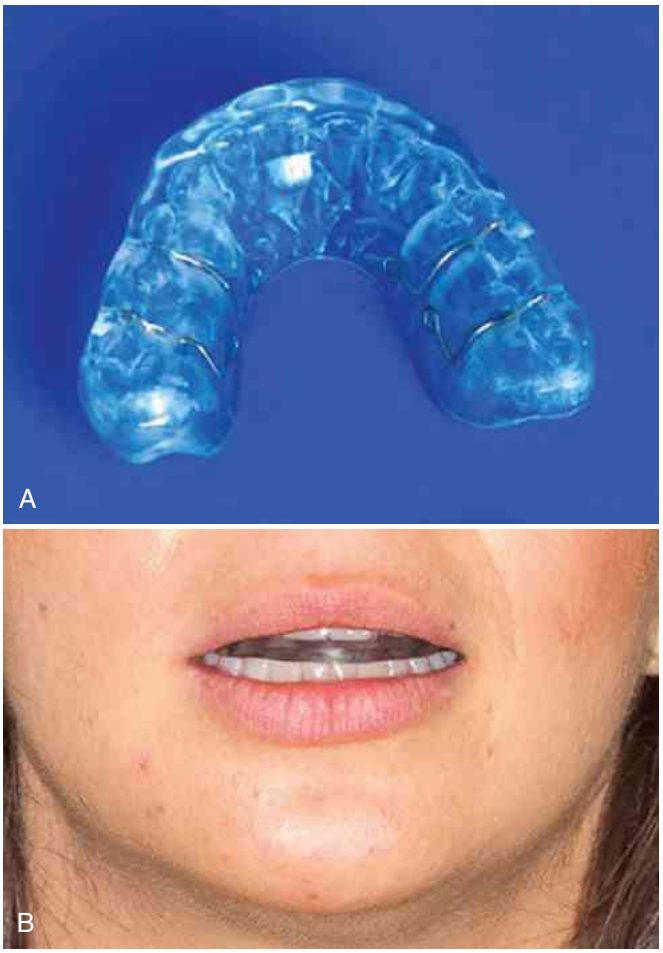


Figure 145-2 A, Example of a hard occlusal stabilization appliance, indicated for the management of sleep bruxism and commonly used to assess the presence and progression of the condition based on wear patterns on the occlusal surface. **B,** In situ on the upper jaw.

Electromyography

From the preceding discussion, it can be gathered that an objective assessment of jaw muscle activities is an essential step in the diagnosis of SB. To that end, electromyography is commonly used. A vast number of electromyographic devices are available, some of which have been developed specifically for the assessment of SB. Although they have various differences in technical and practical characteristics, they all are easy

to use, relatively cheap, and portable—thus allowing recording at home ("ambulatory recording"). A major drawback of ambulatory EMG recorders is that they do not provide the clinician with information about the sleep-wake state, nor do they allow the discrimination between SB activities on the one hand and other sleep-related/nocturnal orofacial activities on the other—thus potentially leading to overscoring of the number of "bruxism" events during the recording time. With that in mind, recent developments that combine EMG recordings with one or more additional leads (e.g., heart rate[30] or multiple leads including pulse wave intervals, actigraphy, and audio-video recordings[31]) may yield the optimal combination between electromyography alone and full polysomnographic recordings for the assessment of SB.

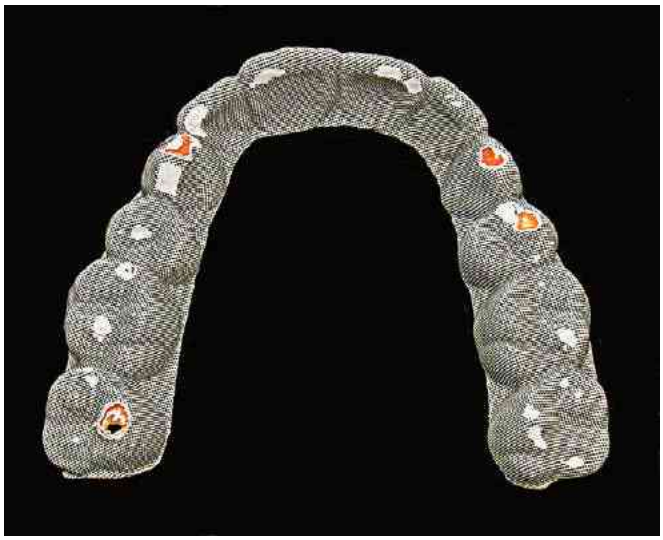


Figure 145-3 Example of a Bruxcore plate (BBMD; Bruxcore, Boston, Mass.), worn on the upper dental arch during the night. Sleep-related grinding activities are shown as wear of the halftone dot screen topmost surface, thus yielding a measure of the amount of sleep bruxism.

Polysomnography

According to the ICSD3,[3] polysomnography is needed for an accurate diagnosis of SB. Pioneer work on this topic has been performed by the research group of Lavigne and collaborators, on whose work most of the following information is based.[32-35] The repetitive jaw muscle activities that are recorded with polysomnography are indicated as rhythmic masticatory muscle activity (RMMA). Traditionally, the characteristics of the RMMA recording are used to define SB events (i.e., bursts and episodes); see Box 145-2. Of note, although separate criteria were previously proposed for ambulatory recordings using one or only a few leads,[36] current equipment usually allows for sufficiently high sample rates to enable application of the criteria as described in Box 145-2 for ambulatory recordings as well.

Full polysomnographic recordings can be obtained with ambulatory equipment in the home environment, but the recordings can also be performed in a highly controlled (but rather unnatural) sleep laboratory environment. Specifically

Box 145-2 SLEEP BRUXISM EVENTS (BURSTS AND EPISODES), DEFINED BY THEIR CONSTITUENT RHYTHMIC MASTICATORY MUSCLE ACTIVITY (RMMA) CHARACTERISTICS

Phasic (Rhythmic) Episodes

At least three suprathreshold EMG bursts* in the masseter and/or temporalis muscles, lasting ≥0.25 second† and <2 seconds and separated by two interburst intervals of <2 seconds‡

Tonic (Sustained) Episodes

One EMG burst of ≥2 seconds

Mixed Episodes

Both phasic and tonic characteristics

Nonclassified Bursts

Either one EMG burst lasting ≥0.25 second and <2 seconds, or two EMG bursts, lasting ≥0.25 second and <2 seconds and separated by one interburst interval of <2 seconds

*EMG threshold can established in different ways: >20% of the maximum voluntary contraction (MVC) level[33]; >10% of the MVC level[34]; or a multiplication (e.g., 3 times) of the background (noise) level of the EMG signal.
†EMG events <0.25 sec are considered twitches or myoclonic activities.
‡When the time interval between two bursts is ≥2 sec, a new event (burst, episode) starts.
EMG, Electromyogram.
Data from Velly Miguel A, Montplaisir J, Rompré PR, et al. Bruxism and other oro-facial movements during sleep. *J Craniomandib Disord* 1992;6:71-81; Reding GR, Zepelin H, Robinson JE Jr, et al. Nocturnal teeth-grinding: all-night psychophysiologic studies. *J Dent Res* 1968;47:786-97; and Ware JC, Rugh JD. Destructive bruxism: sleep stage relationship. *Sleep* 1988;11:172-81.

Box 145-3 OUTCOME MEASURES AND DIAGNOSTIC CUT-OFF CRITERIA FOR SLEEP BRUXISM, BASED ON ANALYSES OF RHYTHMIC MASTICATORY MUSCLE ACTIVITIES (RMMA) DURING SLEEP

Sleep Bruxism Outcome Measures

Classical outcome measures[33]
- Number of sleep bruxism episodes/night
- Number of sleep bruxism episodes/hour (of sleep)
- Number of sleep bruxism bursts/hour (of sleep)
- Number of sleep bruxism bursts/episode

Additional outcome measure[66]
- Total bruxism time by total sleep time: bruxism time index (bti)

Sleep Bruxism Diagnostic Cut-off Criteria

Classical criteria[33]
- >4 episodes/hour, AND
- 6 bursts/episode and/or >25 bursts/hour, AND
- At least 1 episode/night with grinding sounds

Adapted criteria[35,67]
- Low intensity: >1 and ≤2 episodes/hour
- Moderate intensity: >2 and ≤4 episodes/hour
- High intensity: >4 episodes/hour

Data from Lavigne GJ, Rompré PH, Montplaisir JY. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. *J Dent Res* 1996;75:546-52[33]; Rompré PH, Daigle-Landry D, Guitard F, et al. Identification of a sleep bruxism subgroup with a higher risk of pain. *J Dent Res* 2007;86:837-42[35]; van der Zaag J, Lobbezoo F, Wicks DJ, et al. Controlled assessment of the efficacy of occlusal stabilization splints on sleep bruxism. *J Orofac Pain* 2005;19:151-8[66]; and Carra MC, Huynh N, Lavigne G. Sleep bruxism: a comprehensive overview for the dental clinician interested in sleep medicine. *Dent Clin North Am* 2012;56:387-413.[67]

for the assessment of SB, polysomnography includes surface EMG recordings, obtained from at least one jaw-closing muscle (right and/or left masseter and/or temporalis muscle). The use of audio-video recordings is more feasible in the sleep laboratory environment than in the home milieu and contributes to the accuracy of a definite SB diagnosis.[37] These studies help in ruling out other orofacial activities that can be confused for SB events, such as swallowing, yawning, and sleeptalking.

Based on the works of Lavigne and collaborators (as noted previously and described in Chapter 144), research diagnostic criteria have been developed and also have evolved over the years. The classical and current recommendations are given in Box 145-3. An important limitation of the application of those criteria is the fact that the frequency and severity of SB are known to fluctuate considerably over time.[38] Van der Zaag and associates[11] quantified the diagnostic consequences of the time-variant nature of SB by suggesting the use of 95% probability cut-off bands around the previously suggested cut-off points that are included in Box 145-3. For example, with use of a cut-off point of 4 episodes occurring in each hour of sleep to distinguish sleep "bruxers" from "nonbruxers," the 95% probability cut-off band suggests that only patients with at least 7 episodes/hour are likely to be sleep bruxers, whereas persons with one or fewer episodes/hour are likely to be nonbruxers.[11] Of importance, the study authors indicated that because they worked with a mixture of previous data[33] and data that they collected themselves, the suggested cut-off bands should be considered an illustration of principle.

A recent systematic review assessed the diagnostic accuracy of portable instrumental devices for the measurement of SB against polysomnography assumed as the gold standard.[39] Using several databases and a quality assessment tool (i.e., Quality Assessment of Diagnostic Accuracy, version 2), the reviewers identified only four studies evaluating three different devices—Bitestrip, EMG-telemetry recorder, and Bruxoff—that could be included. In a disappointing turn, the validity of the included instrumental diagnostic approaches with respect to polysomnographic recordings was not only scarce but also not solid enough to support use of those techniques as stand-alone tools for disgnosis of SB. The most promising device was the Bruxoff,[30] which assesses not only the increased EMG masticatory muscle events that are characteristic for SB but also electrocardiographic changes that have been shown to occur in association with SB events,[40] thereby improving the diagnostic accuracy of the device over that of the other portable instruments. The outcome of the systematic review by Manfredini and colleagues,[39] however, implies that polysomnography is still the gold standard for diagnosis of SB. On the other hand, the validity of the polysomnographic criteria themselves to detect the pathologic consequences of SB, thus discriminating the "physiologic" phenomenon from the treatment-needing condition, has to be further explored.

DIFFERENTIAL DIAGNOSIS

SB can be confused with certain movement disorders and also may occur as a comorbid condition with other neurologic conditions. In this section, SB is considered in the context of oral movement disorders, rapid eye movement (REM) behavior disorder (RBD), sleep-disordered breathing, and gastroesophageal reflux disease (GERD).

Movement Disorders

Oral movement disorders are common conditions.[41] In some cases, the abnormal oral movement can be considered a focal manifestation of a generalized movement disorder such as Gilles de la Tourette syndrome, Huntington disease, idiopathic torsion dystonia, and Parkinson disease. The disorder also can be a side effect of medication taken for the generalized disorder.[42] In other cases, the oral movement disorder, notably orofacial dyskinesia or oromandibular dystonia, is the sole disorder present.

Orofacial dyskinesia is characterized by involuntary, mainly choreatic (dance-like) movements of the jaw, as well as other structures such as the face, lips, and tongue. The same structures can be affected by oromandibular dystonias, which consist of excessive, involuntary, and sustained muscle contractions. Both disorders have been reviewed extensively.[41,42] As possible causes, loss of inhibitory control of the basal ganglia, certain psychiatric diseases, excessive usage of dopaminergic medications, and the chronic use of antipsychotic drugs (neuroleptics) have been mentioned in the literature. When the disorder is caused by neuroleptics, it is indicated as a "tardive" condition. In addition, some anecdotal evidence from case studies suggests that certain dental conditions such as, for example, edentulism and dentoalveolar trauma, also may play a causal role in the etiology of orofacial movement disorders. Peck and coworkers[43] have formulated a set of diagnostic criteria for this condition; see Box 145-4.

When orofacial movement disorders persist in sleep, they can easily be confused with SB. Notably, orofacial dyskinesias and oromandibular dystonias can be misdiagnosed as grinding

Box 145-4 DIAGNOSTIC CRITERIA FOR OROFACIAL DYSKINESIA AND OROMANDIBULAR DYSTONIA

Orofacial Dyskinesia

History should be positive for the following:
- Neurologic diagnosis of dyskinesia in the orofacial region

Examination by neurologist should be positive for the following:
- Sensory and/or motor nerve conduction deficit, AND
- Central and/or peripheral myopathic disease, AND
- Muscular hyperactivity confirmed by intramuscular EMG

Oromandibular Dystonia

History should be positive for the following:
- Neurologic diagnosis of oromandibular dystonia

Examination by neurologist should be positive for the following:
- Sensory and/or motor nerve conduction deficit, AND
- Central and/or peripheral myopathic disease, AND
- Dystonia confirmed by intramuscular EMG

EMG, electromyogram.
Data from Peck CC, Goulet JP, Lobbezoo F, et al. Expanding the taxonomy of the diagnostic criteria for temporomandibular disorders. *J Oral Rehabil* 2014;41:2-23; and Balasubramaniam R, Ram S. Orofacial movement disorders. *Oral Maxillofac Surg Clin North Am* 2008;20:273-285, vii.[68]

(phasic) or clenching (tonic) behavior, respectively, especially when the disorder is confined to the jaw. SB also needs to be differentiated from faciomandibular myoclonus, parasomnias such as abnormal swallowing, night terrors, and confusional arousals, and—rarely—from sleep-related epilepsy.[3] This underlines the need for proper diagnostic procedures, including an extensive medical and dental history, the more so because the treatment of orofacial movement disorders differs from that of SB, with a primary role for medical specialists instead of dentists.

REM Behavior Disorder

REM behavior disorder is a neurologic disorder that occurs during the REM phases of sleep and features abnormal, powerful body movements mimicking motor behavior, in contrast with the muscle paralysis (atonia) that usually is observed during REM sleep. The disorder may lead to aggressive or complex behaviors, which may be accompanied by loud vocalizations related to the emotions experienced while dreaming. It essentially affects older men and may go undiagnosed for years before medical attention is sought. Diagnosis is confirmed with polysomnography, which typically shows absence of normal REM sleep atonia and sometimes abnormal behaviors. Treatment may be attempted with clonazepam. Because of its frequent association with the alpha-synucleinopathies, such as Parkinson disease, RBD is considered to be an important risk factor for later onset of neurodegenerative disorders, including dementia.[44,45] Pilot case studies reported that patients with RBD also may show tooth grinding.[46,47] In particular, a recent case-control paper comparing data from 28 patients with idiopathic or Parkinson disease–related RBD with those from 9 age- and sex-matched control subjects suggested that in the presence of high-frequency RMMA during sleep, RBD should be suspected.[47] Of interest, on the one hand, these findings support the hypothesis that different sleep disorders may co-occur in the same person as part of complex spectra of alterations of the normal sleep structure; on the other hand, they are in apparent contrast with currently available knowledge on idiopathic SB, which occurs mainly in NREM sleep stages 1 and 2, as described in other chapters.

Sleep-Disordered Breathing

Sleep-related breathing disorders are a group of conditions related with alterations of the normal airflow and respiration during sleep. As described elsewhere in this book, they commonly are divided into five categories: (1) obstructive sleep apnea (OSA) syndromes; (2) central sleep apnea syndromes; (3) sleep-related hypoventilation syndromes; (4) sleep-related hypoventilation resulting from a medical condition; and (5) other sleep-related breathing disorders.[3] In this discussion, the focus is on OSA, which has been called into cause for its possible relationship with SB.[48]

OSA is a breathing disorder that is characterized by the occurrence of apneic-hypoapneic episodes during sleep. It is considered to be a primary sleep disorder, characterized by the collapse of the pharynx over the upper airways, which may cause their total (i.e., apnea) or partial (i.e., hypopnea) obstruction and may lead to oxygen desaturation and arousal—that is, awakening from sleep.[3] Among others, obesity, male sex, menopausal state in women, and individual variability in lung volume and ventilation control have been identified as risk factors for OSA. Nevertheless, as in the case of SB, the pathophysiology of OSA is not yet fully understood, because it is plausible that a combination of neuromuscular and anatomic factors may play a role in the pathogenesis of obstruction.[49]

In the adult population, apneic episodes have been reported in more than 30% of patients with possible bruxism.[50] Because SB is correlated with arousal episodes,[51] and a high number of short arousal episodes also have been observed in patients with OSA, a possible relationship between the two phenomena (i.e., airway obstruction and SB episodes) has been hypothesized.[48] In some polysomnographic studies, an increase in the EMG activity of jaw muscles was observed in 40% to 60% of patients with OSA[52]; in some cases, a tonic or phasic activity of the masseter at the end of the apneic event has been described.[53–57] Findings regarding the temporal relationship between the two phenomena are inconclusive, however, and do not allow determination of important clinical questions: Does bruxism-like jaw muscle activity actually follow or precede the apnea? Do they coincide, or are the events temporally unrelated?

A recent paper suggested that four hypothetical scenarios may be identified: (1) the two phenomena are unrelated with respect to temporal sequence; (2) the onset of the OSA event precedes the onset of the SB event within a limited time span; (3) the onset of the SB event precedes the onset of the OSA event within a limited time span; or (4) the onset of both phenomena occurs at the same moment.[58]

The most plausible hypothesis is that all four of these temporal relationships between an SB event and an OSA event are actually possible, and that the relative predominance of one specific sequence of events may vary from patient to patient. For instance, the role of SB events may range from a protective activity against the apnea, in an attempt to protrude the relaxed mandible and restore airway patency,[48] to an OSA-inducing activity, as a consequence of the airways' mucosal swelling resulting from an SB-induced trigeminal cardiac reflex.[59] Interindividual differences, especially in the anatomic location of the airway obstruction, also may play a role in determining the nature of the SB-OSA relationship.

Gastroesophageal Reflux Disease

Gastroesophageal reflux is the physiologic regurgitation of stomach contents into the esophagus and the mouth, which normally may occur after eating and is asymptomatic. When the frequency and duration of such reflux episodes increase, signs and symptoms may emerge, reflecting the development of pathophysiologic changes leading to GERD. The predominant symptom is known as "heartburn," because of the classic painful chest sensation, and can affect 7 to 10% of people in the general population during waking hours. GERD is more common during sleep owing to the facilitation of reflux in the supine position, but very little is known about its sleep-time prevalence.[60]

With respect to SB, GERD assumes importance because of the need for taking this condition into account during the differential diagnosis when tooth wear (see Figure 145-1) and erosion are detected. Also, despite the absence of sound literature data on the issue, the possible coexistence of the two conditions may increase the risk for severe tooth damage. In particular, an interesting recent hypothesis suggested that SB-like RMMAs can be induced in healthy persons by

experimental esophageal acidification.[61] This suggestion found support in early observations of an increase in the number of SB episodes with lowering of the pH of saliva and esophageal contents,[62] as well as in a recent investigation describing a 73% prevalence of SB in patients with GERD.[63]

FINAL CONSIDERATIONS AND FUTURE DIRECTIONS

SB has been studied with increasing frequency over the past several decades. Notwithstanding the enormous effort, the condition is still not fully understood. An important development has been the formulation of a new definition by an international group of bruxism experts.[1] Because this definition has been widely adopted by leading professional and scientific organizations, at least future studies can be expected to be comparable regarding the description of SB. The main challenge in relation to SB, however, is still its diagnosis. So long as consensus on how to diagnose SB is lacking, studies on this condition will lack comparability and global acceptance. The suggestion from the consensus run by Lobbezoo and colleagues[1] for a diagnostic grading system ("possible," "probable," and "definite" SB, as defined earlier) is only a small first step towards a solution. It should be noted that the diagnostic approach followed depends on the study aim. For example, for a large-scale epidemiologic assessment, a "possible" diagnosis may be sufficient, in that large numbers of participants will compensate for the lower diagnostic accuracy. On the other hand, when associations of sleep-related events with SB phenomena are the aim of a study, a "definite" diagnosis with high accuracy will be necessary, with all of the possible drawbacks of limited availability and high costs attached to it. Of importance, in the absence of full consensus on the diagnosis of SB, a proper differential diagnosis that includes, for example, orofacial dyskinesia and oromandibular dystonia, will be difficult as well. Likewise, purported associations of SB, such as those suggested for RBD, OSA, and GERD, will be hard to interpret unequivocally in terms of their causal implications. It will be a challenge for the nearby future to operationalize the suggestions for diagnostic grading as formulated recently,[1] taking into consideration the important work that has already assessed this topic.[32,33,35,64,65,66,67]

CLINICAL PEARLS

- For a "possible" or "probable" SB diagnosis, self-report and clinical approaches are sufficient, whereas for a "definite" diagnosis, polysomnography, preferably with audio-video recordings, is required.
- SB should be differentiated from several other sleep-related conditions, among which are orofacial movement disorders with manifestations that persist in sleep.
- RBD, OSA, and GERD are important comorbid conditions in patients with SB with possible causal associations.

SUMMARY

SB is a repetitive jaw muscle activity characterized by clenching or grinding of the teeth and/or by bracing or thrusting of the mandible during sleep. For a diagnosis of "possible" or "probable" SB, self-report and clinical approaches constitute adequate confirmation; for a "definite" diagnosis, polysomnography, preferably with audio-video recordings, is required. As yet, the evidence is insufficient to support the use of ambulatory electromyographic devices as stand-alone tools for the diagnosis of SB, as tested against full polysomnographic recordings. This is even more critical in differentiation of SB from several other sleep-related conditions including orofacial movement disorders with manifestations that persist in sleep. RBD, obstructive sleep apnea, and GERD are important comorbid conditions in patients with SB with possible causal associations.

Selected Readings

Abe S, Gagnon JF, Montplaisir JY, et al. Sleep bruxism and oromandibular myoclonus in rapid eye movement sleep behavior disorder: a preliminary report. *Sleep Med* 2013;**14**:1024–30.

Blanchet PJ, Rompré PH, Lavigne GJ, et al. Oral dyskinesia: a clinical overview. *Int J Prosthodont* 2005;**18**:10–19.

Carra MC, Huynh N, Fleury B, Lavigne G. Overview on sleep bruxism for sleep medicine clinicians. *Sleep Med Clin* 2015;**10**(3):375–84.

Koyano K, Tsukiyama Y, Ichiki R, et al. Assessment of bruxism in the clinic. *J Oral Rehabil* 2008;**35**:495–508.

Lavigne GJ, Rompré PH, Montplaisir JY. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. *J Dent Res* 1996;**75**:546–52.

Lobbezoo F, Ahlberg J, Glaros AG, et al. Bruxism defined and graded: an international consensus. *J Oral Rehabil* 2013;**40**:2–4.

Maluly M, Andersen ML, Dal-Fabbro C, et al. Polysomnographic study of the prevalence of sleep bruxism in a population sample. *J Dent Res* 2013;**92**:97S–103S.

Manfredini D, Ahlberg J, Castroflorio T, et al. Diagnostic accuracy of portable instrumental devices to measure sleep bruxism. A systematic literature review of polysomnographic studies. *J Oral Rehabil* 2014;**41**:836–42.

Manfredini D, Guarda-Nardini L, Marchese-Ragona R, Lobbezoo F. Theories on possible temporal relationships between sleep bruxism and obstructive sleep apnea events. *Sleep Breath* 2015 Mar 22 [Epub ahead of print].

Miyawaki S, Tanimoto Y, Araki Y, et al. Association between nocturnal bruxism and gastroesophageal reflux. *Sleep* 2003;**26**:888–92.

Paesani DA, Lobbezoo F, Gelos C, et al. Correlation between self-reported and clinically based diagnoses of bruxism in temporomandibular disorders patients. *J Oral Rehabil* 2013;**40**:803–9.

Rompré PH, Daigle-Landry D, Guitard F, et al. Identification of a sleep bruxism subgroup with a higher risk of pain. *J Dent Res* 2007;**86**:837–42.

van der Zaag J, Lobbezoo F, Visscher CM, et al. Time-variant nature of sleep bruxism outcome variables using ambulatory polysomnography: implications for recognition and therapy evaluation. *J Oral Rehabil* 2008;**35**:577–84.

A complete reference list can be found online at ExpertConsult.com.

Chapter 146

Orofacial Pain and Temporomandibular Disorders in Relation to Sleep-Disordered Breathing and Sleep Bruxism

Gregory K. Essick; Karen G. Raphael; Anne E. Sanders; Gilles Lavigne

Chapter Highlights

- Among chronic orofacial pain conditions, temporomandibular disorders (TMDs) are the most common. The etiology of TMD is multifactorial. The role of sleep-disordered breathing (SDB) and sleep bruxism (SB) in the initiation and persistence of TMD pain is poorly understood but of considerable interest to dental clinicians who treat these patients.
- Evidence for SDB as a causal risk factor for TMD is limited at present to a few cross-sectional studies reporting higher prevalence of SDB or suspected SDB in patients with TMDs, and higher prevalence of TMDs among patients with SDB. One population-based prospective cohort study found that signs and symptoms of SDB predict incident TMD. Moreover, laboratory-based measures of sleep respiration indicate that respiratory effort–related arousals are more frequent in women with TMD than in control subjects. Prevalence of SDB also is greater in patients with fibromyalgia syndrome, a chronic pain condition commonly comorbid with TMD.
- The premise that SDB leads to SB, and in turn to TMD, is not supported by the available scientific evidence. Moreover, the strongest evidence refutes a relationship between SB and TMD. That patients with signs and symptoms of TMD often report grinding their teeth during sleep has been misinterpreted by clinicians as evidence of an association.
- This chapter describes a much more complex set of interrelationships that may explain an association between SDB and TMD. Proposed mechanisms include elevated background levels of activity in the masticatory muscles during sleep and wakefulness, intermittent hypoxia during sleep, altered autonomic nervous system function, oxidative stress, and inflammation.

Temporomandibular disorders (TMDs) are a heterogeneous family of musculoskeletal disorders that represents the most common chronic orofacial pain condition.[1-3] TMD is characterized by persistent pain in the temporomandibular joint, the periauricular region, and the muscles of the head and neck, and by painful chewing and impaired oral function. The *Diagnostic Criteria for Temporomandibular Disorders* (RDC/TMD) protocol, published by the International RDC/TMD Consortium, recently was updated; these critera are used to guide diagnosis of TMD for research purposes.[4,5] Although criteria for TMD have changed over the years, prevalence of TMDs has been estimated historically at between 5% and 12%, with higher rates among women (of reproductive age) than among men.[1,6-9] Estimates suggest that TMD results in almost 18,000,000 lost work days annually for every 100,000,000 working adults in America.[3] In view of the considerable cost to work productivity and to the health care system, much interest is directed at identifying risk factors associated with the onset and maintenance of TMD.[10]

One plausible risk factor for TMD to gain attention in recent years is sleep-disordered breathing (SDB). Some orofacial pain practices report that as many as 75% of patients diagnosed with TMD have clinical characteristics suggestive of SDB.[11] This coexistence of disorders has prompted many dental providers to believe that SDB contributes to sleep bruxism (SB), a disorder characterized by rhythmic grinding of the teeth during sleep. Furthermore, SB has been assumed to cause TMD in susceptible persons through microtrauma to the temporomandibular joint or masticatory muscles from hyperactivity during sleep.[12-15] Against this background, efforts to diagnose and treat SDB, as well as SB, in patients with TMD are assumed to reduce pain and correct dysfunction. However, no clinical trials have been conducted to determine whether treatment of SDB alters the natural course of TMD. For more information on sensory mechanisms, pain, and sleep interaction including also SB and SDB, readers are referred to Chapter 23; for SB etiology and mechanisms and SB diagnosis, they are referred to Chapters 144 and 145, respectively.

This chapter reviews the evidence supporting an association between SDB and TMDs, and between SDB and fibromyalgia syndrome (FMS), a related pain condition. Although

the literature supports a correlation with these pain conditions, efforts to establish causality have been minimal to date. Also addressed in separate sections are the association between SDB and SB and that between SB and TMD. As explained in these sections, however, SB is an unlikely causal risk factor for TMD, and more recent work calls into question the current understanding of the relationship between SDB and TMD. Presented in support of this argument is a summary of recent evidence showing that a sustained increase in background masticatory muscle activity during sleep, rather than SB, may more accurately explain the association between SDB and TMD, although this explanation remains speculative. Finally, alternative, hypothetical biologic mechanisms by which SDB and TMD may be associated are reviewed. Although of scientific interest, supporting evidence for these mechanisms also remains rudimentary at this time.

SLEEP-DISORDERED BREATHING AS A CAUSATIVE CANDIDATE FOR TEMPOROMANDIBULAR DISORDER

Remarkably few studies offer evidence for an association between SDB and TMD. In an early Brazilian study,[16] 87 patients (46 men) with a diagnosis of mild or moderate obstructive sleep apnea (OSA), as indicated by an apnea-hypopnea index (AHI) greater than 5 and less than 30 were evaluated for signs and symptoms of TMD according to the *Research Diagnostic Criteria for Temporomandibular Disorders* (RDC/TMD). Patients with mild OSA (AHI between 5 and 15) were additionally required to have symptoms consistent with OSA. An initial examination revealed that 45 (52%) patients presented with signs or symptoms of TMD for which the investigators gave a tentative diagnosis of TMD. Of the 32 (71%) who returned for the clinical assessment using the RDC, 75% presented with chronic pain associated with TMD, providing an estimated prevalence of TMD of 39% in patients with mild to moderate OSA—which was substantially greater than estimates published for the general adult population. Myofascial pain was the most common diagnosis. The study was limited by the lack of a control group, the relatively small sample size, and selection bias. For example, only patients who were candidates for oral appliance therapy for treatment of sleep apnea were enrolled, and a large number of subjects (29%) did not return to complete the physical RDC assessment for TMD.

Another observational study with an even smaller sample size and no control group was published the same year. Unlike in the earlier study, however, all 53 subjects met RDC/TMD criteria and were evaluated for experimental pain sensitivity and for sleep disorders with polysomnography (PSG).[17] Each subject completed a battery of commonly used pain and sleep questionnaires. Twenty-eight percent (50% of men and 23% of women in the study) met criteria for OSA, and 43% for insomnia signs and symptoms—greatly exceeding prevalence estimates for the general adult population. Most of the subjects with OSA (73%) were classified as having mild disease. For 13 subjects with OSA, the respiratory disturbance index averaged 13.7 events/hour.

Building on these two small studies, a large population-based epidemiologic family of TMD studies named OPPERA (Orofacial Pain: Prospective Evaluation and Risk Assessment) compared the strength of association between signs and symptoms of OSA and chronic TMD in a case-control study (n = 182 cases; n = 1534 control subjects). In a separate OPPERA study—this time a prospective cohort study—these same OSA signs and symptoms were measured at baseline in a TMD-free cohort of subjects (n = 2604) who were followed for a median 2.8 years to investigate the development of incident TMD.[18] Examiners in both OPPERA studies used RDC/TMD diagnostic criteria that required (1) self-report of pain in masticatory structures on 5 days or more per month and (2) examiner-confirmed arthralgia during jaw maneuver or digital palpation of the temporomandibular joint; and/or myalgia during jaw maneuver or digital palpation in at least three of eight muscle groups based on bilateral assessment of temporalis, masseter, lateral pterygoid, and submandibular muscles. For classification as having chronic TMD, subjects had to meet the foregoing RDC/TMD criteria and be symptomatic for at least 6 months. Subjects with at least two of four OSA signs and symptoms were classified as being at high risk for OSA. These four OSA signs and symptoms were taken from a screening questionnaire for OSA that is used to screen patients requiring anesthesia for surgery for risk factors.[19] The four OSA signs and symptoms were snoring, daytime fatigue/sleepiness, observed apnea during sleep, and hypertension. In addition, any OPPERA subject with a physician diagnosis of OSA was classified as being at high risk for OSA.

In OPPERA's chronic TMD case-control study, odds of TMD were elevated more than threefold (odds ratio, 3.6; 95% confidence interval, 2.0 to 6.5) in those at high risk for OSA, independently of demographic, autonomic, and behavioral characteristics. In the prospective cohort, OPPERA subjects at risk for OSA had 1.7 times the incidence of TMD over the median 2.8-year follow-up period, independently of demographic, autonomic, and behavioral characteristics. The association remained statistically significant with further adjustment for subjective sleep quality.

Both OPPERA studies were limited by the absence of PSG data. It is possible that the comparatively young age of OPPERA subjects (80% were younger than 35 years of age) underestimated the strength of the associations in the general population. Furthermore, a majority of OPPERA subjects were female (60%) and nonobese (80%)—the opposite of the characteristics in most SDB studies.[20] It also is possible that many participants classified by questionnaire as having OSA signs or symptoms in fact had milder forms of SDB such as upper airway resistance syndrome (UAR)[21] or respiratory effort–related arousal (RERA).[22]

That TMD may be associated with milder forms of SDB was shown in a recent PSG study of female patients with TMD (n = 124) and demographically matched control subjects (n = 46).[23] As in the foregoing studies, the patients met RDC/TMD criteria. Each participant underwent two PSG studies. Only data from the second night were used, to avoid the so-called first night effect associated with sleeping in a laboratory, a novel environment. Multiple pain ratings were collected for the day before and the day after the PSG studies. The patients with TMD had lighter and more disrupted sleep, a higher percentage of stage N1 sleep, and more sleep instability as seen by a greater number of awakenings and stage shifts toward N1. Furthermore, arousals associated with all respiratory events (apneas, hypopneas, and RERAs) were almost twice as frequent in the patients as in control subjects;

nevertheless, the mean AHI was similar for both groups at less than 4 (events/hour). Accordingly, the respiratory disturbance index showed a trend ($P = .06$) toward being higher in the TMD group (8.1/hour) than in the control group (5/hour) owing to a higher RERA frequency in the former. The increased frequency of RERAs was independent of the measures of usual sleep fragmentation outcome variables such as total number of arousals and percentage of N1 sleep but was directly related to the participant's subjective ratings of myofascial pain during the PSG night. The study may have underestimated the severity of the SDB in patients with TMD because owing to other study aims, approximately 10% of potential TMD group subjects screened were excluded from participation by the presence of severe insomnia or severe OSA or by having reported use of CPAP.

Not all studies investigating the relationship between SDB and TMD find a significant association. For example, an OSA population evaluated for oral appliance therapy had a prevalence of self-reported TMJ symptoms similar to that in historical control subjects.[24] Another study found no association between TMD and SDB as determined by questionnaire.[25] Two studies failed to find differences in respiratory measures from PSG data in RDC/TMD case and control participants.[14,26] Although validated methods for assessing TMD and SDB were employed, as in several studies just described, the sample sizes of 20 and 30 cases were small. Most recently, a study that sought to identify clinical comorbid conditions in TMD failed to detect a significant difference in the proportion reporting sleep apnea in a cohort of persons claiming to be affected by TMDs and musculoskeletal disorders and in a control group.[27] In general, these studies were limited by one or more factors, including the lack of use of recognized RDC assessments in identifying TMD, the use of self-report to identify TMD and/or OSA in the absence of PSG, and rather small samples of participants, with a large variability in methodology and variance in findings.

All considered, strength of the evidence for an association between SDB and TMD is moderate. Studies to date suggest a real or possibly elevated prevalence of SDB associated with TMD and an increased prevalence of TMD associated with SDB. One large prospective cohort study found that signs and symptoms of SDB preceded the onset of TMD in initially TMD-free subjects. Moreover, laboratory-based measures of sleep respiration indicate that RERAs are more frequent in women with TMD than in carefully demographically matched control subjects (Table 146-1). As yet, the

Table 146-1 Observations Supporting an Association between Sleep-Disordered Beathing (SDB) and Temporomandibular Disorders (TMDs) or Fibromyalgia Syndrome (FMS)

Observation	Supporting Study(ies)	Comments	Nonsupporting Study(ies)	Comments
Increased prevalence of TMD in patients diagnosed with OSA	Cunali et al., 2009[16]	Observational study; no control group; small sample size; subject selection bias	Petit et al., 2002[24]	Studies limited by one or more of the following: use of self-report to identify TMD and/or OSA, small samples of participants; large variability in methodology and variance in findings
Increased prevalence of OSA in patients diagnosed with TMD	Smith et al., 2009[17]	Observational study; no control group; small sample size	Collesano et al., 2004[25] Hoffmann et al., 2011[27]	
Increased odds of signs and symptoms of OSA in patients with chronic TMD	Sanders et al., 2013[18]	Large multicenter case-control study; no objective sleep data		
Initially TMD-free subjects with signs and symptoms of OSA more likely to develop first-onset TMD	Sanders et al., 2013[18]	Large multicenter prospective cohort study; no objective sleep data		
Increased frequency of RERAs in patients with TMD compared with control subjects	Dubrovsky et al., 2014[23]	Large sample case-control study	Camparis and Siqueira, 2006[14] Rossetti et al., 2008[26]	
Increased prevalence or severity of SDB in patients diagnosed with FMS	Gold et al., 2004[30]; May, 1993; Shah et al., 2006[33]; Prados et al., 2013[34]	Use of historical or no control subjects; patients selected from specialty clinics or community-based practice	Chervin et al., 2009[35]	Small case-control study; academic setting
Treatment of SDB in patients diagnosed with FMS helps relieve pain and dysfunction	Gold et al., 2004[30]; Sepici et al., 2007[31]; May et al., 1993[32]	No control group, anecdotal observations		

OSA, Obstructive sleep apnea; RERAs, respiratory effort–related arousals.

causal mechanisms underlying this relationship have not been elucidated.

SLEEP-DISORDERED BREATHING AS A CAUSATIVE CANDIDATE FOR FIBROMYALGIA SYNDROME

Further support for an association between SDB and TMD is provided by studies reporting an association between SDB and FMS, a persistent pain condition that is highly comorbid with TMD, exhibits similar characteristics, and is postulated to have similar causes.[28] People with FMS, as with TMD, exhibit increased pain sensitivity, but with FMS the clinical pain is widespread. It is estimated that 24% to 78% of patients with FMS meet criteria for TMD and that 9% to 18% of those with TMD meet criteria for FMS.[28,29]

A number of these studies report an increased prevalence of SDB in patients with FMS. For example, a study evaluated 28 women diagnosed with FMS and 11 women matched for age and obesity diagnosed with UAR.[30] Twenty-seven women in the FMS group exhibited SDB as determined during full-night PSG or during quantitative monitoring of inspiratory airflow and effort upon application of pressures between atmospheric and therapeutic CPAP. Moreover, anthropometric, PSG, and airflow-dynamics data, which included the pharyngeal critical closing pressure, did not differ between the two groups of women, suggesting similarity in the severity of the SDB. After CPAP treatment in 14 of the women with FMS, significant improvements were observed with respect to self-reported fatigue, pain, sleep problems, gastrointestinal symptoms, disability, and rheumatology distress. Although no controlled clinical trials have been conducted to date, anecdotal observations regarding symptomatic improvement in FMS on treatment of SDB with CPAP have been reported in other studies.[31,32] Also, other observational studies have reported a high prevalence of OSA: 83%[33] and 31.8%[34] in females and 44%[32] and 61%[34] in males with fibromyalgia. In a more recent study, however, a comparison of carefully selected female FMS case and control subjects failed to identify differences in standard polysomnographic measures or in measures of daytime sleepiness that would suggest a higher prevalence or severity of SDB in FMS.[35] See also Chapter 131 on FMS for more information on sleep interaction.

SLEEP-DISORDERED BREATHING AS A CAUSATIVE CANDIDATE FOR SLEEP BRUXISM

Bruxism refers to either clenching or grinding of the teeth. It can occur during wakefulness and/or during sleep (more information is available in Chapters 144 and 145). Here we debate the role of SDB in SB; in a later section of the chapter, we address the role of SB in TMD. It is believed and/or proposed by some clinicians that SDB plays a role in SB. Such thinking is well established, particularly for the pediatric population: An association between parent-reported SB (sleep tooth grinding) and SDB has been reported in multiple studies of children. For example, a cross-sectional telephone questionnaire survey of the parents of 3047 children in China revealed a 20.5% prevalence of sleep tooth grinding that was closely associated with habitual snoring.[36] Children with reported sleep tooth grinding were 56% more likely to exhibit habitual snoring than those without grinding. Another survey of parents of 1164 children in Turkey arrived at similar conclusions.[37] Sleep tooth grinding was reported by the parents of 21.6% of children, and habitual snorers were two times more likely than nonsnorers to grind the teeth. In a smaller study of children evaluated by PSG, parent-reported SB occurred in 31.1% of children diagnosed with obstructive sleep disorder, and all children with severe OSA exhibited SB based on parent report.[38] A survey of 69 children, collecting data before and 3 months after adenotonsillectomy for SDB, found a reduction in parental report of the child's bruxism or tooth grinding (from 45.6% to 11.8%) and a complete elimination of SDB symptoms.[39]

An association between self-reports of bruxism/tooth grinding and SDB has been cited for adults as well. A cross-sectional telephone survey of 13,057 respondents in Europe aged 15 years or older found that reported OSA was the highest risk factor (odds ratio, 1.8) for tooth grinding among all sleep symptoms and disorders queried by the investigators.[40] Another cross-sectional questionnaire study done on 1930 residents in Japan found that respondents who reported snoring were 2.6 times more likely to report that they grind their teeth during sleep than those who did not report snoring.[41] These findings, as well as those from studies described in the preceding paragraph, however, have to be interpreted with caution, because they are based solely on the reports of patients' sleep partners or parents.

SLEEP RECORDINGS OF SLEEP BRUXISM IN SLEEP-DISORDERED BREATHING

In contrast with studies that have used surveys and questionnaires to assess the SDB-SB relationship, most studies using overnight sleep studies have found no increase in prevalence of SB among persons with SDB. For example, a study of 12 adults with SDB and 12 age- and gender-matched control subjects found no difference in the frequency of "bruxing" episodes between the two groups, although the frequency of arousals was greater in the SDB group.[42] Bruxing episodes, defined as bursts of masseter EMG activity measured as greater than 40% of maximum during awake clenching, were mostly associated with respiratory events in the SDB group. By contrast, in the control group, bruxing episodes were associated with arousals from limb movements and isolated events. Ten years later, a study of 21 adults reported that 54% of adults with mild and 40% with moderate OSA met empiric criteria for SB based on the presence of two or more of the following: more than 2.5 rhythmic jaw-movement episodes per hour as determined by EMG during polysomnography, clinical observation (attrition, masticatory muscle fatigue, or temporomandibular joint discomfort) or subjective report of tooth grinding or clenching on one or more nights per week.[43] Because the number of rhythmic jaw movement episodes per hour was greater in patients with mild than in those with moderate OSA and the episodes were rarely associated with respiratory arousals, the study authors attributed their presence to the disturbed sleep characteristic of OSA rather than difficulty in breathing.

Similar to the lack of an increased prevalence of SB in adults with SDB, an increased prevalence of SDB has not been observed in adults with SB, as, for example, reported for a small general population sample.[44,45] A recent

comparison of PSG parameters for larger groups of 56 adults with and 569 adults without SB, based on PSG criteria and self-report, also failed to find any group differences in the AHI or nocturnal oxygen saturation.[46] In contrast with adults, a PSG study in children ages 2 to 16 years of age revealed a greater apnea index, AHI, and rapid eye movement (REM) sleep AHI in 70 subjects who met PSG criteria for SB, compared with 49 who did not.[47] An SB episode was defined as occurrence of three or more rhythmic contractions of the temporalis muscle lasting more than 3 but less than 15 seconds.

Collectively, the overwhelming majority of the self-report– or parent-report–based studies support that SB is elevated among both adults and children when SDB coexists as determined by the presence of symptoms suggestive of SDB or by PSG evidence of SDB. However, an increase in prevalence or in severity of SB in persons with SBD, at least in adults, has not been confirmed by PSG data. This lack of evidence brings into question the basis for self-report of SB, as discussed in Chapters 144 and 145.

SLEEP RECORDINGS OF INCREASED MASTICATORY MUSCLE ACTIVITY IN SLEEP-DISORDERED BREATHING

Bearing on a possible solution to this discrepancy between patient reports and objective laboratory findings, some groups of investigators have observed increased masticatory muscle activity in patients with SDB during sleep that is thought to result in forward posturing and stabilization, or in elevation of the mandible to help maintain airway patency. It might be that these muscle activities, even in the absence of tooth contact, are interpreted by adults (or parents of children) with SDB as tooth grinding or clenching during sleep. The earliest studies, conducted some 25 years ago, identified increased EMG activity of the masseter muscles during the inspiratory phase of sleep respiration in patients with OSA, but not in individuals with normal sleep respiration.[48,49] The activity in the masseter was accompanied by similarly timed phasic activity in the submental (genioglossus, geniohyoid, mylohyoid, and digastric) muscles, leading to periodic variations in the jaw gape from end inspiration to end expiration and variations between obstructed and nonobstructive breaths. However, the jaw was more open in patients with OSA than in persons with normal sleep respiration. The study authors surmised that activity in the masseter muscle stabilized the jaw, enabling the submental muscles to more effectively elevate and forwardly posture the hyoid bone, thereby improving airway patency overall in patients with OSA but promoting mouth breathing at the termination of apnea events.

The suggestion that increased activity in the masseter muscles can improve airway patency in patients with SDB received remarkably little attention until recent years.[11,50] In studies reported by one group of investigators, esophageal pressure and multiple channels of EMG data from the masticatory muscles were recorded in addition to the standard PSG channels in patients with mild OSA and UARS. The clinician investigators demonstrated that patients with UARS exhibit frequent periods of tonic contraction of the jaw-closing muscles (including the masseter), upon which esophageal pressure drops appreciably, indicating reductions in the respiratory effort needed to breathe. However, no objective monitoring of jaw position or of tooth contact was performed. Nor was it determined whether the patients met diagnostic criteria for bruxism based on a combination of self-report, clinical, and PSG criteria.

Other investigators have noted that contraction of the masseter and temporalis muscles occurs as part of the arousal response on a third or more of respiratory events in patients with SDB.[51,52] The contractions are dependent on the duration and intensity of the arousals and occur in patients who have no other clinical signs of SB or PSG evidence of SB—that is, of the typical rhythmic masticatory muscle activity (RMMA). It is suggested that these elevations in masticatory muscle activity represent a nonspecific, general motor manifestation of the arousal during sleep and also may contribute to the restoration of a patent upper airway.[51]

SLEEP BRUXISM AS A CAUSATIVE CANDIDATE FOR TEMPOROMANDIBULAR DISORDER

The beliefs that bruxism, sleep and/or awake, plays a role in the onset or maintenance of TMD is well-established clinical dogma (see also Chapter 145). Surveys confirm that dentists with expertise in management of TMD in both the United States and abroad believe that "oral parafunctional habits" are important in the onset of TMD.[53-57] The term "oral parafunction" typically is defined as any oral nonpurposeful activity or behavior involving the masticatory system. The oral parafunction most often linked with TMD is bruxism.[58] Many discussions, reviews, and research studies fail to clearly differentiate whether the bruxism activities are believed to occur during sleep or when awake. Nevertheless, bruxism during sleep is mainly characterized by tooth grinding, and bruxism during wake periods may be characterized mainly by clenching.

The link between SB and TMD has evolved from a long historical tradition, dating back more than half a century. In shifting from earlier mechanistic theories of TMD etiology,[59] the psychophysiologic model of masticatory muscle pain etiology was first proposed by Laszlo Schwartz in the 1950s[12,60] and expanded later by Daniel Laskin.[15] It focused on the muscle rather than the joint as a primary source of pain and integrated psychosocial factors as a cause. The foundation of the psychophysiologic model of TMD is that psychosocial stress initiates a distress response or "tension" in the affected person, causing dysfunctional oral habits such as tooth grinding and clenching. These oral habits are presumed to promote muscle contraction and hyperactivity and subsequent facial pain.

SELF-REPORTS OF SLEEP BRUXISM IN PATIENTS WITH TEMPOROMANDIBULAR DISORDERS

Turning toward research-based evidence, the approach taken in a recent review (described in Chapter 145)[61] is useful in that it separates conclusions about the bruxism-TMD relationship according to method of assessment. Twenty-one studies were identified over a 10-year period, which relied on participant self-report or questionnaire to identify bruxers, sometimes using only a single item. This issue is problematic because, as noted earlier, sleep and awake behaviors may have different consequences for development or maintenance of TMD. An overwhelming majority of the self-report–based

studies reviewed support the common belief that the frequency of some form of bruxism is elevated among patients with TMD.

Of note, in one study, two thirds of subjects with pain reported that it was their dentist who told them about their SB.[62] As identified years earlier, a central problem with the reliance on self-report is that it is tautologic.[63] If the patient's clinician "believes" that bruxism causes TMD pain and therefore tells the patient that he or she is presenting with or suffering from bruxism, studies relying on a patient's self-report will support the prevailing bruxism theory. Updating the self-report or questionnaire literature of SB since 2008, one study found that "dentist-verified" grinding (presumably during sleep) occurred at elevated rates among patients with unexplained facial pain, including those with TMD, compared with those with pain of dental origin.[64] A multisite study, however, reported inconsistent findings between the study sites in the relation between self-reported SB and specific subtypes of TMD or no TMD, concluding that patients may be unable to differentiate between sleep and awake bruxism and that the reliability of self-reported bruxism is questionable in general.[65] The variability in conclusions from even more recent studies undoubtedly reflects the problematic, low-validity assessment of SB based on questionnaire, even when a questionnaire attempts to differentiate between bruxism subtypes.

SLEEP RECORDINGS OF SLEEP BRUXISM IN PATIENTS WITH TEMPOROMANDIBULAR DISORDER

In contrast with the large number of studies examining the SB-TMD relationship using less-than-optimal methods, PSG-based studies are rare. The 10-year review cited only four studies, two of which appear to represent partial overlap in sampling.[61] One study compared groups diagnosed with SB with and without pain,[66] thus begging the question of whether SB is elevated in patients with TMD. Another study reported that subjects without TMD with relatively low levels of SB were more likely to report transient morning masticatory muscle pain than those with no or high levels of SB,[67] as has been reported also more recently.[68] The results, however, did not address the clinical syndrome of TMD or the myofascial pain of TMD, which tends to be worse in the late afternoon or evening.[69]

Two reports of a third study (one being a pilot report for the larger study) were distinguished by standardized definitions of SB events and inclusion of a control group without TMD pain.[26,70] The report with the larger sample[26] but not the smaller sample[70] found significantly higher rates of individuals meeting validated research criteria for SB[71] among participants meeting RDC criteria for TMD than among healthy control subjects. Concern about the study's recruitment methods must be raised, however, because one third of control subjects met research criteria for SB, as did nearly two thirds of subjects with TMD. By contrast, a large population study from Brazil estimated the prevalence of SB as 5.5% using a combination of questionnaires and PSG, or as 7.4% on the basis of PSG alone.[46] Other experts estimate self-reported tooth grinding prevalence as high as 8%.[72] This suggests that the reported[26] finding of unusually high rates of SB in both groups may stem from either use of nonstandard scoring methods for RMMA episodes or oversampling of persons engaging in SB in both case and control samples.

In a study that explicitly sampled patients with TMD and control subjects on the basis of presence or absence of meeting RDC criteria for myofascial TMD, rather than presence or absence of beliefs about engaging in SB, laboratory PSG evaluations found similarly low proportions in both groups of participants meeting PSG-based research criteria for SB (i.e., 10% and 11%, respectively).[73] Of note, as described earlier, the rate of self-reported SB was markedly higher among patients with TMD than among control subjects in this same sample.

Thus, as summarized in Lobbezoo and collaborators in Chapter 145 and in the review by Manfredini and Lobbezoo,[61] the highest level of association between SB and TMD is based on the lowest quality of evidence; on moving to studies with more specific and accurate methods of assessing SB, the association between SB and TMD begins to evaporate. The most parsimonious explanation at present is that no relationship exists between SB and TMD, at least in established TMD cases. The question of whether SB is involved in the initial onset of symptoms cannot be answered by the existing literature, nor is it likely that future research will provide an answer, in view of the low rates of SB in the general population.[46]

Despite this negative conclusion, it may seem paradoxic that a subset of patients with TMD note marked pain on waking. Both clinicians and patients may assume that this is indicative of SB activity during the night. In fact, owing to the lack of valid evidence, the second edition of the *International Classification of Sleep Disorders* (ICSD2) diagnostic and coding manual considered a patient report of "jaw muscle discomfort, fatigue and pain or jaw lock on waking" as one of the two criteria, along with "the presence of regular or frequent tooth grinding sounds occurring during sleep," that must be present to make a diagnosis of SB.[74] In the ICSD3, this was revised; again, such guidelines are intended to help sleep medicine professionals recognize and distinguish SB from other sleep disorders. Overnight attended PSG is recommended to investigate suspected SDB or sleep-related movement disorders.

Indeed, self-reported SB has been significantly associated with pain or jaw tension on waking in the morning in at least one report,[75] but PSG-confirmed RMMA episode frequency is not found to relate to morning muscle symptoms.[76] Also troubling is a finding from the large OPPERA prospective cohort study that a 21-item oral behaviors checklist that included several items on self-reported sleep and awake bruxism ranked fifth in importance, among more than 200 variables examined, for predicting incident TMD.[77] Thus factors related to self-reported oral behaviors seem to prospectively predict TMD onset or persistence, even if PSG-confirmed behaviors do not. Does a patient's being told by a dentist or hearing from the general public that bruxism "causes" TMD explain such reports? Although the basis for the contradiction between self-reports and PSG-confirmed SB is unclear, one must be circumspect about conclusions of future studies that rely on anything but PSG-confirmed SB.

A possible resolution to the contradiction between self-reports versus PSG-confirmed SB may lie in nighttime

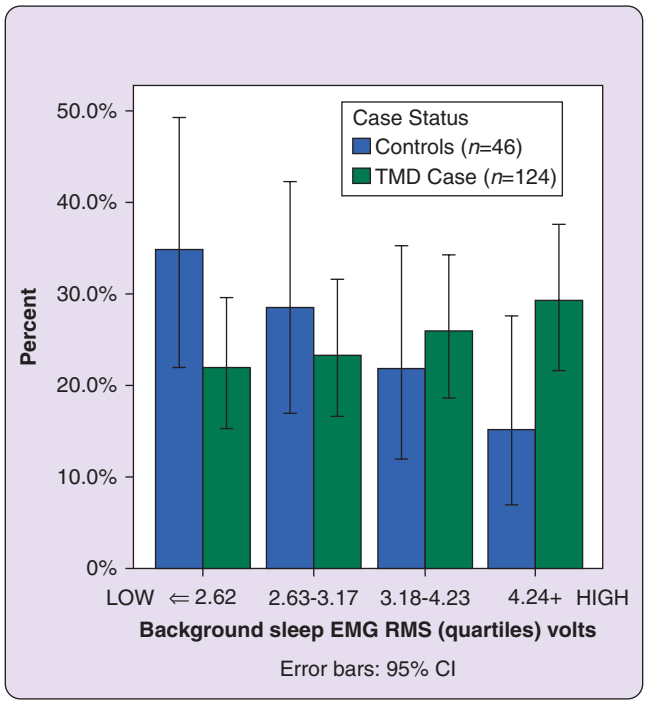


Figure 146-1 Distribution of quartiled background sleep EMG activity (RMS volts) among patients with temporomandibular disorders (TMDs) (n = 124) and control subjects (n = 46). Recent studies are dispelling the widely held view that sleep bruxism (SB) is an important risk factor for TMD development or persistence. It is now shown that background masticatory muscle activity during sleep is elevated in patients with TMDs. The underlying mediators of this relationship are not yet understood. RMS, Root mean square. (Data from Raphael KG, Janal MN, Sirois DA, et al. Masticatory muscle sleep background electromyographic activity is elevated in myofascial temporomandibular disorder patients. *J Oral Rehabil* 2013;40:883–91.)

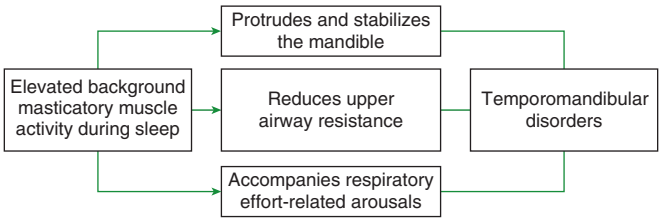


Figure 146-2 Hypothetical contributions of elevated masticatory muscle activity on the electromyogram (EMG) associated with sleep-disordered breathing (SDB) to temporomandibular disorders (TMDs). Some studies have reported increased EMG activity in the masticatory muscles of patients with SDB, and this increased activity could contribute to the elevation observed in patients with TMDs. Studies have suggested that the increased EMG activity in the masticatory muscles of patients with SDB is associated with mechanisms that maintain or restore airway patency or reflect a nonspecific increase associated with respiratory related arousals. However, further research is required to confirm a role for SDB in the increased EMG activity observed in TMD.

activity that is associated with jaw pain on waking, that is, masticatory muscle hypertonicity or elevated background sleep EMG activity occurring during nonbruxism periods. Recent findings indicate that patients with TMD have elevated levels of background sleep EMG activity compared with non-TMD control subjects[78] (Figure 146-1). Moreover, among the patients with TMD, background sleep EMG was positively associated with pain intensity ratings (on a 0 to 10 numerical scale) on morning waking, whereas the frequency of SB event–related EMG was negatively associated with the ratings, as has been observed in other investigations.[67,79] Nevertheless, the elevation in background sleep EMG activity found in patients with TMD persisted after institution of controls for numerous sleep parameters on which the groups did differ (e.g., RERAs, stage N1 shifts).

These observations indicate that patients who have elevated pain in the early morning may falsely attribute their pain to nighttime jaw muscle activity involving SB. This explanation, however, is only partially correct: Although jaw muscle activity may aggravate pain, low-level elevations of activity occurring outside of SB periods are more likely to be responsible for the morning pain aggravation. Of additional note, the increase in root mean square (RMS) EMG activity in the TMD group versus the control group occurring during background sleep EMG (outside of the period with RMMA) was approximately equal to the increase potentially attributable to regular tooth-to-tooth contact (within RMMA). This type of nonfunctional tooth contact during the awake state has been found to occur more often in individuals with TMD than in control subjects.[80-83] It is consistent with a body of experimental studies showing that low-intensity clenching (approximately 10% of maximum bite force) can cause pain and fatigue in the masticatory muscles.[84] Obviously, the clinical significance of and mechanism underlying this finding require further analysis.

Thus, on reviewing the evidence on the relation between SB and TMD pain, the best-quality studies do not support a convincing relationship. Rather, background masticatory muscle activity during sleep, outside of RMMA-SB periods, which is similar to low-intensity clenching or tooth-to-tooth contact during the day, is likely to be of greater pathogenic significance in the persistence of TMD pain. Furthermore, the understanding of any relationship between SDB and TMD cannot be simplified to a simple and single association with SB that is too often determined in clinical milieu. Given that masticatory muscle EMG activity may be elevated during sleep in persons with SDB even in the absence of clinical or EMG evidence of SB,[11,48-52] as described earlier in this chapter, we entertain the possibility that this elevated activity is of pathogenic significance in the persistence of TMD pain (Figure 146-2). This hypothesis is consistent with the stress-hyperactivity model of Ohrbach and McCall[85]:

> It is not the magnitude per se of the muscle activity in pain patients that is responsible for the nociception that leads to pain complaints. Rather, it is a persistent pattern of muscle reactivity, in the form of holding muscles in a particular shortened position, such as a 'braced jaw. ... (page 61).[85]

However, the prevalence of increased background masticatory muscle EMG activity during sleep in patients with SDB is unknown as well as whether patients with increased background EMG activity are more likely to develop or maintain TMD pain than patients who exhibit no increase.

OTHER MECHANISMS LINKING SLEEP-DISORDERED BREATHING TO TEMPOROMANDIBULAR DISORDERS

In addition to increased background masticatory muscle activity, other aspects of SDB may contribute to TMD. One

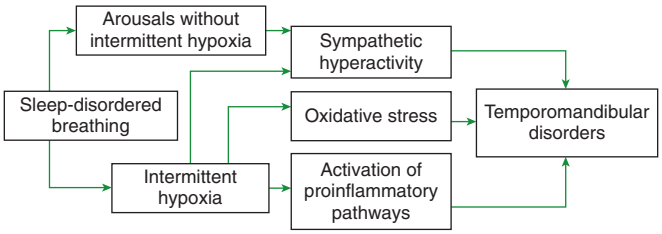


Figure 146-3 Hypothesized biologic pathways leading from sleep-disordered breathing to temporomandibular disorders. One pathway is mediated by arousal and sympathetic hyperactivity in the absence of intermittent hypoxia. A second pathway depicts an effect of intermittent hypoxia stimulating sympathetic hyperactivity, oxidative stress, and activation of proinflammatory pathways. At present, the evidence remains rudimentary and incomplete. Further studies are required to confirm these mechanisms.

innovative suggestion is that chronic intermittent hypoxia mediates the relationship between SDB and TMD. Intermittent hypoxia refers to the brief cyclic bursts of deoxygenation, arousal, and reoxygenation that uniquely characterize OSA or other forms of SDB. Discussed next is how three consequences of intermittent hypoxia—(1) sympathetic hyperactivity, (2) oxidative stress, and (3) activation of proinflammatory pathways—may promote TMD. For simplicity, Figure 146-3 depicts these relationships as unidirectional, whereas in reality some are bidirectional. As yet, these pathways are largely unexplored, and the supporting evidence remains incomplete. For complementary information on sleep related to the cardiac autonomic nervous system, breathing and hypoxia, host defense and immunity, and the endocrine stress response, the reader is referred to Chapters 14 to 17, 19 and 20, respectively, plus Chapter 133 for pain and sleep.

Intermittent Hypoxia, Sympathetic Hyperactivity, and Pain

Intermittent hypoxia increases sympathetic nervous system (SNS) activity,[86] and this response is measurable indirectly by heart rate variability (HRV). Decreased HRV typically is associated with morbidity, including pain disorders (see Chapters 23 and 133 for more information). In FMS, a pain condition characterized by widespread tender spots, decreased pain threshold, fatigue, and unrefreshing sleep (see also Chapter 131 on fibromyalgia in this volume), affected persons have lower HRV than control subjects both during wakefulness and during sleep.[35,87] Sympathetic hyperactivation during wakefulness also is associated with tension-type headache, migraine, sleep-related migraine,[88,89] irritable bowel syndrome,[90] and TMD.[91]

In the OPPERA case-control study of chronic TMD (described earlier), HRV in the TMD group was diminished at rest and in response to orthostatic and psychological challenges.[91] TMD cases also had a higher heart rate, another marker of sympathetic hyperactivity, in response to these orthostatic and psychological activities. The fact that cases showed a decrease in low-frequency band values at rest and during the orthostatic and Stroop activities provides further support of possible SNS dysfunction. Taken together, these autonomic measures imply a shift away from parasympathetic tone toward greater sympathetic tone in TMD cases relative to control data.[91]

The findings build on polysomnographic evidence[92] that subjects with myofascial TMD had a lower RMS successive difference of R-R intervals and lower HRV during sleep than pain-free control subjects. In addition, participants with TMD had lower high-frequency spectral power than control subjects, and the ratio of the low frequency to the high frequency was higher in subjects with TMD, suggesting a sympathetic cardiac dominance. Evidence of diminished nocturnal HRV was independent of sleep continuity and architecture, age, and psychological distress and was consistent with an SNS imbalance characterized as sympathetic hyperactivity.[92]

Hence, although separate lines of evidence indicate that intermittent hypoxia increases SNS activity and that sympathetic hyperactivity is associated with pain conditions including TMD, single studies have yet to combine these findings to show that sympathetic hyperactivity secondary to hypoxia from SDB increases risk of TMD onset or persistence.

Intermittent Hypoxia, Oxidative Stress, and Pain

Oxidative stress is a relatively less-studied mechanism through which intermittent hypoxia may contribute to TMD. Exposure to repeated episodes of hypoxia and normoxia stimulates overproduction of reactive oxygen species that overwhelms the protective capacity of antioxidant mechanisms, giving rise to oxidative stress.[93]

Elevated oxidative load is observed in several pain disorders. Patients with FMS have elevated oxidative stress, greater free radical damage, and lower total antioxidant capacity compared with control subjects.[94-97] Likewise, migraineurs have elevated oxidative stress[98,99] and lower levels of the antioxidant enzyme superoxide dismutase[100] than in control subjects. TMD cases with myofascial pain syndrome have lower total antioxidant capacity, higher total oxidative stress, and higher oxidative stress index values than in healthy control subjects, indicative of an imbalance in oxidative metabolism in the pain group.[101] In a clinical study that compared oxidative stress profiles in patients with TMD and control subjects,[102] the patient group showed higher levels of malondialdehyde, indicative of greater oxidative load, and higher levels of 8-hydroxydeoxyguanosine, indicative of greater oxidative damage. Furthermore, the patients with TMD had lower levels of total antioxidant status, consistent with a reduced antioxidant defense capacity.[102]

The evidence that intermittent hypoxia leads to oxidative stress is well established, whereas evidence linking oxidative stress to TMD and related pain disorders is still developing. Although the developing literature is small, it is noteworthy that studies show directional consistency in finding higher oxidative load in pain disorders. As is the status with SNS hyperactivity, single studies with strong designs are required to demonstrate clear pathways linking sleep disorders to oxidative stress and the development or persistence of pain.

Intermittent Hypoxia, Activation of Proinflammatory Pathways, and Pain

Intermittent hypoxia may influence pain through activation of the inflammatory process (see also Chapter 19, on sleep and host defense). A systematic review[103] of pooled data from 51 studies of 2952 OSA cases and 2784 control subjects found that inflammatory markers of C-reactive protein, tumor necrosis factor alpha, interleukins IL-6 and IL-8, intercellular adhesion molecule, vascular cell adhesion molecule, and selectins were significantly elevated in OSA cases compared with

the control group, with a dose-response relationship of higher levels of inflammation with greater OSA severity.[103]

Several animal and human studies have observed elevated secretion of inflammatory proteins in temporomandibular joint synovial fluid, secondary to sleep disturbance. For example, sleep deprivation in rodents induces activation of nuclear factor kappa B, which in turn leads to increased expression of several inflammatory proteins.[104] Although such reports confirm local inflammation in the temporomandibular joint, they fall short of demonstrating that TMD may be associated with systemic inflammation. Addressing this limitation is one such case-control study of 344 women, in which investigators compared circulating plasma levels of 22 cytokines and activity of 48 transcription factors in the women with TMD and in pain-free control women.[105] Levels of the proinflammatory monocyte chemotactic protein-1 and the antiinflammatory interleukin-1 receptor antagonist (IL-1ra) were significantly elevated in the TMD group.

If inflammation mediates a relationship between SDB and TMD, the finding is clinically important because inflammation can be lowered by treating SDB. A study of patients with OSA[106] showed that nasal CPAP decreased circulating intercellular adhesion molecule-1 and IL-8 levels, suggesting that the putative effects of intermittent hypoxia on inflammation can be reversed by treating OSA. Similarly, treatment of OSA with oral appliance therapy appears to lower circulating levels of inflammatory markers.[107] A logical conclusion of these studies is that lowering inflammation may be therapeutic or protective against pain.

Independent Effect of Intermittent Hypoxia

Intermittent hypoxia may influence pain independently of sympathetic hyperactivity, oxidative stress, and inflammation. Analysis of PSG oxygenation data for 634 Cleveland Family Study participants with SDB showed that a decrease in nadir oxyhemoglobin desaturation was associated with morning headache, headache disrupting sleep, and chest pain in bed, independently of proinflammatory protein levels and sleep fragmentation.[108] This finding is of particular interest and needs to be confirmed. (Additional information on sleep and pain interaction is available in Chapter 133.)

Respiratory Effort–Related Arousals, Sympathetic Tone, and Pain

Arousal from respiratory events may contribute to pain mechanisms independently of any intermittent hypoxia. In fact, even mild levels of airway resistance are sufficient to initiate RERAs in some susceptible people, and these have been associated with TMD and fibromyalgia (see earlier). One interesting suggestion is that apneas-hypopneas and RERAs evoke different responses to autonomic nervous system activity.[109] Although the arousals of apneas-hypopneas increase the ratio between low and high frequencies (in favor of a sympathetic tone dominance), one study found that abnormal inspiratory effort associated with RERAs lead to a decrease in the high-frequency component.[109] This finding is consistent with a stronger sympathetic influence in OSA and a stronger parasympathetic influence in UAR. Again, RERA and UAR are not fully comparable, but both are close proxies for breathing events that disturb sleep continuity.

It is becoming clear that RERAs are associated with pain conditions that overlap with TMD, including headache,[110] fibromyalgia,[30] and irritable bowel syndrome.[111] This correlation was demonstrated in a study, described earlier in this chapter, that compared sleep abnormalities on two-night PSG inpatients with myofascial TMD and healthy control subjects.[23] Although the two groups did not differ in daytime sleepiness or AHI values, the patients had a greater percentage of time in NREM stage 1 sleep and experienced almost twice as many respiratory arousals associated with apnea, hypopnea, and RERA events and a greater frequency of RERA events, independently of body mass index and demographic characteristics.

OTHER CONSIDERATIONS REGARDING AN ASSOCIATION BETWEEN SLEEP-DISORDERED BREATHING AND TEMPOROMANDIBULAR JOINT DISORDERS

The association between SDB and TMD is not limited to factors that might contribute to the initiation or persistence of the pain condition, as discussed previously. For example, insomnia is common in patients with TMD and is characterized by hyperarousability.[17,112] A lowered arousal threshold in the insomniac patient with TMD, coupled with the presence of pain, could evoke respiratory instability during sleep realized as RERAs or OSA.[113] Indeed, in one study described earlier, the frequency of RERAs was found to be directly related to the subjective ratings of myofascial pain during the night of the sleep study by participants with TMD.[23]

Another possibility is that both SBD and TMD are independent outcomes in the setting of unfavorable craniofacial anatomy. Although the correlation is highly controversial, some studies have reported anatomic risk factors for TMD.[114-116] Of interest, the major features identified (e.g., retrognathia, long or hyperdivergent face) also have been associated with OSA.[117-119]

CLINICAL PEARLS

- In clinical populations, a high prevalence of SDB has been observed in patients with TMDs and comorbid chronic pain conditions such as fibromyalgia syndrome. Patients with these idiopathic chronic pain conditions should be screened for SDB and treated as indicated.
- Both patients with SDB and those with TMD may report clenching or grinding their teeth during sleep. However, polysomnography that includes EMG of the masticatory muscles does not confirm the presence of SB in either group.
- Background masticatory muscle activity during sleep, outside of SB periods, which is similar to low-intensity clenching or tooth-to-tooth contact during the day, may be of pathogenic significance in the persistence of TMD pain and thus represent a putative risk factor for TMD.

SUMMARY

The widely held belief that bruxism, either sleep or awake, plays a major role in the development or persistence of TMD pain is not strongly supported by scientific evidence. Moreover, there is little evidence to support the premise of a cascade in time, in which SDB leads to SB, which in turns leads to TMD. Generalization of such beliefs, or a cause-and-effect

relationship, does not stand in large population studies or in studies using optimal assessment methods.

As various lines of evidence unfold, a much more complex set of interrelationships is emerging that may explain an association between SDB and TMD. Pathophysiologic mechanisms may involve elevated background levels of activity in the masticatory muscles during sleep and wakefulness, intermittent hypoxia during sleep, altered autonomic nervous system function, oxidative stress, and inflammation. Readers are cautioned that although the evidence in support of these mechanisms is growing, at this time the evidence base remains rudimentary and causal associations are not yet established. Identification of the roles of stress, anxiety, and/or cognitive arousal in patients with TMDs, and their influences on sleep continuity and breathing, are other research avenues of major interest. Findings not only will affect clinicians' ability to care for patients with TMD but also will strengthen the current understanding of the interaction of pain and poor sleep.

Selected Readings

Balasubramaniam R, Klasser GD, Cistulli PA, et al. The link between sleep bruxism, sleep disordered breathing and temporomandibular disorders: an evidence-based review. *J Dent Sleep Med* 2014;**1**(1):27–37.

Doufas AG, Tian L, Davies MF, et al. Nocturnal intermittent hypoxia is independently associated with pain in subjects suffering from sleep-disordered breathing. *Anesthesiology* 2013;**119**(5):1149–62.

Dubrovsky B, Raphael KG, Lavigne GJ, et al. Polysomnographic investigation of sleep and respiratory parameters in women with temporomandibular pain disorders. *J Clin Sleep Med* 2014;**10**(2):195–201.

Hollowell DE, Suratt PM. Mandible position and activation of submental and masseter muscles during sleep. *J Appl Physiol* 1991;**71**(6):2267–73.

Maluly M, Andersen ML, Dal-Fabbro C, et al. Polysomnographic study of the prevalence of sleep bruxism in a population sample. *J Dent Res* 2013;**92**(7 Suppl.):97S–103S.

Manfredini D, Lobbezoo F. Relationship between bruxism and temporomandibular disorders: a systematic review of literature from 1998 to 2008. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2010;**109**(6):e26–50.

Raphael KG, Sirois DA, Janal MN, et al. Sleep bruxism and myofascial temporomandibular disorders: a laboratory-based polysomnographic investigation. *J Am Dent Assoc* 2012;**143**(11):1223–31.

Raphael KG, Janal MN, Sirois DA, et al. Masticatory muscle sleep background electromyographic activity is elevated in myofascial temporomandibular disorder patients. *J Oral Rehabil* 2013;**40**(12):883–91.

Sanders AE, Essick GK, Fillingim R, et al. Sleep apnea symptoms and risk of temporomandibular disorder: OPPERA cohort. *J Dent Res* 2013;**92**(7 Suppl.):70S–7S.

Schmitter M, Kares-Vrincianu A, Kares H, et al. Sleep-associated aspects of myofascial pain in the orofacial area among temporomandibular disorder patients and controls. *Sleep Med* 2015;**16**(9):1056–61.

Slade GD, Conrad MS, Diatchenko L, et al. Cytokine biomarkers and chronic pain: association of genes, transcription, and circulating proteins with temporomandibular disorders and widespread palpation tenderness. *Pain* 2011;**152**(12):2802–12.

A complete reference list can be found online at ExpertConsult.com.

Chapter 147

Oral Appliances for the Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome and for Concomitant Sleep Bruxism

Christopher J. Lettieri; Fernanda R. Almeida; Peter A. Cistulli; Maria Clotilde Carra

Chapter Highlights

- Oral appliances are an accepted and reliable treatment option for patients with snoring and obstructive sleep apnea–hypopnea syndrome (OSAHS). They are currently indicated for patients with mild to moderate OSAHS; however, they have been shown to be effective in patients with more severe disease. Oral appliances are most effective in younger, thinner patients with mild to moderate OSAHS. They are less likely to be effective in obese patients.
- In general and especially for more severe disease, custom-made titratable devices offer superior treatment and a greater likelihood for successful therapy than other types of oral devices. Successful therapy is more likely to be achieved by performing an at-home progressive titration using a custom-made titratable device. This may be further enhanced by the application of adjunctive means to monitor the therapeutic response.
- Oral appliances are not as effective as positive airway pressure devices for reducing the apnea-hypopnea index and other sleep measures. However, short- and long-term improvements in daytime somnolence, quality of life, neurocognitive function, and cardiovascular outcomes (primarily blood pressure) appear to be similar with both treatments.
- Oral appliances are generally well tolerated, and serious side effects or adverse consequences resulting in discontinuation of therapy are uncommon. Malocclusion is the most common long-term effect. However, this typically does not lead to discontinuation of therapy and may be mitigated by simple exercises each morning after removal of the device.
- Concomitant sleep bruxism (SB) with OSAHS is common. Although the true etiologic nature of these disorders is not fully understood, there appears to be a modest causal relationship. Patients with SB and symptoms of OSAHS should be assessed for underlying sleep-disordered breathing as common SB occlusal splint therapies may, in some cases, worsen obstructive events if they are not recognized.
- Oral appliances appear to provide adequate therapy for both OSAHS and SB alone or in comorbidity.

Currently, positive airway pressure (PAP) remains the most accepted, common, and efficacious management tool for obstructive sleep apnea–hypopnea syndrome (OSAHS). In this chapter, we use the word *treatment*, although we recognize that most devices or appliances do manage sleep-disordered breathing (SDB), in the whole spectrum from snoring to severe OSAHS, whereas some residual apnea and hypopnea, oxygen desaturation, high blood pressure, inflammation, sleepiness, and cognitive alterations, although improved, may persist over time. However, given the ongoing challenges related to the acceptance of and adherence to PAP therapy, there remains an increasing need for reliable and effective treatment alternatives (e.g., appliance, nerve stimulation, sleep positioning devices). Oral appliances (OAs) of various designs are a solid alternative used to manage OSAHS, sleep bruxism (SB), and both when they are concomitant.

The main objective of this review is to describe the use of OAs for OSAHS, with special attention to SB. Other management tools for OSAHS are described in more detail in Chapters 115 and 116 for devices and medications, Chapter 149 for surgeries, and Chapter 150 for other alternatives. SB and orofacial pain are conditions, although they can be concomitant with SDB and obstructive sleep apnea, and are further described in Chapters 144 to 146.

WHY USE ORAL APPLIANCES FOR SLEEP-DISORDERED BREATHING AND OBSTRUCTIVE SLEEP APNEA–HYPOPNEA SYNDROME?

OAs offer effective therapy for many patients with simple (primary) snoring and OSAHS and have become a proven, validated, and accepted treatment option. With the release of

numerous clinical trials and published practice parameters establishing their efficacy and the expanded availability of these devices, OAs have become an increasingly common treatment modality for patients with SDB.

OAs offer several advantages over PAP. They are generally well tolerated in most individuals, and published reports have consistently shown that therapeutic adherence and patient preferences are as good as if not superior to those with PAP.[1,2] In addition, these devices do not require a ready and reliable source of electricity and may be easier to use, especially during travel or for people who live in an area with restricted access to electricity (e.g., sailors, fishermen, humanitarian workers, military).

As published in the American Academy of Sleep Medicine's "Practice Parameters for the Treatment of Snoring and Obstructive Sleep Apnea with Oral Appliances: An Update for 2005," these devices are recommended as primary therapy for patients with snoring and mild to moderate OSAHS or as a reasonable treatment alternative in patients who prefer these devices to PAP or are intolerant of other therapies.[3] An update to these Practice Parameters is scheduled to be published by the American Academy of Sleep Medicine in 2015.

The route to the current best accepted practices encompasses the role of the physician and dentist[4] in the proper selection of the patient and device and the need for an interdisciplinary approach to the appropriate application and ongoing care of OA therapy.

TYPES OF ORAL APPLIANCES FOR SLEEP-DISORDERED BREATHING AND SLEEP BRUXISM

For SDB and OSAHS management, two broad OA types in common clinical use are mandibular repositioning appliances and tongue-repositioning devices. Mandibular repositioning appliances, also known as mandibular advancement splints (MAS) or mandibular advancement devices, are the focus of this chapter as they are more widely used in clinical practice and have a greater evidence base. Numerous design types are available, but these devices generally fall into either one-piece (monobloc) or two-piece (duobloc) configurations. They can differ substantially in size, type of material, degree of customization to the patient's dentition, and coupling mechanisms. In addition, the amount of occlusal coverage, adjustability of mandibular advancement, degree of mandibular mobility permitted (vertical and lateral), and allowance for oral respiration also vary between the different available devices.

This chapter focuses on customized OAs for managing OSAHS. Although prefabricated MAS exist, many of which are available over the counter, there is a paucity of outcome measures validating their clinical use. Whereas these devices may offer a reasonable and cost-effective treatment, particularly for simple snoring, the existing data show limited efficacy and lower acceptance rates compared with customized OAs.[5,6] As such, prefabricated devices are not recommended for clinical use in patients with OSAHS.

Two-piece OAs have become more commonly used in clinical practice and consist of removable upper and lower plates that are coupled together to promote advancement of the mandible and to mitigate subluxation during sleep. There are a variety of modes of coupling between the upper and lower plates, including elastic or plastic connectors, metal pins and tube connectors, hook connectors, acrylic extensions, and magnets. Duobloc splints offer advantages over monobloc devices by allowing adjustability, which facilitates achievement of the most comfortable and efficient position of the mandible and greater degree of lower jaw movements. OAs that permit lateral jaw movement or opening and closing while maintaining mandibular advancement may offer additional advantages by reducing the risk of complications and improving the patient's comfort and acceptance. Because the patient's tolerance to the amount of protrusion increases over time, splints capable of incremental advancement seem to have a clear practical advantage. These adjustable devices also facilitate improved efficacy as they can be titrated to a more optimal setting needed to ablate obstructive events. Although prefabricated appliances ("off the shelf") are commercially available, their efficacy and potential role as a "trial" device have been called into question.[6] The best retention, comfort, and efficacy are achieved with custom-made, titratable OAs. Further comparisons between the different available types of OAs and potential advantages in the management of patients with SDB are discussed later in this chapter.

The most frequently used OA for SB management (i.e., preventing grinding sound and tooth wear and possibly reducing pain) is the occlusal splint (a one-piece monobloc) that can be custom made to fit either the upper or lower jaw. The efficacy in the short term is recognized, but in patients with OSAHS, it is to be used with caution because, as described later, it may exacerbate their condition. In some subgroups of patients, in their lifetime course, it is possible that SB and OSAHS coexist. Although some association was found, as described later and in Chapters 145 and 146, it is premature to conclude that SB causes SDB or the reverse, that SDB causes SB; association is not causality.

ORAL APPLIANCE MECHANISM OF ACTION

Current evidence suggests that OSAHS pathogenesis reflects reduced upper airway size and altered upper airway muscle activity, resulting in diminished patency and airflow obstruction (see Sections 3 and 14 in this volume for more information). Although it has been thought that the primary mechanism of action of OAs arises from the anterior movement of the tongue and consequent increase in the anteroposterior dimensions of the oropharynx, it appears that this is an overly simplistic view. Many studies using a range of imaging modalities, including computed tomography, magnetic resonance imaging, and nasoendoscopy, suggest that OAs induce more complex anatomic changes.[7-10] An increase in airway volume appears to result, in large part, by an increase in cross-sectional area of the velopharynx, in both the lateral and anteroposterior dimensions, and increases in the lateral dimension of the oropharynx. Figure 147-1[10] illustrates the effects of an OA and tongue-stabilizing device on the upper airway. These changes are thought to be mediated through the palatoglossal and palatopharyngeal arches, which link the muscles of the tongue, soft palate, lateral pharyngeal walls, and mandibular attachments. Interindividual variability in the airway configurational changes that occur with mandibular advancement may reflect variations in anatomy, and this is likely to have major relevance to the variable clinical response associated with this treatment modality.

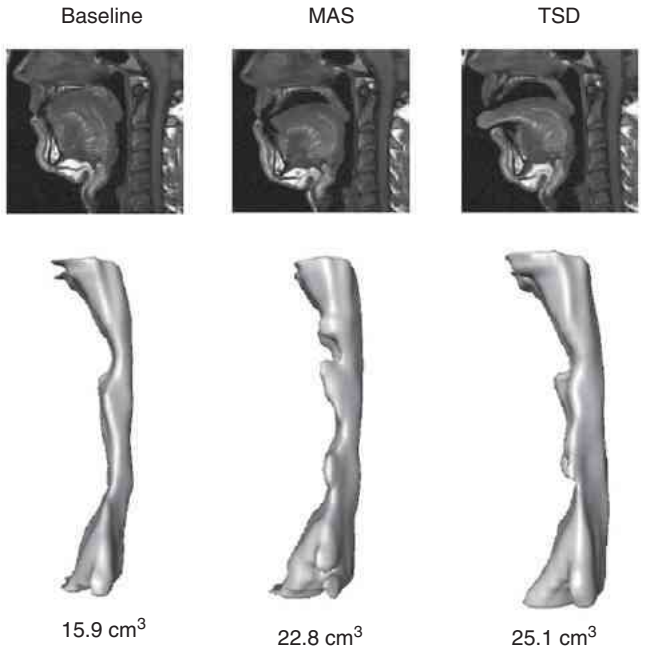


Figure 147-1 Modes of action of mandibular advancement splint (MAS) and tongue-stabilizing device (TSD). Magnetic resonance imaging representation of the increase in upper airway volume while patients have the appliance in place. (From Sutherland K, Deane SA, Chan AS, et al. Comparative effects of two oral appliances on upper airway structure in obstructive sleep apnea. *Sleep* 2011;34:469–77.)

Anatomic imbalance has been proposed as an underlying mechanism in the pathogenesis of OSAHS.[11] In this model, excess tissue within the bony enclosure must be present to generate sufficient tissue pressure to collapse the airway lumen. This extraluminal tissue pressure occurs in the context of either an excess of soft tissue within a normal bony enclosure size or a normal amount of tissue compressed into a reduced bony enclosure. Mandibular advancement delivered by OAs effectively enlarges the bony enclosure and appears to improve anatomic balance.[12]

The effects of OAs on upper airway neuromuscular pathways have not been well studied to date. Whereas some studies indicate that these devices stimulate genioglossus muscle activity,[13,14] studies using inactive "sham" OAs have shown little change in SDB.[15,16] This suggests that mechanical advancement of the mandible is the primary mechanism of action of these devices. This mechanical effect results in greater airway stability, which is evidenced by a reduced upper airway closing pressure during sleep.[17] This effect was demonstrated in a study of anesthetized OSAHS patients by Kato et al,[18] who observed dose-dependent reductions in closing pressure of all pharyngeal segments with progressive mandibular advancement.

CLINICAL OUTCOMES AND MEASURES OF SUCCESS

Since the last published guidelines on the use of OAs in the treatment of OSAHS,[15,16,18-23] there has been a substantial increase in the quantity and quality of clinical trials evaluating the efficacy and effectiveness of OA therapy[2,24-27] (see Figure 147-2). Several trials, including prospective designs using placebo arms and sham comparisons, have produced a high level of evidence and have allowed stronger treatment recommendations. In addition, trials assessing the treatment response between OAs and other primary therapies for OSAHS focusing on clinically important outcomes help further the understanding of appropriate patient selection and the expected therapeutic effect. In short, OAs have been shown to have a good level of efficacy in improving polysomnographic measures, daytime somnolence, quality of life, cardiovascular outcomes, and neurocognitive function for patients with both simple (primary) snoring and OSAHS.

Snoring

Several placebo-controlled trials have found that OAs reduce both subjective and objective measures of snoring frequency and intensity among nonapneic patients with habitual (primary) snoring.[17,19,22,23,28,29] OAs have also been shown to improve quality of life measures in nonapneic snorers. In one trial, the use of OAs produced a mean improvement in Functional Outcomes of Sleep Questionnaire (FOSQ) scores by 3.21 (95% confidence interval, 2.82–3.60; $P < .001$) among patients with habitual snoring.[29]

Polysomnographic Variables and Reduction of the Apnea-Hypopnea Index

OAs have been well established as an effective treatment in the management of OSAHS. There is strong evidence from randomized, controlled trials that these devices produce a significant reduction in the apnea-hypopnea index (AHI) across the full spectrum of OSAHS severities. Abolition of obstructive events to an AHI less than 5 events/hour is achieved in 36% to 70% of patients.[15,16,26,30-32] With use of an AHI threshold of less than 10/hour, OAs provide successful therapy in 30% to 86%.[2,18,30] The likelihood of successful treatment is progressively greater with lower severities of disease. Table 147-1 illustrates success rates with OA therapy according to different disease severity at baseline.

Daytime Somnolence and Quality of Life Measures

OAs have been shown to improve daytime sleepiness and quality of life in patients with OSAHS. In one trial, improvements in daytime somnolence were significantly better with customized OAs compared with nontherapeutic devices.[33] Specifically, the Epworth Sleepiness Scale (ESS) score decreased from 14.7 ± 5.1 to 5.1 ± 1.9 ($P < .05$) in those receiving OA therapy compared with 16.3 ± 2.5 to only 13.6 ± 6.7 (P = NS) in the nontherapeutic group. Similarly, in a long-term follow-up study, OAs produced a decrease in ESS from 13.9 ± 1.3 to 9.3 ± 1.2.[28] OAs have also been shown to improve objective measures of sleepiness. In a placebo-controlled trial, 4 weeks of OA therapy led to an additional improvement in the mean sleep latency on the Multiple Sleep Latency Test of 1.2 minutes compared with an inactive control oral device.[28] Two additional trials evaluating objective measures of wakefulness found that the improvements in Maintenance of Wakefulness Test were similar between both PAP and OA therapy.[1,34] Across numerous published studies, the improvement in both subjective and objective measures of somnolence achieved with OAs appears to be similar to that reported with PAP therapy. However, as with PAP, whereas OAs are shown to improve daytime somnolence,

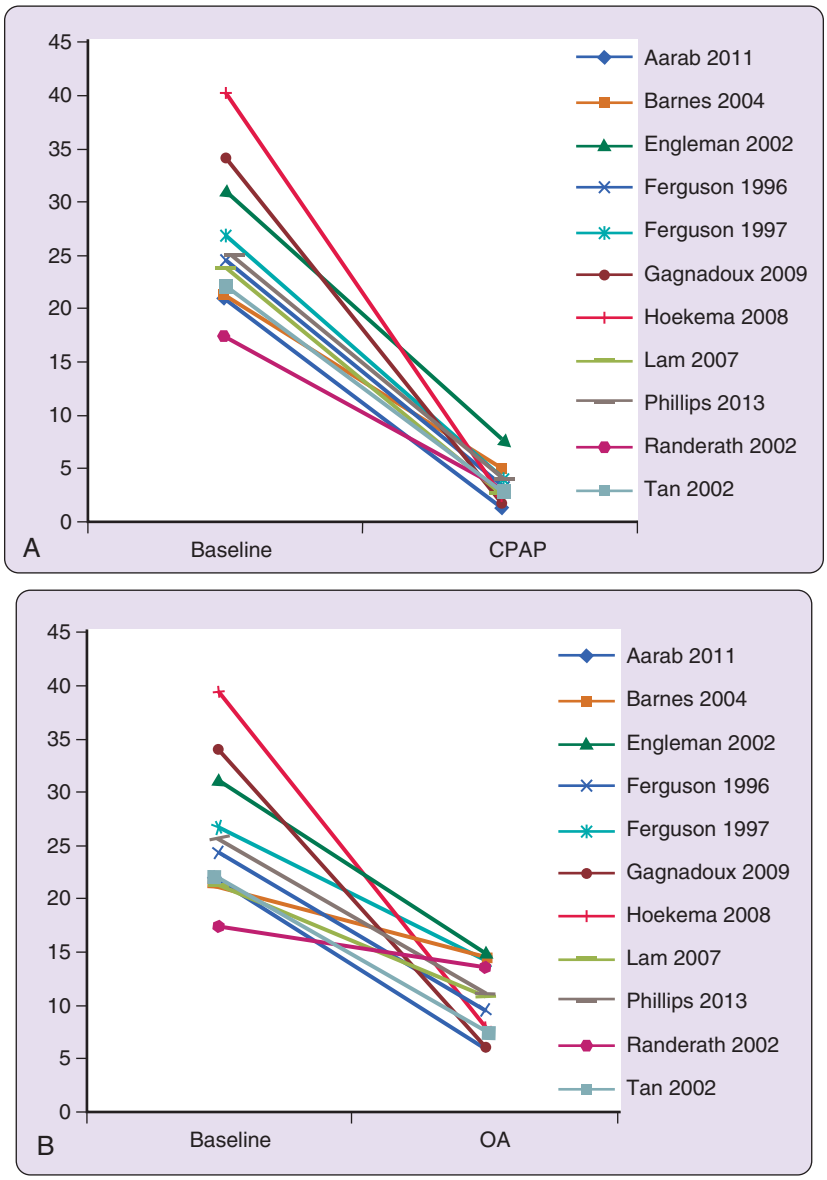


Figure 147-2 Average decrease in apnea-hypopnea index (AHI) from published randomized controlled trials. CPAP, Continuous positive airway pressure; OA, oral appliance.

residual sleepiness may persist despite otherwise adequate therapy.

Compared with no treatment or nontherapeutic (sham) therapy, OAs have also been shown to significantly improve quality of life measures in patients with OSAHS. In one trial, overall FOSQ scores improved by 27.1% from baseline ($P < .001$; effect size, 0.90) in those using a customized OA compared with a –1.7% decline in those using a nontherapeutic sham device.[33] In a randomized, controlled trial comparing OA therapy with a placebo tablet, mandibular advancement produced superior improvements in quality of life as measured by the FOSQ and 36-Item Short Form Health Survey overall health score.[1] The magnitude of improvement was similar to that found with PAP.

Cardiovascular Outcomes

Several studies have addressed the impact of OA therapy on measures of systolic blood pressure.[1,2,29,32,35-39] Overall, OAs were found to lower the systolic, diastolic, and mean blood pressure (Table 147-2). Further, the results from three randomized trials suggest that OAs are as effective as PAP in reducing blood pressure measures.[1,2,32] These studies are largely limited to custom-made, titratable devices. One small, randomized, controlled trial of 36 patients using nontitratable OAs failed to observe significant changes in either the systolic or diastolic blood pressure.[35]

OAs have also been shown to produce significant improvements in endothelial function and markers of oxidative stress similar to those observed with PAP therapy.[1,2,27,32,40] However, as with PAP therapy, the available data suggest that OAs produce only modest reductions in blood pressure recordings and other cardiovascular end points.

Whereas blood pressure and other markers of cardiovascular health have shown small but significant improvements, two studies evaluating cardiovascular morbidity and mortality noted a marked cardiovascular risk reduction with OA

therapy.[41,42] In a study of patients with mild to moderate OSAHS receiving either OA or PAP therapy, there was a marked cardiovascular risk reduction of at least 38%.[42] Recently, a study focused on severe sleep apnea patients described no difference in the cardiovascular death rate of those receiving either treatment.[41]

Despite the limited literature and the presence of residual apneas with OA therapy, the current studies available have shown continuous PAP (CPAP) and OA to be equally effective in the reduction of cardiovascular events.

Neurocognitive Outcomes

OA therapy has been shown to improve neuropsychological functioning. In a trial comparing OAs, PAP, and a tablet placebo for mild to moderate OSAHS, OA therapy improved tension-anxiety, divided attention, and executive functioning compared with placebo.[1] However, PAP was superior to OAs in improving psychomotor speed and mood. Similarly, treatment with an OA was found to significantly improve attention, vigilance, and motor speed, as measured by the distraction–working memory test and continuous performance.[43] In contrast, other trials have failed to demonstrate improved neurocognitive function after OA therapy despite improvements in other measured end points.[1,34,44]

Despite the similarity to PAP, the existing literature shows mixed results with only modest improvements in neurocognitive measures for patients undergoing OA therapy.

Table 147-1 Success Rates with Oral Appliance Therapy according to Different Disease Severity at Baseline

OSA Severity at Baseline	Mild	Moderate	Severe
Success			
Study A	62.3%	50.8%	39.9%
Study B	56.6%	48.1%	21.2%
Partial success			
Study A	13%	25%	24.3%
Study B			
Failure			
Study A	30.4%	26.9%	54.5%
Study B			

The results are expressed in percentage from two trials using custom-made titratable oral appliances: Study A, Holley et al[60] and Study B, Phillips et al.[2] Success is described as an apnea-hypopnea index (AHI) <5 with oral appliance therapy; partial success is at least 50% reduction in AHI but AHI >5; and failure represents <50% reduction in AHI.

OUTCOME COMPARISON BETWEEN TYPES OF ORAL APPLIANCES AND POSITIVE AIRWAY PRESSURE THERAPY

Fixed versus Adjustable Appliances

Customized, individually fabricated MAS are either fixed (nonadjustable) or titratable (adjustable). Nonadjustable devices are monobloc or single-piece appliances in which the degree of mandibular advancement is permanently fixed. Titratable appliances (adjustable or duobloc) allow adjustments in the degree of mandibular protrusion to optimize treatment efficacy. This is an important distinction both in selection of the most effective treatment and in understanding the published literature. Published studies using fixed or single-jaw position appliances may underestimate the impact of OA therapy compared with titratable appliances.[1,25,27,34,45] Titratable or adjustable appliances allow progressive protrusion of the mandible, and the amount of anteroposterior mandibular movement varies considerably among patients. Previous studies have shown that OA efficacy is related to the amount of mandibular advancement.[18,46,47] Determining the optimal degree of mandibular advancement is the most

Table 147-2 Impact of Oral Appliance Therapy on Blood Pressure Measurements in Different Clinical Trials

Author	Mean ΔSBP	Mean ΔDBP	Mean Δ24-h SBP	Mean Δ24-h DBP	Mean ΔNight SBP	Mean ΔNight DBP
Gauthier[149]	−4.3	−10.1				
Trzepizur[32]	−8.8	2				
Andren[36] (3 months)	−14.3	−8.6				
Andren[36] (3 years)	−15.5	−10.3				
Yoshida[39]	−4.5	−2.9				
Gotsopoulos[37]	−4.9	−3.7	−2.3	−1.5	0.5	0.6
Barnes[1]			0.2	0		−2.2
Otsuka[38]	−3.1	−4.2	−4.5	−4.9	−4.7	−4.4
Zhang[151]	−2.1		−2.2	−0.6	−4	−2.7
Lam[27]	−1.2	−2.8				−2.1*
Phillips[2]	0.2	−0.5	0	−0.2	−0.4	−0.2
Phillips[2] (hypertension patients only)	−2.5	−2.4	−2.9	−2.1	−3.4	−1.9

Measurements are in millimeters of mercury. Some studies have only clinical blood pressure, whereas some also describe 24-hour blood pressure. A negative change (Δ) in blood pressure refers to a reduction of the blood pressure.
*Refers to a nighttime clinical blood pressure, not 24-hour assessment.
SBP, Systolic blood pressure; DPB, diastolic blood pressure.

important step in using OA therapy successfully.[48,49] As an analogy, titration of OAs is similar to PAP. The amount of pressure (or degree of mandibular advancement) required for each patient cannot be predetermined on the basis of OSAHS severity or patient-specific characteristics. Objective assessment of the treatment response at progressive increases in the delivered pressure (or advancement) is needed to determine the optimal treatment for each individual patient to optimize the ablation of obstructive events. A systematic review of different types of MAS concluded that there is no one MAS design feature that influences treatment efficacy, although efficacy does depend on the degree of mandibular protrusion and whether the device is a fixed or titratable appliance.[50] The comparison of the efficacy rates between fixed and titratable appliances has demonstrated titratable devices to be superior in their ability to reduce the AHI for patients at all levels of OSAHS severity.[51] Although fixed devices were frequently effective in patients with mild sleep apnea, they had high failure rates in those with moderate or severe disease. These studies reinforce that in future meta-analysis of treatment outcomes, fixed single-jaw positioners should not be evaluated together with titratable devices as these therapies have different treatment outcomes.

Oral Appliances versus Positive Airway Pressure Therapy

Several published randomized controlled trials and crossover studies have compared the efficacy of OAs to PAP in the treatment of OSAHS.[52] As seen in Table 147-3, although both PAP and OAs led to improvements in objective sleep measurements such as AHI, arousal index, and minimum arterial oxygen saturation, the magnitude of improvement in AHI is significantly greater with PAP. Although statistically superior, these differences are not necessarily clinically significant. OAs decrease both subjective and objective measurements of snoring in the majority of the patients.[15,31,53] This reduction is less robust than with PAP, which frequently uses the elimination of snoring as an objective assessment of treatment efficacy. Subjective and objective measurements of sleepiness are similarly improved with both forms of therapy.

Both OAs and PAP have been shown to be superior to placebo or no treatment in improving quality of life measures.[24] With a disease-specific questionnaire, the FOSQ, both treatments similarly improved quality of life. Direct comparative studies (randomized controlled trials) demonstrated that both OAs and PAP are effective in improving quality of life, with each treatment providing certain advantages in specific subcategories.[2] Whether this reflects a true difference between the effects of each treatment or is a product of differing study designs and data collection is uncertain. Regardless, it appears that both forms of therapy are similarly effective, and both are highly dependent on the patient's compliance. Given that adherence with OAs is typically superior to that with PAP, outcomes in clinical practice may favor these devices over PAP, particularly for mild and moderate OSAHS.

Whereas both OAs and PAP significantly reduce the AHI, PAP has been consistently shown to be superior in this measured outcome. Table 147-3 shows the results of randomized controlled trials and their average reduction in AHI. However, there is equivalence in overall health outcomes with both therapies. This may reflect that the greater efficacy of PAP is offset by inferior compliance relative to OAs, resulting in similar effectiveness. In other words, higher adherence rates with OAs likely translate into a similar mean disease alleviation and consequently similar effectiveness compared with PAP.[54] These findings strongly challenge current practice parameters recommending that OA treatment should be considered only in patients with mild to moderate OSAHS or in those who have failed to respond to or refuse PAP treatment. Long-term comparative effectiveness studies of these two treatment modalities are clearly needed.

PATIENT AND DEVICE SELECTION

Proper selection of the patient and device can increase the likelihood of success, improve the therapeutic effect, and enhance outcomes. There are multiple factors that should be considered in determining which patients are ideal candidates for OA therapy. Similarly, there are multiple different types, designs, and brands of OA devices, each with its inherent advantages and limitations. Understanding these factors can aid clinicians in making appropriate treatment decisions.

Indications and Contraindications

The American Academy of Sleep Medicine's Practice Parameters on OAs in the treatment of SDB advocate the use of these devices in patients with mild to moderate OSAHS who prefer this form of treatment to PAP or who do not respond to or are unable to tolerate PAP.[3] Since these guidelines were published, the evidence base supporting the use of OA therapy has grown considerably and suggests that clinicians should use custom-made titratable OAs as first-line therapy or as an alternative to PAP as an effective and reliable means to improve physiologic sleep measures, daytime sleepiness, and quality of life in adult patients with OSAHS.

Selection of appropriate patients for OA therapy, based on the likelihood of successful treatment, remains a somewhat elusive goal at present. Whereas considerable research has attempted to identify the factors that predict a good response, the clinical utility of such approaches remains to be proven. In general, younger, thinner patients with positional OSAHS and an overall lower AHI appear to be the preferred candidates for OA therapy. However, in the absence of clear selection criteria, the clinician should rely on clinical judgment and the patient's preference when choosing the appropriate therapeutic approach.

Dentists experienced in dental sleep medicine have a prominent role in determining whether a patient is an ideal candidate for OA therapy. Patients require a sufficient number and location of healthy teeth to retain the device and to promote mandibular advancement. Specifically, patients require a minimum of eight teeth in the upper jaw and in the lower jaw, with at least two teeth in each quadrant.[55] In addition, the patient should have the ability to protrude the mandible forward to achieve a therapeutic result.

Not all patients are suitable candidates for the use of OAs because of associated medical or dental conditions and factors. In general, PAP affords a more prompt initiation of therapy. A major clinical limitation of OA therapy is in circumstances in which there is an imperative to commence more immediate treatment as there are inherent delays to attaining optimal therapy with use of these devices. This includes situations involving severe symptomatic OSAHS (e.g., concern about

Table 147-3 Summary of the Outcomes of Randomized Controlled Trials Comparing Oral Appliances with Continuous Positive Airway Pressure

Study	Design	No. of Subjects (% Male) [Withdrawals]	Inclusion	Oral Appliance	Treatment [Washout] Duration	Baseline AHI	Treatment AHI		OA vs. CPAP		
							CPAP	OA	AHI	ESS	Patient Preference
Aarab, 2010[27]	Parallel (*placebo group included*)	57 (74%) (20 OA/18 CPAP) [7]	AHI 5–45 + ESS ≥ 10	Customized, two-piece, set 25%, 50%, or 75% advancement, depending on sleep study results at each level	24 weeks	CPAP: 20.9 ± 9.8 OA: 22.1 ± 10.8	1.4 ± 13.1	5.8 ± 14.9	↔ (*P* = .092)	↔	N/A—parallel groups
Barnes, 2004[58]	Crossover (*placebo group included*)	80 (79%) [24]	AHI 5–30	Customized 4-week titration to maximum comfortable advancement	3 × 12 weeks [2 weeks]	21.5 ± 1.6^	4.8 ± 0.5^	14.0 ± 1.1^	CPAP	↔	CPAP
Engleman, 2002[54]	Crossover	48 (75%) [3]	AHI ≥ 5/h + ≥2 symptoms (including ESS ≥ 8)	Customized, one-piece, 80% maximal protrusion, two deigns: (a) complete occlusal coverage or (b) no occlusal coverage, assigned randomly	2 × 8 weeks [not reported]	31 ± 26	8 ± 6	15 ± 16	CPAP	CPAP	↔
Ferguson, 1996[60]	Crossover	25 (89%) [2]	AHI 15–50 + OSA symptoms	Snore-Guard (Hays & Meade Inc), maximum comfortable advancement	2 × 16 weeks [2 weeks]	24.5 ± 8.8	3.6 ± 1.7	9.7 ± 7.3	CPAP	N/A	OA
Ferguson, 1997[50]	Crossover	20 (95%) [4]	AHI 15–55 + OSA symptoms	Customized, two-piece appliance titration starting at 70% maximum advancement over 3 months	2 × 16 weeks [2 weeks]	26.8 ± 11.9	4.0 ± 2.2	14.2 ± 14.7	CPAP	↔	OA
Gagnadoux, 2009[55]	Crossover	59 (78%) [3]	AHI 10–60 + ≥ 2 symptoms, BMI ≥ 35 kg/m^2	AMC (Artech Medical), two-piece, advancement determined by single-night titration	2 × 8 weeks [1 week]	34 ± 13	2 (1–8)#	6 (3–14)#	CPAP	↔	OA
Hoekema, 2008[56]	Parallel	103 (51 OA/52 CPAP) [4]	AHI ≥ 5	Thornton Adjustable Positioner type 1, titratable	8–12 weeks	CPAP: 40.3 ± 27.6 OA: 39.4 ± 30.8	2.4 ± 4.2	7.8 ± 14.4	CPAP	↔	N/A—parallel groups
Lam, 2007[61]	Parallel (*placebo group included*)	101(79%) (34 OA/34 CPAP) [10]	AHI ≥ 5–40 + ESS > 9 if AHI 5–20	Customized, nonadjustable set to maximum comfortable advancement	10 weeks (83%) referred for concurrent weight loss program)	CPAP: 23.8 ± 1.9^ OA: 20.9 ± 1.7^	2.8 ± 1.1^	10.6 ± 1.7^	CPAP	CPAP	N/A—parallel groups
Phillips, 2013[57]	Crossover	108 (81%) [18]	AHI ≥ 10 + ≥2 symptoms	Customized two-piece appliance (SomnoMed), titrated to maximum comfortable limit in acclimatization period before study	2 × 4 weeks [2 weeks]	25.6 ± 12.3	4.5 ± 6.6	11.1 ± 12.1	CPAP	↔	OA
Randerath, 2002[62]	Crossover	20 (80%)	AHI 5-30 + OSA symptoms	IST; Hinz (Herne, Germany) two piece, nontitratable, set to two-thirds of maximum advancement	2 × 6 weeks [not reported]	17.5 ± 7.7	3.2 ± 2.9	13.8 ± 11.1	CPAP	N/A	N/A
Tan, 2002[63]	Crossover	21 (83%) [3]	AHI 5-50	One-piece, 75% maximum advancement and Silensor (Erkodent Gmbh) two-piece, titratable	2 × 8 weeks [2 weeks]	22.2 ± 9.6	3.1 ± 2.8	8.0 ± 10.9	↔	↔	N/A

AHI, Apnea-hypopnea index; CPAP, continuous positive airway pressure; OA, oral appliance; ESS, Epworth Sleepiness Scale; OSA, obstructive sleep apnea; BMI, body mass index; N/A, not applicable, not measured in study.
Data are presented as mean ± standard deviation, unless denoted; ↔ equivalent between treatments; ^mean ± standard error of mean; #median (interquartile range).
From Sutherland K, Vanderveken OM, Tsuda H, et al. Oral appliance treatment for obstructive sleep apnea: an update. *J Clin Sleep Med* 2014;10:215–27.

driving risk or profound daytime impairments) and coexistent medical comorbidities, such as ischemic heart disease. Moreover, this treatment modality has no known role in treating central sleep apnea or hypoventilation states. In addition, some case reports have shown worsening of OSAHS severity in some patients using OAs.[22,56] This, together with the known potential for a placebo response, highlights the need for objective verification of treatment outcome using in-laboratory or home sleep testing with an OA in place at its prescribed, therapeutic position.[15,16] Patients with temporomandibular joint (TMJ) problems may require concomitant use of exercises to be able to adhere to OA therapy.[57] Long-term use of OAs results in few or no TMJ problems.[58] The presence of periodontal disease may promote excessive tooth movement or worsening of dental caries with an OA. These factors tend to limit the scope and application of this form of therapy; it is therefore highly important that a dentist with expertise in sleep medicine assess dental and TMJ health before initiation of therapy.

Predictors of Successful Oral Appliance Therapy

An important and currently unresolved issue limiting the role of OAs for the treatment of OSAHS is the inability to reliably predict an effective response to treatment. Ultimately, the response to therapy is likely to be related to multiple patient factors, device features, and clinical expertise of the treating provider. Patient factors, including anthropomorphic and polysomnographic variables, have been the subject of numerous studies. Clinical features reported to be associated with better outcomes and greater likelihood of success include younger age, lower body mass index, supine-dependent OSAHS, lower AHI, smaller oropharynx, less overjet, shorter soft palate, and smaller neck circumference.[59] In general, it is considered that a good response is more likely in mild to moderate OSAHS, although benefit in severe OSAHS has been reported.[2,60,61] Cephalometric measures, such as a shorter soft palate, longer maxilla, and decreased distance between mandibular plane and hyoid bone, either in isolation or in combination with other anthropomorphic and polysomnographic variables, are thought to provide some predictive power regarding successful therapy with OAs.[16,61]

Physiologic studies indicate that retroglossal rather than velopharyngeal collapse during sleep is highly predictive of OA success.[62] Physiologic measurements during wakefulness, including nasal resistance and flow-volume loops, have been reported to differ between OA responders and nonresponders.[63,64] However, upper airway imaging during wakefulness may aid in predicting the treatment response. Upper airway magnetic resonance imaging studies suggest that although baseline airway and soft tissue anatomic characteristics may not differ between responders and nonresponders, the changes consequent to mandibular advancement do differ such that increases in airway volume are reasonably predictive of a favorable outcome.[65] Furthermore, computational modeling techniques further add to the information that can be derived from sophisticated imaging and appear to enhance these predictions.[66] Whereas such studies are helpful in understanding fundamental mechanisms of airway collapse during sleep and the mechanism of OA therapy, the clinical utility of such approaches is limited by cost and accessibility. Nasoendoscopy offers a more clinically accessible imaging modality. Studies during both wakefulness and drug-induced sleep[67,68] have established the predictive potential of this technique. Lateral widening of the velopharynx during awake endoscopy in the supine position is associated with a higher response rate with OAs.[67] Drug-induced sleep endoscopy has shown good sensitivity for predicting treatment success by allowing visualization of the magnitude and patterns of pharyngeal collapse and identifying patients with greater improvements in pharyngeal patency with mandibular advancement.[69] See Chapter 148 for more information.

Remotely controlled mandibular positioning devices offer a relatively novel approach to identifying individuals who will or will not respond to OA therapy.[70-72] During a single-night titration procedure, these devices use hydraulic or electronic means to incrementally advance the mandible during sleep to determine both the treatment responsiveness and the required degree of advancement to ablate obstructive events. In addition, this may help establish tolerability for OA devices before costly acquisition. A prospective study using a commercially available system demonstrated that it is feasible both to predict treatment outcome and to determine the required "dose" of mandibular advancement during a single-night titration polysomnographic study.[71]

Appliance Selection

It is the sleep dentist's role to determine the most appropriate type of appliance based on specific clinical features to ensure that the patient is provided the most efficacious and cost-effective therapy. Given the wide variability in the reported efficacy across different studies, there is a strong suggestion that OA design has an important influence on treatment outcomes. Duobloc OAs, consisting of upper and lower plates, offer the advantage of a greater degree of mandibular movement (vertical and lateral) and adjustability (advancement), permitting attainment of the most comfortable and efficient position of the mandible. It is generally considered that the best retention is achieved with OAs that are customized and individually fabricated from the patient's dental impressions.[73]

Associated dental conditions, such as bruxism, may influence the choice of appliance design. As discussed later in this chapter, sleep bruxism/tooth grinding is common in patients with OSAHS,[74] although the causal relationship between the two conditions is unclear[31] (see Chapters 144–146). Patients who experience jaw discomfort after wearing a rigid monobloc OA may benefit from using an appliance that allows lateral and vertical jaw movement.

Another important consideration is the vertical dimension of the OA. Minimum vertical opening depends on the amount of overbite. There are conflicting data on the effect of the degree of bite opening induced by OAs on treatment outcomes, although most patients appear to prefer minimal interocclusal openings.[75] In mouth-breathing patients, the selected OA design should have an anterior opening to permit comfortable breathing. In the case of edentulous patients wearing partial dentures, the splint design selected for that patient should adapt to the remaining natural dental structures when the dentures are removed. Whereas tongue-repositioning devices have a limited role in the treatment of OSAHS, they may play a role in individuals with insufficient teeth who fail to respond to or are not tolerant of other therapies.

A number of studies comparing the efficacy of different OA designs have emerged recently.[28,76,77] A retrospective analysis of 805 patients using either an adjustable OA or a fixed device found a higher treatment response rate for adjustable devices (56.8% vs. 47.0%).[77] Two crossover studies have compared two-piece adjustable appliances with different advancement mechanisms and found similar improvements in AHI, symptoms, and side effects.[28,76] One study assessing the addition of tongue protrusion using an anterior tongue bulb on an existing OA device showed further reductions in AHI compared with mandibular advancement alone.[78] These studies suggest that appliance design features are relevant to treatment efficacy, patient tolerance, and use and highlight the need for further research to differentiate the advantages of different OA designs and to aid in the appropriate appliance selection in clinical practice.

OPTIMIZATION OF TREATMENT

Although similar in concept, the optimization of OA treatment is different from that of PAP, which can be performed either during polysomnography or with an automatic/autotitrating PAP platform. One of the main differences is that patients may not be able initially to tolerate the degree of mandibular advancement required to completely relieve obstructive events and may require several months of progressive mandibular advancement to achieve a therapeutic position.[18,46,47,79] Several studies have evaluated whether titration of mandibular advancement during polysomnography could be used to optimize OA treatment, similar to PAP titrations.[47,70,80,81] These protocols have mixed results and only a fair sensitivity in predicting the amount of mandibular advancement needed for successful OA therapy. A remotely controlled mandibular positioner[80] using a success criterion of less than 10 events/hour and 50% improvement with a 4% oxygen desaturation criterion showed a positive and negative prediction value for subsequent successful OA therapy of 94% and 83%, respectively. These studies are promising, but non–industry-supported studies are required to further assess the prediction of treatment outcome with this tool.

The use of a systematic, home-based titration in which patients incrementally advance their devices until subjective improvements in sleep quality and daytime symptoms are achieved is significantly more likely to render more effective OA therapy. Studies have shown that although 55% of patients achieve successful self-titration at home, another 32% can reach success with further polysomnography-guided titration.[48,82] Similarly, the use of home oximetry during titration does help titrate OAs.[49,82] Interestingly, 25% of the patients required additional mandibular advancement because of an abnormal oxygen desaturation index despite resolution of symptoms, whereas 20% of patients required further titration because of persistent symptoms despite a normal oxygen desaturation index.

Similar to PAP, appropriate OA titration is vital to achieve optimum therapeutic efficacy. Successful therapy requires both a dentist with experience in OA treatment to ensure proper adjustments of these devices and a sleep physician who should conduct follow-up evaluations and sleep testing to confirm that effective treatment is achieved. The dentist should observe patients on a yearly basis and may advance/titrate the appliance further if symptoms recur. If maximum mandibular advancement is reached and symptoms are still present, the patient should be referred to the referring physician for further evaluation and consideration of adjunctive or alternative therapies.

The other concept of optimization of treatment is the combination of treatment modalities. One study included patients who were unable to tolerate CPAP because high pressure was required to abolish the apneas. The combination of CPAP with an OA, not interconnected but used simultaneously, resulted in a CPAP pressure reduction of about 2 cm H_2O.[83] Another form of combination is to use the treatments interchangeably. Interestingly, when patients have the opportunity to use either treatment on a regular basis, patients tend to fluctuate between therapies. When patients compared the sleepiness while receiving CPAP only, they showed a significant decrease in their ESS score when they were able to use both treatments. One could hypothesize that patients were less likely to occasionally drop treatment, and therefore the long-term effects on sleepiness were further consolidated.[84] The combination of positional therapy with OA therapy has also been shown to improve clinical outcomes.[85]

Optimization of OA treatment, as described here, is indeed important and can be related to proper titration of the appliance and also the ability to offer combination therapies.

SIDE EFFECTS AND COMPLICATIONS

The primary reasons for discontinuation of OA treatment are an insufficient reduction of snoring, the persistence of apneic events, and the development of treatment-related side effects.[21] The most common reasons that patients completely discontinue OA therapy are device discomfort/cumbersome and limited perceived efficacy, and about 45% of nonadherence occurs within the first 6 months of therapy.[53]

Most side effects caused by OA are typically mild and transient. The most frequently reported events are excessive salivation, dry mouth, mouth or teeth discomfort, muscle tenderness, and jaw stiffness. Device adjustment can decrease short-term side effects by reducing pressure on the anterior teeth and excessive mandibular advancement. The clinical importance of short-term OA side effects has been compared with PAP published observations. By use of the same visual analog scale to categorize side effects of either OA or PAP, the two treatment modalities had a similar side effect score.[25] More persistent and severe side effects, including TMJ dysfunction and dental crown damage, are uncommon.[21] OAs in the titrated position appear to be innocuous to the TMJ in OSAHS patients.[53] Whereas existing TMJ disorders were typically considered a contraindication to OAs, a study of patients with known or prior TMJ dysfunction found that these patients can become eligible for OA therapy after simple physiotherapy exercises.[57,86] Transient and nonserious TMJ pain occurs more frequently with OAs than with PAP, but the risk for development of impairment of the temporomandibular complex is infrequent with long-term MAS use. Therefore, pain related to the initial use of OAs is typically transient and not associated with a significant risk of long-term complications or functional limitations.

Long-term side effects of OA therapy are related to a significant impact on the occlusion. Changes in dentition are

not restricted to OA therapy as the use of nasal masks can also alter the craniofacial structures.[58,87] However, these changes are more prominent in OA users compared with PAP users. Changes observed in craniofacial structures were mainly related to significant dentoalveolar changes (tooth movements).[88-90] There are important concerns about the timing and continuation of dental changes during a long-term period.[91] Importantly, overbite changes were observed to decrease less with time, whereas overjet continuously changed at a constant rate of 0.2 mm per year of OA use. Interestingly, on evaluation of various studies of different appliance designs, such as Herbst, Mobloc, Klearway, SomnoMed, and TAP, it has been shown that the amount of change was related to the duration of therapy and not the type of appliance that was used.[86,89-93]

Although changes in dental occlusion do occur, these are generally not a significant cause of treatment discontinuation. The patient's perceptions do not typically correlate with objective measurements, and occlusal changes often go unnoticed[90]; also, most individuals developed new occlusal contacts resulting from the development of a new occlusal equilibrium over time.[94] In addition, several reports have consistently found that patients perceive dental side effects to be less important than the benefits of improved daytime sleepiness and other sleep apnea symptoms. Therefore, despite the presence of irreversible long-term occlusal changes, OA therapy should be considered a lifelong treatment for patients with OSAHS.

ADHERENCE AND PATIENT PERCEPTIONS

Adherence is defined by the World Health Organization as "the extent to which a person's behavior—taking medications, following a diet, and/or executing lifestyle changes—corresponds with agreed recommendations from a healthcare provider."[95] Numerous factors affect treatment adherence, including social and economic factors, health care system/team, characteristics of the disease, disease therapies, and patient-related factors. The consequences of a nontailored treatment with poor patient adherence are related to poor health outcomes and increased health care costs. As stated before, OSAHS is a chronic disease, and treatment with either PAP or OAs requires cooperation of the patient.

Treatment preference has been correlated to the degree of effectiveness, the impact of treatment on quality of life improvements, and the perceived severity of side effects.[96] Treatment expectations, lifestyle and personality of the patient, marital status, perceived stigma related to the treatment, and cost of treatment may also have an impact on a patient's preferences.[97] It is therefore important to include patients in the decision-making process regarding their treatment choices to promote better acceptance of and adherence with therapy as well as to identify potential barriers of one treatment option over another.

Treatment adherence with OAs appears to be dependent on the type of the appliance, disease severity, and patient supervision.[98,99] For example, adherence rates are greater with custom-made MAS than with tongue-retaining or "boil and bite" type appliances.[100,101] Studies have shown that about 75% of patients remained adherent after 12 months of treatment, which may decrease to 50% after 5 years.[102-105] Table 147-4 summarizes adherence rates in recent studies.[106]

Table 147-4 Adherence Rates of Oral Appliance Therapy in Selected Studies that Evaluated Use after a Minimum of 1 Year

Author	Interval (Months)	No. of Patients	Compliance Rate WCS/BCS
McGown,[99] 2001	22	166	42/56
Dort,[98] 2004	22	110	40/57
Marklund,[103] 2004	12	630	75/76
Almeida,[53] 2005	68	544	30/64
Marklund,[90] 2006	60	450	56/56
Gindre,[150] 2008	17	66	82/82

BCS, Best case scenario, relates to the percentage of patients who responded to the questionnaire only; WCS, worst case scenario, relates to an analysis interpretation of patients who did not return the questionnaires as compliance failures.
From Fleetham J, Almeida FR. Oral appliances. In: McNicholas WT, Bonsignore MR, editors. *European respiratory monograph*. Sheffield, UK: European Respiratory Society Journals; 2010. p. 267–85.

Adherence with OAs is consistently greater than that seen with PAP.[1] Whereas the existing literature strongly suggests that adherence with OAs is superior to that with PAP, most studies are based on subjective reports.[2,25] Recently, devices that provide a means to objectively monitor OA use have been developed.[54] Three microsensors are currently available that can be integrated into OAs: TheraMon (IFT Handels- und Entwicklungsgesellschaft GmbH, Handelsagentur Gschladt, Hargelsberg, Austria), AIR AID SLEEP (AIR AID GmbH & Co KG, Frankfurt, Germany), and Denti-Trac (Braebon Medical Corporation, Kanata, Canada). All three microsensors provide reliability and accurate wear time and can be incorporated into OAs fabricated from different materials.[107]

SLEEP BRUXISM IN PATIENTS WITH OBSTRUCTIVE SLEEP APNEA–HYPOPNEA SYNDROME

Sleep bruxism (SB) is classified as a sleep-related movement disorder in the *International Classification of Sleep Disorders*.[108] For more information on SB, see Chapters 144 to 146.

During sleep, individuals with SB will engage in jaw movements described as tooth clenching and grinding, but on the basis of electromyographic recordings of masseter and temporalis, this has been described as rhythmic masticatory muscle activity (RMMA) to facilitate the recognition of and further refine the specificity of the SB clinical diagnosis.[29,108]

Although the etiology of SB remains unclear, 50% to 80% of SB episodes are associated with sleep arousals in otherwise healthy subjects.[109-114] Sleep arousal precedes the onset of RMMA-SB, and it is characterized by a rise in sympathetic cardiac activation, blood pressure, muscle tone, and breathing amplitude.[111,114-116] Clearly, there is not a single cause for SB, and the role of stress, anxiety, breathing, and cardiac reactivity in relation to wake and sleep hyperarousal is a current question of interest.

SB has been frequently observed with other sleep conditions, such as snoring, upper airway resistance syndrome, and obstructive sleep apnea.[73,113,117-121] It has been estimated that up to 50% of both adults and children with OSAHS may have comorbid SB; in such cases, it is named secondary SB.[73,122,123] In addition, a positive correlation between the severity of OSAHS and the frequency of tooth grinding and clenching has been established.[117,124] A questionnaire-based epidemiologic study by Ohayon et al[120] suggested that OSAHS is a low but significant risk factor for SB (odds ratio, 1.8; 95% confidence interval, 1.2–2.6). All this evidence is derived from questionnaires and lacks power to explain a causative relationship.

A recent population-based community sleep study, which investigated more than 1000 adults using polysomnography, observed no difference in the AHI between individuals with and without SB.[109] This is not surprising; the prevalence of SB, which decreases with age, and the rise in the prevalence of SDB prevent such an association. Furthermore, a recent polysomnographic study failed to show a clear temporal association between RMMA-SB and OSAHS. In one study of OSAHS patients, the majority of SB events were temporally related to apneas and hypopneas, occurring between 0 and 30 seconds after the event. However, 25% of SB episodes showed an opposite temporal relationship (i.e., the SB event occurred before the obstructive event), and 20% had no time correlation with obstructive events.[125] Furthermore, it is possible that SB resulted in some cases from unrecognized or unmeasured respiratory effort–related events. Regardless, this study did demonstrate a significant association of SB in those with SDB and identified a potential dependent relationship between the two conditions in the majority of events.

Given the limited and conflicting existing evidence, the nature of the relationship between SDB and SB remains controversial.

Some authors have suggested that SB-related masticatory muscle activities have a role in reinstating upper airway patency after an episode of respiratory obstruction during sleep.[114,116,125-127] On the basis of this suggestion, it would appear that SB occurs secondary to obstructive events and is a consequence of an OSAHS-related phenomenon similar to sleep arousal or oxygen desaturation. If this is true, treatments of OSAHS would be expected to improve SB.[117,124,128] Indeed, some clinical studies have shown a decrease in SB activity after treatment of OSAHS using different modalities, including adenotonsillectomy, OA, and PAP.[126,129-131] However, as described before, the temporal association between OSAHS events and the masticatory muscle activity has not been confirmed in most of the available studies.[73,125,128,132] Alternatively, SB has been described as masticatory movements required to lubricate the oropharyngeal structures during sleep, which may be particularly dry in snorers and OSAHS patients.[133,134] However, in all cases, the role of sleep arousal and related autonomic sympathetic activation has to be taken into account while investigating the potential relationship between OSAHS and SB.[113,114,116,117,125,132]

The Impact of Bruxism Therapies on the Upper Airway

Management strategies for SB include dental (i.e., OAs and guards), cognitive-behavioral, and pharmacologic approaches (see Chapter 150 for more information).[114] These approaches typically aim to reduce SB detrimental consequences on the teeth (e.g., tooth wear) and dental restorations/prosthesis (e.g., fracture) and to reduce accompanying orofacial complaints (e.g., pain, headache). However, these treatment options largely mitigate the consequences of SB and do not address the SB events or potential underlying etiology or role of comorbidities such as OSAHS.

To protect dental surfaces and to relax the masticatory muscles, various designs and types of OAs are used. The soft vinyl mouth guards or hard acrylic occlusal splints have been extensively used in the management of SB.[114,135-139] These devices are more frequently designed to cover maxillary dentition, and patients are instructed to use them during sleep. The exact mechanism of action is still under debate, and there is no evidence to support their role in halting SB. Their main effect is to protect teeth against damage. Moreover, the lack of well-designed randomized controlled clinical trials and long-term studies makes it difficult to assess their true effectiveness on SB and oropharyngeal functions during sleep.[137] The majority of studies show a decrease in the RMMA-SB index by 40% to 50% during the initial period of treatment (2 to 6 weeks), regardless of the specific design of the occlusal splint.[136,138-140] However, the effect appears to be transitory, with values typically returning to pretreatment levels after a short time, and the outcomes are highly variable between subjects. Indeed, some studies have also reported no effect on or even an increase in electromyographic activities during sleep when an occlusal splint is worn, especially with soft mouth guard designs.[139,141,142]

The Use of Oral Appliances in Patients with Sleep Bruxism and Obstructive Sleep Apnea–Hypopnea Syndrome

Although occlusal splints have been thought to be a safe and conservative option for management of SB, few studies have investigated the effects of OAs designed to treat SB in patients who may have concomitant OSAHS. A pilot study of 10 patients with OSAHS found that a maxillary occlusal splint increased the AHI by more than 50% in half of the patients, likely by reducing the intraoral space and changing the tongue position during sleep.[143] Another group reproduced the direction taken by the pilot study, a risk for exacerbation of breathing in OSAHS patients but with a milder effect due to different morphologic patient characteristics.[144] Although there is a paucity of data, the potential adverse influence of maxillary occlusal splints on snoring and the respiratory disturbance index cannot be ignored, and clinicians should be cautious and consider the potential medical and dental complications of occlusal splints, especially when SB and OSAHS occur in the same patient. Given the potential association between SB and OSAHS in some patients and the potential negative effects of traditional SB therapies with SDB, it is imperative to assess risk of SDB in SB patients with OSAHS mild symptoms to select appropriate treatment options. In general, when SB is concomitant with OSAHS or when SDB is suspected, a mandibular advancement appliance would be preferable.

Some clinical studies have shown a decrease in RMMA-SB activity by using mandibular advancement devices in the absence of OSAHS.[126,145,146] These duobloc titratable OAs appear to be effective in decreasing SB and have been shown

to reduce up to 70% of bruxism events in otherwise healthy young adult subjects.[126,145,146] Duobloc OAs also appeared to relieve morning headaches and snoring concomitant with SB.[126,147] In an experimental study among adolescents, SB-related symptoms were significantly mitigated and sleep structure and quality were preserved by the use of mandibular advancement OAs during sleep.[126] In general, OAs are well tolerated by SB patients, especially if some range of jaw movements is permitted.[126,145-147] The use of mandibular advancement OAs for the treatment of SB seems promising but probably for cases with SDB symptoms. Intuitively, OAs seem ideally suited to treatment of patients with concomitant SB and OSAHS. However, data on the utility of these devices in patients with both conditions are lacking, and the long-term effectiveness of OAs in patients with both SB and OSAHS has not been established.

FUTURE DIRECTIONS

OA therapy has emerged as the main alternative to PAP in the treatment of OSAHS. There is now a strong evidence base demonstrating the benefits of this therapy for improving physiologic sleep measures, daytime sleepiness, and quality of life in OSAHS patients across the spectrum of disease severity. The existing literature suggests that the superior treatment efficacy of PAP is mitigated by inferior compliance relative to OA therapy, resulting in similar health outcomes. Long-term comparative effectiveness studies, using patient-centered and clinically important outcome measures, are required to verify this. The advent of objective compliance monitors for OA therapy is critical to such studies.[148] Cost-effectiveness studies are also warranted to appropriately inform and direct clinical care. In addition, prospective validation studies are required to evaluate predictors of treatment outcome, and more research is needed to determine optimal titration protocols to increase the effectiveness of OA and to decrease the time taken to attain optimal treatment.

Future studies are needed to compare the effectiveness of different types of appliances and different design features (e.g., the amount of vertical opening). Likewise, these studies ideally would help differentiate which patients should receive custom-made, titratable devices and identify those who could be appropriately managed with less expensive, fixed devices. Ongoing refinements and standardization of appliance design may eventually lead to improved outcomes and will be enhanced by a better understanding of the mechanisms of action of OAs.

Snoring and SDB are chronic and progressive conditions, raising the possibility that early intervention of snoring may retard the development of OSAHS. OA therapy would appear to have a major role in such a preventive approach, and further work is warranted to evaluate this possibility.

OSAHS is increasingly recognized as a heterogeneous disorder with multiple pathophysiologic causes. Currently, PAP represents a "one size fits all" solution by preventing upper airway collapse. However, new concepts in OSAHS pathogenesis and phenotypes have recently emerged and could help the field move toward a future "personalized medicine" approach in which treatment is tailored to the patient. Finally, the use of OAs for SDB in the presence of comorbidities, such as SB, headache, pain, gastroesophageal reflux, and opioid use, needs to be better delineated.

CLINICAL PEARL

OAs provide effective therapy for patients with OSAHS and can be used as primary therapy for mild to moderate cases or as an alternative treatment for individuals who fail to respond to, are not compliant with, or are intolerant of PAP and also for managing SB alone or concomitant with SDB. Several different device designs are available, each with inherent advantages and limitations. Types of OA design have to be selected for patient characteristics, dominance of OSAHS or SB, and comorbidity. This makes interpreting the published literature and understanding the role of OAs in clinical practice somewhat difficult. However, the data strongly suggest that custom-made, titratable devices with a small anterior vertical opening offer the most advantageous therapy for most patients with SDB and OSAHS. Appropriate and valid means to determine the required degree of mandibular advancement are not yet available, and much more work is awaited to improve OA efficacy and effectiveness.

SUMMARY

OAs, occlusal splint on single dental arch, can be used to prevent SB consequences on the teeth and pain, but in the presence of OSAHS, they may be contraindicated. Use of OAs with mandibular advancement is the ideal approach to manage OSAHS, conversely to occlusal splint. The putative mechanism is associated with increase in the anterior-posterior diameter of the upper airway space by advancing the oropharyngeal anatomic trio: the mandible, tongue, and soft palate. OAs with such advancement properties also add one important feature: they prevent posterior retrusion of the mandible during sleep, reducing the upper airway collapse. However, there appears to be a more robust and complex mechanism of improved airway patency and stability of upper airway muscle activity provided by these devices.

Although OAs provide reliable and effective treatment for most individuals with OSAHS, especially those with mild and moderate disease, there is a growing body of literature supporting that OAs are effective across a larger range of OSAHS severity. Nowadays, OAs for OSAHS should be considered an additional first-line treatment option. However, as with most appliances, success with OAs is not optimal. Even with titratable devices, a residual AHI is expected in more than 50% of individuals, and much work is expected to achieve higher success on most relevant outcomes, such as optimal sleep continuity, and reversal of signs and symptoms, such as sleepiness, cognitive alteration, blood pressure, and inflammatory reaction linked to SDB.

We recognize that OAs are not as effective as PAP in reducing the AHI, but they perform equally well compared with other measured variables, plus the OAs are more likely to be preferred by patients with better measures of adherence. Treatment with OAs provides a unique opportunity for interdisciplinary collaboration. Diagnosis and follow-up are performed by sleep medicine physicians; OA selection and follow-up are performed by dental sleep medicine professionals. OAs for either SB or OSAHS cannot be made for patients with gum infection, periodontal disease, or jaw pain without dental supervision. OAs may also induce orthognathic and dental changes in some patients that only a dentist can monitor and manage.

Selected Readings

Balasubramaniam R, Klasser GD, Cistulli PA, Lavigne GJ. Link between sleep bruxism, sleep disordered breathing and temporomandibular disorders: an evidence-based review. *J Dent Sleep Med* 2014;**1**: 27–37.

Barnes M, McEvoy RD, Banks S, et al. Efficacy of positive airway pressure and oral appliance in mild to moderate obstructive sleep apnea. *Am J Respir Crit Care Med* 2004;**170**:656–64.

Chan ASL, Cistulli PA. Oral appliance treatment of obstructive sleep apnea: an update. *Curr Opin Pulm Med* 2009;**15**:591–6.

Dieltjens M, Verbruggen AE, Braem MJ, et al. Determinants of objective compliance during oral appliance therapy in patients with sleep-related disordered breathing: a prospective clinical trial. *JAMA Otolaryngol Head Neck Surg* 2015;**24**:894–900.

Krishnan V, Collop NA, Scherr SC. An evaluation of a titration strategy for prescription of oral appliances for obstructive sleep apnea. *Chest* 2008;**133**:1135–41.

Kushida CA, Morgenthaler TI, Littner MR, et al. Practice parameters for the treatment of snoring and obstructive sleep apnea with oral appliances: an update for 2005. *Sleep* 2006;**29**:240–3.

Marklund M, et al. Oral appliance therapy in patients with daytime sleepiness and snoring or mild to moderate sleep apnea: a randomized clinical trial. *JAMA Intern Med* 2015;**175**(8):1278–85.

Phillips CL, Grunstein RR, Darendeliler MA, et al. Health outcomes of continuous positive airway pressure versus oral appliance treatment for obstructive sleep apnea: a randomized controlled trial. *Am J Respir Crit Care Med* 2013;**187**:879–87.

Ramar K, Dort LC, Katz SG, et al. Clinical practice guideline for the treatment of obstructive sleep apnea and snoring with oral appliance therapy: an update for 2015. *J Clin Sleep Med* 2015;**11**(7):773–827.

A complete reference list can be found online at ExpertConsult.com.

Chapter
148

Anesthesia in Upper Airway Surgery for Obstructive Sleep Apnea

David R. Hillman; Peter R. Eastwood; Olivier M. Vanderveken

Chapter Highlights

- This chapter presents an overview of anesthetic considerations relating to upper airway surgery for patients with obstructive sleep apnea.
- Insightful anesthetic management of such cases requires an appreciation of the similarities and differences between the sleep and anesthetic states in relationship to their effects on upper airway and breathing function and on the arousal responses that help protect against disturbance of these essential physiologic functions.
- These issues are discussed in regard to both sedation for preoperative evaluation for obstructive sleep apnea surgery and perioperative management when surgery is undertaken.

Upper airway surgery is an undertaking that requires particularly close cooperation between surgeon and anesthesiologist. Patients presenting for these procedures do so because of compromised upper airway structure or function. In some circumstances, upper airway obstruction is the presenting complaint, as may occur with tumors in the upper airway. In others, the potential for obstruction may be great, as is the case for patients with obstructive sleep apnea (OSA). Managing such airways is challenging in the perioperative period. In anesthesiology terms, they are known as "difficult." A *difficult airway* has been defined by the American Society of Anesthesiologists as one that causes a conventionally trained anesthesiologist difficulty with face mask ventilation delivered to the upper airway, difficulty with tracheal intubation, or both.[1] Patients with OSA are vulnerable in both of these respects. In addition, they are, of course, at increased risk for upper airway obstruction whenever they are asleep or sedated. Whereas the sleeping patient is naturally protected from prolonged asphyxia by a capacity to arouse and reactivate effective breathing function, mild sedation compromises arousal mechanisms, and deep sedation or anesthesia abolishes these responses.[2] Patients with OSA are therefore particularly vulnerable throughout the perioperative period because of their airway compromise, compounded by the effects of anesthetic, opioid, and other sedative drugs and the risk of postoperative hemorrhage or edema associated with upper airway surgery.

In keeping with the orientation of this book toward sleep disorders, the focus of this chapter is on anesthetic considerations relating to upper airway surgery for OSA. Surgery frequently is undertaken to treat OSA in children, in whom tonsillar and adenoidal hypertrophy is a common cause.[3] Upper airway surgery also is regularly undertaken in adults for treatment of OSA, with several specific indications: removal of certain obstructing lesions in the upper airway; correction of abnormalities of the facial skeleton; unacceptability of nonsurgical treatments such as continuous positive airway pressure (CPAP) therapy and oral appliance therapy; and the need to relieve nasal obstruction to allow use of nasal masks for more comfortable, less intrusive delivery of CPAP than that offered by face masks.[4] A variety of procedures are available including nasal, palatal, and tongue base surgery; adenotonsillectomy (if hypertrophy is present); and a variety of orthognathic procedures.[5,6] The choice between them requires careful preoperative evaluation, often involving upper airway endoscopy and radiography including cephalometry, conventional computed tomography (CT), and, increasingly, three-dimensional studies using cone beam CT scans.[7]

Insightful anesthetic management of such cases requires an appreciation of the similarities and differences between the sleep and anesthetic states in relationship to their effect on upper airway and breathing function and on arousal responses that act to protect against disturbance in these functions. One facet of these shared considerations is illustrated by the use of anesthesia to simulate sleep-like conditions during drug-induced sedation endoscopy (DISE), sometimes termed drug-induced sleep endoscopy—a test to evaluate suitability for and type of OSA-related surgery, as described further on. More broadly, anesthetic management of patients with OSA presenting for upper airway surgical procedures must take into account several important factors: the shared influences of sedation, anesthesia, and sleep on upper airway behavior; how surgery can affect this behavior; the potential influences of OSA comorbid conditions including obesity and heart disease; and the early postoperative challenges associated with emergence from anesthesia, pain and its management, and the possibility of airway compromise from edema or hemorrhage.

Children present particular difficulties because of their small airway size and lung volumes relative to body size, predisposing them to upper airway obstruction and to rapid desaturation under such circumstances. Congenital problems associated with airway compromise, such as Down syndrome,

also regularly manifest in childhood, further highlighting the challenges that confront pediatric anesthesiologists.[8] Relatively recently, DISE has been introduced into preoperative evaluation of pediatric patients with OSA, adding to these challenges.[9,10]

PREOPERATIVE EVALUATION OF PATIENTS PRESENTING FOR OBSTRUCTIVE SLEEP APNEA SURGERY

Preoperative evaluation of patients presenting for general anesthesia of any kind should include consideration of the possibility of OSA, in view of its prevalence and the associated perioperative risks.[11] Patients presenting for OSA surgery, however,usually have already had their OSA well characterized by a diagnostic home or laboratory-based sleep study. Often a trial of CPAP to treat the problem may be under way, or CPAP therapy may be a more-or-less established management component. Emergent or subsequent problems with compliance arising from difficulty in accepting or tolerating CPAP, usually because of its intrusive nature, are well documented as a common reason why adults patients with OSA present for upper airway surgery.[12] Other patients may elect nasal surgery to allow easier application of CPAP delivered by means of a nasal mask; otherwise, a more cumbersome face mask is the required interface when there is nasal obstruction.

Often, upper airway surgery for OSA is preceded by a surgeon-initiated investigation to determine the primary site of obstruction (velo-, oro-, or hypopharyngeal) and the nature of the obstructive process (lateral, anteroposterior, or concentric narrowing). This assessment allows the surgeon to determine the best procedure to counteract the problems identified by it.

Apart from radiographic imaging (such as cephalometry and CT scans), such evaluations usually involve endoscopic visualization of the upper airway under conditions conducive to obstruction. Early methods involved the performance of Mueller maneuvers (inspiratory effort against an obstruction—the opposite of the Valsalva manuever) during awake endoscopy or with the patient under mild sedation.[13] However, this approach is far removed from sleep-like conditions, in which the upper airway muscles are relaxed and obstruction occurs without exaggeratedly negative intraluminal pressures, so it has limited validity as an assessment of site and nature of upper airway collapse during sleep.[14]

Drug-Induced Sedation Endoscopy

To better simulate the conditions of sleep, DISE has evolved as a test to help determine suitability for surgery and the type of surgical procedure to be undertaken.[15,16] This procedure directly involves anesthesiologists because drugs are administered (usually intravenous propofol with or without midazolam) to produce sedation and sleep-like muscle relaxation, which induces snoring and upper airway obstruction (partial or complete) in predisposed patients. A recent European position paper on DISE provides an overview of the possible protocols for sedation during DISE, the indications for the procedure, and how the findings might be reported.[17]

Of note, with use of mild (wakeful) sedation, the upper airway muscles remain quite active, and sleep-like conditions cannot be assumed.[18] It is only if sleep itself supervenes or if sedation is deepened to a level at which consciousness is lost through a direct drug effect that behavior of the relaxed upper airway is observed. Indeed, a steplike reduction in phasic genioglossus muscle activity, a major upper airway dilating force, is seen at the transition to unconsciousness at sleep onset and with anesthetic induction.[18,19]

Observations made under the relaxed conditions of DISE will help determine suitability for non-CPAP treatment such as upper airway and skeletal surgery and for oral appliance therapy, as well as guiding the choice of surgical procedure to be undertaken. Resolution of pharyngeal collapse at all levels of the upper airway is necessary to achieve successful treatment in the individual patient. Indeed, it has been demonstrated that multilevel collapse is present in a majority of patients with OSA, and that the prevalence of complete collapse and multilevel collapse increases with increasing OSA severity and overweight and obesity.[20,21]

The use of DISE is based on the concept that upper airway behavior during drug-induced unconsciousness is similar to that in natural sleep, with similar reductions in muscle activation, respiratory drive, and reflex gains at transition to unconsciousness in each state.[18,19] The results of a study evaluating polysomnography with or without propofol administered by continuous target-controlled intravenous infusion demonstrated that although propofol significantly changes sleep macroarchitecture, the main respiratory parameters, apnea-hypopnea index and mean arterial oxygen saturation, remain unaffected.[22]

However, a critical difference that is highly relevant to the considerations that follow (and to recovery after the DISE procedure) is that with sleep, the capacity for spontaneous arousal is preserved, whereas with drug-induced unconsciousness, a dose-dependent depression of arousal responses prevails. This depression persists until physiologic drug elimination allows return of consciousness, so the anesthetized patient is highly vulnerable to asphyxia if upper airway obstruction occurs and is not detected by and dealt with by attending medical or nursing staff.

ANESTHESIA FOR SURGERY TO TREAT OBSTRUCTIVE SLEEP APNEA

Patients with OSA present the anesthesiologist with many challenges. The problems they experience during sleep signify the presence of a narrow airway predisposed to obstruction when the upper airway (and other) muscles relax with onset of anesthesia. Obesity and presence of other comorbid conditions (such as hypertension—systemic and pulmonary, cardiovascular disease, cerebrovascular disease, metabolic syndrome, atrial fibrillation, or heart failure) may be contributing factors. If sufficiently severe, obesity also may predispose the patient to atelectasis under anesthesia and to hypoventilation if spontaneous ventilation is preserved.[23,24]

A reduction in functional residual capacity (FRC) is another phenomenon that occurs with the muscle relaxation (in this case of chest wall muscles) that accompanies onset of anesthesia (and of sleep). The decrease can be profound in the morbidly obese, with FRC dropping to near residual volume.[24] This reduction in FRC is associated with atelectasis in the dependent parts of the lung with consequent shunt and hypoxemia. The loss of lung volume also is a significant contributor (along with upper airway muscle relaxation) to the

Table 148-1 Obstructive Sleep Apnea (OSA) and Perioperative Risk*

Risk Factor	Potential Perioperative Consequences	Identification of Risk	Reducing Risk
OSA	Upper airway obstruction with asphyxiation if unrecognized and untreated	Clinical evaluation Screening tools (e.g., STOP-Bang) Sleep study when indicated	Early identification Preparation for potential difficulties with airway management (including intubation) Techniques to minimize postoperative sedation Perioperative use of CPAP therapy Intensive monitoring until sentient and arousal responses unimpaired Referral for definitive diagnosis and treatment when OSA is first suspected perioperatively
OSA-predisposing factors Obesity Familial Craniofacial Mandibular retrusion Maxillary hypoplasia Other indicators of difficult intubation	OSA-related problems (above) Difficult tracheal intubation Difficult airway management Delayed extubation/ reintubation Atelectasis, hypoventilation when patient is morbidly obese	Clinical evaluation Endoscopy Cephalometry Upper airway imaging (CT, MRI)	Awareness of predisposition and associated risks Monitoring for these risks Preparation for potential risks, including availability of appropriate equipment to manage them Techniques to minimize postoperative sedation
OSA comorbid conditions Hypertension Cardiovascular/ cerebrovascular disease Metabolic syndrome Depression	Worsening of comorbid condition Delayed recovery	Clinical evaluation Biochemical testing when indicated	Optimize control preoperatively Careful perioperative management

*See text for further details.
CPAP, Continuous positive airway pressure; CT, computed tomography; MRI, magnetic resonance imaging.

increased airway collapsibility observed during anesthesia, because of associated loss of longitudinal traction on the upper airway.[25] Conveniently, CPAP counteracts both problems, providing a pneumatic splint to the upper airway and increasing FRC, helping to prevent or recruit atelectatic lung, and offsetting the negative influence of volume loss on upper airway patency.

When hypoventilation is a prominent feature, bilevel ventilatory assistance is a preferred form of positive airway pressure therapy for obese patients, because it combines a background level of pressure to provide the benefits of CPAP with inspiratory pressure support to counteract the inadequate inspiratory flow.[26]

OSA is not restricted to the obese, however, and other pathogenetic factors are involved that are relevant to anesthesia. These include facial skeletal characteristics such as retrognathia and maxillary hypoplasia.[27] Such anatomic configurations, perhaps more than obesity itself, can present substantial challenges to the anesthesiologist. Potential problems include difficulty in performing endotracheal intubation and/or in maintaining airway patency during mask ventilation. Patients presenting such problems are said to have a "difficult airway." The presence of a difficult airway under anesthesia is an indicator of vulnerability to OSA.[28] Conversely, OSA is a risk factor for difficult intubation and/or difficulty with airway maintenance during the anesthetic procedure.[29] Hence the anesthesiologist should approach patients with OSA prepared for a difficult intubation. It is likely that in the assessment of such patients, other factors indicating this possibility will be present, such as oropharyngeal crowding with high Mallampati scores, retrognathia with reduced thyromental distances and/or increased mandibular angulation, or increased neck circumference.[28] Increased neck circumference is a risk factor for OSA that is independent of obesity, although it reflects central fat deposition. Patients with large muscular necks also are at increased risk for OSA and may present additional airway management difficulties under anesthesia.[30] Apart from clinical evaluation, the anesthesiologist also should take advantage of the various radiographic and endoscopic preoperative investigations undertaken by the surgical team to assess upper airway structure and function in preparation for surgery (as outlined previously), to develop greater insights into the challenges ahead (Table 148-1).

Tracheal Intubation

Although some upper airway surgery procedures relevant to OSA can be done using local anesthesia (such as minor nasal surgery undertaken to improve the prospects of successful nasal CPAP delivery in patients with OSA along with nasal obstruction), most cases require general anesthesia. In such cases, the patient is almost always tracheally intubated. The

presence of an endotracheal tube allows the anesthesiologist to step back from immediate proximity to the head, to allow surgical access, while ensuring airway patency and, in addition, ensuring protection of the lower airway from aspiration of blood and other material during the course of the procedure. The principal challenges in dealing with patients with obstruction-prone upper airways lie in placing the tracheal tube at the start of the procedure and in ensuring airway patency after removal of the tube on completion of anesthesia.

Airway Patency after Extubation

Several measures are required to ensure airway patency after extubation. First, apart from adequate reversal of any neuromuscular blockade if such drugs have been used, it is prudent to ensure that the patient has regained consciousness before extubation. Early return of consciousness requires the choice of appropriate anesthetic drugs and titration of doses to facilitate this outcome. Second, extubation in the lateral posture should be considered when practicable, with the patient maintained in this position during sleep for the postoperative period in the hospital. Third, CPAP or other positive airway pressure therapies should be readily available and used during sleep or sedation in patients with threatened or actual upper airway obstructive episodes. This is not a problem with CPAP-compliant patients, who expect the continuation of such therapy. It is more problematic in noncompliant patients, which is a good reason for preoperative induction of CPAP therapy when possible. Although the nose is the favored route of administration of CPAP for a majority of users, a face mask may be required perioperatively, particularly in the presence of edema or other conditions causing nasal obstruction. In the case of nasal surgery, nasal packs should be avoided when possible. Furthermore, an important caveat is that CPAP therapy may be needed in the early postoperative period despite the fact that the aim of surgery was to restore airway patency. This is because edema, secretions, blood, and clots can temporarily worsen airway patency. Although CPAP is a mainstay, use of other devices to improve airway patency, such as mandibular advancement devices, may have a perioperative role. Available literature to guide the perioperative use of these other devices is limited at present, however.

Early Postoperative Care

Finally, the patient must be closely monitored while he or she remains at increased risk. Clinical care thus requires a higher nurse-to-patient ratio than in the general ward and continuous monitoring of oxygenation (oximetry) and ventilation (capnography, oronasal airflow). Such facilities are available in the postanesthetic care unit and other high-dependency areas. Oxygen therapy should be used with caution because it can conceal obstructive events, potentially leading to their prolongation. It is for this reason that monitoring oronasal airflow, either directly or with capnography, adds value to oximetry monitoring.[31] Attending staff members must be expert in the use of CPAP and noninvasive ventilation. Discharge of the patient from such an environment demands return of consciousness and subsequent unimpeded arousal responses to threatened airway compromise. To ensure this outcome, it is highly desirable to avoid opioids and other drugs with sedative potential whenever possible. Alternative analgesic techniques include use of nonsteroidal anti-inflammatory drugs and acetaminophen and analgesic-sparing strategies such as use of corticosteroids. There also is a place for use of topical anesthesia and nerve blocks in the perioperative setting.[32]

Postoperative edema and secretions can temporarily compromise airways in which procedures have been undertaken to ultimately improve patency. Hemorrhage also is a risk, particularly in the early postoperative period.

Postoperative Sleep

Owing to the aforementioned considerations, the perioperative period is one of vulnerability. Added to these potential airway and ventilatory impairments is the disruption to sleep that occurs during the first few days postoperatively. Such sleep impairment increases the risk of postoperative delirium. Furthermore, as this early sleep disruption settles, a night or two of rapid eye movement (REM) sleep rebound may follow, with an increased proportion of sleep spent in the REM stage. This is of relevance because REM is the stage of sleep during which muscle activation and respiratory drive are most depressed and, consequently, the risk of upper airway obstruction and hypoventilation is greatest.[33]

Discharge Requirements

The ideal circumstances for discharge of these patients from high-dependency care and from the hospital are achieved with the following clinical objectives: return of consciousness with no plan for subsequent use of opioids or sedatives that may impair arousal responses; resolution of edema and of the particular risk of hemorrhage; and, when indicated, ability to self-administer CPAP, with willingness to use it when asleep. In some settings, these conditions may be met in the early postoperative period, depending on the nature of the procedure and the anesthetic technique used. If so, and if any comorbid conditions are well controlled, same-day discharge from the hospital after surgery is possible.

These recommendations are consistent with current published guidelines for perioperative care of patients with OSA.[32,34] It must be recognized, however, that this is an evolving area, and such guidelines currently are based largely on expert opinion. An evidence base developed from careful study and analysis of methods to improve perioperative outcomes is slowly accumulating and can be expected to eventually result in well-supported recommendations. Much work in this area remains to be done.

CONCLUSIONS

Anesthesia for upper airway surgery to treat OSA is challenging. The surgery often is undertaken in patients who have been unable or unwilling to tolerate CPAP therapy or in those who have surgically correctable upper airway or craniofacial abnormalities. Sometimes it is undertaken to improve CPAP compliance by relieving nasal obstruction. Anesthetic management requires careful preparation, allowing for the possibility of difficult intubation, and careful management of extubation and emergence to ensure maintenance of airway patency particularly in the period that precedes full return of consciousness and ability to promptly arouse from subsequent sleep in response to hypopneic or apneic events.

CLINICAL PEARLS

- The similar effects of anesthesia and of sleep in reducing muscle activation and ventilatory drive mean that vulnerability to upper airway obstruction in one state indicates high risk of problems in the other.
- The anatomic features (e.g., retrognathia) that predispose to airway obstruction also increase difficulties with tracheal intubation and mask ventilation performed with the patient under anesthesia.
- A key difference between the states is suppression of the arousal responses that protect the sleeping patient by anesthetic agents, with increased risk of asphyxia in those with obstruction-prone airways, such as patients with OSA. Postoperative use of opioid and sedative drugs may extend these risks beyond the immediate postoperative period, and caution with use of these agents is mandatory.
- Postoperative edema further compounds potential problems for patients who have undergone surgery for OSA.
- Close peri- and postoperative monitoring and management must take these matters into account.

SUMMARY

Anesthesia for upper airway surgery to treat OSA is challenging. Often such surgery is undertaken in patients who have been unable or unwilling to tolerate CPAP therapy. Sometimes it is undertaken to improve CPAP compliance by relieving nasal obstruction. Preoperative surgical evaluation often entails DISE to assess suitability for and determine the optimal type of surgery. This procedure requires an appreciation of the relationships between drug-induced sedation and sleep in terms of muscle relaxation, ventilatory depression, and reflex suppression to ensure that airway behavior is observed under sleep-like conditions. Anesthetic management for the upper airway surgery requires careful preparation, allowing for the possibility of difficult intubation, and careful management of extubation and monitoring of emergence to ensure maintenance of airway patency, particularly in the period that precedes full return of consciousness and the capability for prompt arousal from subsequent sleep in response to hypopneic or apneic events.

Selected Readings

Brown KA. Outcome, risk, and error and the child with obstructive sleep apnea. *Paediatr Anaesth* 2011;**21**:771–80.

Chung F, Subramanyam R, Liao P, et al. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. *Br J Anaesth* 2012;**108**: 768–75.

De Vito A, Carrasco Llatas M, Vanni A, et al. European position paper on drug-induced sedation endoscopy (DISE). *Sleep Breath* 2014;**18**(3): 453–65.

Hillman DR, Walsh JH, Maddison KJ, et al. Evolution of changes in upper airway collapsibility during slow induction of anesthesia with propofol. *Anesthesiology* 2009;**111**:63–71.

Hiremath AS, Hillman DR, James AL, et al. Relationship between difficult tracheal intubation and obstructive sleep apnoea. *Br J Anaesth* 1998; **80**:606–11.

Leach J, Olson J, Hermann J, Manning S. Polysomnographic and clinical findings in children with obstructive sleep apnea. *Arch Otolaryngol Head Neck Surg* 1992;**118**:741–4.

Lee CH, Kim DK, Kim SY, et al. Changes in site of obstruction in obstructive sleep apnea patients according to sleep position: a DISE study. *Laryngoscope* 2015;**125**(1):248–54.

Pelosi P, Croci M, Ravagnan I, et al. The effects of body mass on lung volumes, respiratory mechanics, and gas exchange during general anesthesia. *Anesth Analg* 1998;**87**:654–60.

Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology* 2014;**120**(2):268–86.

Rosenberg J, Wildschiodtz G, Pedersen MH, et al. Late postoperative nocturnal episodic hypoxaemia and associated sleep pattern. *Br J Anaesth* 1994;**72**:145–50.

Vroegop AV, Vanderveken OM, Boudewyns AN, et al. Drug-induced sleep endoscopy in sleep-disordered breathing: report on 1,249 cases. *Laryngoscope* 2014;**124**:797–802.

A complete reference list can be found online at ExpertConsult.com.

Chapter 149

Upper Airway Surgery to Treat Obstructive Sleep-Disordered Breathing

Olivier M. Vanderveken; Aarnoud Hoekema; Edward M. Weaver

Chapter Highlights

- Management with continuous positive airway pressure (CPAP) remains the treatment of choice for moderate to severe obstructive sleep apnea (OSA). Some patients cannot or will not accept CPAP therapy, or CPAP may not be effective in certain cases, so use of other therapies may be indicated.
- Objective sleep testing is always required before the decision for surgical treatment, to identify those patients at risk for complications of OSA, to guide selection of appropriate management, and to provide a baseline to establish the efficacy of treatment.
- Surgical treatments for OSA fall into the following categories: surgical upper airway modifications including tongue suspension techniques, upper airway neurostimulation therapy, tracheostomy, skeletal modifications, and bariatric surgery.
- Objective sleep testing is always required after surgical treatment, to document efficacy.

OVERVIEW AND BACKGROUND

Management with continuous positive airway pressure (CPAP) remains the treatment of choice for moderate to severe obstructive sleep apnea (OSA).[1] Adequate CPAP treatment improves blood pressure to combat hypertension and reduces the risk of nonfatal and fatal cardiovascular events, and successful CPAP treatment has been shown to prolong survival.[2-5]

The clinical effectiveness of CPAP often is hampered and limited by important patient-specific factors: low rate of acceptance, poor tolerance, and suboptimal adherence.[6] Therefore many patients with OSA remain inadequately treated owing to inconsistent CPAP adherence,[6,7] with consequent reduced clinical effectiveness of therapy.[5,6,8-12] When the response to CPAP is considered inadequate as indicated by both symptom evaluation and objective monitoring data, despite intensive efforts to improve adherence to the CPAP regimen including trials of combination therapies, non-CPAP alternatives for the management of OSA need to be considered.[13,14] At this stage, proper and careful selection of the right patient for the right treatment option(s) is of utmost importance.[13]

The focus of this chapter is on upper airway surgery to treat OSA. Sometimes surgery is most helpful to facilitate CPAP or oral appliance therapy, such as with surgical correction of nasal obstruction. Alternatively, surgical management may be indicated in place of these other therapies—for example, when factors including unfavorable anatomy or patient intolerance preclude their use. It is important to recognize that often multimodality therapy is the best strategy, and surgery can be combined with any and all other OSA treatments including CPAP and non-CPAP approaches when indicated.[15]

Surgical modifications for treating OSA fall into five main categories: upper airway bypass procedure, surgical upper airway modifications including tongue suspension techniques, upper airway neurostimulation therapy, skeletal modifications, and bariatric surgery.

Objective sleep testing is required before initiation of surgical treatment, to identify those patients at risk for complications of OSA, to guide selection of appropriate management, and to provide a baseline to establish the physiologic efficacy of subsequent treatment.[14] Follow-up sleep studies are routinely indicated to monitor the response to surgery for OSA.

PREOPERATIVE EVALUATION

The comprehensive diagnostic workup of the patient with OSA should include a complete medical and thorough sleep history. Previous OSA treatments and their clinical and polysomnographic results also should be evaluated and documented in this workup, as described in later chapters (see Section 14 and Chapter 143 for orthodontic management). Before proceeding with surgery, the medical workup should identify complicating factors that can increase the medical, anesthetic, and surgical risk of the surgical procedure. A thorough upper airway examination should be performed, including assessment of the anatomy and functional status of the nose, mouth, pharynx, larynx, and related structures (e.g., craniofacial, dental, and airway structures). The OSA examination often includes upper airway endoscopy to visualize all the levels of the pharynx, larynx, and related structures—the palate, tongue base, the lateral pharyngeal walls, the vallecula, the pyriform sinuses, the larynx, and the epiglottis. The diagnostic pretreatment assessment needs to include a baseline sleep study. The polysomnographic recording should be recent

enough that it represents the current status of the OSA and ideally consists of a full-night diagnostic study.

The option of surgical interventions for OSA is based in part on the demonstration that the upper airway is significantly smaller in patients with OSA than in normal subjects, especially at the retropalatal and retroglossal levels.[16,17] The upper airway is described as a structure with a rigid support in its proximal (laryngeal) and distal (nasal) segments, with a collapsible portion in between. This collapsible anatomic portion extends from the soft palate to the epiglottis, with the size of its lumen subject to the influence of surrounding anatomy, presence of fat tissue, pressures gradients, tissue laxity, and dilator muscle activity. The potential loss of upper airway patency during sleep in patients with OSA is therefore usually localized to this so-called collapsible segment (Figure 149-1).

In the early days of OSA therapy, tracheostomy, or "upper airway bypass surgery," offered the only effective management, and it often was used in patients with profound OSA.[18] Tracheostomy, however, does not address the collapse directly in the vulnerable segment of the upper airway but instead successfully bypasses the obstruction, thereby moderating OSA. Tracheostomy is thus a highly effective treatment for OSA. On the other hand, tracheostomy is a physically and socially invasive procedure that may substantially reduce perceived quality of life.[18] The more commonly used upper airway surgical procedures nowadays have greater acceptance but are less effective.[19]

The decision process to select the appropriate upper airway surgical procedures for any specific patient are determined by the site(s), the degree, and the pattern of upper airway obstruction[13] (Figure 149-2). Indeed, most upper airway surgical approaches, apart from surgical treatments with a global effect (i.e., tracheostomy bypass, bariatric surgery, and maxillomandibular advancement [MMA]) (Table 149-1), ideally strive to intervene at the specific anatomic region(s) that become obstructed during sleep. In addition to the history, examination, and in-office awake upper airway endoscopy, a variety of other modalities may be used to assess sites, degree, and pattern of obstruction. Pharyngeal pressure measurements, imaging and endoscopic techniques can be used to gain insight into the pathogenesis of this disease, as well as to identify appropriate therapeutic options.[20-23] Imaging techniques available for the preoperative evaluation include radiography, fluoroscopy, conventional computed tomography

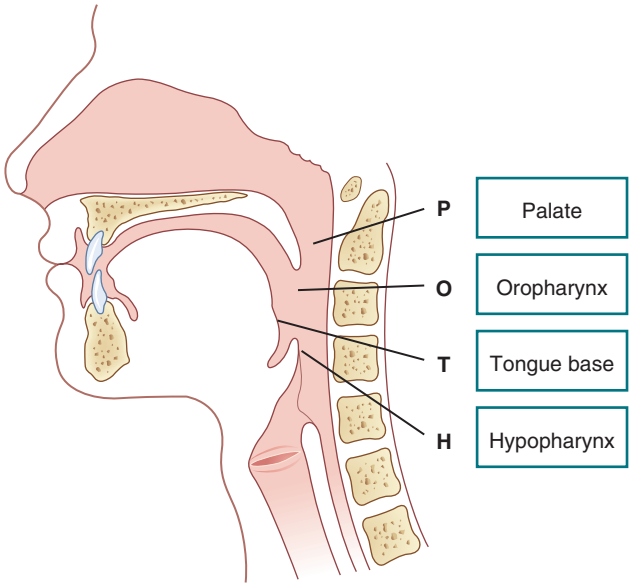


Figure 149-1 Sagittal cross section of the upper airway with the different levels prone to upper airway collapse during sleep within the "collapsible segment": the velopharynx with the palate as the most relevant structure; the oropharynx, including both the palatine tonsils and the lateral oropharyngeal walls; the level of the tongue base, including the lingual tonsils; and the hypopharynx, including the epiglottis and the hypopharyngeal lateral walls. The upper airway has rigid support in its proximal and distal segments but has a collapsible portion, the so-called "collapsible segment" extending from the soft palate to the epiglottis, with the size of its lumen subject to the influence of surrounding pressures and the activity of dilator muscles.

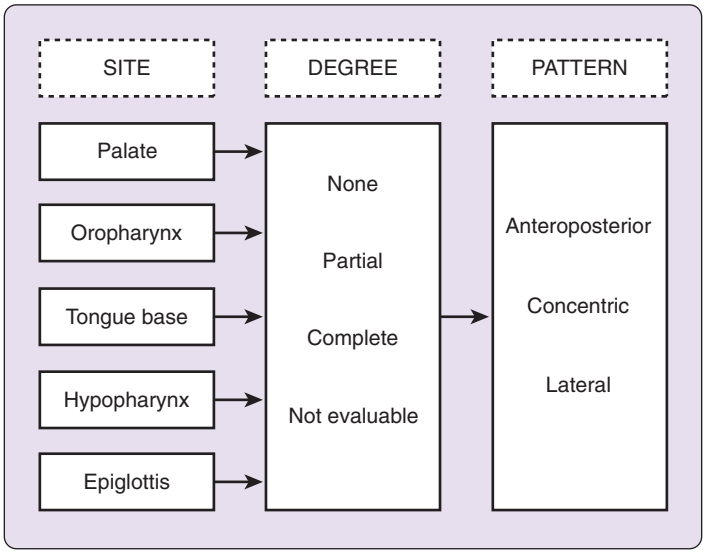


Figure 149-2 Scoring form for drug-induced sedation endoscopy (DISE): Reporting of the assessment of the site, the corresponding degree, and pattern of upper airway collapse. (From Vroegop AV, Vanderveken OM, Wouters K, et al. Observer variation in drug-induced sleep endoscopy: experienced versus nonexperienced ear, nose, and throat surgeons. *Sleep* 2013;36:947–53.)

Table 149-1 Treatment Options Other than Continuous Positive Airway Pressure (CPAP) for Patients with Obstructive Sleep Apnea

Category	Intervention/Procedure
General measures	Avoidance of sedatives Avoidance of alcohol Sleep hygiene Weight loss in case of overweight or obesity Avoidance of supine position during sleep in case of positional sleep apnea
Specific therapeutic options	Oral appliance therapy Surgery Surgical upper airway modifications Upper airway neurostimulation Skeletal modifications including maxillomandibular advancement Upper airway bypass: tracheostomy Bariatric surgery

Table 149-2 Sample List* of Site-Specific Surgical Upper Airway Modifications for Treatment for Obstructive Sleep Apnea

Site-Specific Category	Procedures
Nasal surgery	Septoplasty, septorhinoplasty Turbinate reduction Nasal valve surgery
Pharyngeal procedures	Adenoidectomy, tonsillectomy, lingual tonsillectomy Uvulopalatopharyngoplasty (UPPP) and variations (e.g., expansion sphincter pharyngoplasty, many others) Tongue base reduction Radiofrequency tongue reduction Transoral robotic surgery Coblation ablation or excision CO_2 laser excision Epiglottoplasty Genioglossus advancement Hyoid suspension Tongue suspension

*Not exhaustive.

(CT), magnetic resonance imaging (MRI), and three-dimensional studies of the upper airway using cone beam CT (CBCT) scans.[24-28] In addition, cephalometric analysis can be performed.[28-30,30a] More recently, computational models of the upper airway based on CT or MRI have been introduced to simulate the effects of upper airway manipulations and to predict success of various treatment options.[17,31-34]

The different techniques available for investigation of the upper airway all have their specific advantages and limitations.[20] Differences among the various techniques include aspects on invasiveness, exposure to radiation, costs related to the examination, and potential side effects.[20,22,35] In addition, a crucial question of whether the upper airway behaves differently during wakefulness versus during natural sleep remains unanswered to date.[36] Inasmuch as OSA occurs exclusively during sleep, it would be ideal to evaluate the upper airway during a full night of natural sleep; such investigation during natural sleep, however, is of limited feasibility, owing to the associated sleep disruption and the labor-intensive nature of such studies.

The technique of drug-induced sedation endoscopy (DISE), first described as "sleep nasendoscopy" by Croft and Pringle in 1991, has emerged as an alternative method to dynamically investigate the upper airway before non-CPAP treatment selection in patients with OSA.[13,36-38] Drugs typically used for DISE include propofol, midazolam, dexmedetomidine, and/or ketamine.[36,39] In a comparison of DISE with awake endoscopy, identical findings were observed in only 25% of the cases.[40] In this particular study reporting on 250 patients with OSA, the discrepancies between awake endoscopy and DISE involved the oropharynx in 33%, the hypopharynx in 50%, and the larynx (e.g., epiglottis) in 33% of the patients.[40] These findings have important implications for surgical treatment choices, especially with operations involving the tongue base.[41]

Recent studies indicate that both inter- and intra-rater reliability of DISE are reasonably good but that experience in performing DISE is needed in order to obtain reliable observations.[42-44] Much investigation, standardization, and optimization of DISE techniques are needed regarding factors such as method of sedation, depth of sedation, patient position, use of airway anaesthetics, duration of observation, and others.[36] In the literature, no consensus exists on a standard DISE classification system or quantification of degrees of collapse or obstruction; in Figure 149-2, a generic scoring form is depicted.[44]

Several studies confirm that the most frequent site of upper airway obstruction in patients with OSA is the palatal level[23,38,45]; however, a majority of patients with OSA have other sites of obstruction as well (multilevel collapse).[23,27] The probability of multilevel collapse increases with increasing OSA severity and with an increased level of overweight or obesity.[23] Logically, resolution of pharyngeal collapse at all levels involved in the upper airway collapse during sleep will be necessary to achieve successful management in the individual patient with OSA undergoing surgery.

Surgical Upper Airway Modifications

A schematic overview of several different site-specific techniques available for surgically modifying the upper airway in patients with OSA is provided in Table 149-2.

Nasal Surgery

Depending on the anatomic or functional abnormalities at the level of the septum, turbinates, and nasal valves, nasal surgery can be indicated to correct nasal obstruction. However, nasal surgery including septoplasty, septorhinoplasty, turbinate reduction, and nasal valve surgery usually is insufficient to treat OSA and cannot be recommended as a stand-alone procedure for the management of OSA in most patients.[46] On the other hand, nasal surgery might be performed in patients with nasal obstruction that experience problems with the use of CPAP or oral appliance therapy because of these nasal

complaints in order to improve adherence to treatment.[47] Successful nasal surgery may also be associated with a reduction in therapeutic CPAP pressure required to alleviate OSA.[48,49] Accordingly, a recent systematic review with meta-analysis of the existing literature on the relationship between nasal surgery and its effect on therapeutic CPAP pressure has been performed by Camacho et al.[50] The authors concluded that, indeed, isolated nasal surgery in patients with OSA associated with nasal obstruction reduces therapeutic CPAP pressures and that the data consistently suggest that isolated nasal surgery in these selected patients also increases CPAP use.[50]

Pharyngeal Surgery

The most commonly used surgical treatments for OSA are aimed at the pharyngeal airway where collapse and obstruction occur. It is important to recognize that an isolated single-level pharyngeal procedure would not be expected to solve multilevel obstruction. In cases of multilevel obstruction, isolated procedures might have a partial treatment effect or be one part of a multilevel treatment plan (e.g. staged surgery or multiple procedures). Adenotonsillar hypertrophy is the most common cause of OSA in children but a rare cause of OSA in adults.[46] As a consequence, adenotonsillectomy is the first-line therapy for nonobese children with OSA. In adult patients with OSA, (adeno)tonsillectomy can be recommended in the presence of (adeno)tonsillar hypertrophy with an expected reduction in sleep respiratory disturbances and improvement in architecture.[50a]

Uvulopalatopharyngoplasty (UPPP) is the most commonly performed pharyngeal surgical procedure for the treatment of OSA.[51] The classical UPPP procedure consists of partial excision and closure of the soft palate, uvula, and pharyngeal pillars, as well as tonsillectomy, if not previously performed[51] (Figure 149-3). A pooled data analysis of the literature (predominantly case series) published in 1996 revealed that UPPP had an overall success rate (i.e., reduction of AHI by more than 50%, with less than 20 events per hour of sleep, or reduction of apnea index by more than 50%, with less than 10 events per hour of sleep) of only 41% in unselected patients with OSA, but selecting patients with appropriate anatomy had an important impact on polysomnography outcomes.[18] When DISE is added to the diagnostic workup and preoperative evaluation for patient selection, it has been demonstrated that the success rate of UPPP increases over that for historical control data.[52] Recently, a randomized controlled trial comparing a modified UPPP with surgery delay showed a major reduction in OSA severity in the UPPP group compared with the control group (mean AHI reduction of 60%, versus 11%).[53] More important, UPPP improves clinical outcomes such as mortality risk, cardiovascular disease risk, motor vehicle crash risk, symptoms, and quality of life.[4,9,10,54-58]

The many variations of UPPP published in the literature aim to improve effectiveness, reduce surgical morbidity, and target specific patterns of collapse or obstruction. Examples are the uvulopalatal flap (UPF) technique, expansion sphincter pharyngoplasty, transpalatal advancement pharyngoplasty, and lateral pharyngoplasty.[59-61] The advantage of the UPF is that it is a potentially reversible flap that can be taken down if necessary to reduce the risk of nasopharyngeal incompetence; the technique comes with less postoperative side effects as there are no sutures and scar along the free edge of the palate[60] (Figure 149-4). The expansion sphincter pharyngoplasty (ESP) addresses lateral oropharyngeal or velopharyngeal collapse[62] (Figure 149-5). The early results with ESP are promising, and minimal complications have been reported with the procedure.[62] Transpalatal advancement pharyngoplasty is aimed at addressing anterior-posterior narrowing of the velopharynx, especially in concert with a long hard palate and vertically oriented soft palate.[59] The technical details and the wide array of UPPP variations are beyond the scope of this discussion.

Laser-assisted uvuloplasty (LAUP) is an office-based surgical procedure that progressively shortens and tightens the uvula and palate through a series of CO_2 laser incisions and vaporizations. This technique has not demonstrated consistent effect or benefit on OSA severity or symptoms or on quality of life measures and therefore, also in accordance with the practice parameters of the American Academy of Sleep Medicine (AASM), LAUP is not recommended for the management of OSA.[46,63]

Radiofrequency ablation (RFA) can be used to reduce or stiffen soft tissues. RFA of the palate is considered to be a minimally invasive surgical outpatient clinic procedure that decreases socially disturbing snoring as reported by the partner

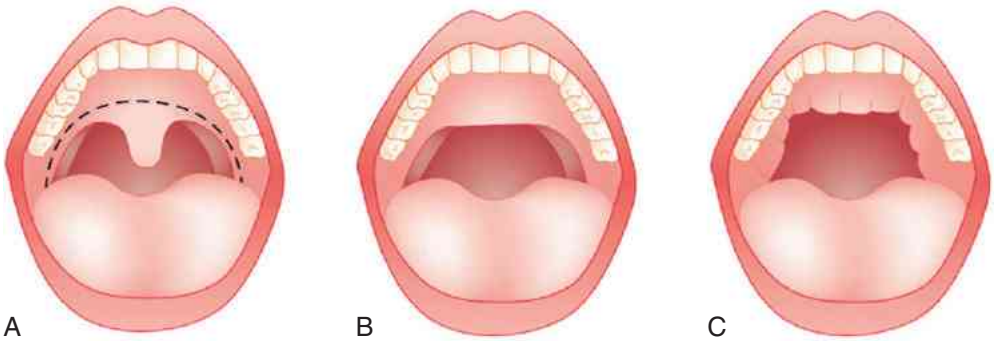


Figure 149-3 Classical uvulopalatopharyngoplasty technique. **A,** Redundant soft palate and pharyngeal pillar mucosa are outlined. **B,** Tonsils, pharyngeal pillar mucosa, uvula, and soft palate have been excised. The extent of soft palate excision is determined by placing traction on the uvula and noting the position of the mucosal crease. **C,** Mucosal flaps of the lateral pharyngeal wall and palatal muscle are advanced and closed with absorbable suture. (From Troell RJ, Strom CG. Surgical therapy for snoring. *Fed Pract* 1997;14:29–52.)

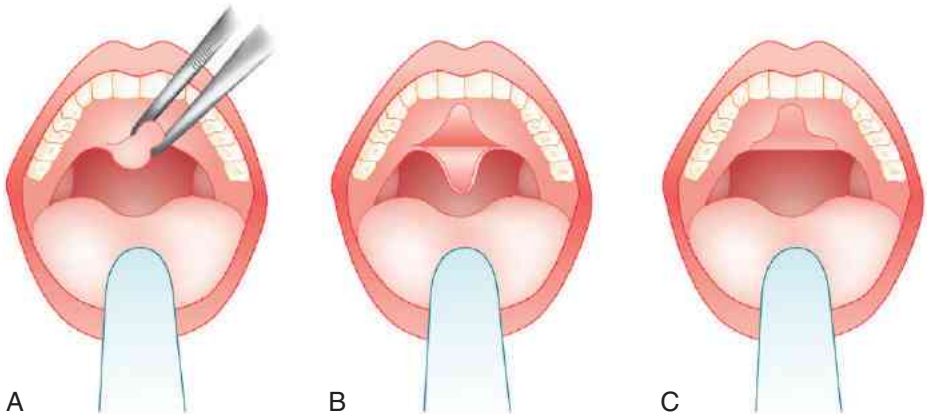


Figure 149-4 Uvulopalatal flap (UPF) technique. **A,** Uvula is reflected to identify mucosal crease of muscular sling. **B,** Knife removes mucosa on proposed flap site. **C,** Wound is closed with a half-buried suture of braided 3-0 Vicryl at the tip of the uvula and simple interrupted sutures along the mucosal closure. (From Troell RJ, Strom CG. Surgical therapy for snoring. *Fed Pract* 1997;14:29–52.)

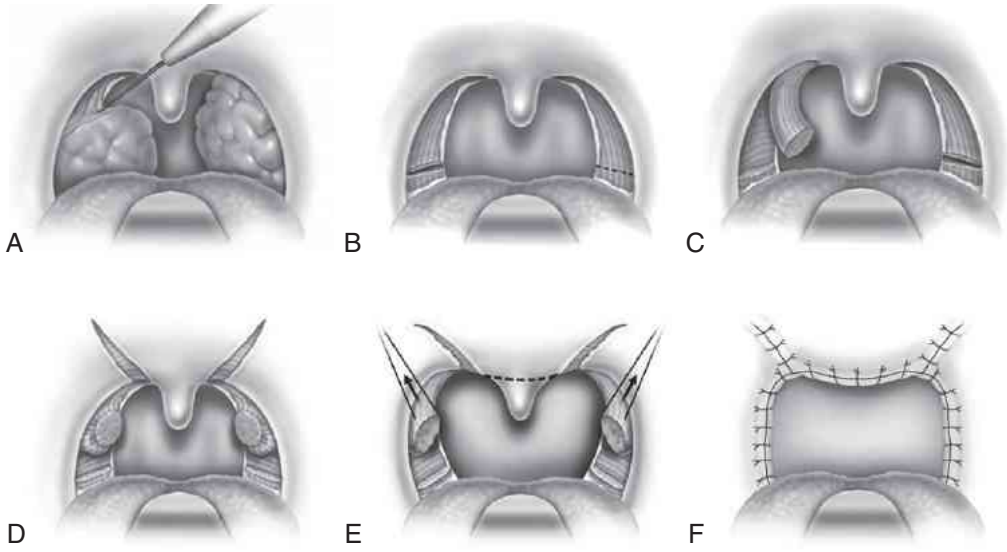


Figure 149-5 Expansion sphincter pharyngoplasty (ESP) technique. **A,** Tonsillectomy is performed. **B,** Horizontal incision made to divide the middle part of the palatopharyngeus muscle. **C,** The palatopharyngeus muscle is mobilized, although not completely, with care taken to leave its fascia attachments to the deeper horizontal constrictor muscles. **D,** Superolateral incision made on the soft palate, revealing the arching fibers of the palatini muscles. **E,** Vicryl sutures are used to hitch up the palatopharyngeus muscle to the soft palate muscles superolaterally. **F,** Closure of the palatal incisions. (From Pang KP, Woodson BT. Expansion sphincter pharyngoplasty: a new technique for the treatment of obstructive sleep apnea. *Otolaryngol Head Neck Surg* 2007;137:110–14.)

of the patient. RFA of the palate has minimal effect on OSA severity, however, so it is not recommended as a single-stage procedure for the management of OSA.[46,64-67]

Taken together, the available evidence on pharyngeal procedures that address the palatal level suggests that UPPP and its variations can be applied in well-selected patients, whereas both LAUP and RFA of the palate are not recommended for the management of OSA.[52,53,57,63,64]

RFA tongue reduction is a minimally invasive technique for decreasing tongue volume when tongue hypertrophy contributes to the OSA.[68] When combined with RFA of the palate in patients with mild to moderate OSA, the procedure had a significant effect on airway volume, apnea index, and quality of life in a randomized sham placebo-controlled trial.[69] However, the magnitude of treatment effect is modest, especially in patients who are obese or have more severe OSA.[70,71] Consequently, RFA tongue reduction usually is performed as a valuable adjunctive surgical procedure, and not a primary procedure, in the management of OSA.[72,73]

In patients with OSA primarily related to hypertrophy of the tongue base, various techniques of partial glossectomy have led to significant reduction in OSA severity. The original studies describing partial glossectomy for OSA used laser techniques.[74,75] More recently, investigators have obtained good success, with less patient morbidity, using an irrigating, suctioning bipolar device instead of the laser.[76-79,79a] Tongue

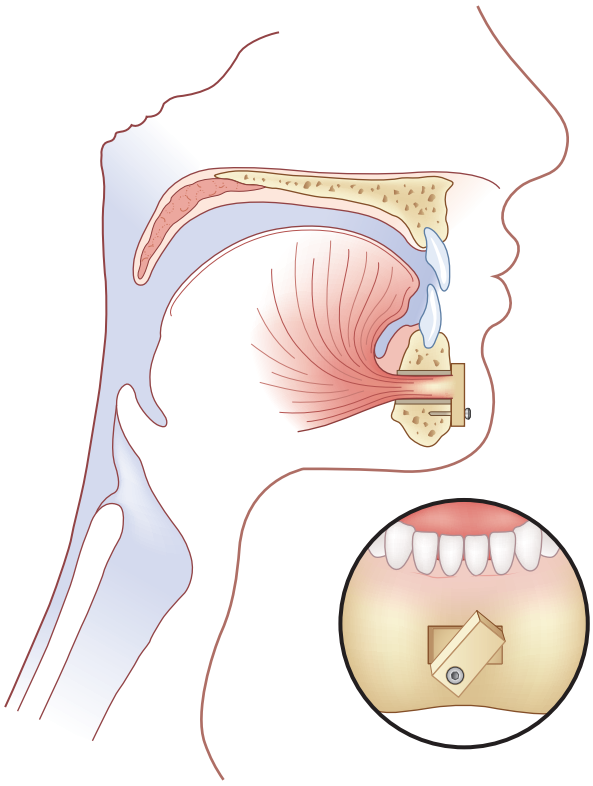


Figure 149-6 Genioglossus advancement. In this procedure, the tongue is placed under anterior traction by performing a limited mandibular osteotomy, with subsequent advancement of the genial tubercle–genioglossus muscle complex. After removal of the buccal cortex and medullary bone, the lingual cortex of the advanced bony segment including the genial tubercle is fixated in its new anterior position (*inset*). (From Riley RW, Powell NB, Guilleminault C. Obstructive sleep apnea and the hyoid: a revised surgical procedure. *Otolaryngol Head Neck Surg* 1994;111:717–21.)

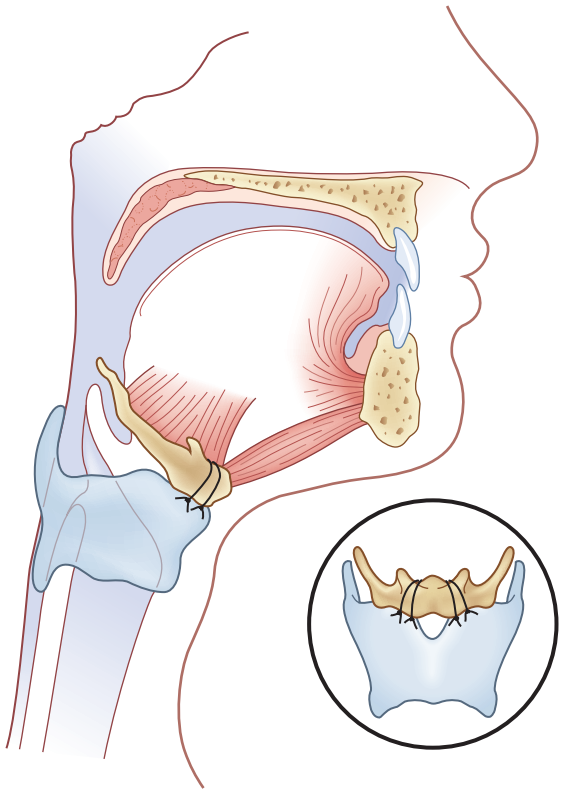


Figure 149-7 Modified hyoid myotomy and suspension (HMS) procedure. The hyoid bone is isolated, the inferior body is dissected clean, and the major portion of the suprahyoid musculature remains intact. The hyoid is advanced over the thyroid lamina and immobilized with sutures placed through the superior aspect of the thyroid cartilage (*inset*). (From Riley RW, Powell NB, Guilleminault C. Obstructive sleep apnea and the hyoid: a revised surgical procedure. *Otolaryngol Head Neck Surg* 1994;111:717–21.)

reduction achieved using transoral robotic surgery (TORS) techniques also has been proved to be feasible, low-risk, and well tolerated.[80-82] As with most surgical upper airway modification techniques, it has been shown that the preoperative BMI predicts treatment success using TORS for OSA management.[83]

Another treatment approach to manage tongue obstruction is to advance or stabilize the tongue using various techniques. Genioglossus advancement uses the forward placement of the geniotubercle to place sufficient tension on the tongue, preventing it from collapsing (Figure 149-6). Hyoid suspension or hyoidthyropexia for the treatment of OSA consists of securing the hyoid arch anteroinferiorly to the thyroid lamina, with or without hyoid myotomy (Figure 149-7). Alternatively, the hyoid bone can be suspended from the mandible using any of various methods. Genioglossus advancement and hyoid suspension are rarely performed in isolation, so assessment of isolated effectiveness is difficult. Suture-based tongue suspension procedures aim to tether the tongue to the mandible by means of sutures, ribbons, or barbs.[84] Adjustable tongue advancement, effected through placement of a tissue anchor in the tongue base and an adjustment spool at the mandible, and of a tetherline to suspend the tongue, was reported to be feasible and well tolerated, but further research is needed on the efficacy of this novel procedure.[85,86] Tongue suspension should be considered in patients with OSA who demonstrate tongue base obstruction. As a stand-alone procedure, its success rate is only 36.6%,[84] possibly related to the multilevel obstruction common in OSA. In addition, tongue suspension is effective and safe as part of a multilevel surgical approach for patients with OSA.[84]

Various case series have shown mixed results with the various hypopharyngeal procedures, but most demonstrate at least partial effectiveness.[84,87-89] As a result, consensus is lacking regarding which procedure is best used to address hypopharyngeal obstruction in OSA, because individual anatomy and obstruction patterns are indications for different types of procedures. For example, tongue hypertrophy might best be treated with tongue reduction (e.g., partial glossectomy), whereas tongue collapse might be more readily treated with tongue advancement or stabilization (e.g., genioglossus advancement, hyoid suspension, tongue suspension). The previously described surgical procedures are site-directed to specific areas of the upper airway. Often these procedures are combined in various ways to address multilevel obstruction. Other surgical procedures also have a multilevel effect on the airway and are indicated in specific circumstances. As described next, neuromuscular upper airway stimulation therapy, especially with electrical stimulation of the hypoglossal nerve, is a new multilevel treatment that activates pharyngeal airway dilator muscles to enlarge the lower pharynx and indirectly can stabilize the upper pharynx.[90-92] MMA, also described

further on, has a multilevel effect by enlarging the anterior-posterior dimensions of the upper and lower pharynx and stabilizing lateral wall collapse.[93]

Upper Airway Neurostimulation Therapy

Direct electrical stimulation of the hypoglossal nerve, which innervates the intrinsic and extrinsic muscles of the tongue, during sleep, to restore or maintain upper airway patency, is an experimental therapy for OSA with a history of 20 years of research and development.[90,94-98] From a pathophysiologic standpoint, selective stimulation of the hypoglossal nerve during sleep may provide an interesting approach to management of OSA based on the restoration or improvement of upper airway dilator muscle activity during sleep.[90,91]

The feasibility of and therapeutic potential for chronic hypoglossal nerve stimulation using a first-generation system (Inspire 1, Medtronic, Inc., Minneapolis, Minnesota) in patients with OSA were investigated in a multicenter study.[97] Eight patients with OSA received this first-generation, fully implantable pulse generator (IPG) triggered by a pressure sensor, located at the sternum, detecting respiratory effort. Electrical stimulation was sent from the IPG to the hypoglossal nerve by way of a stimulation lead containing a half-open cuff around the nerve.[97] Throughout the entire study, unilateral hypoglossal nerve stimulation decreased the severity of OSA. After a minimum follow-up period of 6 months, all patients tolerated long-term stimulation, and no adverse effects were noted.[97]

On the basis of this first experience, the safety and therapeutic feasibility as well as the efficacy of three different second-generation hypoglossal nerve stimulation devices have been explored in recent clinical studies. Three types of systems were analyzed in these studies: the HGNS (Hypoglossal Nerve Stimulation) system (Apnex Medical, Inc., St. Paul, Minnesota), the Aura6000 system (ImThera Medical, Inc., San Diego, California), and the Inspire II Upper Airway Stimulation (UAS) device (Inspire Medical Systems, Inc., Maple Grove, Minnesota). Early results of these studies have been published and confirm the safety and feasibility of the second-generation implantable systems for upper airway neurostimulation therapy for OSDB.[94,95,98] The primary component in all three systems is an IPG. Furthermore, the main difference between the three systems is that with the Aura6000 system, the stimulation onto the body of the proximal hypoglossal nerve from the multielectrode lead is continuous, obviating the need for respiration-sensing leads, whereas with the two other systems, the HGNS system and the Inspire II UAS device, the hypoglossal nerve stimulation will be intermittent and synchronized with the signal of the respiration-sensing leads that measure the respiratory cycle.[94,95,98] Among other differences, during the surgical technique for the Inspire II UAS system, the cuff section of the stimulation lead needs to be placed on the medial division of the distal hypoglossal nerve, thereby aiming at selective stimulation of the protrusor muscles of the tongue only.[90,92] Subsequently, appropriate placement of the stimulation lead needs to be confirmed by observing tongue protrusion during stimulation and by electromyographic monitoring during surgery.[90] The identification of patients with OSA who are more likely to benefit from upper airway neurostimulation therapy has been an important part of the research agenda.[98,99] Recent data suggest that responders to this therapy have a body mass index (BMI) of 32 kg/m^2 or less and AHI of 50 or below (i.e., fewer than 50 events per hour of sleep).[98] In addition, the absence of complete concentric collapse at the level of the palate as documented during DISE may predict therapeutic success with implanted upper airway neurostimulation therapy (Figure 149-8).[98,99] Therefore DISE can be recommended as a patient selection tool for use of implanted upper airway stimulation therapy for OSA.[99]

Although the early experience with stimulation is promising, the study results have been mixed. A prospective randomized trial of upper airway neurostimulation used the HGNS system, but the trial was discontinued in 2013 because the

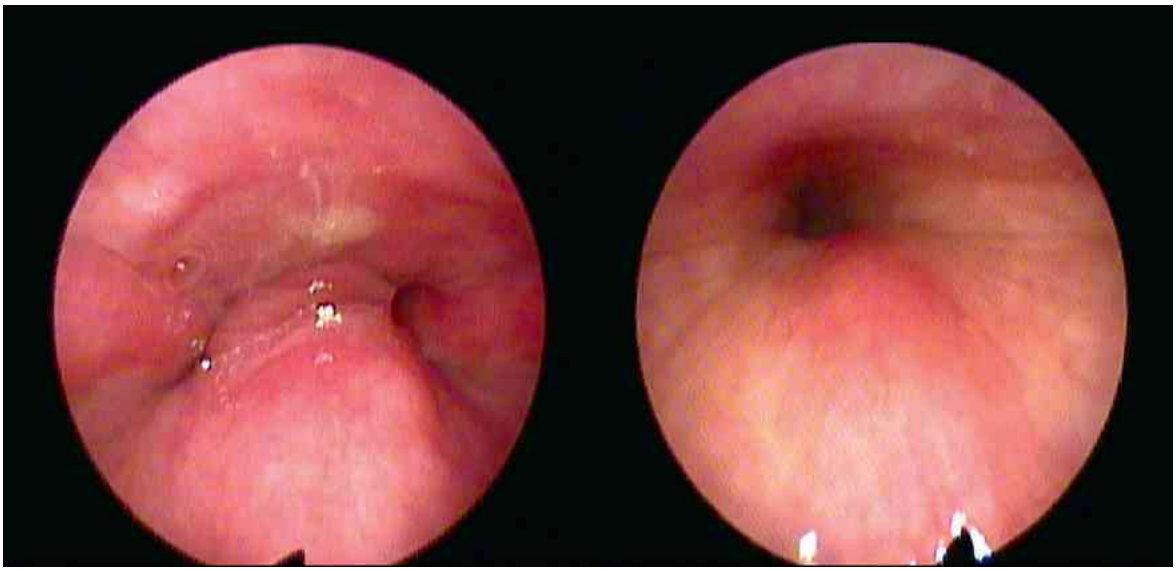


Figure 149-8 Example of anteroposterior (*left*) versus concentric (*right*) collapse at the palatal level during drug-induced sedation endoscopy (DISE). (From Vanderveken OM, Maurer JT, Hohenhorst W, et al. Evaluation of drug-induced sleep endoscopy as a patient selection tool for implanted upper airway stimulation for obstructive sleep apnea. *J Clin Sleep Med* 2013;9:433–8.)

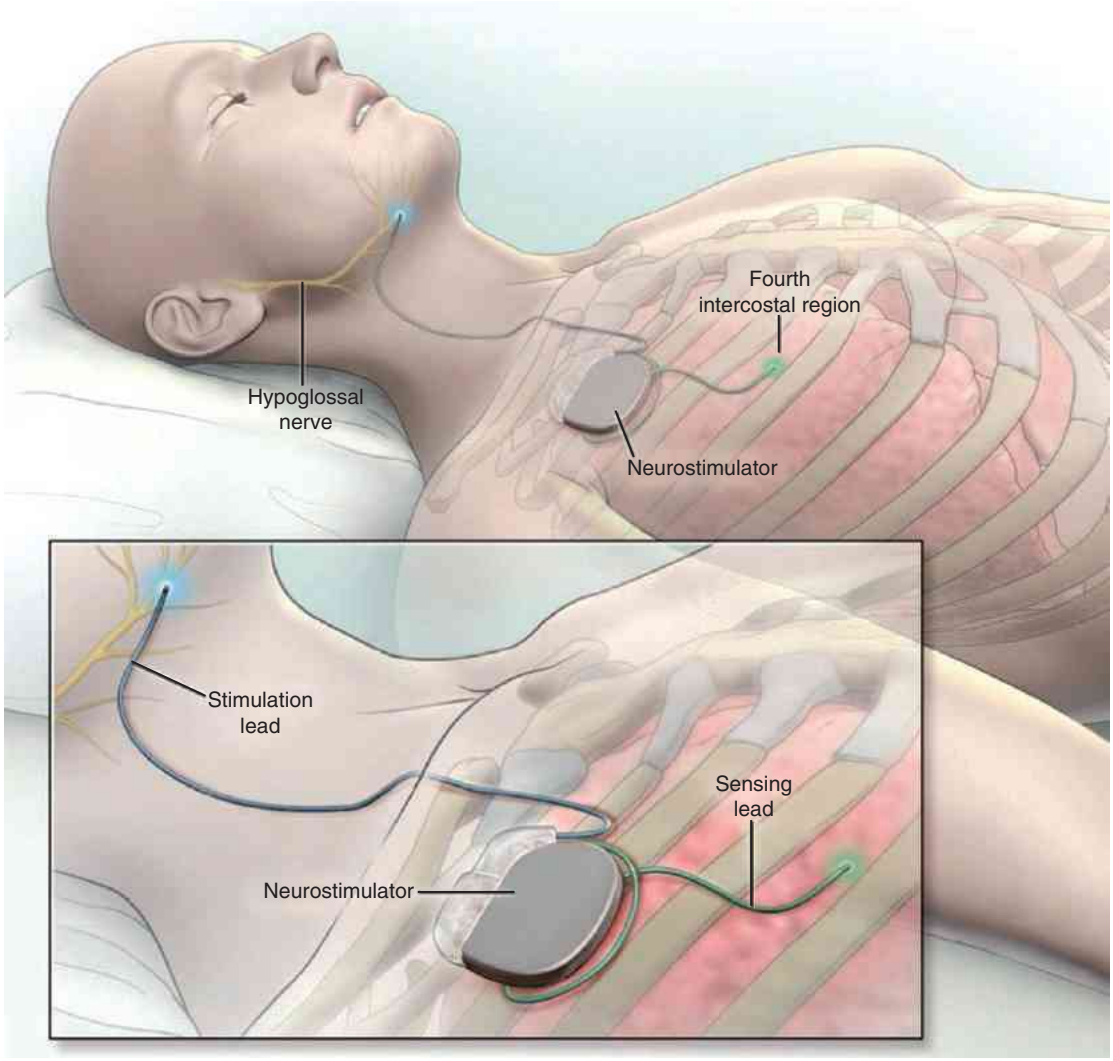


Figure 149-9 Upper airway stimulation using Inspire 2 implant (Inspire Medical Systems). The neurostimulator delivers electrical stimulating pulses to the protruding branches of the hypoglossal nerve through the stimulation lead; the stimulating pulses are synchronized with ventilation detected by the sensing lead. (From Strollo PJ Jr, Soose RJ, Maurer JT, et al. Upper-airway stimulation for obstructive sleep apnea. *N Engl J Med* 2014;370:139–49.)

interim analysis showed that it was unlikely to meet the primary efficacy end point. Apnex Medical, the producer of HGNS and the trial's sponsor, ceased its activity in 2013 because of the nonviable results of that trial (available at ClinicalTrials.gov—https://clinicaltrials.gov/ct2/show/NCT01446601).

More recently, a large multicenter, prospective case series assessed the safety and effectiveness of upper airway neurostimulation therapy using the Inspire II UAS device (Figure 149-9) in 126 well-selected patients with OSA.[90] The results of this pivotal study demonstrated that hypoglossal nerve stimulation led to significant improvements in polysomnographic parameters in 66% of patients, with associated improvements in sleepiness, snoring, and quality of life, in this sample.[90] Unfortunately, exploratory analysis did not yield predictors of failure in the 34% of nonresponders. Serious adverse events (e.g., reoperation) were uncommon, and the side effects, such as tongue irritation, were not bothersome or else resolved in most patients.[90] In a minority of patients, a tooth guard was needed to permit healing of tongue soreness or abrasion related to the overnight tongue protrusion induced by the stimulation.[90] In addition, the results of a randomized, 1-week therapy-withdrawal trial among a subset of responders showed that continuation of stimulation was necessary to maintain the effect (i.e., withdrawal of stimulation led to worsening OSA almost to pretreatment levels).[90] A randomized withdrawal trial in responders, however, should not be confused with a prospective randomized trial like the HGNS trial. The promising results of this multicenter case series assessing the Inspire II UAS device led the U.S. Food and Drug Administration (FDA) to approve the Inspire UAS device for treatment of selected patients with moderate to severe OSA.

The results of a recent systematic review and metaanalysis of the literature indicate that hypoglossal nerve stimulation for OSA is reported to carry low risk of serious complications.[91] The investigators concluded that the available data in literature show high rates of therapy adherence and stable outcome results over 12 months of follow-up in selected patients with OSA.[91,96] Endoscopic findings in a subset of

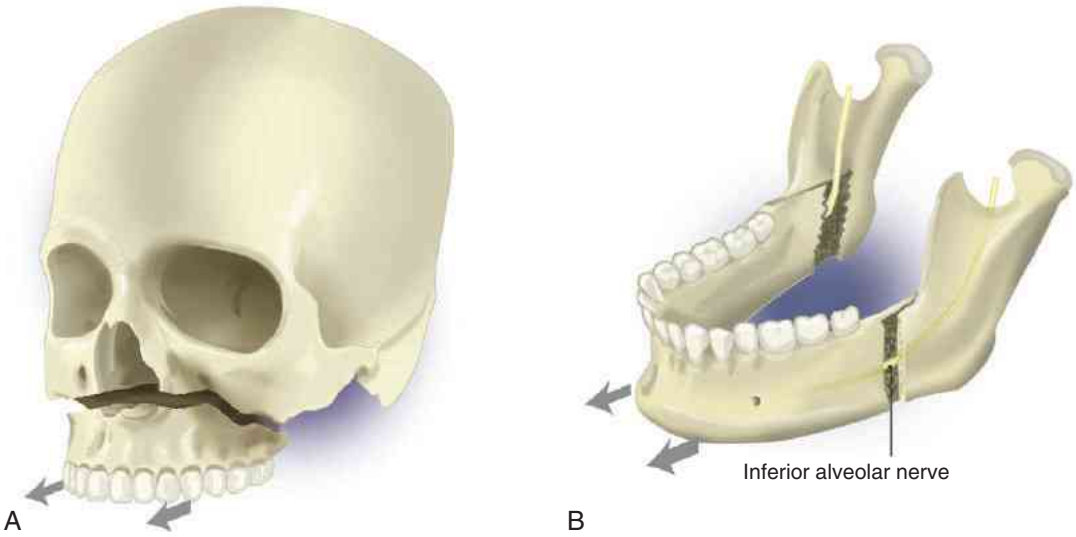


Figure 149-10 Maxillomandibular advancement (MMA) surgery for obstructive sleep apnea syndrome provides enlargement of the upper airway by means of a Le Fort I advancement osteotomy of the maxilla **(A)** and a bilateral sagittal split advancement osteotomy of the mandible **(B)**. (From Rosenberg AJ, Damen GW, Schreuder KE, Leverstein H. [Obstructive sleep-apnoea syndrome: good results with maxillo-mandibular osteotomy after failure of conservative therapy]. *Ned Tijdschr Geneeskd* 2005;149:1223–6.)

patients who underwent upper airway neurostimulation therapy using the Inspire II UAS device revealed that responders had greater retropalatal enlargement with stimulation than nonresponders, and that the neurostimulation increased both the retropalatal and retrolingual cross-sectional areas.[100] This observation of multilevel enlargement induced by upper airway neurostimulation therapy may explain the sustained reductions in OSA severity in two thirds of the selected patients receiving UAS therapy.[90,100] It is important to reiterate that patient selection appears to be critical, and even among carefully selected patients (e.g., BMI of less than 32 kg/m^2, no complete circumferential velopharyngeal collapse, and other criteria), the nonresponder rate is still significant, with no clear predictors.[96] Further research is ongoing and will be necessary to elucidate the optimal roles for this exciting therapy.

SKELETAL MODIFICATIONS

Orthognathic surgery for the treatment of OSA was first described when mandibular advancement surgery was reported to reverse the symptoms of sleep apnea.[101] In 1986, Riley and coworkers were the first to describe the combination of advancement of both maxilla and mandible to improve airway patency in patients with OSA.[102] Although trial-based evidence is still scarce, MMA currently is regarded as a highly effective and safe surgical modality for treatment of OSA.[93] MMA surgery in patients with OSA consists of a bilateral sagittal split osteotomy (BSSO) of the mandible with advancement and a Le Fort I osteotomy of the maxilla with advancement (Figure 149-10). In patients with OSA, MMA surgery generally requires a minimum advancement of the mandible by 10 mm to achieve optimal effectiveness.[103,104] As a consequence of the skeletal advancement, several upper airway muscles and ligaments are repositioned anteriorly,

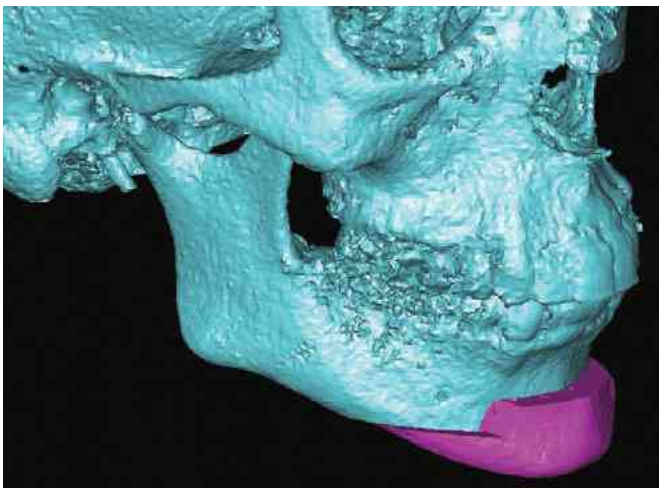


Figure 149-11 Modified genioplasty. In this procedure, the tongue is put under anterior traction by performing a trapezoid-shaped osteotomy with advancement of the chin (*purple*) and the genial tubercle–genioglossus muscle complex.

including the anterior belly of the digastric, mylohyoid, genioglossus, and geniohyoid muscles. The advancement of the maxilla pulls the soft tissue of the palate forward, tightens the palatoglossal and palatopharyngeal muscles, and increases tongue support. Moreover, adding the maxillary advancement also increases the amount of mandibular advancement that can be accomplished with surgery. To achieve additional improvements in oro- and hypopharyngeal airway patency, MMA surgery can be combined with a genioglossus advancement or a modified genioplasty[105,106] (Figure 149-11; see also Figure 149-6). To decrease the cervical fat mass, as appropriate, and to further improve airway patency, cervicomental liposuction may be added to the surgical plan in selected

cases.[106] MMA results in structural enlargement of the nasooropharyngeal airway and enhanced tension and decreased collapsibility of the pharyngeal airway. When these surgical techniques are applied only for OSA and not for dentofacial abnormalities per se, the procedure is referred to as "telegnathic surgery" instead of "orthognathic surgery." For parallel information on orthodontic therapy for management of OSA, see Chapter 143.

General Outcomes of Maxillomandibular Advancement Surgery for Obstructive Sleep Apnea

OSA management after MMA surgery generally is successful in a high proportion of patients. A recent metaanalysis of data for 627 patients from 22 studies demonstrated that the median rate of surgical success, defined by a postoperative AHI less than 20 (i.e., fewer than 20 events per hour of sleep), with greater than 50% reduction overall, is 86%.[93] Surgical cure, defined more stringently by a postoperative AHI less than 5, was observed in 43% of patients in this metaanalysis. After a mean follow-up period of 5 months, a statistically and clinically significant reduction in the mean AHI, from 63.9 to 9.5, has been observed.[93] Although the number of studies evaluating long-term outcomes of MMA surgery is still limited, long-term results appear to be relatively stable over follow-up periods exceeding 2 to 5 years.[103,104,107-109] In addition, several studies found statistically and clinically significant improvements in blood pressure after MMA surgery in patients with OSA.[110,111] Finally, evidence for the medium- to long-term stability of the mandibular or maxillary advancement with MMA surgery in patients with OSA has been corroborated by several cephalometric studies.[104,112-114]

In evaluations of subjective outcomes after MMA surgery for OSA, most patients report improvements in snoring, witnessed apneas, excessive daytime sleepiness and quality of life, morning headaches, memory loss, and impaired concentration.[93,110,115,116] In most cases, CPAP can be discontinued after MMA surgery, with patients overall reporting that treatment was worthwhile and recommendable to others.[115]

The specific effects of MMA surgery on upper airway dimensions and surrounding structures have been extensively studied using different techniques, including cephalometry (Figure 149-12) and three-dimensional CT studies (Figures 149-13 to 149-15).[30a,116-118] On the other hand, up to this date, the available data are insufficient to support a relationship between OSA improvements such as reductions in AHI and changes in the upper airway and its surrounding bony structures.[117]

Simultaneous advancement of the maxilla and mandible changes the skeletal framework of the face, thereby resulting in a possible rejuvenation of the middle and lower third of the face. This concept of a "reverse face lift," with positive effects on facial aesthetics, after MMA surgery in patients with OSA is observed in a majority of cases (Figure 149-16). One study found that at 6 months after surgery, 50% of patients reported a younger and 36% reported a more attractive facial appearance.[105,107] Of note, however, 9% of patients in this study reported a less attractive facial appearance after surgery. Conversely, in another study, patients with OSA indicated that they were not bothered by their appearance after MMA surgery.[119] Although patients seeking treatment for OSA generally do not desire an aesthetic facial improvement, it is mandatory to communicate the anticipated facial changes before surgery. Indeed, although 50% of patients undergoing MMA are satisfied with the aesthetic result after the surgery, 30% are indifferent to their postoperative appearance, whereas up to 13% and 5% are disappointed and unsatisfied,

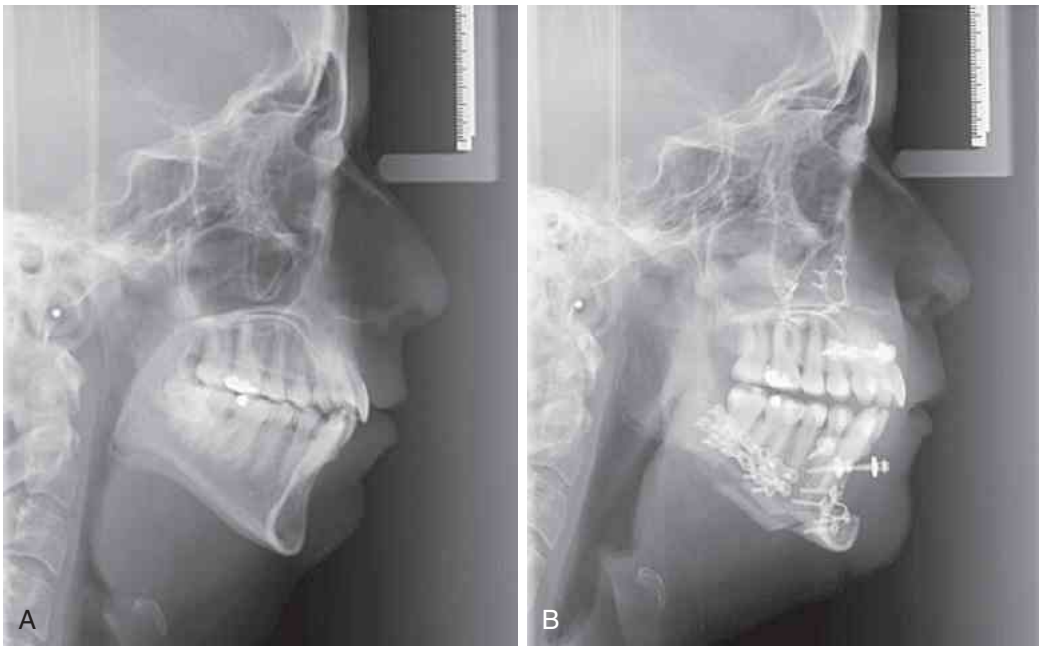


Figure 149-12 Two-dimensional airway changes after maxillomandibular advancement surgery. Preoperative **(A)** and postoperative **(B)** lateral cephalometric radiographs indicating the sagittal changes in upper airway space after maxillomandibular advancement surgery combined with a modified genioplasty and cervicomental liposuction in a patient with severe obstructive sleep apnea syndrome. (From Doff MH, Jansma J, Schepers RH, Hoekema A. Maxillomandibular advancement surgery as alternative to continuous positive airway pressure in morbidly severe obstructive sleep apnea: a case report. *Cranio* 2013;31:246–51.)

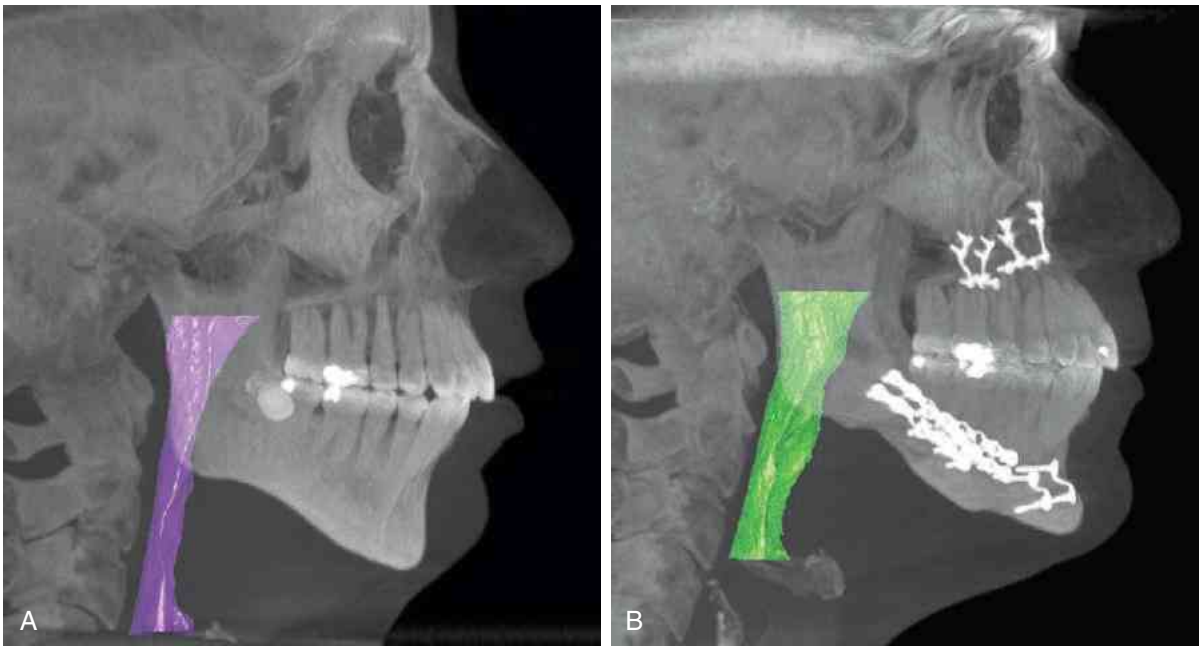


Figure 149-13 Three-dimensional airway changes after maxillomandibular advancement surgery. In *purple* the preoperative **(A)** and in *green* the postoperative **(B)** cone beam computed tomography (CBCT) three-dimensional reconstruction of the upper airway indicating changes in upper airway space after maxillomandibular advancement surgery combined with a modified genioplasty and cervicomental liposuction in a patient with severe obstructive sleep apnea syndrome.

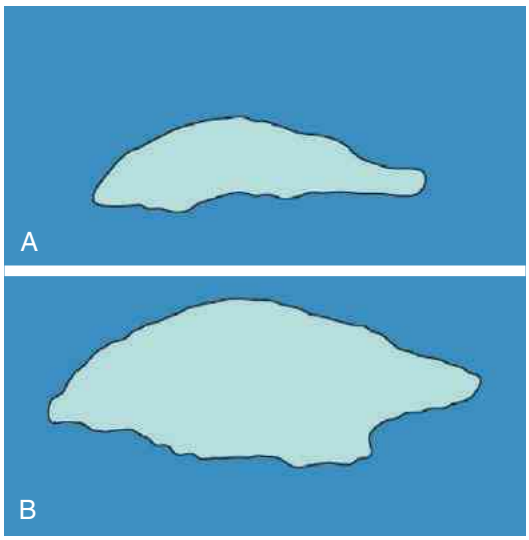


Figure 149-14 Velopharyngeal airway changes after maxillomandibular advancement surgery. Enlargement of the minimal cross-sectional area of the velopharyngeal airway before **(A)** and after **(B)** maxillomandibular advancement surgery combined with a modified genioplasty and cervicomental liposuction in a patient with severe obstructive sleep apnea syndrome. (From Doff MH, Jansma J, Schepers RH, Hoekema A. Maxillomandibular advancement surgery as alternative to continuous positive airway pressure in morbidly severe obstructive sleep apnea: a case report. *Cranio* 2013 31:246–51.)

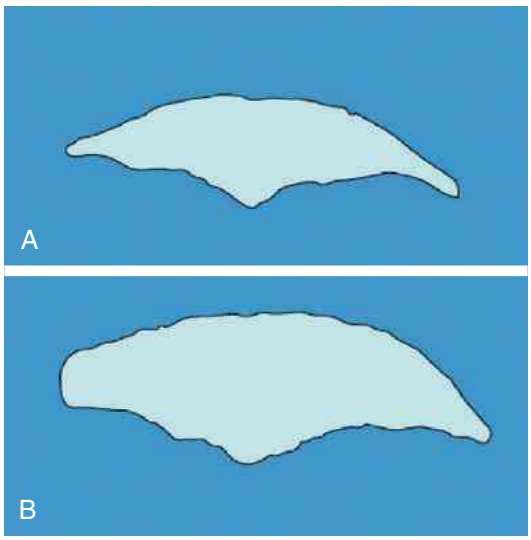


Figure 149-15 Oropharyngeal airway changes after maxillomandibular advancement surgery.Enlargement of the minimal cross-sectional area of the oropharyngeal airway before **(A)** and after **(B)** maxillomandibular advancement surgery combined with a modified genioplasty and cervicomental liposuction in a patient with severe obstructive sleep apnea syndrome. (From Doff MH, Jansma J, Schepers RH, Hoekema A. Maxillomandibular advancement surgery as alternative to continuous positive airway pressure in morbidly severe obstructive sleep apnea: a case report. *Cranio* 2013;31:246–51.)

respectively.[120] A surgical technique involving a so-called counterclockwise rotation of the occlusal plane, which previously has been used in correcting severe "bird-face" deformity, may be used both to achieve aesthetic goals and to fulfill the main objective in the treatment of patients with OSA—an optimal increase in airway patency.[121]

Major complications associated with MMA surgery in patients with OSA are rarely reported.[93] Individual and nonfatal cases of postoperative cardiac arrest or dysrhythmia have been reported.[122] One study described a case of a life-threatening airway obstruction after extubation that required reintubation.[123] However, no immediate postoperative deaths

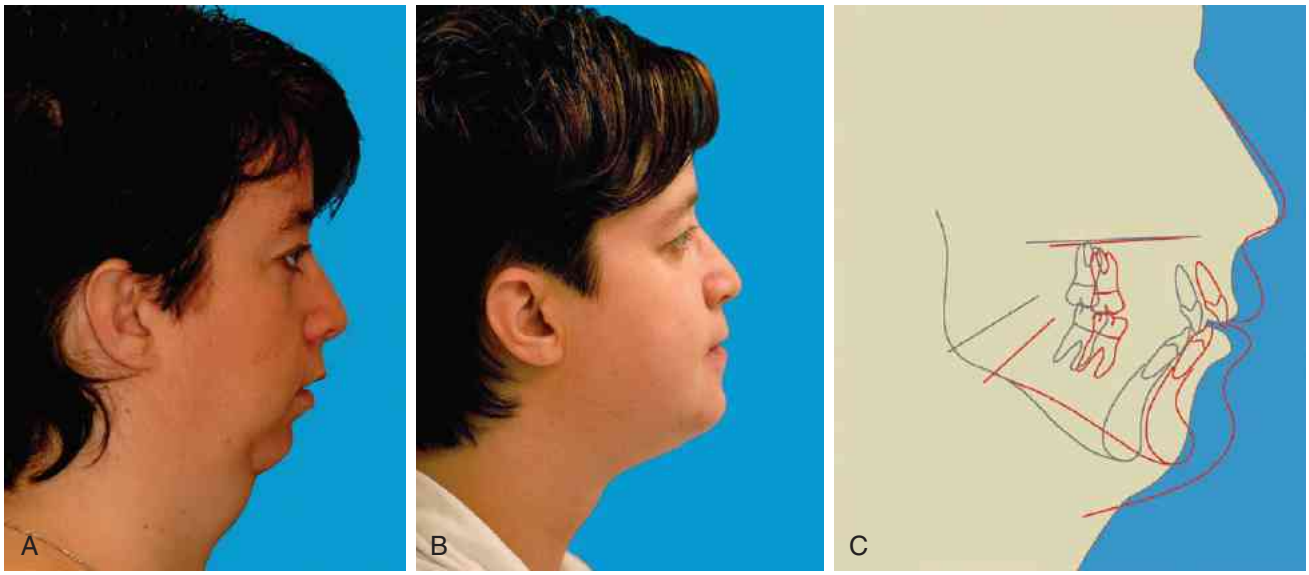


Figure 149-16 Effects on facial aesthetics of maxillomandibular advancement surgery. Preoperative **(A)** and postoperative **(B)** photographs illustrating the rejuvenation of the middle and lower thirds of the face after maxillomandibular advancement surgery combined with a modified genioplasty and cervicomental liposuction in a patient with severe obstructive sleep apnea syndrome. The effect on the patient's profile of profound advancement of the lower third of the face can be appreciated when the pre- and postsurgical cephalometric radiographs are superimposed **(C)**. (From Doff MH, Jansma J, Schepers RH, Hoekema A. Maxillomandibular advancement surgery as alternative to continuous positive airway pressure in morbidly severe obstructive sleep apnea: a case report. *Cranio* 2013;31:246–51.)

have been reported after this type of surgery in patients with OSA.[93] (See also Chapters 121 and 148.)

Minor complication rate has been reported to be approximately 3%.[93] Minor complications include hemorrhage or local infections that generally are cured with either antibiotics or surgical drainage. The presence of postoperative malocclusions or facial paresthesia is not included in this minor complication rate. Facial paresthesia is present in almost all patients after surgery, but resolution has been reported in approximately 85% of patients at the 1-year follow-up evaluation.[93] By contrast, some studies indicate that one half of the patients treated report a persistence of facial paresthesia.[120] Although malocclusions may be seen in up to 44% of patients, they generally can be resolved with (prosthetic) dental treatment or equilibration of the dental occlusion.[122] However, when patients are orthodontically prepared for MMA surgery, malocclusions generally pose no long-term problem with postoperative orthodontics. Some studies report a trend for poor bone healing and increased foreign body reactions after MMA surgery in patients with OSA.[124]

At 3.5 days, average hospital stay is slightly longer after MMA surgery in patients with OSA than in "conventional" orthognathic patients.[93] Most patients with OSA are able to return to full-time work within 2 to 10 weeks after their surgery.[110,125]

The most relevant patient characteristics and clinical factors predictive of a favorable outcome after MMA surgery in patients with OSA include younger age, lower preoperative BMI, and less severe OSA.[93] Moreover, an increased postoperative posterior airway space, determined from lateral cephalometric radiographs, appears to be the only relevant cephalometric variable of predictive value for a successful outcome of MMA surgery.[93] In addition, the amount of advancement of the maxilla appears to correlate with the degree of reduction in AHI.[93,116]

Preoperative Evaluation for Maxillomandibular Advancement

In general, prerequisites for MMA surgery include clinically "significant" OSA that is not amenable to conservative management (e.g., CPAP), a medically and psychologically stable condition, and the patient's desire for surgery and informed consent before the procedure.[126] A lateral cephalometric head film is mandatory to plan MMA surgery, as well as a preoperative polysomnographic recording. Surgery also may be planned using three-dimensional imaging techniques such as (cone beam) CT. Subsequently, virtual surgical planning can be conducted, which offers the surgeon valuable information on the anticipated skeletal, airway, and facial aesthetic changes (Figure 149-17).[127] Fiberoptic nasopharyngoscopy also is recommended before surgery. This imaging procedure can further help to identify nasal, retropalatal, or tongue base pathology that may affect the outcome of MMA surgery (e.g., lingual tonsillar hypertrophy).

Finally, the response to an oral appliance also may be used to select suitable candidates for MMA surgery.[128] Patients demonstrating a substantial reduction in baseline AHI (i.e., to less than 50%) with oral appliance therapy appear to be especially good candidates for MMA surgery.[128] Despite the variety of treatment protocols, a precise treatment algorithm for choosing MMA surgery in OSA management has yet to be established, and the decision depends in large measure on patient's preference.

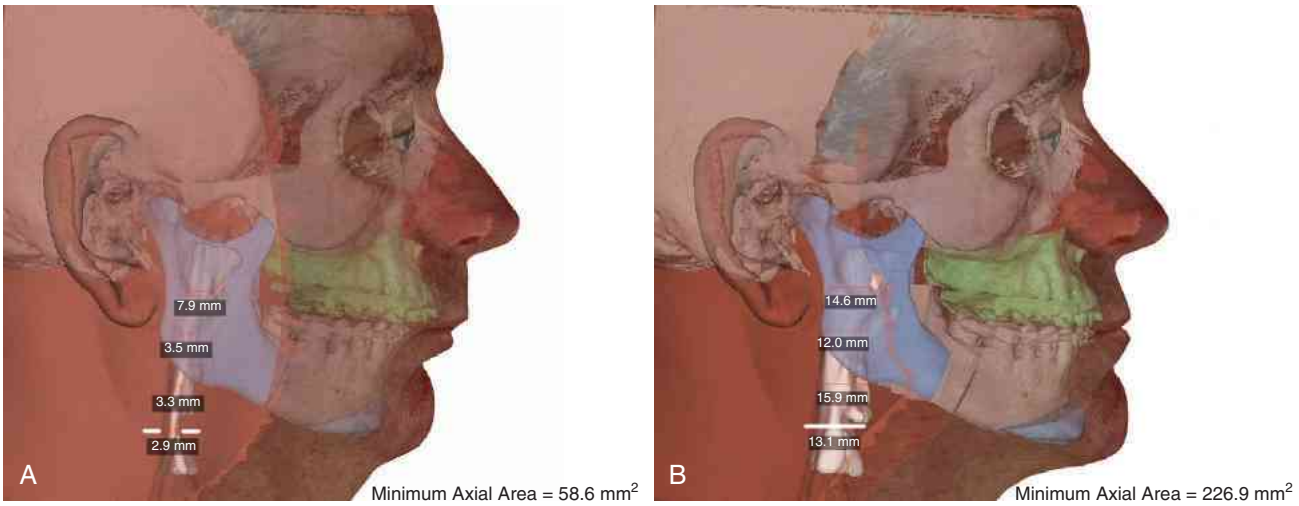


Figure 149-17 Three-dimensional planning of maxillomandibular advancement surgery. Preoperative **(A)** and postoperative **(B)** "morphs" illustrating the anticipated skeletal, airway and facial aesthetic changes after maxillomandibular advancement surgery with a genioplasty in a patient with sleep-disordered breathing. (Courtesy D. Brock; 3D Systems, Inc., Rock Hill, S.C.)

Postoperative Management after Maxillomandibular Advancement

Medical surgical management of the postsurgical patient with OSA is more complicated than with conventional orthognathic patients, despite the profound and immediate postsurgical improvement in pharyngeal airway patency. If recovery is sufficient, discharge is up to the surgeon and patient but requires proper pain control and capability of oral intake. Because younger patients tend to recover more quickly, discharge usually is earlier in this patient category.

Follow-up evaluation depends on the surgeon's individual protocol and the specific patient characteristics. Of note, postoperative edema usually is maximal 72 hours after surgery. Frequent postoperative follow-up visits are recommended until full recovery has been achieved. A polysomnographic follow-up study at 6 months or later after MMA is recommended.[129] As with all other surgical options for OSA treatment, it should be stressed to the patient that weight loss is an important part of the pre- and posttreatment OSA management, because even modest weight change will affect the outcome.

Upper Airway Bypass: Tracheostomy

Tracheostomy remains an important treatment option for OSA even though it is used by only a small minority of patients. The tracheostomy provides an alternative airway that bypasses the upper airway obstruction, and it is highly effective for controlling the OSA.[25] It becomes an important treatment option in patients with severe OSA and serious comorbid illnesses that render other surgical treatments too risky.

Tracheostomy has several important advantages. The surgery is relatively simple and short compared to what would otherwise typically be required to address severe OSA. The major treatment effect is immediate. Patients function normally with eating and speaking by capping the tracheostomy cannula during waking hours and then simply uncap the cannula for sleep. A tracheostomy is reversible in case it is somehow intolerable to the patient, which is unlike almost every other surgical treatment of OSA. Patients typically become very skillful with management of the tracheostomy cannula. Specific disadvantages of a tracheostomy have been recognized. The cannula requires daily maintenance with cleaning and/or changes, although this process is simple. Some patients are self-conscious about the visible cannula. The stoma can accumulate granulation tissue or scar tissue that may require intermittent procedures to clear. Local infections can occur, requiring antibiotics or topical stoma treatment. Cannula dislodgement can be problematic in rare cases, and even life-threatening if the cannula dislodges from the the normal stoma position into the trachea. In addition, patients must avoid submerging the neck under water lest water aspiration occur, even if the cannula is capped.

Bariatric Surgery

Bariatric surgery is indicated in morbidly obese patients (BMI of 40 kg/m^2 or greater) with OSA in whom significant weight loss cannot be achieved through conservative measures such as hypocaloric diet and exercise training[46] (see also Chapter 121). In case of significant obesity-related comorbidity such as hypertension or diabetes, bariatric surgery should be considered starting from BMI of 35 kg/m^2 or greater. It has been clearly demonstrated that the surgical weight loss induced by bariatric surgery can result in significant decrease in OSA severity.[130,131] However, OSA of considerable degree can persist in some patients even after substantial weight loss.[130,132,133] Therefore, a follow-up sleep study is recommended after bariatric surgery to check for residual OSA, especially in patients with low minimum SaO_2 levels and high supine AHI preoperatively.[130,131] If necessary, retitration of CPAP settings may be scheduled, which may lead to higher treatment compliance afterward.[131] CPAP optimization and vigilance may be necessary to treat residual OSA when present and to help maintain weight loss, because there are limited data to suggest an association between postoperative CPAP use and weight loss outcomes.[133]

Combination Therapy Including Multilevel Surgery

To reach the therapeutic target, preferably an alleviation of the disease, it may be necessary to prescribe two or more therapies, with adjunctive modalities used as needed to supplement the primary treatment options.[14] In the management of OSA, however, combining treatment options has been somewhat undervalued. A combination approach is especially appropriate in this setting because as described earlier in the chapter and elsewhere in this book, no single therapy (surgical or nonsurgical) is universally effective. Many of the surgical treatments are site-specific, so combinations of procedures may be necessary to address all sites of obstruction.

Concerning the nonsurgical treatment options for OSA, several possible combinations have been reported, such as combining oral appliance therapy with sleep positioning therapy (SPT) or the combination of CPAP with oral appliance therapy (see Chapters 116, 147 and 150).[134-136] In addition, positional therapy may also be an adjunctive treatment in patients on CPAP who require high pressure levels in the supine position to improve adherence to the CPAP regimen.[46] The addition of oral appliance therapy has been shown to be an effective mode of combination therapy to control OSA after UPPP failure.[137] The combination of positional therapy and surgical upper airway modifications can result in a significant decrease in OSA severity in patients in whom surgery converted non-positional OSA into positional OSA.[138,139] Similarly, the combination of two or more surgical techniques, performed either simultaneously or staged, or so-called *multilevel surgery*, can be regarded as combination therapy for OSA. For example, combining UPPP (or anatomically-directed variants) with hypopharyngeal procedures results in a significant decrease in OSA severity, a significant improvement in daytime sleepiness, and a high degree of patient satisfaction, with an acceptable complication rate.[140,141,142,143] With currently available upper airway surgery protocols to address all relevant areas of obstruction can achieve clinical outcomes (e.g., reduction in symptoms and improvement in quality of life) comparable to CPAP therapy.[144]

CLINICAL PEARLS

- CPAP remains the standard modality to treat OSA. The clinical effectiveness of CPAP treatment, however, often is limited by poor patient (and partner) acceptance, leading to suboptimal adherence. Consequently, an urgent need for non-CPAP treatment options is well recognized.
- To move away from a "trial and error" empirical clinical paradigm, there has been increasing interest in drug-induced sedation endoscopy (DISE), as part of the therapeutic decision-making process regarding upper airway surgery.
- Many surgical options are available, and each has its indications. Nasal airway surgery complements most other surgical and nonsurgical treatments when the nose is obstructed. Site-directed upper airway procedures can be very effective, when combined appropriately, for addressing many combinations of obstruction. Upper airway neurostimulation appears promising for selected patients with multilevel soft tissue collapse and possibly in patients with particularly collapsible tissue (as opposed to structural abnormalities). MMA is an especially useful option in patients with malocclusion, which can be addressed simultaneously, or with facial skeletal compromise. Tracheostomy is an important option for patients with severe comorbid illnesses who cannot tolerate CPAP or more extensive surgery. Bariatric surgery can allow weight loss for significant clinical improvement in patients presenting with severe obesity associated with OSA or other obesity-related comorbidity.

SUMMARY

OSA has major socioeconomic consequences and should be approached as a chronic disease requiring long-term, multidisciplinary management. Although CPAP is overall the most successful treatment for moderate to severe OSA when used properly and consistently, its clinical effectiveness often is limited by poor patient and partner acceptance, which leads to suboptimal adherence. Because many patients with OSA remain inadequately treated owing to inconsistent levels of adherence to CPAP, and because mild OSA deserves treatment in those with problematic symptoms, there is a genuine need for non-CPAP treatment options (Table 149-1). Additionally, it may be necessary to combine two or more of the available treatment options to achieve a successful outcome targeting disease alleviation.

The spectrum of OSA is quite diverse in terms of nature and severity, so the selection of the right treatment regimen for each individual patient will be of utmost importance. Among the different techniques that can be used for the preoperative upper airway assessment, DISE is increasingly performed for dynamic upper airway evaluation to select the proper non-CPAP treatment for patients with OSA.

Surgical treatment of upper airway abnormalities, craniofacial deformations, or obesity may be applied in selected patients with OSA. In general, sleep surgery procedures are directed at specific collapsible upper airway structures, so preoperative upper airway investigation may add to proper selection of a specific surgical procedure for an individual patient. Taking into account that in a majority of patients with OSA, a multilevel collapse is observed within the upper airway, the treatment plan should aim at resolution of the anatomic compromise at all levels involved in the obstructions occurring during sleep.

The results of upper airway neurostimulation therapy in carefully selected patients with OSA are promising. Its safety, combined with associated high rates of therapy adherence and sustained reductions in OSA severity, should sustain further effectiveness research in this particular area.

MMA surgery plays an important role in the correction of OSA that is refractory to noninvasive therapies. MMA surgery in patients with OSA generally is more complex than "conventional" orthognathic surgery because OSA patients require large advancements, are usually older, and often have other comorbidities. With proper precautions, however, it is a safe and highly effective treatment modality for OSA. MMA surgery probably is the most effective surgical intervention besides tracheostomy in patients who are skeletally compromised (e.g., retrognathia or bimaxillary retrusion). These patients should therefore be informed about MMA surgery as one of the primary treatment modalities.

Tracheostomy remains an important treatment option for patients with severe OSA and serious comorbid conditions that might limit the other surgical treatment options. Bariatric

surgery is a valuable treatment option to treat obesity-related OSA and other obesity-related diseases and conditions in patients who are unable to lose sufficient weight by conservative measures.

Finally, combining different treatment options for the alleviation of OSA is undervalued and underinvestigated. Further research on the possible combinations is strongly needed, because no single treatment can adequately manage all patients with OSA.

Selected Readings

De Vito A, Carrasco Llatas M, Vanni A, et al. European position paper on drug-induced sedation endoscopy (DISE). *Sleep Breath* 2014;**18**:453–65.

Denolf PL, Vanderveken OM, Marklund ME, Braem MJ. The status of cephalometry in the prediction of non-CPAP treatment outcome in obstructive sleep apnea patients. *Sleep Med Rev* 2015;**27**:56–73.

Eastwood PR, Malhotra A, Palmer LJ, et al. Obstructive sleep apnoea: from pathogenesis to treatment: current controversies and future directions. *Respirology* 2010;**15**:587–95.

Epstein LJ, Kristo D, Strollo PJ Jr, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;**5**:263–76.

Holty JE, Guilleminault C. Maxillomandibular advancement for the treatment of obstructive sleep apnea: a systematic review and meta-analysis. *Sleep Med Rev* 2010;**14**:287–97.

Murphey AW, Kandl JA, Nguyen SA, et al. The effect of glossectomy for obstructive sleep apnea: A systematic review and meta-analysis. *Head Neck* 2015;**153**:334–42.

Murphey AW, Kandl JA, Nguyen SA, et al. The effect of glossectomy for obstructive sleep apnea: a systematic review and meta-analysis. *Otolaryngol Head Neck Surg* 2015;**153**:334–42.

Senchak AJ, McKinlay AJ, Acevedo J, et al. The effect of tonsillectomy alone in adult obstructive sleep apnea. *Otolaryngol Head Neck Surg* 2015;**152**:969.

Strollo PJ Jr, Soose RJ, Maurer JT, et al. Upper-airway stimulation for obstructive sleep apnea. *N Engl J Med* 2014;**370**:139–49.

Vanderveken OM. Combination therapy for obstructive sleep apnea in order to achieve complete disease alleviation: from taboo to new standard of care? *J Dental Sleep Med* 2015;**2**:7–8.

Weaver EM, Woodson BT, Yueh B, et al. Studying Life Effects & Effectiveness of Palatopharyngoplasty (SLEEP) study: subjective outcomes of isolated uvulopalatopharyngoplasty. *Otolaryngol Head Neck Surg* 2011;**144**:623–31.

Woodson BT, Soose RJ, Gillespie MB, et al; STAR Trial Investigators. Three-year outcomes of hypoglossal cranial nerve stimulation for obstructive sleep apnea: the STAR trial. *Otolaryngol Head Neck Surg* 2016. [In press].

A complete reference list can be found online at ExpertConsult.com.

Chapter
150

Pharmacotherapy, Complementary, and Alternative Medicine for Sleep Bruxism

Ephraim Winocur; Luis Buenaver; Susheel P. Patil; Michael T. Smith

Chapter Highlights

- Sleep bruxism and obstructive sleep apnea are two common disorders involving oral and/or oropharyngeal dysfunction that (1) can be independent or concomitant and (2) can be difficult to successfully manage by means of mainstream, conventional medical approaches.
- Although conventional management approaches are efficacious, such as with dental appliances to prevent tooth wear from repeated bruxism, they often are poorly tolerated, and alternative pharmacologic and complementary approaches frequently are sought by the patient.
- This chapter reviews emerging pharmacotherapies, behavioral, and complementary and alternative medicine approaches for the management of sleep bruxism.
- Also given appropriate consideration are issues of importance regarding clinical application of these approaches: the preliminary nature of the evidence base, the need for systematic investigation of promising interventions, and potential side effects and the minimal or modest efficacy of many of these approaches that limit their use at this time.

This chapter reviews the evidence supporting the use of pharmacologic therapies, psychological/behavioral interventions, and complementary and alternative medicine (CAM) to manage sleep bruxism (SB). Orthodontic, oral appliance, and surgical treatments for sleep-disordered breathing (SDB) are covered in Chapters 143, 147, and 149, respectively. Detailed information on SB definition, prevalence, etiology, and differential diagnosis is specifically found in Chapters 144 and 145.

SLEEP BRUXISM

Bruxism is a repetitive jaw-muscle activity characterized by clenching or grinding of the teeth and/or by bracing or thrusting of the mandible. Bruxism has two distinct circadian manifestations: It can occur either during sleep—*sleep bruxism*(SB)—or during wakefulness—*awake bruxism* (AB).[1] This distinction between AB and SB is made because the etiologies are increasingly believed to be distinct and the interrelationships between the two entities are unclear.[2]

SB is broadly conceptualized as a sleep-related movement disorder, but its etiology remains poorly understood. As opposed to earlier conceptualizations of SB, which had little empirical support and emphasized the role of malocclusion and peripheral nervous system mechanisms, modern evidence-based theories emphasize central nervous system dysfunction and autonomic arousal as primary to the pathophysiology of SB. Multiple intrinsic factors, including genetics, aberrant autonomic sympathetic-cardiac activation, and neurotransmitter system dysfunction, have been implicated in the disorder.[3] Extrinsic factors such as medication side effects and tobacco use also may induce and/or exacerbate AB and/or SB.[4] In view of the lack of clear understanding of SB and the likelihood of a heterogeneity of pathophysiologic mechanisms, it is not surprising that a definitive cure for SB is lacking.

Current conventional treatments focus on managing the harmful consequences of SB and protecting orofacial structures. The most frequently used interventions are based on use of dental appliances to prevent tooth wear. These approaches, however, do not treat the underlying pathophysiology, and the devices often are poorly tolerated by the patient.[5] Medications, behavioral interventions, and CAM approaches for treatment of SB have received some preliminary research attention, but decisions to intervene with these approaches must be considered carefully, owing to reported side effects, relatively sparse scientific evidence regarding their use, and a potentially unfavorable cost-benefit ratio. Presented next is an overview of pharmacologic approaches, behavioral interventions, and CAM for SB management.

EMERGING AND ALTERNATIVE APPROACHES TO MANAGEMENT OF SLEEP BRUXISM

Beyond conventional medical treatments, additional management strategies for use in SB fall into three categories: (1) pharmacotherapies, (2) behavioral interventions, and (3) CAM. Of note, the studies reviewed in the subsequent discussions of these categories are not intended to exhaustively represent these three modalities. The emphasis is on empirical studies published in the English language, specific to subjects diagnosed with SB. Also mentioned are some

approaches that may be widely used but have minimal empirical support.

Pharmacotherapies for Sleep Bruxism

Because SB is widely believed to be centrally mediated, ideal treatments presumably would be neurotropic medications that may regulate cerebral and autonomic hyperarousal (which can be a state of sustained hypervigilance and/or repetitive physiologic reactivity). A critical review intended to assess the exacerbating and ameliorating effect of medication on bruxism in humans, however, found relatively more drugs that cause or aggravate SB than drugs that effectively reduce bruxism.[6] At present, strong evidence–based data from which to draw firm clinical conclusions concerning the effects of medication on SB are lacking.[7] The literature remains controversial and is based largely on anecdotal case reports or experimental studies targeting SB mechanisms.[6] Accordingly, owing to lack of efficacy and efficiency data, no definitive recommendations can be made about the pharmacologic treatment of SB. Table 150-1 summarizes selected studies from the relevant literature.

Sedative, Anxiolytic, and Antidepressant Drugs

Psychosocial stress and psychopathologic conditions, particularly anxiety, have traditionally been conceptualized as playing a pathophysiologic role in SB, although this perspective has

Table 150-1 Drug Therapies for Sleep Bruxism

Reporting Study	Study Design	Drug (Generic Name)	Influence on Sleep Bruxism Muscle Activity Index	Comments
Saletu et al. 2010[9]	Single-blind, placebo-controlled, nonrandomized, cross-over. (polysomnographic study)	Clonazepam	↓	Possible serious adverse side effects
Montgomery et al., 1986*	Open trial (conducted with portable EMG)	Diazepam	↓	Possible serious adverse side effects
Winocur et al., 2003[6]	Case reports (described in critical review)	Buspirone	Mostly ↓ *In a few cases*: ↑ or ↔	Possible serious adverse side effects Relieve SSRI-induced SB
Janati et al., 2013[14]	Single case report	Baclofen	↓	Possible serious adverse side effects
Kast, 2005[15]	Case report	Tiagabine	↓	Possible serious adverse side effects
Ghanizadeh and Zare, 2013[17]	Randomized double-blind, placebo controlled clinical trial (questionnaire only)	Hydroxyzine	↓	Pediatric study
Huynh et al., 2006[10]	Randomized *experimental* controlled crossover studies with placebo and active treatments. (polysomnographic study)	Clonidine	↓	Possible serious adverse side effects
Winocur et al., 2003[6]	Cross-over, double-blind study (described in critical review)	Bromocriptine	↓ or ↔	Lobbezoo (1997) = ↓ 4 out of 6 patients dropped out due to serious adverse side effects Lavigne (2001) and Nishioka (1989) = ↔
Lee et al., 2010[21]	Randomized clinical trial	Botulinum toxin	↓	Reserved only for severe cases of SB owing to occasional complications and high cost
Tan and Jankovic, 2000[20]	Open-label prospective study	Botulinum toxin	↓	
Shim et al., 2014[22]	Open-label prospective study (some polygraphic methods)	Botulinum toxin	↓	Amplitude of contractions reduced, not frequency or duration of muscle activity; no action on central pattern generator of SB activity

One study reported equipotnet effect of gabapentin vs. occlusal splint at 2 months: Madani AS, Abdollahian E, Khiavi HA, et al. The efficacy of gabapentin versus stabilization splint in management of sleep bruxism. *J Prosthodont* 2013;22(2):126–31.

**Complete source information*: Montgomery MT, Nishioka GJ, Rugh JD, et al. Effect of diazepam on nocturnal masticatory muscle activity. *J Dent Res* 1986;65:9 [abstract].

SB, Sleep bruxism; ↓, ameliorate; ↔, no effect; ↑, exacerbate.

been questioned owing to a lack of clear evidence linking these phenomena specifically to SB. AB has been more clearly linked with anxiety and psychosocial stress.[8] Nevertheless, benzodiazepines have been used clinically for SB because of their anxiolytic, muscle relaxant, and hypnotic properties. A placebo-controlled polysomnographic study found clonazepam to significantly improve both the SB index and subjective sleep quality[9]; diazepam has been reported (only in an abstract format) to significantly reduce nocturnal masseter electromyogram (EMG) activity in patients suffering from clinical symptoms of masticatory hyperactivity.[10] Caution is warranted in prescribing benzodiazepines, especially for long periods, because of the potential for various adverse side effects, including respiratory depression, tolerance, dependency, abuse, seizures on abrupt withdrawal, somnolence, and occasionally muscular hypotonia and coordination disturbances. Respiratory depression is a particular concern in patients with comorbid sleep apnea. Dependency and abuse are among risk factors that prevent general use of clonazepam for SB management.

Buspirone, an atypical anxiolytic, has been reported to relieve selective serotonin reuptake inhibitor–induced bruxism ("secondary bruxism") in some patients, to have no such effect in others, and to possibly cause bruxism in another study.[10] It is not a medication likely to be selected for primary SB management.

Use of sedating tricyclic antidepressants such as amitriptyline has been suggested as a treatment for SB, on the basis of presumed associations between depression and SB. However, negative findings have been reported (i.e., no reduction in SB motor index), and insufficient evidence-based data have been accumulated to support the foregoin assumption.[6,11] Moreover, a recent comparative study that aimed to investigate the short-term effects of occlusal splint therapy versus tricyclic antidepressants (amitriptyline) found that occlusal splint therapy may be more effective than these drugs in the management of bruxism.[12] In general, available evidence suggests that drugs prescribed to treat anxiety and depression are relatively poor candidates for SB management—particularly selective serotonin and serotonin-norepinephrine reuptake inhibitors, which in turn also can exacerbate AB and/or SB.[13]

GABAergic Medication

Baclofen,[14] a gamma aminobutyric acid (GABA) agonist, and tiagabine,[15] a GABA reuptake inhibitor, were found to effectively decrease SB. These studies suggest that SB, which has been categorized since 2005 by the *International Classification of Sleep Disorders* (ICSD2)[16] as a movement disorder, might be successfully treated with the inhibitory neurotransmitter GABA, as with other sleep movement disorders such as restless legs syndrome and periodic limb movement disorder. This assumption regarding pathophysiology requires further investigation in well-designed studies to support such medication use in SB management.

Antihistaminic Medication

A recent, randomized placebo-controlled clinical trial investigated the efficacy of hydroxyzine for treating parent-reported SB in 30 children.[17] Hydroxyzine is a histamine H_1 receptor antagonist with sedating properties that typically is used to treat pruritus (itching) and anxiety in children. Compared with placebo, hydroxyzine decreased the parental self-reported bruxism score without serious adverse effects. Although the mechanism of action is not known, the investigators speculated that the effects may be due to increased sleep depth, decreased anxiety, and muscle relaxation. More research with objective measures on SB muscle index (as described in Chapters 144 and 145) and assessing side effects (such as school-time sleepiness in children) is warranted before following such avenues of investigation in SB management.

Sympatholytic Medications

Sympatholytic medications have been suggested as having the potential to reduce SB by dampening sympathetic nervous system arousal during sleep. A promising experimental randomized control study of sleep bruxers found that clonidine, an $\alpha 2$ adrenergic agonist, but not propranolol, an β_2-adrenergic receptor antagonist, reduced SB activity by decreasing sympathetic tone in the 60-second period preceding the onset of SB events. The investigators concluded that the diminution of paroxysmal sympathetic activation, which typically is observed before an SB event, was associated with reduced subsequent motor activation during SB.[10] Caution is indicated, however, in prescribing clonidine for SB because of its reported serious adverse side effects, including hypotension, syncope, bradycardia, atrioventricular block, somnolence and fatigue, headache, sexual dysfunction, and others. Coordinated collaboration with the patient's primary physician is especially important in prescribing these medications, at the lowest dose as possible. It is not yet a safe route for managing SB in the general population.

Dopamine-Related Agents

A relationship between dopamine, the immediate precursor of norepinephrine, and SB has been hypothesized on the basis of the role of dopamine in movement disorders such as Parkinson disease. An early neuroimaging study explored possible functional abnormalities of the central dopaminergic system in humans with SB.[3] This double-blind, placebo/experimentalcontrolled trial investigated the effects of bromocriptine, a dopamine agonist, on SB. Unfortunately, four of six subjects discontinued participation owing due to severe adverse side effects. The two participants who completed the trial, however, demonstrated a 20% and 30% reduction, respectively, in number of bruxism episodes per hour (rhythmic masticatory muscle activity [RMMA]) of polysomnographically measured sleep compared with a placebo.[18] A subsequent study by the same group[11] and an investigation by another group,[19] however, failed to demonstrate any significant effects of bromocriptine on SB. Furthermore, an experimental controlled trial showed a mild beneficial effect of L-dopa, a catecholamine precursor, on RMMA (20% to 30% reduction of SB as reported by Lobbezoo and coworkers[6]).

The literature on the effects of dopamine antagonists on SB in humans also is mixed. Some studies report exacerbating effects of haloperidol and other antagonists, one study reported no impact, and at least one study found that risperidone was associated with clinical improvement in patients with SB.[6] Thus the effect of dopamine-related medications on bruxism remains unclear. More controlled, evidence-based research that identifies potential subgroups of responders and those at risk for dopamine\-related exacerbation is needed.

Botulinum Toxin Injections

Botulinum toxin (BT) injections commonly are used to treat cervical dystonia, blepharospasm, hemifacial spasm, tardive

dyskinesia and sometimes in dentistry for severe oromandibular dystonias, including bruxism. BT is a neurotoxin produced by the anaerobic bacterium clostridium botulinum, which prevents the release of acetylcholine from presynaptic vesicles at the neuromuscular junctions, resulting in the blockade of motor fibers. Clinical effects are a temporary muscle contraction weakness, typically lasting 3-4 months.[13] Studies of BT injections for SB are limited, including a handful of case reports, two open-label prospective studies, and at least one small-sample-size randomized, controlled clinical trial.

In the first open-label study, 18 subjects diagnosed with severe bruxism were treated with a mean dose of approximately 62 mouse units (MU) per masseter muscle over a 3-year period.[20] Eighty-nine percent of the patients in this self-report study indicated a marked relief of self-awarenss of tooth grinding (a debated marker if based on self-report) and functional improvement in chewing, swallowing, and/or speaking. The investigators concluded that subjects required BT injections approximately every 5 months for effective, sustained relief. A lack of placebo control and objective EMG measurement, however, limits conclusions that can be drawn from this study.[20]

In a small-sample-size randomized, placebo-controlled trial (n = 12), 80 MU of botulinum toxin A (Dysport; Ipsen Ltd., Wrexham, United Kingdom) diluted in 0.8 mL of saline significantly reduced EMG-measured SB compared with placebo saline injections at 4, 8, and 12 weeks. No subjects withdrew from the study, and no adverse events were reported.[21]

A more recent open-label study of patients with SB (n = 20) whose bruxism was refractory to oral splint treatment evaluated the effects of bilateral BT injections to either the masseter or the masseter plus temporalis muscles.[22] A comprehensive PSG assessment of SB activity conducted 4 weeks later demonstrated that participants in both treatment groups exhibited significant reductions in the peak amplitude of rhythmic masticatory muscle activity but no differences in the frequency, number, or duration of bursts.[22] A majority of the patients also reported a significant reduction in morning jaw muscle stiffness. Of note, 70% of subjects reported localized discomfort associated with a decrease in the sensation of masticatory force, but this generally was well tolerated over the 4 weeks. Fifteen percent of the affected subjects, however, complained that muscle weakness contributed to masticatory difficulties.

Although the available data suggest that BT injections may be an efficient treatment for severe oromandibular dystonias, the long term consequence and risk of BT in SB patients need to be balanced with SB severity. Furthermore, the small number of studies and unresolved issues including occasional complications related to local muscle weakness and the high cost of treatment indicate that BT injections should be reserved for extreme cases of SB in which all other therapies have failed and the adverse clinical consequences of SB, including significant tooth wear, interference with dental rehabilitation, jaw muscles and/or temporomandibular joint pain, headaches, and social /marital conflicts, are evident. In the presence of secondary AB or SB due to psychotic medication for mental health disorders, BT use is to be recommended with caution, because BT action is temporary (i.e., it lasts only a few months).

Behavioral Interventions for Sleep Bruxism

The studies on behavioral interventions for SB reviewed here derive from mainstream psychological theories and practice, some of which are more relevant to AB and were conducted when the distinction between AB and SB was not well established or made. AB and SB can overlap in a significant number of subjects but with a different time course over the lifespan.

Massed Practice Therapy and Habit Reversal for Bruxism

Massed negative practice and habit reversal are related behavioral therapies developed to reduce unwanted daytime behaviors, habits, compulsions, and tics. Early clinical approaches to bruxism failed to appreciate that sleep and wake bruxism might be different problems requiring unique treatments. Both forms of bruxism were conceptualized as a ticlike habit,[23] which potentially could be reduced or even extinguished by enhancing the noxious aspects of the behavior through conscious repetition of "bruxing" (voluntary repetitive rapid and brief contraction) or clenching (voluntary sustained contraction) to induce muscle fatigue.[24] One early study that compared such massed negative practice with relaxation training found that neither approach significantly reduced bruxism.[25] Some case reports, however, suggest a possible benefit of massed practice for AB.[23,26,27]

Habit reversal is a related behavioral strategy designed to decrease maladaptive behaviors by increasing awareness of grinding and clenching and substituting a competing behavior such as jaw muscle relaxation, which is subsequently reinforced. Habit reversal was developed by Azrin and Nunn in 1973[28] and has been effective in treating motor disorders (e.g., thumb sucking, nervous tics, nail biting), many of which are oral in nature (as reviewed by Miltenberger and associates).[29] A handful of studies of temporomandibular joint disorder (TMD) have found that habit reversal reduces facial pain[30-34] and associated maladaptive oral habits, which include AB and clenching.[30,32] There is minimal evidence, however, documenting that either massed practice or habit reversal interventions practiced during the daytime can generalize to or be effective for SB.

Arousal with Overcorrection

We are aware of at least one study that combined an arousal procedure with a behavioral intervention known as *overcorrection* to reduce SB behavior.[35] Overcorrection involves repeated practice of a positive substitute behavior (e.g., massage, teeth brushing, flossing, mouth rinsing, and so on) after punishment (awakening) when the targeted behavior is expressed. Arousal with overcorrection is an intervention based on *operant conditioning,* a principle positing that behavior is largely controlled by contingencies; punishment (i.e., the arousal) and reinforcement (i.e., the positive behavior). Arousal with overcorrection has been successfully used to treat nocturnal enuresis.[36]

In an early application of arousal with overcorrection for SB (n = 2),[35] each subject completed multiple baselines over time, alternated by two experimental conditions: (1) arousal alone (spouses awakened patient for 15 to 120 seconds during an SB episode) and (2) arousals with overcorrection. In the overcorrection condition, patients were awakened and instructed to complete a 10-minute procedure that included face and hand washing, brushing and flossing teeth, rinsing

the mouth with water, then use of mouthwash, and repeating. The arousal with overcorrection condition was moderately effective in reducing the number of sleep bruxing episodes (lasting 15 seconds or longer) as observed and recorded by the subject's spouse during the 2-hour period immediately after sleep onset. Similar to other aversive conditioning studies, SB activity returned to baseline levels following cessation of treatment. Limitations of this study include requiring implementation with a partner, the lack of objective measures, and the ability to target only audible bruxing. Although the approach is promising, replication is needed with better objective measures and larger sample sizes, before the efficacy of overcorrection for SB can be determined.

Nocturnal Biofeedback and Aversive Conditioning

Aversive conditioning is a type of behavioral conditioning in which a noxious stimulus is repeatedly paired with a maladaptive behavior targeted for extinction. EMG activity–activated alarms have been investigated as a form of aversive conditioning to treat SB. Typically, an audible tone is triggered during sleep "bruxing" episodes with the purpose of waking a subject from sleep. Under this paradigm, for patients to avoid the noxious feedback and achieve consolidated sleep, they must learn to sleep without bruxing. Early studies using masseter and/or temporalis EMG activity thresholds to induce arousals consistently demonstrated reductions in SB duration, but not in frequency.[37,38] Moreover, most studies found that SB activity returns to baseline or is increased on discontinuation of biofeedback.[39,40]

Investigators seeking sustained treatment effects upon discontinuation of EMG feedback have increased the aversive nature of the intervention by requiring that subjects become fully awake after an alarm. Two small (n = 6 and n = 10) studies employed an arousal task—one requiring subjects to perform a 3- to 5-minute task and the other requiring subjects to get out of bed, cross the room, and record the time and sleep quality—after auditory feedback. Both studies found significant reductions in SB lasting for up to 2 weeks on discontinuation of the feedback.[41] A more recent single case study reported that reductions in bruxism frequency were maintained 6 months after cessation of treatment.[42] Several additional one- and two-subject case studies of EMG feedback plus arousal, however, have yielded mixed results.[43-46]

More contemporary approaches have sought improve biofeedback approaches by (1) delivering a stimulus that does not disrupt sleep and (2) improve upon SB detection algorithms to differentiate true bruxism episodes from benign parafunctional activities. A five-night study (n = 7) explored afferent stimulation of the maxillary division of the trigeminal nerve using mild electrical stimulation of the lip during SB episodes.[47] The investigators found significant reductions in SB event duration but not in event frequency. Subjects did not report being awakened by the stimuli. Four small studies (sample sizes ranging from 11 to 19 subjects) have attempted to classically condition electrical stimulation to grinding and clenching behavior in order to extinguish SB. These studies, termed *contingent electrical stimulation* (CES), used a portable EMG device with advance signal processing features designed to differentiate grinding and clenching activity from other benign oral motor movements.[48-51] Patients precalibrate the device while awake by engaging in a variety of oral motor activities (e.g., grimacing, swallowing) and are able to adjust the level of the electrical stimulation so that it does not induce frank awakenings. Only one of the studies, however, used PSG to demonstrate that CES had no apparent effects on sleep architecture or continuity.[49] These studies suggested that the device could discriminate between common parafunctional jaw muscle activity and SB and that feedback significantly reduced the number of SB episodes per hour. A recent systematic review of biofeedback treatments for SB, however, concluded that much of the literature is of poor quality and subject to bias.[52] This report also found that the pooled CES outcomes data did not show improvement in EMG measures of SB episodes compared with control conditions. At this time, it is unclear whether CES is effective for SB and even less clear whether the effects persist on discontinuation of the device. Newer biofeedback approaches appear to have potential but require larger and well-designed studies.

Complementary and Alternative Medicine Approaches

Hypnotherapy

Although several of the studies and case reports just described are inconclusive, preliminary explorations with varying degrees of promise, they are all grounded in well-established behavioral learning theories, typically considered outside the realm of CAM. Hypnotherapy might arguably be better characterized as a form of CAM. Hypnosis has been described as a state of focused attention involving concentration and inner absorption with a relative suspension of peripheral awareness.[53] The intervention often involves facilitating the patient's tendency to become engrossed in a perceptual or creative experience, the promotion of dissociation (i.e., mental separation of elements of experiences that typically would be processed together), and suggestibility (i.e., responsiveness to social cues, leading to a greater likelihood to comply with hypnotic suggestions).[53] Hypnotic techniques often are used in pain management, the treatment of phobias, and depression, with some empirical support

There are a handful of published papers investigating the use of hypnosis in treating SB.[54-57] Most are case studies that used unstandardized and poorly defined methodology, self-report measures, and combined hypnotherapy with other intervention elements, preventing conclusions regarding efficacy to be drawn. One uncontrolled study in eight subjects, however, incorporated EMG monitoring and found that SB activity was significantly reduced after a suggestive hypnotherapy intervention.[56] EMG data were not collected at follow-up assessments, so it is unclear whether the effects had any durability. Overall, owing to the lack of methodologic rigor, standardization, and objective measurement, the efficacy of hypnotherapy for SB is unclear.

Acupuncture

Acupuncture often is mentioned as a treatment approach for bruxism. Acupuncture has been used to treat and ameliorate chronic facial pain related to TMD and mandibular dysfunction,[58-64] but the available evidence evaluating or demonstrating an effect of acupuncture on SB is minimal.

Nutritional Supplements

Studies investigating the efficacy of certain nutritional supplements (i.e., magnesium, calcium, potassium, and vitamins) are lacking and of insufficient scientific quality to support their use in SB management at this time.

Conclusion

At this time, except for oral appliances, few treatments for SB have been adequately studied to accumulate a scientific knowledge base supporting their use.[65] See selected readings for more information.

CLINICAL PEARLS

- In general, there is insufficient evidence supporting any pharmocotherapy for SB, but some medications, such as GABAergic agents, α2 adrenergic agonists (e.g., clonidine) and antihistamines warrant further research to assess efficacy and safety.
- If used in patients with SB, clonidine and clonazepam should be prescribed in low doses, in exceptional and rare circumstances, for short periods with medical supervision, owing to side effects including hypotension or other cardiovascular problems for clonidine, and to the risk of dependency and abuse or adverse effect on sleep breathing for clonazepam.
- The available evidence suggests that drugs prescribed to treat anxiety and depression are relatively poor candidates for treating SB, particularly selective serotonin and serotonin-norepinephrine reuptake inhibitors, which typically can exacerbate SB.
- BT injections may be efficacious for some cases of SB (for reduced amplitude of muscle activity, but not its frequency or duration). However, this approach should be reserved only for severe cases in which all other therapies have failed and when the harmful consequences of SB outweigh the risks of BT adverse effects.
- Many behavioral/psychosocial treatment approaches to SB are based on outdated conceptualizations of the pathophysiology of SB and generally lack strong support for their efficacy. However, because SB etiopathophysiologic mechanisms can be multiple and are not fully known, treatment planning should not automatically exclude behavioral/psychosocial treatment in the era of personalized medicine.
- Integrated medicine requires further development for managing SB.
- Studies on other CAM approaches to management of SB, such as hypnotherapy, acupuncture, and nutritional supplements in particular, are of limited and insufficient scientific quality to support their use.
- Currently, EMG biofeedback devices to manage SB vary in quality and capability. Some studies, however, support the efficacy of some specific biofeedback devices, with or without arousal procedures. There is an obvious need for larger randomized controlled trials and for effectiveness ("real-world") clinical studies using standardized outcome measures before final conclusions and recommendations can be given.

SUMMARY

In light of the fact that the existing mechanical treatment modalities for SB often are associated with poor adherence rates and/or are lacking in long-term outcomes data, more effort should be made by the biobehavioral and integrative medical communities to develop and test new pharmacologic, behavioral, and CAM treatments for SB. Unfortunately, the evidence base for these approaches remains relatively weak and underdeveloped. None of the reported modalities can be fully recommended at this time. Some approaches, however are quite promising and may have benefit as adjuvants to conventional treatments. As the pysiologic, behavioral, and psychological factors of these diseases become better understood, novel pharmacologic, behavioral, and CAM interventions can be expected to be more readily developed and rigorously tested.

Selected Readings

Carra MC, Huynh N, Lavigne GJ. Sleep bruxism: a comprehensive overview for the dental clinician interested in sleep medicine. *Dent Clin N Am* 2012;**56**:387–413.

de-la-Hoz JL. Sleep bruxism: review and update for the restorative dentist. *Alpha Omegan* 2013;**106**:23–8.

Huynh N, Lavigne GJ, Lanfranchi PA, et al. The effect of 2 sympatholytic medications—propranolol and clonidine—on sleep bruxism: experimental randomized controlled studies. *Sleep* 2006;**29**:307–16.

Klasser GD, Rei N, Lavigne GJ. Sleep bruxism etiology: the evolution of a changing paradigm. *J Can Dent Assoc* 2015;**81**:f2.

Lee SJ, McCall WD Jr, Kim YK, et al. Effect of botulinum toxin injection on nocturnal bruxism: a randomized controlled trial. *Am J Phys Med Rehabil* 2010;**89**:16–23.

Manfredini D, Ahlberg J, Winocur E, Lobbezoo F. Management of sleep bruxism in adults: a qualitative systematic literature review. *J Oral Rehabil* 2015. doi: 10.1111/joor.12322.

Saletu A, Parapatics S, Anderer P, et al. Controlled clinical, polysomnographic and psychometric studies on differences between sleep bruxers and controls and acute effects of clonazepam as compared with placebo. *Eur Arch Psychiatry Clin Neurosci* 2010;**260**:163–74.

Valiente López M, van Selms MK, van der Zaag J, et al. Do sleep hygiene measures and progressive muscle relaxation influence sleep bruxism? Report of a randomised controlled trial. *J Oral Rehabil* 2015;**42**(4):259–65.

Winocur E, Gavish A, Voikovitch M, et al. Drugs and bruxism: a critical review. *J Orofac Pain* 2003;**17**:99–111.

A complete reference list can be found online at ExpertConsult.com.

Section
19

Sleep in Older Adults

Chapter
151

Psychiatric and Medical Comorbidities and Effects of Medications in Older Adults

Steven R. Barczi; Mihai C. Teodorescu

Chapter Highlights

- Insomnia in older adults fits the typical profile of a "geriatric syndrome" as characterized by high prevalence, multifactorial etiology, frequent involvement of multiple organ systems, and ultimately, adverse effects on the older person's quality of life.
- Late-life insomnia frequently coexists with other health problems that may disturb sleep, with a preferred nosology of comorbid insomnia. Because these syndromes cross organ systems and transcend discipline-based boundaries, they challenge traditional ways of planning and delivering care.
- Management of comorbid insomnia requires a multifaceted approach that addresses the associated medical and psychiatric conditions, optimizes medications, intervenes on concurrent primary sleep disorders, and also targets perpetuating factors that may reinforce poor sleep behaviors during and/or after the initial physiologic insult.

SLEEP DISTURBANCES IN THE OLDER ADULT

Epidemiology

The number of Americans older than 65 years of age has been steadily increasing, reaching approximately 40 million people at present.[1] It is estimated that by 2050, the U.S. population will more than double (to 89 million). The prevalence of sleep difficulties rises with increasing age. Sleep is considered to be reflective of overall health and is susceptible to the influence of many endogenous and exogenous factors. Changes in sleep in older adults appear to be more strongly associated with psychosocial and health factors than with aging itself.[2,3] In a cohort of more than 300 community-dwelling subjects with a mean age of 72 years, followed for 3 years, sleep disturbances appeared to be related to physical, environmental, and health factors rather than to age-dependent sleep changes.[4]

Health problems appear to have additive effects with respect to the likelihood of concomitant sleep complaints. The 2003 "Sleep and Aging" survey (part of the "Sleep in America" poll) reported that 36% of people 65 years of age and older without comorbid illnesses had sleep problems, 52% with one to three comorbid conditions had sleep disturbances, and 69% of those with four or more comorbid illnesses had disturbed sleep.[2] In addition, an inverse proportion was noted between the self-perceived quality of the respondents' sleep and the number of their comorbid conditions.[2]

Approximately two thirds of aged Medicare beneficiaries have multiple chronic conditions.[3] The simultaneous occurrence of several medical conditions in the same person constitutes the concept of *multimorbidity*. The prevalence of multimorbidity increases significantly with age in both men and women. The specific prevalence of individual conditions varies depending on the age group, gender, and population

sampled (e.g., community-based versus clinic cohort). Most commonly defined chronic conditions in people 70 years or older include osteoarthritis, hypertension, heart disease, diabetes, obstructive lung disease, and cancer in approximate descending order of frequency, from 55% to 10%.[5,6] In a community-based longitudinal study of more than 1200 elderly persons followed for 2 years, the baseline number of physical disorders was associated with prevalence, incidence, and persistence of insomnia. Insomnia also was associated with incident depression and an increase in reported physical disorders.[7]

Population-based surveys reported that between 89% and 94% of those older than 65 years of age take prescription medications, with nearly 40% taking more than 5 medications and 12% taking more than 10 medications.[8,9] *Polypharmacy* generally is defined as taking more than 5 medications, with incremental increases corresponding to the number of coexisting age-associated diseases.[9] Existing practice guidelines recommend use of multiple medications for a number of chronic diseases (e.g., congestive heart failure [CHF], diabetes). Consequences of polypharmacy include adverse drug effects, drug-drug interactions, and medication cascade effects.[9] The *cascade effect* refers to the use of medications to treat the side effects of other medications.

Nosology

In older adults with chronic illnesses, the terms *primary insomnia* and *secondary insomnia* do not adequately represent the complex interplay of cause and effect in determining etiology (Figure 151-1). Sleep difficulties in older persons can be conceptualized as a geriatric syndrome, because they are attributable to multiple and interdependent predisposing, precipitating, and perpetuating factors.[10] Medical or psychiatric conditions may manifest with precipitating events that produce symptoms of discomfort and emotional distress, potentially leading to increased sympathetic drive, hypothalamic-pituitary-adrenal axis activation, and, ultimately, recruitment of neural systems to produce arousal.[11,12] In acute situations, pathophysiologic changes such as hypoxemia, metabolic derangements, fever, or systemic inflammatory responses also can lead to delirium with characteristic alterations in sleep-wake patterns.[12] Psychological factors of hyperarousal, stress response, predisposing personality traits, and maladaptive attitudes can perpetuate sleep changes seen during illness in later life.[13] In many situations, the interplay between underlying illnesses and their treatments, sleep hygiene, medications, and the sleep environment all are contributing to the problem. The precise relationship between an illness and the neurophysiologic, biochemical, and hormonal sequelae that contribute to sleep disruption remains uncertain in most cases of insomnia in older adults. For these reasons, establishing a clear, consequential relationship becomes a clinical challenge. Therefore the term *comorbid* is believed to more adequately describe insomnia. The 2005 National Institutes of Health State-of-the-Science Conference on Insomnia proposed using the diagnosis of "comorbid insomnia" when an illness coexists with sleep changes but the dynamics of cause and effect are not proved.[11]

Management

The approach to the management of comorbid insomnia is evolving. Historically, if a sleep problem was attributed to an associated physical or mental illness, then the primary focus was on optimizing treatment of that underlying health concern. Intervention consisted of correction of neurochemical, metabolic, and physiologic derangements in illness with the hope that sleep would improve. More recently, a series of controlled trials have demonstrated the efficacy of cognitive-behavioral therapy (CBT) in managing insomnia complaints in a variety of different illnesses. On these grounds, a growing trend is for all comorbid or secondary insomnia to be treated the same as for primary insomnia—with CBT and, when indicated, adjunctive hypnotic therapy.[14] This approach is supported by a study of CBT for insomnia in older adults with osteoarthritis, coronary artery disease, or pulmonary disease. Ninety-two participants (with a mean age of 69 years) were randomly assigned to receive classroom CBT or stress management and wellness training (placebo condition). CBT participants demonstrated larger improvements on 8 of 10 self-report measures of sleep; the type of chronic disease had no impact on these outcomes. These results challenge the clinical distinction between primary and secondary insomnia and suggest that psychological factors are likely to be involved in insomnias that are presumed to be secondary to medical conditions.[15] Although more data are needed to determine the most effective approach, a multifaceted intervention appears to be most logical, including optimization of comorbid health issues and treatment of insomnia using CBT and/or pharmacotherapy.

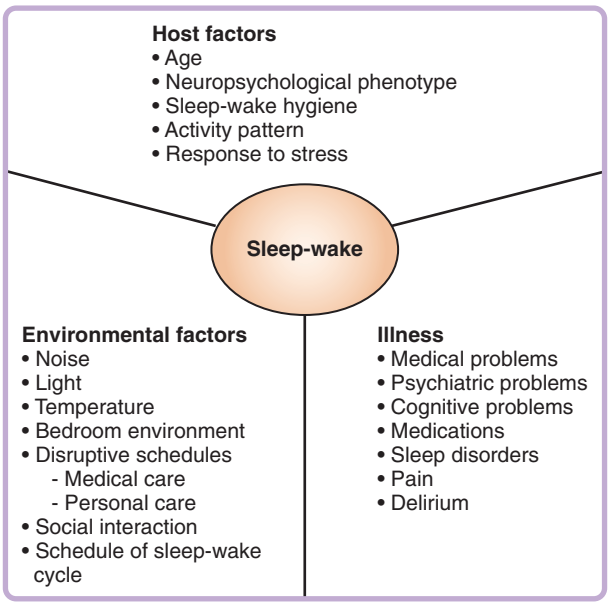


Figure 151-1 Sleep-wake symptomatic status depicted as a syndrome that results from dynamic interactions among genetic and phenotypic characteristics, illness as a function of comorbid conditions and polypharmacy, and environment, to include modulating nociceptive interactions with disruptors and zeitgeber qualities. This model could further identify predisposing, precipitating, and perpetuating factors, but it underscores the limited ability to establish a clear, consequential relationship in the context of the ever-changing interplay of these factors.

PSYCHIATRIC CONDITIONS AND SLEEP IN THE OLDER ADULT

Mental health conditions are prevalent in the geriatric population and often are associated with complaints of difficulties

related to sleep. Indeed, sleep disruption is a part of the diagnostic criteria of many psychiatric disorders. Therefore, to be an independent diagnosis, insomnia has to constitute a distinct complaint and be more prominent than that typically associated with mental disorders.[16] Psychiatric and somatic symptoms, as well as the level of physical activity, significantly and independently increase the risk of insomnia in older clinic patients.[17] Acutely ill psychiatric patients show reduced sleep efficiency, prolonged sleep latency, decreased total night sleep, and increased arousals.[18] Conversely, sleep difficulties in the general population are associated with a higher prevalence of psychiatric disorders and symptoms than in persons without sleep complaints.[19-21] Part of this clinical correlation may reflect the high degree of overlap of medical conditions with psychiatric conditions.

Depressive Disorders

Depression is associated with sleep disturbances in both elderly general populations[22] and clinical populations, with insomnia in the primary care setting showing a stronger association with depression than with any other medical disorder.[23] In a large cross-sectional study (comprising more than 700 subjects, with a mean age of 80 years), approximately one third of participants experiencing depression symptoms (Geriatric Depression Scale score greater than 5) reported moderate to severe sleep onset or sleep maintenance difficulties, whereas almost another third reported mild sleep problems.[24] Older adults with a history of depression show impairments in sleep quality and lower levels of health functioning; these impairments were on a gradient, with declines in those with current depression.[25]

Sleep complaints also foreshadow the onset of depression in older adults, with insomnia a strong risk factor for future depression in elderly patients not currently depressed.[26] Insomnia reported at baseline and 1 year later increased the risk of development of depression eightfold[27] in one elderly cohort. Clinical sleep disturbance, defined as a Pittsburgh Sleep Quality Index (PSQI) score higher than 5, was found to be a powerful predictor of depression recurrence (adjusted hazard ratio, nearly 5) in a 2-year longitudinal study of community-dwelling older adults.[28]

Objective and subjective sleep findings in depression often have different implications and may not always correlate with each other.[29,30] From 60% to 80% of depressed adults and older adults complain of difficulty falling or staying asleep, or being tired in the day.[31,32] These subjective sleep complaints may persist beyond the resolution of depressive symptoms.[33] Prevalence of symptoms is comparable between genders.[34] An apparent association has been noted between greater level of depressive symptoms and worse subjective sleep quality and more subjective daytime sleepiness, as well as objectively measured wake (time) after sleep onset (WASO) and wake episodes longer than 5 minutes.[35]

Objective sleep findings in older adults with depression include prolonged time to sleep onset, decreased sleep efficiency, poor sleep continuity, and increased early-morning awakening. Additionally, total sleep time is reduced, with relative increases in stage 1 and stage 2 sleep and corresponding decreases in slow wave sleep. Other findings include shortened rapid eye movement (REM) sleep latency, a longer first REM sleep period, and increased total REM sleep and density of the rapid eye movements.[18,36,37] Although these patterns of sleep change are well described in depressed persons, no single sleep variable currently is specific enough to distinguish depression from other psychiatric disorders.[18] Furthermore, longitudinal studies of EEG sleep profiles in depression show state and trait characteristics, meaning that some of these characteristics can change over time.[38] Such differentiation may play a role in identifying persons at high risk for depression and in predicting future episodes of depression, as well as in assessing the response to pharmacologic and behavioral treatments.

Bipolar Affective Disorder

Only limited data are available with regard to sleep in older adults with bipolar affective disorder. Overall, compared with normal subjects, patients with bipolar spectrum disorders experience significantly more sleep loss and social rhythm disruption after both minor and major life events, leading to disrupted biologic rhythms and affective symptoms.[39] Manic episodes usually involve a prolonged period of increased activity with minimal sleep. Insomnia can precede manic episodes as a first symptom in up to 77% of cases.[40] Although first-onset mania can occur in elderly persons, this feature warrants suspicion for medical illness or drug effects as contributors. Of people with initial onset of mania in later life, many have a history of depression.[41] Moreover, sleep loss may trigger manic episodes, emphasizing a need to clinically monitor sleep changes and to address sleep-disrupting factors in an attempt to decrease such exacerbations.[40]

Bereavement and Grief

A systematic review looking at late-life spousal bereavement found a strong relationship between bereavement and nutritional risk with involuntary weight loss, and between evidence for impaired sleep quality and increased alcohol consumption.[42] Levels of grief usually are highest within the first 4 months of bereavement, with recovery periods varying with the stress of the bereavement or coping skills and extending to 2 to 3 years. In 38 bereaved seniors assessed at more than 2 months after the event, questionnaire- and diary-based measures of sleep correlated with the level of depressive symptomatology. Sleep duration correlated with the level of grief severity, although no sleep measure correlated with days since loss.[43] In a cross-sectional survey of 170 women (mean age, 66 years) assessed at an average of 26 months from the loss event, insomnia was reported by 13% of subjects. Other sleep-related symptoms included an irregular sleep pattern, nightmares, and sleeping excessively.[44]

Sleeping 6.5 to 9 hours per night at baseline predicted better social functioning, better emotional health, and more energy in older persons who have suffered bereavement.[45] Baseline complicated grief scores of recently widowed elderly persons were significantly associated with sleep difficulties at an 18-month follow-up evaluation.[46] In a study of 28 spousally bereaved older adults assessed at 4 months from the loss event or later, polysomnographic data demonstrated mildly prolonged latency to sleep and REM sleep and mildly reduced sleep time and sleep efficiency. More profound grief tended to be associated with less time spent asleep and reduced alertness at 8 PM.[47] In summary, problems with sleep during bereavement may predict current and future depression and offer a potential target for treatment.

Anxiety Disorders

Common anxiety conditions in older adults include generalized anxiety disorder (GAD) and panic disorder. Most anxiety conditions in the elderly are continued disorders from earlier in their life, with the exception of some increased agoraphobia. Although they cannot be said to suffer from a specific anxiety disorder, many older adults can experience increased anxiety when faced with chronic health conditions, functional limitations, or concerns about issues such as safety or finances, which in addition to the effects of certain medications used to treat other conditions may contribute to neurochemical changes and/or increased adrenergic drive. This association is reflected in the finding that even subclinical anxiety symptoms are associated with higher levels of wake after initial sleep onset as measured by actigraphy and sleep log.[44]

GAD, a condition of overall hyperarousal, is the most prevalent condition among subjects complaining of insomnia who have a mental health diagnosis in the adult population.[48] Sleep disturbance is a core symptom in the diagnosis of GAD and is reported by appproximately two thirds of patients with this diagnosis.[43] In a cross-sectional sample derived from the Einstein Aging Study (comprising 702 participants with an average age of 80), approximately 30% of subjects scoring in the anxiety range (Beck Anxiety Inventory score greater than 11) indicated moderate to severe sleep onset or sleep maintenance difficulties, whereas approximately 50% reported mild difficulties.[24] Objective findings include more awakenings, longer time to fall asleep, and decreased sleep efficiency and total sleep time, with less consistent findings of more stage 2 and decreased slow wave sleep.[18,49]

Approximately 80% of older adults with GAD also have a depressive disorder, making specific objective sleep findings difficult to interpret. Anxiety symptoms, however, were associated with poor sleep efficiency and more time spent awake after sleep onset (i.e., sleep fragmentation) in a population of elderly women, even after accounting for significant depressive symptoms.[50] These findings suggest that untreated anxiety symptoms might account for poor sleep quality in older women that persists even after treatment of depressive symptoms.

Panic disorder also is associated with sleep complaints, including trouble falling sleep and disturbed and restless sleep, as well as nocturnal panic attacks. Nearly one fourth of patients with panic disorder report either severe sleep restriction (to 5 hours or less) or increases in sleep duration (to 9 hours or longer).[51] Nocturnal panic attacks occur at least weekly in 18% to 45% of affected patients.[52,53] These attacks are characterized by similar symptom severity and duration of daytime attacks and occur usually in the first few hours of sleep in transition from stage 2 to slow wave sleep.[54] People with nocturnal panic attacks have higher rates of insomnia and depression than people with panic disorder without nighttime episodes.[55,56] Sleep panic attacks can manifest with aspects of the clinical history similar to that for sleep apnea, parasomnias, gastroesophageal reflux disease (GERD), and posttraumatic stress disorder (PTSD)—a consideration during the workup for sleep complaints.

In older adults with PTSD, many had symptoms for years, with this condition continuing in later life. It appears that although the severity of PTSD seems similar in younger and in older adults, the elderly report less reexperiencing but more symptoms of hyperarousal compared with younger people.[57] Sleep disturbance is common in PTSD, with reported rates of 44% to 91%, including insomnia and nightmares.[58] The most specific findings in PTSD are recurrent awakenings and excessive body movement.[59] A meta-analysis of 20 polysomnographic studies comparing sleep in people with and without PTSD reported more stage 1 sleep, less slow wave sleep, and greater density of rapid eye movements; older patients with PTSD had less slow wave sleep and more REM sleep than age-matched control participants, with possible confounding by the time elapsed since traumatic experiences.[60]

Elderly persons also may be more predisposed to exacerbations or new onset of symptoms in the context of worsening health conditions or on being confronted with the death of friends or spouses, as well as their own mortality. In addition, previous mechanisms for coping may now be affected by physical (e.g., as in a man who habitually works himself to fatigue, or runs to relieve tension, but now has retired or has pain that limits his exercise ability) or cognitive limitations (e.g., regular inhibition of intrusive thoughts leading to more effective coping may be compromised in a person with executive dysfunction such as mild cognitive impairment), which are more likely to occur with increased age.

MEDICAL CONDITIONS AND SLEEP IN THE OLDER ADULT

A specific medical illness may have several mechanisms that interfere with sleep, and different effects on sleep architecture may be seen in its acute versus its chronic state. The immediate physiologic derangements or distress attributable to an illness may transiently disturb sleep. Chronic pain, cardiovascular disease, pulmonary disease, chronic kidney disease, gastrointestinal conditions, and endocrine and genitourinary conditions all have been associated with poor sleep. Sleep disturbance may worsen symptoms in these disorders or even worsen the prognosis.

Pain

Pain is a common contributor to comorbid insomnia. The relationship between pain and sleep is complex: Pain can disrupt sleep, and poor sleep may increase perceived pain intensity.[61] Studies suggest that 25% to 50% of community-dwelling older persons have important pain problems.[62] In a population of adults aged 55 to 84, 19% reported that pain disrupted their sleep at least a few nights per week, and 12% reported almost nightly sleep fragmentation due to pain.[2] In referral populations of patients with chronic pain, prevalence rates of insomnia can range between 50% and 70%.[63] In a polysomnographic study of older adults with chronic pain, afflicted persons spent significantly longer time in bed and had worse sleep onset latency, sleep latency to N2, sleep efficiency, wake time after sleep onset, and number of awakenings compared with control group subjects; sleep duration and time spent in each sleep stage did not differ between the two groups. Investigators also found that older people with chronic pain had lower intensity in the delta frequencies (0.5 to 1.99 Hz and 2 to 4 Hz) throughout the night, especially in the first 6 hours.[64] Actigraphy recordings and sleep diary data seem to confirm these findings and showed subjects with

chronic pain spending significantly more time in bed, with consequent lower sleep efficiency.[65]

Arthritis

As a condition producing pain and disability, arthritis increases with advancing age.[66] Evidence suggests that as many as 60% of those with arthritis experience pain during the night, mediating a substantial amount of the relationship between arthritis and sleep problems. Adults older than 65 with knee arthritis have been observed to have problems initiating sleep (31%), problems maintaining sleep (81%), and a tendency to awaken early in the morning (51%).[67] In a sample of approximately 600 older adults with osteoarthritis (with a mean age of 78), correlates of poor sleep were greater arthritis severity, presence of three or more comorbid conditions, depressed mood, and restless legs symptoms; poor sleep was significantly associated with greater fatigue.[68]

Patients with self-reported arthritis-related sleep disruption are more likely than those without sleep disturbance to pursue multiple sources of self-care and medical care.[69] Cognitive-behavioral approaches demonstrated efficacy in improving self-reported measures of sleep.[15] At least one third of respondents in a recent survey had clinically moderate to severe levels of pain and sleep symptoms, significant enough to lead almost half of the affected patients to ultimately participate in a randomized trial comparing three group-format behavioral interventions.[70]

Gastroesophageal Reflux Disease

Insomnia frequently is associated with GERD.[71] Up to 70% to 90% of persons with GERD report nighttime symptoms and sleep disruption.[72,73]

The prevalence of daily reflux symptoms among those older than 50 years has been reported at 10%.[74] A survey including more than 14,000 respondents reported a higher prevalence of heartburn in elderly persons (approximately 62%) than in younger adults (59%), and the prevalence of frequent symptoms (more often than twice per week) also was higher among elderly subjects (approximately 31%).[75] Sleep disturbance was present in 29% of the older adults, compared with 19% of younger ones. The relationship between disturbed sleep and GERD is likely to be bidirectional: Sleeping increases the likelihood of reflux, and reflux episodes often awaken the patient.[76,77]

Patients with nighttime acid reflux may underestimate the degree of sleep disruption that occurs when objective measurements of pH and electroencephalogram (EEG)-confirmed arousal are compared against patient recollection the next morning.[76] On ambulatory 24-hour esophageal pH monitoring in 54 of 313 consecutive patients (older than 62 years of age) from a primary care setting, 20% (11 of 54) of this subgroup demonstrated increased acid contact time; only 6 patients (11%) exhibited both symptomatic and objective indications of acid reflux.[78] A pilot study enrolling 16 subjects with chronic insomnia found silent reflux in 4 (25%), as evidenced by abnormal results on 24-hour pH testing. Aggressive treatment for reflux resulted in normalizaton of sleep efficiency in 3 of the 4 affected subjects.[79]

Reviewing the clinical history for nocturnal cough or wheezing as a surrogate for reflux also is important, because not all patients with overnight reflux will experience classic chest pain, but their sleep may be disrupted nevertheless. Published evidence demonstrates that acid suppression therapy helps control nighttime heartburn symptoms, reduces the number of GERD-associated sleep disturbances, and improves subjective sleep quality and next-day work performance.[80] Finally, a mechanical link between the phrenoesophageal ligament and the lower esophageal sphincter may explain the coexistence of sleep apnea and reflux.[77,81]

Heart Disease

Coronary artery disease and CHF together constitute a leading cause of morbidity and mortality in older adults. Many relationships exist between cardiac disease and sleep. Nocturnal ischemia, nighttime arrhythmias, and sleep-disordered breathing all are linked to altered sleep in underlying heart disease. A well-described circadian pattern of myocardial ischemia or infarction occurs in early to mid-morning and is ascribed to the catecholamine surge that accompanies awakening and upright status. In a retrospective analysis of data on more than 3300 adults presenting with acute coronary syndrome (ACS), 26% of the subjects were awakened from sleep[82]; older age and lower left ventricular ejection fraction were independent predictors of nocturnal ACS in this cohort. Chronic problems with sleep initiation correlate with an increased risk of death from coronary artery disease in male patients.[83] In study of more than 1200 women experiencing initial myocardial infarction, one half reported new onset of or worsening sleep disturbance before myocardial infarction, with similar prevalence rates across racial groups. Women reporting prodromal sleep disturbance were more likely to be older, to be heavier, and to report cognitive changes, new or increasing anxiety, and unusual fatigue.[84]

Moderate or severe insomnia was reported by 37% of patients experiencing an ACS during hospitalization and was associated with 76 minutes more awake time after initial sleep onset as measured by home polysomnography (PSG). Although depression and insomnia were strongly associated, approximately 1 in 4 patients with insomnia did not report significant depressive symptoms.[85]

Finally, coronary artery bypass surgery is associated with protracted sleep disturbance up to 2 years after the procedure.[86] The mechanism for this sequela is unclear, with occult heart failure, secondary mood issues, or brain microvascular ischemic changes as possibilities.

With increasing average life span and improvements in the management of acute coronary ischemia, it is projected that the incidence and prevalence of CHF will continue to rise. In a cross-sectional study of 223 elderly patients with New York Heart Association class II to IV heart failure, consistent difficulties maintaining sleep were reported by 23% of men and 20% of women, and 25% of the subjects were awake 1 to 3 hours per night.[87] In a study of more than 600 older adults (mean age, 78), cardiopulmonary symptoms (i.e., dyspnea and nighttime palpitations) and pain led to significant direct associations with sleep disturbances.[88] Presence or absence of heart failure did not alter this association. In a multisite randomized, controlled trial including subjects with a heart failure–related hospitalization (in the previous 6 weeks), 45% of participants reported poor sleep (PSQI of 5 or higher).[89] Compared with that in control subjects, the PSQI global score improved with twice-weekly structured exercise training. Improved sleep quality correlated with improved exercise capacity and reduced depressive

symptoms, but not with changes in body mass index or resting heart rate.

The classic sleep maintenance disturbance associated with CHF includes orthopnea, paroxysmal nocturnal dyspnea, nocturia, and sleep-disordered breathing. More than 50% of patients with moderate to severe CHF experience periodic breathing or Cheyne-Stokes respiration. The altered breathing may lead to increased sleep fragmentation, with consequent increase in daytime sleepiness.[90,91] Comorbid depression also was identified as a factor that increasingly contributes to sleep disturbance in elderly patients with CHF.[92]

Nocturia is common and often severe in patients with stable CHF. In a cross-sectional observational study of 173 patients (with a mean age of 60 years and a left ventricular ejection fraction of 32%) with stable chronic heart failure, a third of patients awakened three or more times per night to void.[93] This group exhibited a nearly sevenfold increase in the frequency of reported insomnia symptoms. Other findings included decreases in sleep duration and efficiency, REM and stage 3 sleep duration, and physical function and increases in the percentage of wake time after sleep onset, insomnia symptoms, fatigue, and sleepiness across levels of nocturia severity.

Chronic Lung Disease

Chronic obstructive pulmonary disease (COPD) contributes to poor sleep continuity as well as to increased daytime sleepiness.[94] In a study of correlates of insomnia in 142 Chinese patients with COPD aged 60 years of age or older, compared with sex- and age-matched control subjects, frequency of insomnia was 47.2% in the patient group and 25.7% in the control group, with higher rates in frequency of early, middle, and late insomnia.[95] Another study of 183 patients with COPD showed higher likelihood of insomnia in current tobacco users (odds ratio [OR], 2.13) and in those with frequent sadness or anxiety (OR, 3.57); oxygen use was associated with lower risk of insomnia (OR, 0.35).[96]

Reported sleep changes in COPD include increases in sleep stage shifts, decreased total sleep time, and increased number of arousals.[97] Even persons with mild to moderate COPD have lower sleep efficiency, a lower total sleep time, and lower mean overnight oxygen saturation compared with control subjects.[98] The sleep disturbance frequently is related to nocturnal cough, wheezing, and shortness of breath due to worsening of pulmonary mechanics and gas exchange during sleep. Hypoxemia, which is common in COPD during REM sleep, correlates with an increase in arousal and excessive daytime sleepiness. Although the use of oxygen therapy frequently corrects the underlying hypoxemia, it does not appear to improve sleep quality.[97] Use of inhaled ipratropium bromide improves sleep quality and duration, presumably through improved airflow.[99] However, use of a long-acting bronchodilator, such as formoterol, although overall beneficial for the lung process, may result in insomnia.[100]

The prevalence of insomnia symptoms in a sample of more than 1800 subjects with asthma, part of a 25,000-subject questionnaire, was significantly higher among asthmatics than nonasthmatics (47% versus 37%).[101] In the subgroup reporting asthma and nasal congestion, 56% had insomnia symptoms. The risk of insomnia increased with the severity of asthma; nasal congestion (OR, 1.50), obesity (OR, 1.54), and smoking (OR, 1.71) were other factors associated with increased risk. Nocturnal asthma symptoms resulting in nighttime awakenings may occur in more than 70% of persons with asthma. Other reported sleep symptoms include difficulty falling asleep, difficulty maintaining sleep, early-morning awakenings, and daytime sleepiness. Asthma control correlates with quality of sleep.[102]

Diabetes and Endocrine Disorders

Diabetic patients may have disrupted sleep as a consequence of advanced age, obesity, and treatments for and complications of common comorbid diseases (e.g., depression, cardiovascular disease). Diabetes-specific complications, such as neuropathy, may directly interfere with sleep or contribute to restless legs symptoms and nocturnal leg cramps. In almost 10,000 adults participating in the National Health and Nutrition Examination Survey in the period 2005 to 2008, diabetes was associated with increased risk of inadequate sleep, frequent daytime sleepiness, restless legs symptoms, sleep apnea, and nocturia.[103] All of these showed higher risk with increasing severity in graded fashion. Diabetes duration was significantly associated with the same problems: Risk increased 20% to 30% per 10 years since diagnosis.

In a large study of more than 13,000 adults with diabetes mellitus, 24% reported insomnia.[104] In the more than 200 subjects older than 60 years of age, evaluated at less than 2 years before death, the prevalence was approximately 36%. Nonetheless, insomnia was more prevalent in younger (i.e., younger than 60 years of age) subjects surviving more than 24 months (compared with those 60 years of age or older).

A growing body of literature indicates a link between sleep and glycemic control. Thirty percent of diabetic patients demonstrate sleep maintenance disturbances, with the severity of disruption correlating with the degree of hyperglycemia.[105,106] Of interest, the likelihood of being insulin resistant increases linearly with concurrent sleep complaints.[107] Consequently, high HbA1c was found to be associated with difficulty in maintaining sleep but also early-morning awakenings.[108] In an ancillary study to the Coronary Artery Risk Development in Young Adults, among the 40 subjects with diabetes, 10% higher sleep fragmentation was associated with a 9% higher fasting glucose level, a 30% higher fasting insulin level, and a 43% higher insulin resistance estimated using the homeostatic model assessment method. Insomnia was associated with a 23% higher fasting glucose level, a 48% higher fasting insulin level, and an 82% higher homeostatic model assessment insulin resistance level.[109]

One third of patients with diabetes have problems with sleep fragmentation, with nocturia, leg cramps, leg pain, and cough as contributing factors. Likewise, patients with diabetes have an increased prevalence of both restless legs syndrome (RLS) and periodic limb movements of sleep.[110] A growing body of epidemiologic and experimental evidence links sleep apnea and disorders of glucose metabolism; however the cause and effect relationship remains to be determined.[111]

Thyroid disease also may influence sleep architecture. Among 6000 subjects (older than 65 years) participating in the Osteoporotic Fractures in Men study, no difference in sleep quality was found between subclinically hypothyroid and euthyroid men.[112] Compared with euthyroid men, subclinically hyperthyroid men had lower mean actigraphic total sleep time, lower mean sleep efficiency, higher mean

wake time after sleep onset, and increased risk of sleep latency of 60 minutes or longer.

Renal and Urologic Diseases

Sleep disruption is common in patients with urologic and kidney disease. Benign prostatic hyperplasia and prostate cancer typically are diseases of the aging male, steeply increasing in frequency with age. Overactive bladder is characterized by urinary urgency, frequency, nocturia, and sometimes incontinence. It increases markedly with advancing age in both men and women.[113]

Nocturia is a well-recognized etiologic factor in sleep maintenance disturbance in later life, and nighttime urination often is associated with poor quality of sleep and increased fatigue in the daytime.[114] It is reported to be an independent predictor both of self-reported insomnia (75% increased risk) and reduced sleep quality (71% increased risk) based on a survey of more than 1400 elderly persons.[115] In approximately 16,000 participants in the Third National Health and Nutrition Examination Survey, prevalence of nocturia, defined as two or more voiding episodes nightly, was approximately 16% for men and 21% for women. This study showed that nocturia is a strong predictor of mortality, with a dose-response pattern in increased mortality risk with increasing number of voiding episodes nightly.[116] The severity of nocturia increased with advancing age. Urinary incontinence, recurrent cystitis, and diabetes mellitus were the strongest associated factors for nocturia of any degree.[117] Other etiologic factors include polyuria, low bladder capacity, sleep apnea, excessive fluid intake before bedtime, alcohol, caffeine, diuretics, medical disorders such as hypertension, CHF, and prostatic disease. The *nocturnal polyuria syndrome* is characterized by an inappropriate nocturnal urine output, often an undetectable plasma ADH during the night, and increased thirst, particularly at night[118]; 24-hour urine output is normal or only moderately increased. In older adults with concurrent insomnia and nocturia, brief behavioral treatment for insomnia (consisting of instructions on reducing time in bed and setting a regular sleep schedule) also may produce symptomatic improvement in patients with self-reported nocturia.[119]

Approximately 40% of community-dwelling older adults report sporadic or chronic urinary incontinence,[120] with the most common cause documented as the overactive bladder syndrome. Urine loss is considered a geriatric syndrome in that it leads to a spectrum of physical, psychological, and social consequences that can be a detriment to the older person's function and quality of life. The effects of incontinence on sleep are best studied in the nursing home population with both polysomnographic and actigraphy samples demonstrating sleep disruption with nocturnal wetting episodes. Of interest, 51% of these episodes occurred during or within 60 seconds of an abnormal sleep breathing event.[121] In people who are chronically incontinent, the relationship between urine loss and awakening is not as tightly linked.[122]

Rates of chronic kidney disease (CKD) and end-stage renal disease show a steady increase beyond the age of 60, with a striking rise in prevalence after age 75.[123] Fifty-seven percent of patients with end-stage renal disease report sleep maintenance problems, and 55% report early-morning awakening. There are marked abnormalities seen on the sleep EEGs of patients with CKD, including overall reduction in total sleep time, decreased sleep efficiency due to wakefulness after sleep onset, and reduced total REM sleep.[124] Overall prevalence of insomnia in patients managed with hemodialysis (HD) was described to be between 45% and 86%.[125] In a longitudinal study (up to 2 years) of patients with CKD not on dialysis and those on HD, disruption of sleep was independent of several risk factors.[126] Among the patients on HD, sleep disruption was of much greater severity than among those with CKD not on dialysis. Missing a dialysis session or shortening the prescribed duration of dialysis appeared to be associated with greater severity of sleep disturbance.

Patients on HD have a 50% to 80% rate of sleep-wake complaints and a higher prevalence of OSA (which lessens in severity after dialysis), RLS, periodic limb movements of sleep (PLMS), early insomnia, and excessive daytime sleepiness.[124,127,128] Quality of sleep was associated with hemoglobin level, serum albumin, and depression in 89 patients on HD with a mean age of 60 years. In a study involving patients on HD with a mean age of 65 years, treatment of kidney disease with erythropoietin improved sleep quality by polysomnographic and subjective measures and reduced the number of periodic limb movements.[129] More than 50% of patients on HD report chronic pain, and this complaint is thought to be significantly associated with insomnia and depression in this condition.[130]

Cancer

Cancer frequently is a disease of older persons, with greater than 60% of all cancers diagnosed after the age of 65 years. The prevalence of sleep problems in persons with cancer is difficult to determine with wide variance based upon type and stage of cancer. Large epidemiologic studies suggest that sleep problems are very common, affecting 55% to 87% of patients[131,132]; approximately one half of the patients reported onset within the period 6 months before diagnosis to 18 months after diagnosis.[133] In a study of 867 older adults (46% of whom were women) who were newly diagnosed with breast, colorectal, lung, or prostate cancer and followed at four points during the year after diagnosis, insomnia remained present in 23.2% of patients 1 year after their cancer diagnosis. The study authors reported a lessening of reports of pain, fatigue, and insomnia over time; however, a high attrition rate was noted.[134]

Persons with cancer may have a baseline history of insomnia or a primary sleep disorder, or they may have sleep effects from the cancer, its treatment, or the psychological response to the diagnosis.[132] In a meta-analysis of sleep across chemotherapy treatment regimens, subjective and objective sleep quality was found to be poor, with a high frequency of nocturnal awakenings. Daytime sleepiness increases in the active phase of chemotherapy, and insomnia symptoms are common before and after chemotherapy sessions. In women with recurrent or metastatic breast cancer, difficulty falling asleep, nocturnal awakenings, difficulty awakening, and daytime sleepiness are problematic at different points in the course of chemotherapy.[135]

Most of the studies reported in the literature report on sleep changes in early-stage cancer.[133] Forty-four percent of hospitalized cancer patients are prescribed hypnotic therapy.[136]

In a large sample of approximately 2000 cancer patients, approximately 23% reported taking hypnotic medications. Factors associated with a greater utilization of hypnotic

medication were older age, greater difficulties initiating sleep, more stressful life events experienced in the past 6 months, higher levels of anxiety, past or current psychological difficulties, greater use of opioids, and past or current chemotherapy treatments.[137]

Trials of CBT in cancer patients suggest significant improvement in their sleep with this therapy.[131] In a recent study of 111 cancer subjects with insomnia, although mindfulness-based stress reduction protocols produced a clinically significant change in sleep and psychological outcomes, CBT for insomnia (CBT-I) was associated with rapid and durable improvement and is still regarded as the best choice for nonpharmacologic treatment of insomnia.[138]

MEDICATIONS IN OLDER ADULTS AND IMPACT ON SLEEP

The use of prescription drugs and over-the-counter (OTC) drugs is common in older persons. A multitude of both OTC and prescription medications are known to influence the sleep-wake cycle and to produce comorbid insomnia. Their adverse effects are diverse but can be broadly categorized as those that produce drowsiness or daytime somnolence, those that are activating or stimulating to the brain, those that interfere with sleep by indirect mechanisms, those that may directly exacerbate primary sleep disorders, and those that influence sleep architecture through other effects. Adverse side effects and drug-drug or drug-disease interactions are more likely to occur in older adults. These effects can then lead to direct or indirect effects on overnight sleep quality and quantity. Many of these drugs can alter patterns of sleep and wakefulness both during periods of administration as well as during withdrawal. It is especially important to avoid using hypnotics or stimulants as agents in the cascade effect (use of medications to treat sleep-related side effects of other medications) unless all other attempts at medication adjustment have been considered.

Medications that Promote Daytime Sleepiness

Drowsiness is an extremely common drug side effect, with close to 600 medications cited as causing drowsiness in the side effects index of the *Physicians' Desk Reference*.[139] Excessive daytime sleepiness is a frequent sleep complaint in the elderly population, with a baseline tendency among older adults to have shorter sleep latency during the day.[140] Many medications have the capacity to interfere with acetylcholine or histamine, both regulatory neurotransmitters for wakefulness. These anticholinergic agents are known to have somnogenic effects—drowsiness while awake—as well as negative cognitive, affective, and quality-of-life outcomes in older adults.[141] Antihistaminergic drugs have variable central nervous system (CNS) penetration and binding. Histamine H_1 receptor antagonists such as diphenhydramine are much more likely to produce sedation and cognitive impairment than tertiary antihistamines.[142] General classes of agents with these effects are antihistamines, antispasmodics, antipsychotics, antiemetics, and antiparkinsonian drugs. Notable specific examples include tricyclic antidepressants such as amitriptyline, doxepin, imipramine; cimetidine, mirtazapine, and oxybutynin.

Medications also can produce sleepiness by other mechanisms. In persons taking levodopa or dopamine agonists, an increased prevalence of excessive daytime sleepiness and sleep attacks has been noted.[143] Anticonvulsant agents such as gabapentin, lamotrigine, tiagabine, and levetiracetam frequently produce sleepiness in older adults. Morphine and other opiate analgesics can contribute to daytime somnolence and decreased alertness, as well as disrupting overnight sleep efficiency and architecture.

Medications that Activate the Central Nervous System

A large number of medications disturb sleep through excitation or activation of the CNS. Sleep quality may be affected if these agents are taken before the patient's bedtime or have a sustained half-life that extends into the typical sleep period. Common OTC products for cold and flu contain pseudoephedrine, ephedrine, or other sympathomimetics. OTC analgesics used for headache therapy may contain caffeine. Agents used to manage chronic lung disease such as inhaled and oral beta agonists, corticosteroids, and theophylline can contribute to sleep disruption. Activating antidepressants such as desipramine, bupropion, venlafaxine, reboxetine, and most selective serotonin receptor inhibitors (SSRIs) can sometimes adversely affect sleep initiation and maintenance. Insomnia has been reported as a frequent side effect with SSRI use, with a prevalence of 16.4% with sertraline, 15% with fluoxetine, and 14% with paroxetine.[144] With use of SSRIs, although patients may perceive an improvement in subjective sleep quality, objective sleep often worsens.[29] Activating medications such as methylphenidate, selegiline, and modafinil often are seen on geriatric medication lists. Careful evaluation of doses and times of administration for such medications may be considered to avoid interference with the desired sleep period.

Medications that Affect Sleep by Worsening other Conditions

Medications can sometimes interfere with sleep by worsening an underlying medical or psychiatric condition, which then adversely affects sleep. Medications that worsen heart failure such as nonsteroidal antiinflammatory drugs, calcium channel blockers, or sodium-complexed antibiotics have the potential to cause central sleep apnea, nocturia, or other sleep problems seen in this condition. Medications including nitrates and calcium channel blockers can decrease lower esophageal sphincter tone with resultant nocturnal gastroesophageal reflux. Amitriptyline and other anticholinergic medications, although potentially helpful with sleep because of their sedative effects, also can contribute to confusion as well as urinary retention, with subsequent arousals from delirium or nocturia. Late-afternoon or evening diuretic treatment may cause nocturia and sleep fragmentation. Many antipsychotic medications, which may be used for various symptoms in the elderly, have the potential to produce parkinsonian features, with the sleep difficulties commonly associated with these conditions. Quetiapine may be one of the least common offenders in this category of medications. Hypoglycemic agents, if they produce nocturnal hypoglycemia, can increase nocturnal arousals.

Medications that Can Exacerbate Primary Sleep Disorders

A number of medications have been reported to exacerbate primary sleep disorders. Nocturnal movement disorders

such as RLS and PLMS can worsen in the setting of treatment regimens featuring a number of antidepressant medications. In a study of 274 consecutive patients on antidepressants, as compared with 69 control subjects, those taking SSRIs or venlafaxine were five times more likely to have an elevated PLM index (greater than 20 movements per hour), whereas those taking bupropion had a risk of PLMS similar to that for control subjects.[145] Tricyclic antidepressants and lithium also are associated with a greater prevalence of nocturnal movement disorders. Caffeine, antihistamines, alcohol, and benzodiazepine withdrawal all can worsen RLS. Antipsychotic therapies are associated with greater PLMS prevalence.

As indicated by findings in a series of small studies and case reports, use of opiate analgesics, especially sustained-release or long-half-life formulations, is associated with increased central apneas, sustained hypoxemia, and prolonged duration of the abnormal breathing events.[146] These changes occur in the context of the well-established acute respiratory depressant effects. Hypnotics such as benzodiazepines also may worsen sleep-disordered breathing by lowering the arousal threshold.

Medications that Affect Sleep Architecture by Other Mechanisms

Certain medications may directly affect the sleep architecture. Beta blockers frequently are prescribed in older adults for the management of hypertension and heart disease. The more lipophilic agents such as propranolol and some of the newer-generation beta blockers have been shown to suppress melatonin, increase sleep fragmentation, and increase nightmares in some people.[147] Other agents such lithium, benzodiazepines (on withdrawal), benzodiazepine receptor agonists, and gamma-aminobutyric acid hydroxybutyrate are associated with a worsening of disorders of non-REM parasomnias. Tricyclic antidepressants, monoamine oxidase inhibitors, venlafaxine, and mirtazapine all have been documented to induce REM sleep behavior disorder, a parasomnia very specific to older adults. Oxybutynin is the most studied of the anticholinergic agents. Night terrors have been associated with use of this agent in case reports, in addition to mild sedation. Polysomnographic changes include a decrease in the amount of REM sleep of approximately 15%, increased REM sleep latency, and power reduction in theta, alpha-1, alpha-2, and beta-1 bands on the electroencephalogram.[148]

Reported effects of donepezil include an increase in REM sleep with a decrease in slow frequencies in REM sleep, decrease in stage 1, and increase in stage 2.[149] An increase in REM sleep density and a decrease in REM sleep latency and nightmares have been reported in healthy elderly volunteers.[150] Changes observed with galantamine in healthy volunteers include a decrease in REM sleep latency, an increase in REM sleep density, and a decrease in slow wave sleep.[151]

SUBSTANCE ABUSE IN OLDER ADULTS AND RELATIONSHIP TO SLEEP

Legal substances of abuse (alcohol, nicotine, and caffeine) as well as illicit ones (stimulants, marijuana, opiates) can disrupt sleep. A high rate of substance use has been documented among older adults with comorbid psychiatric conditions and health problems, which further contributes to and complicates sleep disturbances and warrants clinical assessment and treatment as a potential route of addressing sleep complaints. Among persons with a mental disorder, the lifetime prevalence of an addictive disorder not including nicotine or caffeine dependence is 29%, mostly accounted for by alcohol.[152]

Alcohol Dependence

The direction of sleep disruption and alcohol dependence is likely to be bidirectional, with 50% of alcoholic persons reporting sleep problems before the onset of alcohol dependence.[153] Quantity of drinking and depression predicted insomnia severity in a sample of more than 360 alcohol-dependent patients.[154] Sleep problems may increase the risk of an older adult's developing alcohol problems.[155] Alcoholics report high rates of insomnia, ranging from 36% to 72% in inpatient and outpatient studies. Community rates of alcohol dependence in the geriatric population are 2% to 3% for men and 1% for women.[156] In elderly clinical cohorts, rates are higher (4% to 23%), probably as a consequence of the increased comorbid medical disorders that accompany alcohol dependence.[157] Elderly alcoholics are quite likely to have other substance-related comorbid conditions, which may further impair their sleep, including nicotine dependence in 50% to 70% and dependence on prescribed sedatives, anxiolytics, and opioid analgesics (2% to 14%).[157]

Although acute ingestion may decrease the sleep latency, alcohol-induced sleep may be of poorer quality and shorter duration. Increased arousals occur as the blood alcohol level falls in the hours after alcohol intake, with resultant sleep maintenance insomnia.[158] Acute alcohol intake affects multiple neurotransmitter systems, including acetylcholine, glutamate, gamma-aminobutyric acid, norepinephrine, dopamine, and adenosine. Acute alcohol intake also disrupts the circadian cycle and body temperature, stimulates release of cortisol, and interferes with nocturnal release of growth hormone. Alcohol decreases pharyngeal muscles tone and lowers the arousal threshold, increasing the risk of apneas and more severe oxygen desaturations during apneas.

Chronic consumption results in long-term alterations of the neurotransmitter systems affected by alcohol. A majority of alcohol-dependent patients report disrupted sleep symptoms at the time of entering treatment to stop drinking.[159] Although sleep complaints vary, difficulty falling asleep is reported as the most significant factor associated with substance use.[155] Alcoholics older than 55 had higher sleep latency, lower sleep efficiency, and decreased delta sleep activity on the electroencephalogram when compared with younger alcoholics and with nonalcoholics of both age groups.[159]

An older person with alcohol dependence is likely to still have disrupted sleep even after quitting alcohol and maintaining abstinence, because alcohol dependence may result in permanent effects on the brain, as supported by animal studies that reveal alterations in the *PER* gene expression involved in circadian rhythm after alcohol consumption.[160] In acute withdrawal, PSG findings include increased sleep latency, decreased total sleep time, and increased REM percentage. Alcoholics who have been abstinent for 3 to 6 weeks have worse sleep than nonalcoholics by polysomnographic assessment,[161] with changes persisting at 2.5 years.[155,162] During abstinence, insomnia and sleep architecture changes are significant factors predicting alcoholism relapse.[159,163] In

abstinent male alcoholics, 75% of those older than 60 years of age had sleep-disorered breathing, compared with 25% for those in the 40- to 59-year-old age group and 3% for those younger than 40.[164] Higher rates of PLMS in abstinent alcoholics also have been observed.[165]

Caffeine

Consumption of caffeine in coffee, tea, and sodas, as well as medications, is common in the elderly and may contribute to sleep disruption. In a study of 1528 older adults, participants in the 60- to 69-year-old age group reported consuming 17.6 cups of caffeinated coffee per week, with a subsequent reduction over the next two decades of life to 12.8 cups of caffeinated coffee per week.[166] Caffeine has stimulatory effects on the cerebral cortex and medullary centers, with blood levels peaking 15 to 45 minutes after intake and a half-life of 3 to 7.5 hours. Because of the greater proportion of adipose tissue to lean body mass, a dose of caffeine determined as mg/kg of total body weight may result in higher plasma and tissue concentrations in elderly than in younger persons. Age and plasma caffeine concentrations significantly predict a poorer-quality sleep.[167] In adult populations, daytime sleepiness and insomnia were associated with high daily caffeine consumption.[168,169] Changes related to caffeine intake include a decreased total sleep time, increase in sleep onset latency, and increased number of awakenings. Intake shortly before sleep time has been shown to disrupt sleep in adults on objective and subjective measures.[170,171] Adults older than 67 years of age on caffeine-containing medications report significantly more trouble falling asleep, after institution of controls for multiple factors.[172] Hospital-dwelling elderly patients with higher serum caffeine concentrations reported sleep problems more often than did those with lower levels.[167]

Tobacco and Nicotine

Habitual tobacco use constitutes a chronic disease with multiple relapses, especially in older adults. Nearly 11% of the population 65 years of age and older are smokers.[173] Several studies have suggested that insomnia and sleep apnea are more common in smokers, but establishing a causal relationship has been difficult.[174] Current smoking was independently associated with an increased likelihood of snoring, short sleep, and poor sleep among U.S. adults.[175] In a cohort of almost 500 women studied over 25 years, chronically heavy-smoking subjects were more likely to report insomnia at a mean age of 65 years (adjusted odds ratio, 2.76)[176] Nicotine enhances acetylcholine neurotransmission in the basal forebrain and dopamine release, potentially influencing sleep-wake control mechanisms toward wakefulness.[177] A dose-dependent reduction in REM and slow wave sleep, along with an increase in wakefulness and total sleep time, has been observed with nicotine use.[178] Elderly smokers report more difficulties with sleep onset and staying asleep than nonsmokers. In all age groups, smokers drink more caffeine and are more likely to be depressed.[178] The overlap of nicotine dependence with psychiatric conditions is an important issue across all age groups. In clinic and community populations of depressed patients, 40% to 60% smoke, with even higher rates in those affected by PTSD and schizophrenia.[179] Quit rates also are much lower in people with psychiatric disorders. When offered the tools they needed, older smokers quit smoking at rates comparable to those for younger smokers.

SPECIAL CONDITIONS

Caregiving

Elderly caregivers represent 13% of all caregivers and are more likely to be caring for a spouse, often with dementia. Caregivers or a spouse of a person with dementia or disability are at increased risk for sleep problems. In a study of community-dwelling spousal caregivers of patients with Alzheimer disease, caregivers objectively slept less than older noncaregivers and subjectively reported more sleep problems and functional impairment as a result of poor sleep.[180]

Up to two thirds of older adult caregivers are afflicted with some form of sleep disturbance.[181] Predisposing factors for changes in the sleep of a caregiver include increasing age, female gender, and higher caregiver burden.[182] Caregiver burden encompasses the spectrum of physical, psychological or emotional, social, and financial problems experienced by the care provider. A study of the sleep quality of caregivers of persons with dementia demonstrated links between poor sleep and fluctuations in the status of the care recipient, the need for vigilance to safeguard the care recipient at night, and worry about current and future events, which caused rumination.[183] Caregivers identified barriers to health promotion activities, including lack of time, decreased energy, and additional costs for providing care for the care recipient.

The caregiver's situation is analogous to that of a rotating shift worker, who must be alert both at night and during the day, often on an inconsistent schedule.[181] The combination of depression and high-stress situations (e.g., caring for a spouse or a person with dementia, or living with the care recipient) increased the likelihood of sleep problems.[184] In addition, by virtue of their age, older caregivers are at risk for developing many of the chronic illnesses that can directly affect sleep, as described in previous chapters.

Sleep problems in a caregiver may be precipitated by changes in the nighttime routines of the person for whom they are caring, including agitation, sundowning, wandering, and sleep-disordered breathing. In the case of the vulnerable caregiver, it can be very difficult to fall back asleep after being awakened by a care recipient who needs assistance with toileting, administration of medication, redirection back into bed, orientation, or emotional reassurance, particularly when these nocturnal interactions are prolonged or emotionally charged. This situation puts the caregiver at risk for long-term partial sleep deprivation that may partially explain why one of the strongest factors in deciding to institutionalize a spouse or family member with dementia is poor overnight sleep of the affected person.[185]

Geriatric Syndromes

Falls and Mobility

Falls and gait problems are a common and serious problem facing older adults. Approximately one third of those aged 65 and older living in the community fall at least once a year. This proportion increases to 1 in 2 for persons aged 80 and older.[186] Falls typically are the result of multiple, interacting etiologic factors. It is well acknowledged that older persons with insomnia on hypnotic therapy are at higher risk for falls. Prevailing evidence suggests that this risk is due in part to the adverse effects of benzodiazepines and other hypnotics on balance, cognition, and reaction time,

with withdrawal of these agents reducing the risk for future falls.[187]

Increasing evidence suggests that it may be sleep itself, with its neurocognitive consequences, that predispose the older person to falls in selected populations such as in the nursing home.[168] Sleep quality has been linked to physical function. In a study of more than 2800 men, total sleep time less than 6 or more than 8 hours, sleep efficiency less than 80%, WASO of 90 minutes or longer, RDI of 30 or greater, and hypoxia were found to be associated with poorer physical function.[188] In a similar study, this time evaluating more than 800 women over a period of 5 years, shorter sleep duration, greater WASO, and lower sleep efficiency were risk factors for functional or physical decline.[189] A cross-sectional study of more than 2800 men and women (older than 55 years of age), after adjustment for lifestyle factors and diseases, indicated that longer sleep (9 hours or more) was associated with a decreased walking speed in women aged 65 or more years and shorter sleep (6 hours or less) with a higher likelihood of mobility limitation in women 65 years of age or older and in men 55 to 64 years of age.[190] A discrete sleep disorder or insomnia was independently associated with both decreased walking speed and mobility limitation in men 55 years of age or older but with only mobility limitation in women 65 years of age or older.

Consensus statements from expert panels advise a multifaceted approach to reduce future falls, including behavioral interventions for sleep, drug review, adjustment of psychotropic and hypnotic agents as possible, an exercise program, and evaluation and treatment of other comorbid medical problems as appropriate.

Delirium

Delirium is a serious acute neuropsychiatric syndrome with core features of inattention and cognitive impairment and associated features including changes in arousal, altered sleep-wake cycle, and fluctuating changes in mental status. Delirium is common in critically ill and palliative care patients and probably encompasses adverse effects of stress response pathways, induced by aberrations in the normally adaptive systemic and CNS responses to stressors. Sleep disturbances appear to have a bidirectional relationship with delirium and are part of many delirium assessment instruments, such as the Intensive Care Delirium Screening Checklist.[191] Such disturbances are defined as "sleeping less than 4 hours or waking frequently at night, sleeping during most of the day." In a series of 600 consecutive patients admitted to the intensive care unit for more than 24 hours, sleep-wake cycle disturbances occurred in almost 70% of patients with delirium, compared with 12% of patients without delirium, and were highly discriminating for the diagnosis of delirium.[192] Another study of 100 cases of delirium in a palliative care inpatient service reported sleep disturbances in 97% of patients.[193]

In a study of 13 patients (6 with delirium and 7 without delirium) employing actigraphy, delirious patients showed fewer nighttime minutes resting, fewer minutes resting over 24 hours, greater mean activity at night, and a smaller amplitude of change in activity from day to night. Rest and activity consolidation were significantly reduced in delirious patients, as was the amplitude of day-night differences in rest and activity.[194]

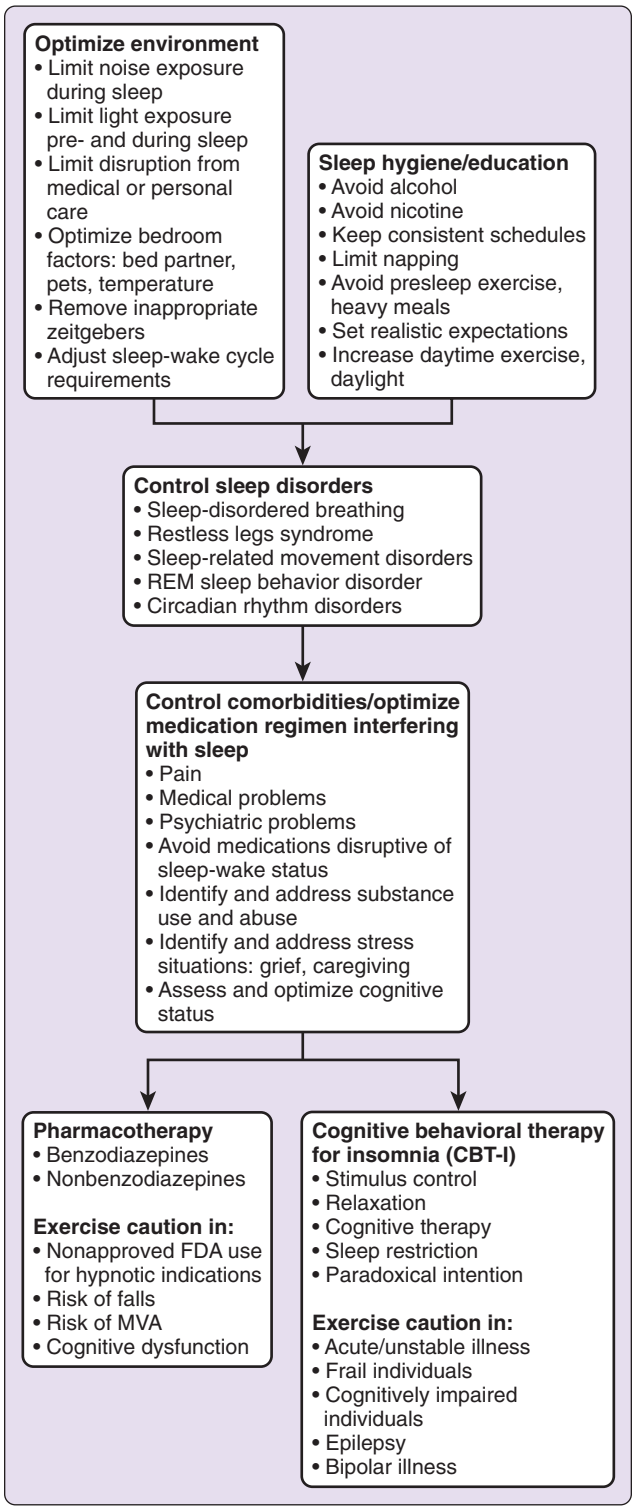


Figure 151-2 Interventions to be considered in comorbid insomnia. A multifactorial targeted intervention appears to be most appropriate in accordance with current concepts (see Figure 151-1) and as indicated by the available evidence. Considering all potential predisposing, precipitating, and perpetuating factors can increase the effectiveness of the therapeutic act while ensuring optimal control. Interventions should be balanced with regard to potential adverse consequences. MVA, Motor vehicle accident.

CLINICAL PEARLS

- Sleep complaints often are comorbid with chronic health issues, as well as the associated changes in the patient's lifestyle, sleep hygiene, and medication regimens that are secondary to the illness. Furthermore, these health problems rarely exist in isolation and instead frequently coexist with psychiatric diagnoses, neurologic diseases, and primary sleep pathology.
- In view of the multiple pathways of disrupted sleep, a multifaceted management approach is best in order to improve symptoms. This strategy includes optimizing control of the underlying illness and the environment, adjusting or eliminating potentially offending medications, using cognitive-behavioral approaches, and employing judicious hypnotic therapy based on current evidence.
- Older adults often are in special life situations (caregiving, hospitalization, end of life, abstinence after alcohol dependence) that make them particularly vulnerable to incident sleep disturbances. Heightened awareness and recognition of these situations are essential for optimal management.
- Many medications can alter patterns of sleep and wakefulness both during periods of administration and during periods of withdrawal. It is especially important to avoid using hypnotics or stimulants to treat sleep-related side effects of other medications unless all other options have been considered.

SUMMARY

As the field of medicine continues to advance, people are living longer, with more comorbid medical and psychiatric conditions. This higher burden of illness and the sheer number of medications used to treat these conditions play an important role in the quality and quantity of sleep in older adults. In approaching sleep complaints in geriatric patients, it is essential that practitioners recognize the multidimensional mechanisms by which illness affects sleep. Equally important, a balanced management approach (Figure 151-2) that includes optimizing the underlying illness, adjusting medications, using cognitive-behavioral approaches, and offering judicious hypnotic therapy seems justified as indicated by the current evidence.

Selected Readings

Benca RM, Obermeyer WH, Thisted RA, et al. Sleep and psychiatric disorders. A meta-analysis. *Arch Gen Psychiatry* 1992;**49**:651–68.

Bloom HG, Ahmed I, Alessi CA, et al. Evidence-based recommendations for the assessment and management of sleep disorders in older persons. *J Am Geriatr Soc* 2009;**57**:761–89.

Campos Costa I, Nogueira Carvalho H, Fernandez L. Aging, circadian rhythms and depressive disorders: a review. *Am J Neurodegener Dis* 2013;**2**:228–46.

Gupta R, Zalai D, Spence DW, et al. When insomnia is not just insomnia: the deeper correlates of disturbed sleep with reference to DSM-5. *Asian J Psychiatry* 2014;**12**:23–30.

Lichstein KL. Secondary insomnia: a myth dismissed. *Sleep Med Rev* 2006;**10**:3–5.

Pagel JF. Medications and their effects on sleep. *Prim Care* 2005;**32**: 491–509.

Quan SF, Zee P. A sleep review of systems: evaluating the effects of medical disorders on sleep in the older patient. *Geriatrics* 2004;**59**:37–42.

State-of-the-Science Panel. National Institutes of Health State of the Science conference statement: manifestations and management of chronic insomnia in adults, June 13-15, 2005. *Sleep* 2005;**28**:1049–57.

Stepanski E, Rybarczyk B. Emerging research on the treatment and etiology of secondary or comorbid insomnia. *Sleep Med Rev* 2006;**10**:7–18.

Taylor DJ, Mallory LJ, Lichstein KL, et al. Comorbidity of chronic insomnia with medical problems. *Sleep* 2007;**30**:213–18.

Vaz Fragoso CA, Van Ness PH, Araujo KL, et al. Age-related differences in sleep-wake symptoms of adults undergoing polysomnography. *J Am Geriatr Soc* 2015;**63**(9):1845–51.

Wu JQ, Appleman ER, Salazar RD, et al. Cognitive behavioral therapy for insomnia comorbid with psychiatric and medical conditions. A meta-analysis. *JAMA Intern Med* 2015;**175**(9):1461–72.

A complete reference list can be found online at ExpertConsult.com.

Chapter
152

Obstructive Sleep Apnea in Older Adults

Barbara A. Phillips

Chapter Highlights

- The prevalence of obstructive sleep apnea (OSA) increases with age, but the peak prevalence of clinically diagnosed OSA is in middle age. Differences in the clinical presentation, severity, and manifestations of obstructive sleep-disordered breathing between older and younger persons probably account for much of the gap in diagnosis.
- More than half of older adults report sleeping difficulty.[1-5] Because sleep complaints are common in the older age group, clinicians may discount them. Sleep complaints in seniors, however, correlate with health complaints, depression, and mortality.[1-6] Undiagnosed OSA probably accounts for some of the sleep complaints voiced by older adults. Indeed, OSA leads both to sleep disturbance and to increased mortality and is likely to account for much of the association between sleep complaints and adverse outcomes in older people. Despite this, sleep-disordered breathing is woefully underdiagnosed and undertreated in the older adult population.
- Accumulating data indicate that continuous positive airway pressure use is associated with reduced cardiovascular morbidity, cognitive dysfunction, and all-cause mortality in seniors; clinicians caring for geriatric patients should consider sleep-disordered breathing in the older adult with sleep complaints. This chapter focuses on OSA in the older patient. For discussion of central sleep apnea, see Chapters 109 and 110.

EPIDEMIOLOGY AND DEFINITIONS

Although studies of clinical populations identify peak prevalence of clinically significant sleep-disordered breathing in middle age, population-based studies have shown that sleep-disordered breathing increases with age.[7,8] Longitudinal and cross-sectional studies also have shown that the prevalence of sleep apnea increases with increasing age[9-13] (Table 152-1).

Estimates of the prevalence of obstructive sleep apnea (OSA) in older people depend on how it is defined, so prevalence estimates for OSA in older people are moving targets. Indeed, there has been considerable variation in the definitions and measurements of respiratory events (such as apneas, hypopneas, and respiratory effort–related arousals used to identify sleep-disordered breathing, even since the last edition of this text. Furthermore, which respiratory events to "count" toward the threshold used to define "sleep apnea" has also varied. The apnea-hypopnea index (AHI) typically includes only apneas and hypopneas, but the respiratory disturbance index (RDI) may include other events, such as respiratory effort–related arousals, or may be defined on the basis of recording time rather than sleep time. Finally, the demarcation between "normal" and "abnormal" also has been somewhat fluid. In population-based studies, approximately one third of those older than 65 years of age have AHIs of 5 or more events per hour of sleep,[14,15] and some two thirds have RDIs of 10 or more events per hour.[15,16] Although measures of sleep-disordered breathing events alone do not establish a diagnosis of OSA, the classically associated symptoms of the disorder—sleepiness, hypertension, cognitive dysfunction—increase in prevalence with aging. Thus a majority of older persons who meet laboratory-defined numeric criteria for OSA also will have a clinical symptom that is commonly associated with the syndrome; for this reason, defining clear-cut criteria for sleep apnea in older persons is difficult. However, the Centers for Medicare and Medicaid Services, the primary provider of health care coverage for people older than 65 in the United States, defines OSA as an AHI 15 or greater, or an AHI of greater than 5 plus hypertension, stroke, sleepiness, ischemic heart disease, or mood disorder.[17] The Centers for Medicare and Medicaid Services now covers continuous positive arway pressure (CPAP) therapy as indicated by results of portable monitor testing and requires objective documentation of use for payment beyond the first 90 days.

Regardless of how sleep apnea is defined, its prevalence increases progressively from age 18 to approximately 70, when it may plateau.[18] On the basis of these findings, it is likely that the underdiagnosis of sleep apnea in older people is even more common than in younger people, and this may be especially true in minority populations[18]

Diagnosis of sleep-disordered breathing is evolving rapidly, with revised, more liberal diagnostic criteria[19] and the

Table 152-1 Differences between Younger (<60 years) and Older Patients with Obstructive Sleep Apnea (OSA)

Risk Factor	Older Patients	Younger Patients
Male gender	1:1	2:1
Obesity	Unimportant	Very important
Clinical features		
Witnessed apneas	Witnessed apneas rarely reported	Witnessed apneas strongly predictive
Snoring	Infrequently reported	Frequently reported
Prevalence		
AHI >5	30%-40%	9% for women, 24% for men
RDI >10	62%	10%
Consequences	Death, cardiovascular disease, stroke, nocturia, impaired cognition, atrial fibrillation	Death, ischemic cardiac disease, hypertension, cerebrovascular disease, depression, metabolic disturbances
Treatment	May require lower CPAP pressures No difference in tolerance or adherence	May require higher CPAP pressures No difference in tolerance or adherence

AHI, Apnea-hypopnea index; CPAP, positive airway pressure; RDI, respiratory disturbance index.

recognition that oximetry alone can be quite predictive of important outcomes. For example, in a study of 100 patients with a mean age of 62 years, Ohmura and associates reported that sleep-disordered breathing, as determined by predischarge pulse oximetry (based on oxygen desaturation index of 4% [i.e., ODI = 4]), was associated with significantly increased risks for necessity for readmission and death, independent of other risk factors.[20] In a study of patients with OSA and matched control subjects, oxygen saturation predicted cognitive function, but AHI did not.[21] In the Spanish Sleep Network study, oxygen desaturation predicted cancer risk, but again, AHI did not.[22] Similarly, time elapsed with oxygen saturation below 90% predicted 3-year mortality in older persons with cardiovascular disease. It also predicted self-reported insomnia.[23] These and other findings are likely to change, yet again, the diagnostic criteria for significant sleep apnea, at least for third-party payers, who are increasingly data-driven.

CLINICAL MANIFESTATIONS AND PRESENTATION

Most studies of the clinical presentation and manifestations of OSA have focused on the middle-aged. Reports derived from clinical populations tend to include people whose mean age is approximately 50. As more data accumulate about sleep-disordered breathing in older people, it is becoming increasingly clear that the phenotype of OSA can be quite different in younger and in older populations.

Perhaps most striking is the change in gender as a risk factor for sleep-disordered breathing with aging. Prospective data from the Wisconsin Sleep Cohort have demonstrated male sex is no longer an important risk factor for OSA after the age of approximately 50 years,[24] confirming work from other studies.[8] At least part of the reason for this phenomenon is that the prevalence of OSA rises strikingly for women as they go through menopause, which occurs at approximately age 50.[25–27] As a consequence, some investigators have reported a male-to-female ratio of 1:1 for older people.[24]

In addition to the loss of effect of male gender as a risk for OSA with aging, there is reduced importance of obesity as a risk factor. Beginning at approximately 60 years of age, obesity is no longer a statistically significant risk factor for sleep-disordered breathing.[24,26] These observation are of particular interest, in view of known decreases in obesity with increasing age.[28] Some data suggest that obesity is a more important risk factor for men than for women, but that aging, perhaps specifically achieving menopause, is a more important risk factor for women than for men.[24-27,29,30] However, in an 18-year follow-up study of 427 community-dwelling elderly persons, Ancoli-Israel and associates found that observed changes in RDI were associated only with changes in BMI and were independent of age[31]; they pointed out that this finding underscores the importance of managing weight for older adults, particularly those with hypertension.

In a study of nearly 100 community-dwelling adults aged 62 to 91 years, Endeshaw found that almost one third (equally divided between men and women) had an AHI of 15 or more events per hour of sleep, and that the "traditional" risk factors such as snoring, body mass index, and neck circumference were not significantly associated with OSA in this group.[32] Rather, those with an AHI of at least 15 were more likely to report not feeling well rested in the morning and to have higher Epworth Sleepiness Scale scores and a greater frequency of nocturia.[32] These findings confirm earlier work from the Sleep Heart Health Study, which reported that witnessed apneas are much less frequently reported in older patients than in younger ones.[33]

Studies of OSA in older people have tended to report "milder" disease, with lower AHIs, and better-preserved oxygen saturation than in younger adults.[9,14,26]

In short, the "classic" clinical presentation of OSA is uncommon in older adults, which may account in part for the reduced prevalence of clinical diagnosis of the disorder in this population.

PATHOPHYSIOLOGY

The pathophysiology of OSA may be different in older persons from that in younger people. Chapter 111 outlines the pathophysiology of OSA in adults. With aging, loss of tissue elasticity also may contribute to airway collapse. For older women, declining levels of sex hormones appear to be partly responsible for increased collapsibility of the posterior oropharynx.[34,35]

CLINICAL CONSEQUENCES

Overview

A majority of studies of the consequences of OSA have been undertaken in clinical samples of middle-aged people. Data specifically focusing on the consequences of OSA in older

persons are limited. OSA has long been associated with increases in the risk of death in younger populations, but early studies suggested that it was not associated with increased mortality in older groups.[36,37] A recent well-done study from the Spanish Sleep Network, however, clearly demonstrated a twofold increase in risk of death in a group of patients with severe OSA (AHI of 30 or greater) whose mean age was 71 years over that in the control group, after initiation of controls for age, body mass index, preexisting cardiovascular disease, smoking, diabetes, sleepiness, gender, dyslipidemia, and/or respiratory failure. Furthermore, CPAP use reduced risk of all-cause mortality, as well as cardiovascular death and death from stroke and heart failure, but did not reduce the risk of death from ischemic disease in this cohort. Indeed, even those seniors 75 years of age and older had reduced risk of cardiovascular death with CPAP use, and continuous adherence was associated with reduced risk of cardiovascular death as a continuous variable.[38] The inclusion of a large cohort with severe sleep apnea (AHI of 30 or higher) probably partly accounts for the ability of this study to demonstrate a mortality effect with sleep apnea and with CPAP. In a meta-analysis of prospective cohort studies of OSA and risk of cardiovascular disease, Wang and colleagues demonstrated a "dose-response" relationship between OSA severity and cardiovascular outcomes and calculated a 17% greater risk of cardiovascular disease for each 10-unit increase in AHI.[39] It is possible that older people tolerate milder degrees of sleep-disordered breathing better than do their younger counterparts, and that part of the past difficulty in demonstrating an association between OSA and adverse outcomes in older patients was because few patients with moderate to severe disease were included in earlier studies.

Other plausible explanations have been proposed for the inability to readily demonstrate the impact of sleep-disordered breathing on cardiovascular outcomes in older people. Lavie and Lavie, for example, speculated that the reduced effect of OSA on mortality in older people is because of ischemic preconditioning resulting from the nocturnal cycles of hypoxia-reoxygenation and pointed out that in patients with sleep-disordered breathing, there is an association of ischemic preconditioning with increased levels of vascular endothelial growth factor and increased production of oxygen reactive species, heat shock proteins, adenosine, and tumor necrosis factor alpha.[40]

Among the most striking manifestations of sleep-disordered breathing in aging populations are nocturia, cognitive dysfunction, and cardiac disease.

Nocturia

Nocturia is a particularly troublesome symptom of aging and appears to be related to the severity of sleep-disordered breathing. Because older adults with significant sleep-disordered breathing may not manifest classic symptoms of OSA, the presence of nocturia in the older patient should heighten clinical suspicion for OSA. Indeed, nocturnal urination more than three times nightly had positive and negative predictive values of 0.71 and 0.62, respectively, for severe OSA in one study.[41]

The postulated mechanism of nocturia in OSA is that the negative intrathoracic pressures resulting from occluded breaths cause distention of the right atrium and ventricle. This right-sided cardiac distention results in release of atrial natriuretic peptide, which inhibits the secretion of antidiuretic hormone and aldosterone and causes diuresis through its effect on glomerular filtration of sodium and water.[42] Several studies have demonstrated symptomatic improvement in patients with nocturia with use of CPAP.[43-45] The mechanism of CPAP-related improvement may involve promoting the normal nocturnal rise in antidiuretic hormone, resulting in increased resorption of sodium and water from the collecting tubules and production of lower volumes of more concentrated urine.[46] In a retrospective review of data on 196 patients whose mean age was 49 years, predictors of nocturia included increasing age and diabetes mellitus. Although a complaint of nocturia was equally likely to occur in patients with and without OSA, nocturic frequency was significantly related to age, diabetes, and severity of sleep-disordered breathing in those patients who had OSA. Furthermore, patients with OSA and nocturia who were treated with CPAP experienced significant reductions in the frequency of nocturnal voiding.[47] In a study of 21 women with a mean age of 65 years, the same group of investigators reported that OSA is present in a majority of women with nocturia, and that the presence of diluted nighttime urine in a patient with nocturia is a sensitive marker for OSA.[48]

Impaired Cognition

Impaired cognition, including sleepiness, impaired vigilance, worsened executive function, and dementia, increases in prevalence with aging. Neuropsychological assessment of patients with OSA demonstrates decline in cognition similar to that with aging. For example, patients with OSA experience sleepiness[49] and demonstrate impaired executive function,[50] working memory,[51] alertness,[52] and attention.[53] In general, the association between sleep-disordered breathing and impaired cognition has been best studied in middle-aged patients. Because of the association with aging itself on impaired cognition, the effects of OSA on cognitive function in older people have been difficult to tease out. In a small study with subjects older than 55 years of age who had OSA, Aloia and associates found that the degree of sleep-disordered breathing, especially oxygen desaturation, was associated with delayed verbal recall and impaired constructional abilities. After 3 months, subjects who were compliant with CPAP showed greater improvements in attention, psychomotor speed, executive functioning, and nonverbal delayed recall than in those who were not compliant.[54] Despite the logical notion that cognitive impairment associated with OSA in older people might be more severe than in younger people because of cumulative effects of age and sleep-disordered breathing, Mathieu and colleagues were unable to demonstrate any group-by-age interaction for any neuropsychological variable; in a study of matched older and younger patients both with and without OSA, they found that performance on most tasks deteriorated with advancing age in both control subjects and patients with OSA without evidence of a compounded effect.[55] Persons at high risk for OSA based on Berlin questionnaire scores (57% of whom were female) had lower cognitive function scores than those without, but the risk was most pronounced during middle age and was attenuated after age 70.[56]

The evidence indicates age- and gender-dependent relationships as well. In a large study of patients with OSA who were at least 40 years old, the risk of developing dementia

within 5 years of diagnosis was 1.70 times greater than in age- and sex-matched subjects who did not have OSA, after appropriate adjustment for some potential confounders. In this study, men aged 50 to 59 had a sixfold increased risk of developing dementia compared with matched control subjects, but women 70 years of age or older had a threefold increased risk of developing dementia. One large study reported that sleep-disordered breathing severity was not associated with indices of sleep-related symptoms or sleep-related quality of life in community-dwelling older women, suggesting that this group may be resistant to the adverse effects of OSA on cognition.[57]

In addition to age and gender differences in susceptibility to dementia in patients with OSA, genetic predispositions also are likely. Data from the Wisconsin Sleep Cohort suggest that apolipoprotein E epsilon 4 genotype (APOE4)-positive persons with sleep apnea of moderate severity have impairment of cognition and executive function comparable to that in persons without this genotype.[58] Preliminary work suggests an association between sleep-disordered breathing and Alzheimer disease biomarker in the cerebrospinal fluid of cognitively normal persons.[59] These early data point to genetic influences on the propensity to develop dementia in patients with SDB.

It is likely that many other factors affect susceptibility to cognitive dysfunction in patients with OSA. For example, Alchantis and coworkers have proposed that high intelligence may protect against cognitive decline caused by sleep-disordered breathing, perhaps as a consequence of increased cognitive reserve.[60]

The mechanism by which sleep-disordered breathing impairs neurocognitive function remains incompletely understood. Some investigators have suggested that sleep fragmentation is the primary culprit,[61] whereas others maintain that hypoxemia is the primary cause. It is likely that the functions are affected by hypoxemia differ from those affected by sleep deprivation. As suggested by Sateia, "Disturbances in general intellectual function and executive function show strongest correlations with measures of hypoxemia. Not unexpectedly, alterations in vigilance, alertness, and, to some extent, memory seem to correlate more with measures of sleep disruption."[62] However, in a small study of matched patients with OSA, AHI did not predict or correlate with cognition, but mean oxygen saturation correlated with executive functioning and access to long-term memory.[21]

Kim and associates demonstrated that moderate to severe OSA is an independent risk factor for white matter change in more than 500 people (mean age 59 ± 7.48 years) and presented an excellent discussion of potential mechanisms, including hypoxemia and hypercapnia during apneic events, which activate arousal- and chemoreflex-mediated increases in the cerebral circulation and activation of oxidative and imflammatory processes.[63]

Information about the effects of CPAP treatment on cognition in older people is limited, and data about CPAP's effects on cognition in older persons even more so. In a small group of patients whose mean age was 56 years, CPAP resulted in normalization in attentive, visuospatial learning and in motor performances after 15 days, but no further improvement was observed after 4 months of treatment. CPAP did not improve performance on tests evaluating executive functions and constructional abilities.[64] A meta-analysis of randomized, placebo-controlled crossover studies of CPAP treatment involving 98 patients with sleep apnea demonstrated mostly trends for better performance on CPAP than on placebo.[65] In a group of middle-aged adults with significant OSA, Zimmerman and associates demonstrated that memory normalized for the group that used CPAP at least 6 hours a night[66] but did not improve in those who were not adherent to CPAP treatment. In a review of CPAP adherence and benefits among older people, Weaver and Chasens noted that in general, older adults demonstrate increased alertness; improved neurobehavioral outcomes in cognition, memory, and executive function; and decreased sleep disruption.[67] These investigators also noted that older persons may require lower CPAP pressures than younger ones and tolerate CPAP well, with similar rates of adherence. Thus, despite differences in the clinical presentation and impact of sleep apnea in the elderly population, CPAP treatment is likely to be well tolerated and beneficial in symptomatic patients.[68,69]

With regard specifically to Alzheimer disease, the prevalence of OSA is higher among patients with this disorder than among nondemented seniors, and sleep-disordered breathing is believed to contribute to cognitive dysfunction in those with Alzheimer disease. A randomized double-blind, placebo-controlled crossover trial of CPAP in patients with OSA and Alzheimer disease demonstrated a significant improvement in cognition after 3 weeks of CPAP.[70] In addition to improving cognition, CPAP treatment may reduce sleepiness in patients with Alzheimer disease and OSA.[71]

In view of the facts that improvement in cognition is likely to depend on CPAP adherence, that gender, age, genetic, and intelligence are probable influences on susceptibility to impaired cognition, and that most studies addressing this issue have not included geriatric patients with OSA or objectively measured adherence, firm conclusions about the reversibility of cognitive deficits in older patients woth OSA are impossible at present.

Cardiovascular Disease

In addition to nocturia and cognitive impairment, sleep-disordered breathing also is strongly linked to *cardiovascular disease,* including hypertension, congestive heart failure, stroke, cardiac arrhythmias, ischemic events, and pulmonary artery hypertension.[72] Few studies of the relationship between OSA and cardiovascular disease are prospective and control for confounding variables such as obesity. Even fewer studies have been conducted in older adults. Hypertension, atrial fibrillation, and stroke, however, are particularly relevant comorbid conditions associated with OSA in older patients, because of their higher prevalence and clinical importance in that population.

Hypertension

That sleep-disordered breathing causes *hypertension* has been demonstrated by multiple studies, including prospective and CPAP sham–controlled work.[73-79] Recent data also have established that severe sleep apnea (AHI of 30 events or more per hour) is an independent risk factor for incident hypertension in older people (mean age, 68.2 years).[80] In general, CPAP has modest effects on blood pressure in patients with OSA but is most effective in those who have significant hypertension and are most adherent with its use.[73,81,82]

Atrial Fibrillation

Atrial fibrillation is strongly associated both with aging and with OSA.[72,83] The Sleep Heart Health Study investigators found that persons with severe sleep-disordered breathing had double or quadruple the risk of complex cardiac arrhythmias compared with those with no sleep-disordered breathing, after institution of controls for multiple relevant confounders.[83] In this study, atrial fibrillation was the arrhythmia most strongly associated with sleep-disordered breathing. Gami and coworkers reported that both obesity and nocturnal oxygen desaturation were independent predictors of incident atrial fibrillation, but only in subjects younger than 65 years of age.[84] However, Ganga's group noted that the presence of overlap syndrome (OSA combined with chronic obstructive pulmonary disease) is associated with a marked increase in new-onset atrial fibrillation in elderly patients over that associated with the presence of either OSA or chronic obstructive pulmonary disease alone.[85]

In a small, retrospective study of patients with atrial fibrillation, some of whom had either treated or untreated sleep apnea and some of whom did not have sleep apnea, the patients with untreated OSA had a higher recurrence of atrial fibrillation after cardioversion than that for the patients without sleep apnea.[86] Furthermore, treatment with CPAP in the sleep apnea group was associated with lower recurrence of atrial fibrillation at 1-year follow-up evaluation. This study is particularly relevant for the management of older patients because the mean age of the study population was approximately 66 years. A study of patients undergoing pulmonary vein isolation reported that the 32 patients who had OSA and used CPAP were less likely to have atrial tachyarrhythmias, use of antiarrhythmic drugs, and need for repeat ablations were compared to the 30 patients who did not use CPAP during a follow-up period of 12 months; the difference in atrial fibrillation–free survival rate was 71.9% versus 36.7% ($P = .01$).[87]

As in younger adults, sleep-disordered breathing is associated with subtle measures of myocardial injury. In a large study of patients whose mean age was 62.5 years, OSA severity correlated with measures of high-sensitivity troponin T and N-terminal pro–B-type natriuretic peptide, and high-sensitivity troponin was related to risk of death or heart failure in all categories.[88] Elderly patients with OSA exhibit cardiac structural changes and diminished left ventricular function compared with that in control subjects who do not have OSA.[89] In a nonrandomized, retrospective review of data on 130 patients aged 65 to 86 years, Nishihata and associates demonstrated that those with untreated sleep apnea had increased likelihood of cardiovascular death and hospitalization due to cardiovascular disease, including heart failure, over a follow-up period of approximtely 3 months. Furthermore, adequate CPAP treatment improved the cardiovascular outcomes in this cohort.[90]

Stroke

The prevalence of stroke increases with age, but untreated sleep apnea appears to impose an additive risk. In a 6-year follow-up study of more than 1000 patients whose mean age at enrollment was approximately 60, Yaggi and coworkers found that OSA was a risk factor for stroke, controlling for other important variables.[91] Treatment did not affect the risk of either stroke or death in this study.

A review of the findings in 10 reports that included 1203 patients who had experienced stroke or transient ischemic attacks noted a dose-response relationship between severity of sleep-disordered breathing and the risk of recurrent events and all-cause mortality; 3 of the studies included information about patients who received CPAP, but the data were too limited for a compelling argument that CPAP improves outcome in patients with stroke/transient ischemic attack.[92] In a study of patients with acute cerebral ischemia who underwent polysomnography, Kepplinger and colleagues showed that sleep apnea is associated with clinically silent microvascular brain tissue changes such as leukoaraiosis (white matter hyperintensities) and lacunar infarcts.[93]

In summary, OSA is strongly associated with cardiovascular disease in middle-aged populations, and the accumulating evidence suggests that sleep-disordered breathing increases the risk of cardiovascular disease and stroke in the elderly population. The best-proven association and evidence for benefit is with hypertension, for which the data are derived largely from middle-aged populations.

Other Effects

Sleep-disordered breathing is a systemic problem with systemic consequences. In addition to the adverse outcomes already noted, OSA in elderly persons is associated with multiple potential consequences.

In middle-aged men, OSA is associated with increased health care costs that decrease after treatment. CPAP treatment is cost-effective for treatment for severe sleep apnea in middle-aged people.[94] Within the sleep apnea population, expenditures for health care in older patients are approximately twice as high as they are in middle-aged patients. After adjustments for age, body mass index, and AHI, cardiovascular disease and use of psychoactive drugs were important determinants of health care costs for older patients with sleep apnea in one study.[95] An enormous study of elderly veterans documented that 4.4% (in all likelihood, representing only the "tip of the iceberg") were diagnosed with OSA, and these patients had many more comorbid conditions and much higher health care utilization than those who did not carry the diagnosis.[96]

Sleep apnea also was associated with an increased risk of pneumonia (the "old man's friend") in a large study.[97]

Nocturnal hypoxemia also is associated with increased risk of falls in older men[98] but paradoxically is associated with preserved bone mineral density in elderly men and women, even after adjustment for sex, BMI, metabolic values, and hypertension.[99]

Gender is likely to influence the effects of sleep-disordered breathing in older people, just as it does for younger persons. For example, a significant relationship between sleep-disordered breathing and hypertension, history of diabetes, and low high-density lipoprotein cholesterol has been reported in women older than 65 years of age, but these effects were not demonstrable in older men.[100]

TREATMENT OF OBSTRUCTIVE SLEEP APNEA IN OLDER ADULTS

Continuous Positive Airway Pressure

As with younger adults, CPAP is the treatment of choice in older patients. Because most studies of effects of CPAP have

been done in clinical (e.g, middle-aged) populations, evidence for the benefits of CPAP in older persons is not yet robust.

Complex sleep apnea appears to be more prevalent in older than in younger people. *Complex sleep apnea* is characterized as OSA in which central apneas and periodic breathing develop when CPAP is applied.[101,102] This phenomenon appears to be more prevalent in older men with congestive heart failure, and its clinical significance is unclear. In many cases of these "treatment-emergent central apneas," the central apneas will simply resolve over time. In a convenience sample of a variety of sleep-disordered breathing syndromes resistant to CPAP, the mean age of 72 years was much higher than typically observed in clinical populations of sleep apnea patients. In that cohort, adaptive servoventilation appeared to be effective and well tolerated in approximately half.[103] Because complex or treatment-emergent central apnea appears to be more prevalent in older adultss, formal, in-lab titrations may be more important for this group. (For a more detailed discussion of complex sleep apnea, see Chapter 110.) Adherence to CPAP therapy in older patients may be impaired by factors such as cognitive impairment, medical and mood disturbances, nocturia, lack of a supportive partner, and impaired manual dexterity. Older age in itself, however, does not affect adherence to CPAP treatment,[67,104,105] and behavioral interventions can improve CPAP adherence in the elderly.[106] The major predictors of CPAP nonadherence in older patients with sleep apnea are nocturia, current cigarette smoking, lack of symptom resolution, and advanced age at time of diagnosis.[107] Older men with nocturia may find CPAP particularly confining and may be particularly likely to have difficulty with its use, although CPAP may actually help with the nocturia.[104,107]

Patients with OSA who have dementia, including Alzheimer disease, have been demonstrated to tolerate CPAP, although depressive symptoms appear to predict worsened adherence in demented seniors with sleep apnea.[108]

Oral Appliances

Oral appliances are effective in treating snoring and mild to moderate OSA.[109-112] Although not as effective as CPAP, these agents improve sleep-disordered breathing, sleepiness, nocturnal oxygen saturation, and blood pressure. There are two basic types of oral appliances:

1. Mandibular repositioners, which pull the mandible (and with it, the tongue) forward
2. Tongue-retaining devices, which adhere to the tongue by suction and pull it forward. Because these are not approved by the U.S. Food and Drug Administration for treatment of sleep apnea, they are used much less commonly in clinical practice.

See Chapter 147 for a detailed discussion of the use of oral appliances.

Common side effects of oral appliances include dry mouth, increased salivation, tooth soreness, and jaw muscle or jaw joint discomfort. Pain occasionally can be severe enough that patients discontinue the use of the appliance.[112] Bite changes (e.g., the inability to close on the back teeth) combined with heavy contact of the front teeth on removal of the appliance in the morning also are reported, but these changes generally resolve on removal of the device.

Oral appliances can be made to fit over false teeth, although this is not optimal. The use of oral appliances in people who are edentulous may be attempted with a tongue-retaining device, but these devices are not U.S. Food and Drug Administration-approved, and the efficacy of this approach is unknown. A small prospective study of factors associated with efficacy of oral appliances has suggested that age older than 55 years may be associated with reduced efficacy.[113]

Surgery

As with younger adults, upper airway surgery is not particularly effective treatment for OSA for older patients and may be associated with especially high morbidity in the elderly.[114] However, as in younger patients with significant sleep-disordered breathing in the context of obesity, bariatric surgery, specifically laparoscopic-adjustable gastric banding, has been shown to be well tolerated and reasonably effective in resolving or reducing sleep apnea in people older than 70.[115]

Pharmacologic Treatment

Several medications have been applied to the treatment of sleep-disordered breathing. In general, no drug is effective enough to recommend for use in first-line treatment. Antidepressants, nasal steroids, hormone replacement therapy (HRT), and modafinil all have been studied in younger patients.

More than two decades ago, protriptyline was demonstrated to show modest efficacy in treating apnea, probably because it reduces REM sleep, when apnea is worst.[116] There is some early experimental work with the selective serotonin reuptake inhibitors (SSRIs) in the treatment of sleep apnea, but results have not been promising in humans.[117] SSRIs can suppress REM sleep, but not as much as that seen with the tricyclic antidepressants.

Nasal steroids have been demonstrated to have modest efficacy in the treatment of sleep-disordered breathing.[118]

In the Sleep Heart Health Study, women who were on HRT were less likely to have sleep apnea, but overall lifestyle and health care are significant confounders in drawing conclusions about the efficacy of HRT for OSA.[119] Although estrogen is an option to consider, it would need to be discussed carefully with the patient because of subsequently recognized complications of HRT.

Body Position

The supine position predisposes the sleeper to airway collapse and to reduced lung volume and has long been known to exacerbate OSA; indeed, some affected persons experience obstructive events exclusively when sleeping on their backs.[120-122] Upper airway size decreases with increasing age in both men and women, and upper airway collapsibility with supine positioning increases with age.[122]

In my own clinical experience, position-related obstructive apnea is relatively common among older individuals. Position therapy has not been well studied for any group of patients but shows promise as treatment for the older patient with mild disease.[123]

DRIVING AND THE OLDER PATIENT WITH OBSTRUCTIVE SLEEP APNEA

Untreated OSA is a well-established risk factor for involvement of drivers in motor vehicle crashes (see Chapter 114) and might be expected to affect older drivers as well. In a review of conditions increasing crash risk in older drivers,

Marshall found that several conditions were believed to be associated with increased risk of crash in older persons, including alcohol abuse and dependence, cardiovascular disease, cerebrovascular disease, depression, dementia, diabetes mellitus, epilepsy, use of certain medications, musculoskeletal disorders, schizophrenia, vision disorders, and, finally, OSA. He noted that these "conditions can serve as potential warnings for reduced fitness to drive, but many persons with these medical conditions would still be considered safe to continue driving."[124]

CLINICAL PEARLS

- The clinical presentation of OSA in older adults differs from that in their middle-aged counterparts, so the disorder may be overlooked by clinicians when it manifests in this age group.
- Female gender and obesity are less important risk factors in older people than in younger people.
- Symptoms of sleep apnea change with aging: Whereas the classic symptoms of OSA are witnessed apneas and sleepiness, older patients are more likely to present with sleep complaints, nocturia, and cognitive dysfunction.
- Moderate to severe OSA is associated with increased risk of cardiovascular morbidity and mortality as well as cognitive dysfunction, and CPAP treatment is associated with reduced risk.

SUMMARY

OSA is prevalent and potentially deadly in older people and may be overlooked by clinicians because the clinical presentation is different from that in younger people. Seniors with OSA tend to be "thinner" and are more likely to be female and less likely to report classic symptoms of witnessed apnea, snoring, and fatigue. Because CPAP use is associated with reduced morbidity and mortality in older (as well as younger) people, clinicians need to consider the possibility of OSA in older patients with sleep complaints.

Selected Readings

Ancoli-Israel S, Gehrman P, Kripke DF, et al. Long-term follow-up of sleep disordered breathing in older adults. *Sleep Med* 2001;**2**:511–16.

Chang WP, Liu ME, Chang WC, et al. Sleep apnea and the risk of dementia: a population-based 5-year follow-up study in Taiwan. *PLoS ONE* 2013; **8**:e78655.

Guillot M, Sforza E, Achour-Crawford E, et al. Association between severe obstructive sleep apnea and incident arterial hypertension in the older people population. *Sleep Med* 2013;**14**:838–42.

Holmqvist F, Guan N, Zhu Z, the ORBIT-AF Investigators. Impact of obstructive sleep apnea and continuous positive airway pressure therapy on outcomes in patients with atrial fibrillation-Results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). *Am Heart J* 2015;**169**:647–54.

Jennum P, Tønnesen P, Ibsen R, Kjellberg J. All-cause mortality from obstructive sleep apnea in male and female patients with and without continuous positive airway pressure treatment: a registry study with 10 years of follow-up. *Nat Sci Sleep* 2015;**7**:43–50.

Kushida C, Nichols DA, Holmes TH, et al. Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: The Apnea Positive Pressure Long-term Efficacy Study (APPLES). *Sleep* 2012;**35**:1593–602.

Martinez-Garcian MA, Campos-Rodruigez F, Catalan-Serra P, et al. Cardiovascular mortality in obstructive sleep apnea in the elderly: role of long-term continuous positive airway pressure treatment. *Am J Respir Crit Care Med* 2012;**186**:909–16.

McMillan A, Bratton DJ, Faria R, et al. A multicentre randomised controlled trial and economic evaluation of continuous positive airway pressure for the treatment of obstructive sleep apnoea syndrome in older people: PREDICT. *Health Technol Assess* 2015;**19**(40):1–188.

Peppard PE, Young T, Barnet JH, et al. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol* 2013;**177**:1006–14.

Russo-Magno P, O'Brien A, Panciera T, et al. Compliance with CPAP therapy in older men with obstructive sleep apnea. *J Am Geriatr Soc* 2001;**49**:1205–11.

Yaggi HK, Concato J, Kernan WN, et al. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* 2005;**353**:2034–41.

A complete reference list can be found online at ExpertConsult.com.

Chapter 153

Insomnia in Older Adults

Tamar Shochat; Sonia Ancoli-Israel

Chapter Highlights

- Insomnia is a complaint of dissatisfaction with sleep characterized by difficulty initiating and/or maintaining sleep and/or nonrestorative sleep, resulting in significant daytime symptoms and functional impairments. Chronic insomnia affects approximately 10% of the adult population.
- Although late-life insomnia usually is attributed to medical and psychiatric comorbidity rather than to age-related changes per se, underlying age-related physiologic changes in sleep-wake regulation—that is, circadian rhythms and sleep homeostasis—have been identified. Other factors associated with insomnia in the elderly population include use of certain medications or other substances, psychosocial issues, and primary sleep disorders.
- Key elements in appropriate evaluation and management include considering the type of insomnia complaint, assessing sleep patterns including daytime napping, daytime health and wellness consequences, and comorbidity. Behavioral treatment should be considered first, and when necessary, the newer hypnotic medications may be added.

OVERVIEW

According to the American Psychiatric Association, as defined in the most recent *Diagnostic and Statistical Manual of Mental Disorders* (DSM5), insomnia is a complaint that involves dissatisfaction with sleep quality and/or quantity, difficulty initiating sleep, difficulty maintaining sleep, waking up too early, and/or nonrestorative or poor sleep, with a negative impact on daytime functioning and occurring at least three nights a week for more than 3 months.[1] Insomnia is a chronic condition, often lasting for several years. In the elderly population, chronic insomnia is common and often coexists with a medical or psychiatric condition, polypharmacy, and/or another sleep disorder. Growing evidence suggests that insomnia also is a potential contributor to subsequent morbidity.[2,2a] Yet owing to the widespread notion that insomnia is an inevitable consequence of aging, the problem often is ignored and not properly treated. It is therefore imperative to increase awareness in both clinicians and older adults of the significance of insomnia as a health issue, and of the importance of timely and appropriate interventions for successful treatment, to improve sleep and maximize wellness in this population.

EPIDEMIOLOGY AND RISK FACTORS

The prevalence of insomnia in the general adult population in Norway based on DSM5 criteria is estimated at 7.9%, controlling for sociodemographic and comorbidity factors.[3] In a U.S. study, the prevalence of nighttime symptoms and nonrestorative sleep complaints combined was 5%, whereas the prevalence of any sleep symptom was 18%.[4]

Prevalence of insomnia increases with age and is higher among women.[3-6,7] Age and gender differences may be related to differences in modes of measurement. In a sample of more than 5000 middle-aged and older adults from the Sleep Heart Health Study, older age was significantly related to poor sleep in men on the basis of polysomnographic recordings, particularly reduction in slow wave sleep and increased sleep stages 1 and 2, whereas in women, older age was related to subjective sleep complaints, particularly difficulty falling asleep.[8] In men and women, older age was associated with shorter sleep, lower sleep efficiency, and more arousals based on polysomnography (PSG), yet with fewer daytime complaints of unrest and sleepiness.

Furthermore, medical and psychiatric comorbid conditions as well as sociodemographic and lifestyle factors overshadow the risks of age.[6,9] In a study of more than 13,500 participants aged 47 to 69 years, 22% complained of difficulty falling asleep, 39% complained of difficulty staying asleep, and 35% complained of nonrestorative sleep.[5] Depression and heart disease were associated with all three complaints. Other related factors were medical illnesses, lower socioeconomic status and education level, and unhealthy behaviors such as alcohol use and cigarette smoking. Findings from the National Health and Nutrition Examination Survey (NHANES) of nearly 11,000 adults aged 20 and older found distinct risk factors for nighttime insomnia symptoms versus the nonrestorative sleep complaint.[4] Whereas nighttime insomnia symptoms were associated with increased age, lower income and educational level, and presence of cardiovascular disease, nonrestorative sleep was associated with primary sleep disorders such as sleep apnea, respiratory disease, thyroid disease, and cancer, as well as increased inflammation.

Other studies have examined the prevalence of insomnia and related risk factors in the elderly population. In a sample of more than 9000 participants aged 65 and older from the National Institute of Aging's Established Populations of Epidemiological Studies of the Elderly, more than 50% reported at least one sleep complaint, and 35% to 40% reported difficulty in initiating and/or maintaining sleep on a chronic basis.[10] At a 3-year follow-up evaluation, an annual insomnia incidence of 5% was reported.[11] Rates were highest in those with chronic medical conditions such as heart disease, stroke, and diabetes. Remission occurred in nearly one half of those with insomnia at baseline and was related to improvements in perceived health. Similarly, in a representative sample of general medical practice patients aged 65 years and older, the annual incidence of late-life insomnia was 3.1%.[9] Significant and independent risk factors in this sample were depressed mood, poor physical health, and low physical activity.

As indicated by the National Sleep Foundation survey from 2003, depression, heart disease, bodily pain, and memory problems were the disorders most commonly associated with insomnia.[12] In a more recent sample of 2000 South Korean community-dwelling older adults (older than 65 years of age), 29% reported nighttime insomnia symptoms and 17% also reported daytime impairments.[13] For all insomnia symptoms, prevalence was higher among women as well as among people with no education, those living alone, and persons with a diagnosis of restless legs syndrome or depression and/or a lifetime history of physical illness. Older age was again associated with increased nighttime symptoms but with decreased complaints of nonrestorative sleep. These findings lend further support to the epidemiologic evidence demonstrating that geriatric sleep complaints and disorders are not the result of age per se but rather cosegregate with medical and psychiatric disorders and related health burdens.[14]

Insomnia in elderly persons often is unrecognized by physicians.[2] In a study of older adults in primary care practices in the midwestern United States, 69% of patients confirmed the presence of at least one sleep problem, 40% had two or more sleep problems, and 45% experienced symptoms of insomnia, yet these complaints were reported in the medical charts only 19% of the time.[15] The two questions that best identified persons with poor sleep and those at risk for medical and psychiatric problems were "Do you feel excessively sleepy during the day?" and "Do you have difficulty falling asleep, staying asleep, or being able to sleep?" These two questions could easily be integrate into the standard patient history at clinical intake assessment.

HEALTH AND WELLNESS CONSEQUENCES

Studies have assessed the health consequences of insomnia as well as the effects of insomnia on physical functioning and performance.[2a,15a] Symptoms of insomnia, daytime sleepiness, and sleep medication use all predicted 4-year incidence of depressive symptoms in nearly 4000 older adults who were free of depression symptoms at baseline.[16] In a study of several thousand older men, lighter, fragmented sleep, particularly shorter total sleep time, sleep efficiency below 80%, and more than 90 minutes of wake time after initial sleep onset, was associated with poorer physical performance, particularly in age-adjusted models.[17] In the Study of Osteoporotic Fractures, poor sleep (i.e., fewer than 5 hours of nightly sleep and sleep efficiency of 70% or lower), but not benzodiazepine use, was an independent risk factor for falls in elderly women.[18,19]

The impact of insomnia on cognitive function in the elderly population is less clear. In an 8-year longitudinal study of nearly 5000 older adults in France, cognitively intact at baseline, excessive daytime sleepiness but not insomnia predicted global cognitive decline.[20] Similar findings were reported in longitudinal studies in the United Kingdom[21] and the United States.[22] Furthermore, in a 2-year prospective study assessing relationships among snoring, sleep duration, and sleep difficulties with cognitive functioning in elderly women, short sleep duration (5 hours or less) and insomnia complaints were related to lower scores on cognitive tests at baseline but not at 2-year follow-up evaluation.[23] Finally, in a 3-year longitudinal study, chronic insomnia was found to be an independent risk factor for cognitive decline in older men, but not in older women,[24] suggesting that cognitive decline related to insomnia may be gender-dependent. Alternatively, daytime sleepiness may be the underlying factor for cognitive decline, rather than insomnia.

On the other hand, cross-sectional analysis in the Study of Osteoporotic Fractures of approximately 3000 women 70 years of age or older showed cognitive decline associated with poor sleep (based on actigraphically measured sleep efficiency of 70% or less, long sleep latency, and increased wake time after sleep onset).[25] The study investigators also reported that preclinical cognitive decline predicted increased sleep disturbance (i.e., reduced sleep efficiency, longer sleep latency, and increased wake time after sleep onset), suggesting reverse causality.[26]

Population-based studies on the consequences of insomnia in middle-aged and older adults have identified additional negative functional outcomes. In a metaanalysis, symptoms of insomnia (except difficulty falling asleep) and short sleep duration were associated with an increased risk for hypertension.[27] In a longitudinal analysis of more than 80,000 postmenopausal women of the Women's Health Initiative who did not have heart disease at baseline, insomnia was associated with increased risks for incident coronary heart disease and incident cardiovascular disease in fully adjusted models.[28]

Longitudinal studies based from the Korean Genome and Epidemiology Study demonstrated that persistent insomnia predicted increased risks for depression and suicidal ideation and reduced physical and mental measures of health-related quality of life.[29,30] Similarly, disorders of initiating sleep but not other insomnia symptoms predicted incident depression in a nationally representative sample of older adults in Japan.[31] In a cross-sectional study of postmenopausal women, being unemployed or on a disability support pension, sedentary lifestyle, and reduced physical and mental measures of quality of life were strongly associated with insomnia symptoms and accounted for 28% of the explained variance.[32]

Insomnia also increases the risk of mortality. In a study of older adults followed for five years, those with initial sleep latencies longer than 30 minutes or sleep efficiency lower than 80% had close to twice the risk for mortality.[33] In the prospective MrOs study in 2500 older men (65 years of age or older) who were not frail at baseline, poor subjective sleep quality, more nighttime wakefulness, and greater nocturnal hypoxemia

were associated with increased risks for incident frailty or mortality at follow-up evaluation.[34]

ETIOLOGY

Underlying factors involved in the development of late-life insomnia include age-related changes in homeostatic and circadian sleep-wake regulation, psychiatric and medical comorbidity, use of certain medications and other substances, and primary sleep disorders. Evidence for each of these factors is reviewed next.

Age-Related Changes in Sleep Regulation

Developmental changes in elderly sleep are characterized by advanced sleep phase, including both earlier bedtimes and wake times, reduced sleep consolidation, and altered sleep architecture indicating a transition to lighter sleep. To understand the basis for these changes, it is necessary to understand some of the basic mechanisms of human sleep regulation.

Sleep regulation is based on an interaction between the homeostatic pressure for sleep and the output of the circadian pacemaker. Homeostatic sleep pressure reflects the increasing need for sleep which accumulates during the waking hours and then dissipates during sleep, as marked by increased slow wave sleep at the beginning of nocturnal sleep which gradually decrease throughout the night. The endogenous circadian pacemaker, located in the suprachiasmatic nucleus of the hypothalamus, regulates the synchronous timing of several physiologic variables including core body temperature and sleep/wake states. Regulation of the timing of sleep and wake states is achieved by promoting a signal of increased wakefulness throughout the day, and a signal of increased sleep consolidation throughout the night.[35,36] In young adults, daytime wakefulness and nighttime sleep consolidation are high owing to both of these bioregulatory mechanisms. However, both homeostatic and circadian mechanisms change with age. Advanced age has been associated with a marked reduction in slow wave sleep and an increase in lighter sleep. This reduction in slow wave sleep indicates weaker homeostatic sleep pressure in elderly persons.[37]

Under entrained conditions, habitual wake times are advanced to an earlier hour in the elderly population, and the amplitude of the circadian rhythm of core body temperature is decreased, indicating a reduced circadian signal, which would otherwise promote sleep in the early morning hours.[38] The circadian signal for wakefulness also is reduced, as reflected in sleep episodes and reports of sleepiness in the early evening hours.[39] In summary, reductions in both the homeostatic drive for sleep and the strength of the circadian signal for sleep in the early morning hours and for wakefulness in the early evening hours have been implicated as underlying factors for reduced sleep consolidation, advanced sleep phase, and early-morning awakenings in the elderly.

Medical and Psychiatric Comorbid Conditions

Numerous chronic medical conditions and illnesses are known to be associated with sleep disruption. Arthritis, angina pectoris, congestive heart failure, coronary artery disease, chronic obstructive pulmonary disease, end-stage renal disease, diabetes, asthma, stroke, gastroesophageal reflux disease, dementia/Alzheimer disease, Parkinson disease, cancer, and menopause all have been implicated.

In a representative U.S. survey of 1500 adults 50 years or older, those with heart disease, lung disease, depression, obesity, and bodily pain were more likely to report difficulties with sleep.[12] In addition, multiple medical conditions were associated with more sleep complaints and with poor self-perceived sleep quality. In another survey examining sleep and health in 1500 adults 60 years of age or older in 11 primary care offices, complaints of poor sleep and excessive daytime sleepiness were associated with both poor physical and poor mental health–related quality of life status.[15] Similar findings were reported in a longitudinal investigation of over 90,000 women who participated in the Women's Health Initiative (WHI), whereby incident and persistent insomnia were associated with increased risks of physical, mental, and mixed impairments.[2a] Yet another cross-sectional survey in six clinical centers with more than 3000 older men as participants demonstrated increased risks for poor sleep in men with depression. In multivariable-adjusted models, men with depression were nearly four times more likely to report poor sleep quality and nearly twice as likely to spend at least 1 hour falling asleep.[40]

Medical and psychiatric comorbid diseases may be considered as both risk factors and consequences of insomnia, and it often is difficult to determine the direction of causality. Accordingly, *comorbid* insomnia has been suggested as the appropriate term for insomnia associated with other, concurrent diseases, replacing the earlier designation of *secondary* insomnia.[41] Such a distinction has important implications not only in reassessing causal relationships between insomnia and comorbidity but also for considering treatment strategies. Treatment and management should be directed not only at comorbid illnesses but should also focus on insomnia as a discrete entity.

Medications/Substances

Polypharmacy is a serious problem in older adults, and the use of multiple medications also contributes to insomnia. Prescription medications known to be related to insomnia include antidepressants, such as selective serotonin reuptake inhibitors and serotoninergic-noradrenergic reuptake inhibitors, and numerous agents commonly used in the treatment of various medical conditions, such as bronchodilators, beta blockers, corticosteroids, decongestants, central nervous system stimulants, gastrointestinal drugs, and cardiovascular drugs.[2] Adjustment of the timing or amount of dosing and eliminating or substituting other medications with recognized contraindications or drug-drug interactions in elderly persons may lead to improvements in their sleep.

Other substances known to be related to insomnia in older persons include alcohol, caffeine, and nicotine. In a sample of more than 6000 adults 50 years of age or older, occasional and frequent binge drinking was associated with an increased risk for insomnia after controlling for demographic and clinical factors; this increase in risk was partially mediated by cigarette smoking.[42] Of note, however, in a study of nearly 10,000 adults 65 years of age or older in France, moderate alcohol and caffeine consumption reduced the risk for insomnia symptoms in women, demonstrating a protective effect.[43] Furthermore, in a study assessing sleep hygiene patterns in four sleep subgroups of older adults (i.e., with or without insomnia and with or without sleep complaints), there were no differences between the subgroups for alcohol, nicotine, and

caffeine use, and those who abstained entirely, indicating that lifestyle changes in sleep hygiene practices may not be an effective therapy in this age group.[44]

Primary Sleep Disorders

Insomnia often is comorbid with primary sleep disorders common in the elderly population, such as sleep-disordered breathing (SDB), periodic limb movement disorder, and restless legs syndrome.

Sleep-Disordered Breathing

SDB (see also Chapter 152) is characterized by episodes of partial (hypopnea) to complete (apnea) airway collapse, causing pauses in breathing during sleep and occurring repeatedly during the night. These respiratory events reduce blood oxyhemoglobin saturation and terminate in partial arousals. Clinical manifestations of SDB include excessive daytime sleepiness and loud snoring. The prevalence of SDB increases considerably with age, to as much as 62% for a respiratory disturbance index of 10 or greater in adults 65 years of age or older (70% in men, 56% in women).[45]

In a more recent study of older adults (60 years of age or older) in Korea, prevalence was 36.5% for an apnea-hypopnea index of 15 or higher (52.6% in men, 26.3% in women).[46]

Despite its high prevalence, the consequences of SDB are considered more benign in older than in younger adults, suggesting that considerations regarding management should take into account the severity of the disorder, its impact on functioning, and implications of therapeutic intervention.[47]

By some estimates, approximately one half of the patients with SDB also may have insomnia.[48] Cooccurrence of these disorders is more common in women and is associated with lower sleep quality and a higher rate of psychiatric disorders than in patients with SDB alone. In the elderly population, cooccurrence of insomnia and SDB has been associated with worse daytime dysfunction than that observed for either sleep disorder alone, suggesting an additive effect on functional impairment.[49] Although the underlying relationship between these two common sleep disorders is not clear, it has been suggested that hypoxia may be a mediating factor between SDB and insomnia in older adults with cardiovascular disease.[50]

An important consideration is that SDB may be veiled in older patients with insomnia. In a sample of 80 adults with insomnia older than 59 years of age in whom initial clinical intake evaluation showed no traditional signs of SDB, subsequent rigorous screening identified that 29% had an apnea-hypopnea index of greater than 15, indicating significant SDB.[51] These findings confirm that clinical interview alone may not suffice for identifying veiled SDB in older adults and stress the importance of referring patients with insomnia who demonstrate reduced functional capabilities for further evaluation in the sleep clinic.[49]

Periodic Limb Movements in Sleep and Restless Legs Syndrome

Periodic limb movements in sleep (PLMS) is a disorder characterized by involuntary leg jerks during sleep appearing in repetitive clustered episodes, often leading to brief awakenings from sleep. PLMS traditionally has been associated with insomnia and/or excessive daytime sleepiness. Clinical diagnosis of PLMS is indicated when the PLM index (PLMI), the number of limb movements per hour of sleep, is greater than 5. Although survey data do not demonstrate an increased prevalence of PLMS with age,[52] prevalence rates based on home PSG in community-dwelling women were high, with 66% for a PLMI of 5 or greater and 27% for PLMI 5 or greater accompanied by arousals.[53] In men and women, PLMS, particularly when accompanied by arousals, was related to polysomnographic indicators of disturbed sleep that are associated with insomnia.[53,54]

Despite its high prevalence, long-term follow-up evaluation of a cohort of elderly persons showed no change in PLMS severity with increasing age[55]; yet a study on the pathologic significance of PLMS in this age group has shown that in older men, increased frequency of PLMS both with and without arousals was associated with incident cardovascular disease, particularly in those who were nonhypertensive.[56]

Restless legs syndrome is a neurologic disorder characterized by dysesthesia in the legs and an irresistible urge to move them in order to relieve the discomfort. Symptoms of this disorder increase during the evening and at night, usually when the patient is in a relaxed or restful state, resulting in sleep disturbance. Based on survey data, the prevalence of restless legs syndrome increases with age, from 2.7% in teens to 8.2% in adults 80 years of age or older.[52] PLMS is a common finding in 80% of cases of restless legs syndrome, yet PLMS and related sleep disturbance also may appear without restless legs symptoms.[57] Both disorders are prevalent in the elderly population and are strongly related to insomnia. Accordingly, proper identification and management are warranted.

Pharmacologic treatment studies for PLMS and restless legs syndrome have mostly studied middle-aged to older adults. Management strategies for both conditions are similar, and treatment should be limited to symptomatic persons with comorbid insomnia and/or daytime sleepiness. Medications approved for the treatment of PLMS and restless legs syndrome may increase daytime sleepiness in the older adult, so monitoring and follow-up are recommended.[58]

EVALUATION AND DIAGNOSIS OF INSOMNIA

Despite its widespread prevalence, insomnia is underrecognized and often inadequately treated in health care systems.[15,59] Guidelines for the evaluation of insomnia in the adult population, based on the Standards of Practice Committee of the American Academy of Sleep Medicine (AASM),[60] call for a thorough sleep history and determination of specific sleep complaints, sleep-wake schedules, history of psychological symptoms such as depression or anxiety, comorbid conditions such as medical or psychiatric illness, use of certain medications and other substances, and other primary sleep disorders.

The assessment of insomnia does not necessitate overnight polysomnographic monitoring; however, symptoms and signs of comorbid sleep disorders, especially sleep-disordered breathing, warrant a full PSG study.[49,60] In accordance with AASM standard guidelines, the use of actigraphy is indicated in patients with insomnia, including the elderly and institutionalized elderly subgroups, for the assessment of circadian rhythms and sleep-wake patterns and their disturbances, as well as for monitoring treatment for insomnia and circadian rhythm disturbances.[61] Actigraphy is considered a

Table 153-1 Nonpharmacologic Therapies for Late-Life Insomnia

Therapy	Description
Stimulus control	Patient is instructed to go to bed only when sleepy. If sleep is not obtained in 20 minutes, patient leaves bedroom and returns to bed only when sleepy. This process is repeated as needed until sleep is obtained. Wake time is normal, and naps are not allowed.
Cognitive-behavioral therapy for insomnia (CBT-I)	Maladaptive beliefs and attitudes regarding the distinction between normal and abnormal sleep are identified, addressed, and redefined, to induce positive changes in sleep-related cognition and associated behavioral and emotional outcomes.
Relaxation training	Various relaxation techniques including meditation, muscle relaxation, and biofeedback for the reduction of somatic and cognitive arousal.
Sleep restriction	Time in bed is restricted to self-estimated total sleep time. Sleep efficiency (ratio between time asleep and time in bed) is assessed weekly. Time in bed may gradually be increased on confirmation of increases in sleep efficiency. No naps are allowed.
Paradoxical intention	Patient is instructed to lie in bed and attempt to remain awake with eyes open for as long as possible. Bedroom should be dark and quiet, and patient must not engage in other activities other than the continuous endeavor to remain awake.
Sleep hygiene education	On the basis of the sleep history, ineffective habits and behaviors related to poor sleep are targeted and clear guidelines for better sleep are provided, such as creating a stable sleep-wake schedule and maintaining a sleep-inducing bedroom environment, minimizing napping, and avoiding sleep-inhibiting substances and behaviors at night, including caffeine, nicotine and alcohol, heavy meals and high fluid intake, worrisome thoughts, and clock viewing.
Bright light	Patients are exposed either to a bright light fixture or to natural outdoor light in the evening hours.*

*Indicated for advanced sleep phase and early-morning awakenings.

more reliable measurement tool than sleep logs (although sleep logs often are used to complement actigraphy) and is particularly useful in populations with poor tolerance of PSG, such as institutionalized elderly persons.

TREATMENT OF INSOMNIA

The effectiveness of pharmacologic and nonpharmacologic treatments has been demonstrated for late-life insomnia. The evidence for their efficacy is reviewed next, followed by a summary of specific considerations for their use in the elderly population.

Nonpharmacologic Treatments

Nonpharmacologic treatments that have been investigated for insomnia in older adults include psychological, cognitive-behavioral, and bright light treatments (Table 153-1). In an update of the AASM practice parameters for psychological and behavioral treatments of insomnia,[62] stimulus control, relaxation training, and cognitive-behavioral therapies were recommended treatments for chronic insomnia with a high level of evidence from clinical trials. Treatments reaching moderate levels of evidence for efficacy included sleep restriction therapy, multicomponent therapy (without cognitive therapy), biofeedback, and paradoxical intention. These treatments have been found to be effective in older adults and in chronic hypnotic medication users. Insufficient evidence was available to determine the efficacy of sleep hygiene, imagery training, and cognitive therapy.

Cognitive-behavioral therapy for insomnia (CBT-I) has proved to be successful for older adults with primary and comorbid insomnia, and for those with hypnotic medication dependency.[62a] For example, in a clinical ambulatory PSG trial with 6-week and 6-month follow-up periods, CBT-I was compared with zopiclone and placebo in elderly patients with chronic insomnia. For the CBT-I group, wake time was significantly reduced, and both sleep efficiency and slow wave sleep were significantly increased at both 6-week and 6-month follow-up assessments, compared with the zopiclone and placebo groups.[63] In a sample of older adults with sleep maintenance insomnia, a brief (4-week) group-based CBT-I intervention showed significant improvements in sleep measures at completion of treatment and at a 3-month follow-up evaluation, including reduced wake time after sleep onset, increased sleep efficiency, and reduced time in bed, as reported in sleep diaries.[64] At the conclusion of treatment and at the 3-month follow-up assessment, 46.5% and 42.2%, respectively, were no longer in the diagnostic range for insomnia based on the Insomnia Severity Index (ISI).

In another study on late-life insomnia, comparing CBT-I with a sedative hypnotic medication (temazepam) and with both modalities combined (CBT-I and temazepam) for the treatment of insomnia, positive short-term efficacy was reported for all three regimens compared with placebo, with a small increased benefit for the combined regimen.[65]

However, at a 2-year follow-up evaluation, sustained long-term gains were achieved only in the CBT-I groups, especially in the CBT-I–only group, indicating that CBT-I alone is superior to combined therapy. As indicated by this evidence, CBT-I provides a particularly promising alternative for this age group. In the 2005 National Institutes of Health State of the Science Conference on Insomnia, CBT-I was found to be as effective as prescription medications for the brief treatment of chronic insomnia, with indications that in contrast with medications, the beneficial effects of CBT-I may last well beyond termination of treatment.[41]

Evening bright light exposure has been indicated for the treatment of advanced circadian sleep phase syndrome and insomnia characterized by early-morning awakening, both of which are common among elderly persons. The practical usefulness of bright light treatment in late-life insomnia, however, generally has not been supported by research findings.[66] In view of the importance of implementing feasible therapeutic alternatives to medication in this population, further investigation, possibly targeting specific subtypes of insomnia that may benefit from bright light therapy, is warranted.

Pharmacologic Treatments for Insomnia

Commonly used hypnotic prescription medications for the treatment of insomnia include traditional benzodiazepine sedative-hypnotics (temazepam, estazolam, flurazepam, quazepam, triazolam), newer selective nonbenzodiazepine sedative "Z-drug" hypnotics (eszopiclone, zaleplon, zolpidem, and zolpidem MR), the melatonin receptor agonist ramelteon, and low-dose doxepin. Other sedating antidepressants (e.g., trazodone), antipsychotics (e.g., quetiapine), and antihistamines (diphenhydramine) also are used off label, despite limited efficacy data and often considerable safety risk. Self-treatments for insomnia include over-the-counter medications and herbal or dietary supplements. In the 2005 National Institutes of Health State of the Science Conference on Insomnia, the panel concluded that the risks associated with off-label drugs outweigh the benefits; therefore it did not recommend the use of antidepressants, antihistamines, or antipsychotics, or of over-the-counter or herbal supplements, for the treatment of insomnia.[41]

In a representative sample of more than 32,000 noninstitutionalized American adults (i.e., the NHANESs) aged 20 to older than 80 years, the prevalence of medications commonly used for insomnia in the preceding month was 3%, with a significant increase between 1999–2000 and 2009–2010.[67] The most commonly used medications were Z-drugs and trazodone, and the frequency of concurrent use of two or more sedatives was high (55% for at least two and 10% for at least four concurrent sedatives). The likelihood of consuming sedatives for insomnia after multivariable adjustments was highest in older adults, those seeing a mental health provider, and those reporting polypharmacy in sedative use.

Although the Z-drugs are considered safer and more effective than the older benzodiazepines,[41] in a meta-analysis of data submitted to the U.S. Food and Drug Administration[69] assessing the effects of Z-drugs on objective (PSG) and subjective sleep latency in the adult population, small but significant improvements were found, yet these improvements were comparable to placebo effects and were more pronounced in younger persons, in females, and at higher drug doses. In view of the typically higher rates of sleep maintenance insomnia and the increased risk for adverse effects, particularly with higher doses, in older adults, the true efficacy of these sedatives in the elderly population remains questionable.

Hypnotic medication use is highest and most chronic in older adults, yet it has been suggested that the benefit of such medications is lower in older than in younger adults.[68,69] General guidelines for the use of sedative-hypnotics in the elderly population include selection of the appropriate drug (short- versus long-acting), determining the type of insomnia complaint (sleep onset versus sleep maintenance insomnia), starting with a low dose and increasing as needed, recognizing the risk of specific drug-drug interactions, and monitoring for adverse effects, particularly residual effects of daytime sleepiness, impaired cognitive performance, and increased risk of falls.[68] It is important to keep in mind that the adverse effects of sedative-hypnotic use may outweigh the benefits, particularly in persons with baseline cognitive impairment.

Nevertheless, use of selected nonbenzodiazepine hypnotics has been investigated in the older population. In an open-label 6- to 12-month investigation in older adults with chronic insomnia, long-term use of 5 and 10 mg of nightly zaleplon was found to be safe and effective, demonstrating improvements in subjective sleep measures with no rebound insomnia on discontinuation.[70] The efficacy of eszopiclone was evaluated in nearly 400 older adults 65 to 85 years of age who were randomly assigned in a double-blind trial to receive eszopiclone (2 mg/night) or placebo.[71] The eszopiclone group demonstrated significant improvements on subjective measures of sleep, including a mean increase of more than 60 minutes in total sleep time and a mean decrease of 36 minutes in wake time after initial sleep onset. Improvements were maintained for 3 months, and no rebound effects occurred on discontinuation.

Ramelteon is a selective agonist for melatonin subtype 1 and 2 receptors. Assessment of the efficacy of ramelteon 8 mg/night versus placebo for more than 5 weeks in 800 older adults with chronic insomnia has demonstrated significantly decreased sleep latency, with no withdrawal or rebound insomnia after discontinuation based on patient reports.[72] These findings were supported by a polysomnographic repeated measures design, with two treatment days for each treatment phase (8 mg and placebo) and an intervening washout period.[73] PSG results demonstrated reduced latency to persistent sleep and increased total sleep time and sleep efficiency for both doses compared with placebo. No residual effects were found for next-day cognitive and psychomotor performance.

Nevertheless, it is worthy of emphasis that although some evidence has demonstrated the safety and efficacy of the newer sedatives in older adults with chronic insomnia, the overall clinical significance of outcomes has been questioned, and these medications are not risk-free.

CLINICAL PEARL

Insomnia is very common in older adults and generally is related to medical or psychiatric illness, use of certain medications, circadian rhythm and sleep homeostatic changes, or other primary sleep disorders, rather than to aging per se. Clinicians should routinely screen for sleep problems in this age group and initiate appropriate treatment.

SUMMARY

Symptoms of insomnia are very common in the elderly population, and the problem is largely underrecognized and undertreated. Increasing evidence of the significant adverse

consequences of insomnia warrants careful assessment and appropriate treatment strategies for this age group. Proper treatment of insomnia in this age group is effective and may improve overall physical and mental health, subjective sense of well-being, and quality of life in the older patient. Further research is required in selected older adult cohorts to elucidate the mechanisms underlying the development of insomnia and its related comorbid conditions and consequences, and to determine how these processes can be effectively treated or prevented.

Selected Readings

Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep* 2014;**37**(2):343–9.

Bloom HG, Ahmed I, Alessi CA, et al. Evidence-based recommendations for the assessment and management of sleep disorders in older persons. *J Am Geriatr Soc* 2009;**57**(5):761–89.

Ensrud KE, Blackwell TL, Ancoli-Israel S, et al. Sleep disturbances and risk of frailty and mortality in older men. *Sleep Med* 2012;**13**(10):1217–25.

Grandner MA, Martin JL, Patel NP, et al. Age and sleep disturbances among American men and women: data from the U.S. behavioral risk factor surveillance system. *Sleep* 2012;**35**(3):395–406.

Jaussent I, Bouyer J, Ancelin ML, et al. Excessive sleepiness is predictive of cognitive decline in the elderly. *Sleep* 2012;**35**(9):1201–7.

Karlin BE, Trockel M, Spira AP, et al. National evaluation of the effectiveness of cognitive behavioral therapy for insomnia among older versus younger veterans. *Int J Geriatr Psychiatry* 2015;**30**:308–15.

Lovato N, Lack L, Wright H, Kennaway DJ. Evaluation of a brief treatment program of cognitive behavior therapy for insomnia in older adults. *Sleep* 2014;**37**(1):117–26.

National Institutes of Health. National Institutes of Health State of the Science conference statement on manifestations and management of chronic insomnia in adults, June 13-15, 2005. *Sleep* 2005;**28**(9): 1049–57.

Phillips B, Mannino DM. Correlates of sleep complaints in adults: the ARIC study. *J Clin Sleep Med* 2005;**1**(3):277–83.

Reid KJ, Martinovich Z, Finkel S, et al. Sleep: a marker of physical and mental health in the elderly. *Am J Geriatr Psychiatry* 2006;**14**(10): 860–6.

Sands-Lincoln M, Loucks EB, Lu B, et al. Sleep duration, insomnia, and coronary heart disease among postmenopausal women in the Women's Health Initiative. *J Womens Health (Larchmt)* 2013;**22**(6):477–86.

Song Y, Dzierzewski JM, Fung CH, et al. Association between sleep and physical function in older veterans in an adult day healthcare program. *J Am Geriatr Soc* 2015;**63**:1622–7.

Unruh ML, Redline S, An M, et al. Subjective and objective sleep quality and aging in the Sleep Heart Health Study. *J Am Geriatr Soc* 2008; **56**(7):1218–27.

Yokoyama E, Kaneita Y, Saito Y, et al. Association between depression and insomnia subtypes: a longitudinal study on the elderly in Japan. *Sleep* 2010;**33**(12):1693–702.

Zaslavsky O, LaCroix AZ, Hale L, et al. Longitudinal changes in insomnia status and incidence of physical, emotional, or mixed impairment in postmenopausal women participating in the Women's Health Initiative (WHI) study. *Sleep Med* 2015;**16**:364–71.

A complete reference list can be found online at ExpertConsult.com.

Chapter
154

Circadian Rhythms in Older Adults

Gregory J. Tranah; Katie L. Stone; Sonia Ancoli-Israel

Chapter Highlights

- Aging is associated with altered circadian rhythms, the intrinsic physiologic cycles of approximately 24 hours that are critically involved in control of sleep-wake cycles and numerous physiologic processes.
- Wrist actigraphy can be used to measure 24-hour recordings of activity, from which sleep and wake can be scored. Activity patterns correlate strongly with entrained endogenous circadian phase.
- Older women and men with weak and shifted circadian activity rhythms have higher mortality rates.
- Older women with weak and late-shifted circadian activity rhythms are at higher risk for development of mild cognitive impairment and dementia.
- Older men with decreased circadian activity rhythm robustness are at increased risk for cardiovascular events, primarily those related to coronary heart disease or stroke.
- Older women with depression have more desynchronized circadian activity rhythms.
- If the benefits of interventions (e.g., physical activity, bright light exposure) that regulate circadian activity rhythms are confirmed in other populations, studies will be needed to test whether such interventions will improve health outcomes in elderly persons.

CIRCADIAN RHYTHMS AND AGING

Circadian rhythms are intrinsic physiologic cycles of approximately 24 hours that are critically involved in control of sleep-wake cycles and numerous physiologic processes. Several biologic functions are under circadian control, including release of certain hormones, body temperature, blood pressure and heart rate, bone remodeling, and sleep and activity cycles (rest-activity rhythm). In mammals, the suprachiasmatic nucleus (SCN), located in the anterior hypothalamus, receives light signals from the retina through the retinohypothalamic tract and serves as the primary circadian pacemaker, setting the phase and period of biologic rhythms throughout the organism. The SCN also regulates pineal melatonin biosynthesis to produce rhythmic melatonin levels that regulate sleep-wake cycles. Although the SCN is crucial for the generation of biologic rhythms throughout the organism, a series of peripheral oscillators also are distributed among the cells and tissues, including the liver, endocrine tissues, heart, and skeletal muscles.[1-3] These peripheral oscillators are similar to the SCN in terms of basic molecular organization[4] in that the individual cells generate self-sustained molecular and physiologic oscillations.[5,6] The sleep-wake cycle is synchronized to the time of day by a number of cues, the strongest of which is the environmental rhythm of light and darkness.

Aging is associated with alterations in circadian activity rhythms, including decreased amplitude (height or strength of rhythm),[7] fragmentation or loss of rhythms (weakening of rhythmic pattern),[8-14] shortened natural free-running period (period of rhythm without environmental cues), a tendency toward internal desynchronization,[15-17] and decreased sensitivity to phase-resetting signals, including light and those induced by sleep medications.[16] The timing of peak rhythm activity also typically advances with age, resulting in an earlier onset of sleepiness in the evening and an earlier morning waking time.[9] The disruption of circadian rhythms during aging is paralleled by decreased photic input secondary both to age-related losses in circadian photoreception[18] and to typically reduced light exposure in older people.[19] It has been hypothesized that disruptions in these rhythms may occur in part as a result of age-related deterioration of the SCN.[20] Animal studies suggest that aging also is associated with a number of molecular changes in the circadian system, including changes in expression patterns of clock genes and changes in the neurochemistry of the SCN.[16,21-23]

Little is known concerning the causes of age-related changes in circadian patterns and the subsequent effects of these changes on health and well-being. Circadian and sleep-wake rhythm abnormalities have been observed among older adults with a wide variety of medical conditions. For example, activity phase abnormalities in older adults with dementia have been shown to predict a shorter survival.[24] Disturbances of the sleep-wake cycle, which are reflected in poor activity rhythms, are particularly pronounced in Alzheimer disease[25-27] and are hypothesized to be one of the primary causes of institutionalization[28,29] of persons with this disorder, along with shifts in daytime activity patterns.[30] Patients diagnosed with Alzheimer disease also exhibit reduced amplitudes and phase delay of circadian variation in core body temperature and activity.[31] In patients with metastatic colorectal cancer,

two-year survival was five times higher among those with stronger circadian activity rhythms than among those with rhythm abnormalities.[32] Activity phase abnormalities in elderly persons with dementia have been shown to predict shorter survival.[24] Furthermore, those with more daytime than nighttime activity had an apparent better quality of life.[32,33]

It is not clear whether activity rhythms direct influence morbidity and mortality or whether they represent biomarkers of advanced physiologic aging that provide additional information on risk beyond that of traditional covariates. Indeed, van Hilten and colleagues[34] examined the influence of age on nocturnal behavior in 100 healthy older adults and concluded that in the absence of illness, age itself has only marginal effects on the sleep-wake cycle.

This chapter reviews the latest evidence from large, population-based cohorts of elderly participants that alterations in circadian activity rhythms are associated with increased mortality rates,[35,36] dementia and mild cognitive impairment (MCI),[37] cardiovascular disease (CVD),[38] and depression.[14,39]

MEASUREMENT OF RHYTHMS

A variety of methodologies have been used for analyzing circadian activity rhythms,[14,35,36,40-45] but no single standard approach has been accepted. *Actigraphy* records limb movements and allows for 24-hour recordings of activity from which wake and sleep measures can be scored. Traditionally, the actigraph is placed on a wrist, although sometimes activity from the leg or waist is recorded. Once actigraphic data are collected, the signal is displayed on a computer and examined for patterns of activity-inactivity and analyzed for wake-sleep patterns. Wrist actigraphy has the advantages of being cost-efficient, allowing the monitoring of sleep behavior in natural environments, tracking both daytime and nighttime activity, and recording for long time periods. Although actigraphy is not a replacement for electroencephalogram (EEG) or polysomnography (PSG) recording of sleep behavior or circadian measures based on 24-hour melatonin or cortisol, it provides several clear-cut advantages for data collection: Actigraphy is a valid marker of entrained PSG sleep phase and correlates strongly with entrained endogenous circadian phase.[46] It also is useful in identifying sleep that has been disturbed by circadian rhythm changes. Actigraphy allows the study of rhythms occurring over many days; it is therefore well suited to the study of circadian rest-activity rhythms. Actigraphy also is a useful adjunct to the routine evaluation of circadian rhythm disorders. Another advantage is that actigraphy can be deployed in large clinical studies to examine the role of activity rhythms in health and disease, as well as for tracking interventions targeting rhythm disturbances.

Use of the wrist actigraph may be particularly valuable for studying people who have difficulty sleeping in a sleep laboratory, or with the wires used for traditional PSG, such as those with insomnia, children, and demented elderly persons. With actigraphy, patients are studied in their natural environment. Mormont and colleagues[33] used actigraphy to study circadian rhythms and sleep-wake cycles in patients with cancer, as a preliminary step toward the advancement of chronotherapy. Studies have also been done to examine sleep schedules of adolescents,[46] shift workers,[47,48] and in-flight crews.[49]

Three large, population-based studies of elderly persons have measured activity rhythms using wrist actigraphy: the Study of Osteoporotic Fractures (SOF), the Osteoporotic Fractures in Men (MrOS) study, and the Rotterdam Study (RS). The SOF and MrOS studies collected activity data with a sleep watch–styled actigraph (SleepWatch-O, Ambulatory Monitoring, Inc., Ardsley, New York) and looked at activity rhythm association with increased mortality,[35,36] dementia and MCI,[37] CVD,[38] and depression.[39] In these two studies, an extension to the traditional cosine model was used to map the circadian activity rhythm to the activity data.[45] The following activity rhythm parameters were calculated from the extended cosine curve: *amplitude,* an indicator of the strength of the rhythm, is the peak-to-nadir difference in activity (measured in arbitrary units of activity [counts/minute]); *mesor,* mean level of activity (measured in arbitrary units of activity [counts/minute]); *pseudo–F value,* a measure of model fit, with smaller values indicating a less robust rhythmic pattern in rest-activity and hence overall reduced circadian rhythmicity; *acrophase,* timing of peak activity measured as time of day; *beta* (β) *statistic,* a measure of steepness of the curve, in which larger values represent more square-shaped waves, which would suggest a more constant level of daytime activity; *alpha* (α) *statistic,* a measure of peak-to-trough width, in which small values would represent curves where the troughs are narrower than the peaks, suggesting greater a daytime to nighttime activity ratio; and *minimum,* the lowest-level modeled activity in which large values may be indicative of greater nighttime activity. The RS collected activity data with the Actiwatch model AW4 (Cambridge Technology, Cambridge, United Kingdom) and examined several correlates of activity rhythms including depression.[14] The following activity rhythm parameters were calculated in the RS: *M10 onset,* onset of 10 most active hours (proxy for acrophase); *L5 onset,* onset of 5 least active hours; *amplitude,* difference between average activity level in M10 and L5 periods; *relative amplitude,* amplitude divided by the sum of M10 plus L5; *average,* average of activity over the day (proxy for mesor); *interdaily stability,* degree of regularity in rhythm over days; and *interadaily variability,* fragmentation of the rhythm.

FACTORS STUDIED IN ASSOCIATION WITH CIRCADIAN RHYTHMS

Mortality

A disrupted or less robust circadian activity rhythm has been associated with medical illnesses such as dementia and cancer. Indeed, disturbances of the sleep-wake cycle, which are reflected in poor activity rhythms, are hypothesized to be a major factor contributing to institutionalization.[28,29] Whereas the association between circadian activity rhythms and illness is fairly strong, evidence for an association between disrupted activity rhythms and ensuing death is limited.[24,32,33] Tranah and associates[35] examined whether circadian activity rhythms are associated with mortality in 3027 community-dwelling women from the SOF cohort (mean age, 84 ± 4 years; range, 77 to 99 years). Activity data were collected for a minimum of three 24-hour periods, and vital status, with cause of death determined from death certificates, was prospectively ascertained. Over an average follow-up period of 4.1 years, a total of 444 deaths (15%) occurred. Those women with lower peak (amplitude) and mean (mesor) levels of activity and less

robust rhythms had the shortest overall survival, and all-cause mortality risk increased from the highest to the lowest quartile of amplitude. Relative to that for women in the highest quartiles, approximate doubling of the adjusted higher all-cause mortality risk was observed for those in the lowest quartiles of amplitude (hazard ratio [HR], 2.18; 95% confidence interval [CI], 1.63 to 2.92), mesor (HR, 1.71; 95% CI, 1.29 to 2.27), and rhythm robustness (HR, 1.97; 95% CI, 1.50 to 2.60) after multivariate adjustment for age, clinic site, race, body mass index (BMI), cognitive function, sleep performance, exercise, instrumental activities of daily living impairments, depression, medications, alcohol, smoking, self-reported health status, marital status, and comorbid illness. An increased risk of death from atherosclerotic disease also was observed for the lowest quartiles of amplitude (HR, 1.81; 95% CI, 1.13 to 2.90) and mesor (HR, 1.61; 95% CI, 1.02 to 2.54) activity levels as well as the lowest quartile of rhythm robustness (HR, 2.31; 95% CI, 1.45 to 3.68). The relationships with atherosclerotic mortality were largely driven by associations between circadian activity rhythms and either coronary heart disease or stroke-related mortality, which accounted for 34% and 32% of atherosclerotic disease–related deaths, respectively. Increased risk of "other-cause" mortality was observed in the lowest quartiles of amplitude (HR, 3.11; 95% CI, 1.93 to 5.00), mesor (HR, 2.09; 95% CI, 1.33 to 3.28), and rhythm robustness (HR, 2.17; 95% CI, 1.43 to 3.31), compared with the highest quartiles. Further analysis excluding the two most common causes of "other cause" mortality (pulmonary and cognitive causes) produced similar results, suggesting that these two causes of death do not explain this association.

In this study, acrophase was examined in terms of the deviation from the population mean. The investigators identified three categories based on having a peak time of more than 1.5 standard deviations (SDs) above and below the mean for the study population. Phase-advanced participants were defined as having an acrophase of earlier than 12:50 PM (–1.5 SD from the mean), and phase-delayed participants were defined as having an acrophase of later than 4:33 PM (+1.5 SD from the mean). The mean peak range was 12:50 PM to 4:33 PM. Acrophase deviation was not associated with all-cause or "other-cause" mortality. However, older women with a delayed acrophase were at significantly increased risk for cancer-related (HR, 2.09; 95% CI, 1.04 to 4.22) and stroke-related (HR, 2.64; 95% CI, 1.11 to 6.30) death when compared with those in the mean peak range.

Paudel and colleagues[36] examined whether circadian activity rhythms are associated with mortality in 2964 community-dwelling men from the MrOS cohort (mean age, 76 ± 6 years). Activity data were collected for a minimum of five 24-hour periods, and vital status, with cause of death confirmed with death certificates and additionally with medical records when available, was prospectively ascertained. Over an average follow-up perod of 3.5 years, 233 deaths (8%) occurred. After adjustment for multiple potential confounders (age, age^2 [modeling non-linear associations], race, alcohol use, smoking status, caffeine use, education, self-reported health status, IADL impairments, cognitive impairment, depression, and number of medical conditions), men in the lowest quintile of rhythm robustness had a 57% higher mortality rate (HR, 1.57; 95% CI, 1.03 to 2.39) compared with men in the highest quintile. This association among those in the lowest quintile of rhythm robustness was even stronger with increased risk of CVD-related mortality (HR, 2.32; 95% CI, 1.04 to 5.22). Additionally, men in the latest quintile of acrophase (later than 3:09 PM) had a 2.8-fold higher rate of CVD-related mortality (HR, 2.84; 95% CI, 1.29 to 6.24) when compared with those with mean acrophase timing (1:58 PM to 2:32 PM). In general, amplitude, mesor and rhythm robustness steadily declined with advancing age. However, there was no association between amplitude and mesor and risk of CVD-related, cancer-related or other causes of mortality.

Dementia and Mild Cognitive Impairment

Circadian rhythms play an important role in rhythms associated with cognitive processing including alertness, learning, and memory. During waking, rhythms in synaptic plasticity,[48,49] focused attention, and behavioral flexibility have been confirmed.[50,51] Little is known concerning the causes of age-related changes in circadian activity patterns in nondemented older adults and the subsequent effects of these changes on health and well-being. Tranah and associates[37] examined whether circadian activity rhythms were associated with incident MCI and dementia in 1282 women from the SOF cohort, using both actigraphy data and findings on a follow-up visit that included an expanded cognitive assessment and adjudication of presence of MCI-dementia by an expert panel. After a mean follow-up period of 4.9 years from the collection of actigraphy data, 195 (15%) women were found to have dementia and 302 (24%) had MCI. An approximately 50% adjusted higher likelihood of developing dementia or MCI relative to that in the cohort without any dementia or MCI was observed for subjects in the lowest quartiles (n = 320) of amplitude (odds ratio [OR], 1.54; 95% CI, 1.07 to 2.21) and rhythm robustness (OR, 1.55; 95% CI, 1.08 to 2.22) when compared with those in the highest quartiles (n = 322).

Acrophase was examined in terms of the deviation from the population mean. Three categories were defined on the basis of a peak time of more than 1 standard deviation (SD) above and below the population mean for the study population. Phase-advanced participants were defined as having an acrophase of earlier than 1:34 PM (–1 SD from the mean), and phase-delayed participants were defined as having an acrophase of later than 3:51 PM (+1 SD from the mean). Acrophase deviation was associated with increased odds of developing dementia or MCI. A delayed acrophase, after 3:51 PM (n = 173, 13%), was associated with a significant increase in odds of developing dementia or MCI (OR, 1.83; 95% CI, 1.29 to 2.61) when compared with the mean peak range of 1:34 PM to 3:51 PM (n = 927, 73%). An advanced acrophase, before 1:34 PM (n = 182, 14%), was associated with an elevated but nonsignificant odds of developing dementia or MCI (OR, 1.36; 95% CI, 0.96 to 1.92) when compared with the mean peak range. All analyses were multivariate-adjusted for age, clinic site, race, education, BMI, walking for exercise, functional status, depression, benzodiazepine and antidepressant use, sleep medication use, alcohol use, caffeine intake, smoking, self-reported health status, and previous medical conditions. Results for all models were unchanged after further adjustment for sleep efficiency.

Cardiovascular Disease

Previous studies are consistent in suggesting that disruptions in circadian activity rhythms are associated with

cardiovascular death, such as increased risk of death from CVD, atherosclerotic disease, or stroke,[35,37] as well as short-term consequences in metabolic measures.[52] Although disrupted circadian activity rhythms and CVD events both are more common in the elderly population, little is known regarding associations between circadian activity rhythm variables and risk of CVD-related events in older adults. Paudel and associates[38] examined whether circadian activity rhythms are associated with risk of cardiovascular events in 2968 men from the MrOS cohort. Over an average follow-up period of 4.0 years, 490 cardiovascular events (16.5%) occurred. Overall, increased risk of such events was observed among participants with reduced amplitude (HR, 1.31; 95% CI, 1.01 to 1.71 for quartile 2 versus quartile 4) and greater minimum (nadir) (HR, 1.33; 95% CI, 1.01 to 1.73). In addition, reduced amplitude (HR, 1.36; 95% CI, 1.00 to 1.86) and greater minimum activity counts (HR, 1.39; 95% CI, 1.02 to 1.91) were associated with increased risk for coronary heart disease–related events. Reduced rhythm robustness (HR, 2.88; 95% CI, 1.41 to 5.87) also was associated with an increased risk of peripheral vascular disease events when the lowest quartile was compared with the highest quartile. Men in the lowest quartile of β (indicating less constant levels of daytime activity) had a 78% increased risk of incident stroke compared with men in the highest quartile in multivariable-adjusted models (HR, 1.78; 95% CI, 1.01 to 3.14). Additional analyses excluding men who reported having had prevalent CVD or models adjusting for traditional cardiovascular risk factors such as diabetes, blood pressure, total cholesterol, and HDL did not affect the statistical significance of results.

Depression

A considerable body of evidence supports a role for circadian rhythm disturbances in the pathophysiology of depression.[53-55] Circadian rhythm disturbances have been observed in unipolar depression, bipolar depression, and seasonal affective disorder.[56,57] In unipolar depression, multiple rhythms are disrupted including those governing secretion of prolactin, cortisol, growth hormone, and melatonin, along with body temperature and the sleep-wake cycle.[58] At the neuroanatomic level, depressed patients have been shown to have different patterns of regional brain glucose metabolism throughout the day compared with normal patients, and the expression of enzymes that control catecholamine metabolism may be regulated by clock genes.[59,60] Although aging promotes disruption in circadian rhythms, the relationship between depression and circadian rhythm disruption in older adults remains largely unexplored. Maglione and associates[39] examined the relationship between depressive symptoms assessed with the Geriatric Depression Scale (GDS) and circadian activity rhythms in older adults among 3020 women (mean age, 83.6 ± 3.8 years) from the SOF cohort. Higher levels of depressive symptoms on the GDS were associated with greater desynchronization (decreased amplitude, rhythm robustness, and mesor) of circadian activity rhythms in linear regression models. In addition, greater levels of depressive symptoms were associated with a later average time, becoming active in the morning but not with acrophase. When categorizing participants by GDS score as "normal" (0 to 2) (referent group, $n = 1961$), "some depressive symptoms" (3 to 5) ($n = 704$), or "depressed" (6 or higher) ($n = 355$), Maglione's group[39] also assessed the odds of falling into the lowest quartile for activity rhythm variables. Risk of falling into the lowest quartile of amplitude was elevated both for participants defined as having "some depressive symptoms" (OR, 1.32; 95% CI, 1.04 to 1.67) and for those designated "depressed" (OR, 1.51; 95% CI, 1.11 to 2.05). In addition, the likelihood of falling into the lowest quartile of mesor was elevated both for the "some depressive symptoms" group (OR,1.55; 95% CI, 1.24 to 1.94) and for the "depressed" group (OR, 1.62; 95% CI, 1.20 to 2.18). Additional adjustment for total sleep time and sleep efficiency did not affect the statistical significance of results.

Luik and coworkers[14] examined activity rhythm stability and fragmentation in 1734 middle-aged and elderly participants (mean age, 62 ± 9.4 years) from the Rotterdam Study cohort. Activity data were collected for a minimum of four 24-hour periods and investigated age, sociodemographics, mental health, lifestyle (coffee use, alcohol use, and smoking), and sleep characteristics as determinants of activity rhythms. The results of this study indicate that older age is associated with a more stable 24-hour activity rhythm but also with a more fragmented distribution of periods of activity and inactivity. Both BMI and smoking were associated with less stable and more fragmented activity rhythms. In concordance with the results from the SOF cohort,[39] more depressive symptoms were related to less stable and more fragmented activity rhythms.

PATHOPHYSIOLOGIC MECHANISMS

In interpreting the foregoing results, two possibilities should be considered. First, activity rhythms may directly influence morbidity and mortality in older adults independent of other features of aging. In support of this association, emerging animal and human data show the existence of both central and peripheral (e.g., in the liver, pancreas, and other organs) circadian rhythms, with evidence that misalignment of internal rhythms may be a predisposing factor for impaired glucose tolerance and alterations in immunologic and inflammatory processes. Previous studies suggested that some but not all peripheral circadian oscillators exhibit age-related changes in rhythmicity[61] and that some of these tissues retained the capacity to oscillate but were not being appropriately driven in vivo (e.g., by physical activity or feeding).[62] The presence of arrhythmic peripheral tissues may be due to weakened behavioral and physiologic rhythms that provide less effective signals to the peripheral oscillators.[54] The change in phase relationships of behavioral and physiologicl rhythms, therefore, may not be due to age-related changes in the entrained phase of the SCN itself but rather may arise from age-related alterations in other rhythmic components of the circadian system.[39] Evidence of an age-related phase advance is clear from studies involving body temperature, sleep-wake cycle, melatonin, and cortisol,[37] in which a phase difference of approximately 1 hour typically is found between young and old individuals. Age-related phase advances also are found in the circadian rhythms of blood pressure, levels of iron and magnesium, and numbers of neutrophils and lymphocytes.[38] Acrophase deviations from the mean may represent an altered phase relationship between the circadian activity rhythm and the light-dark cycle.

For example, the circadian timing system most likely affects memory, cognitive function, and behavior through a variety of neuroanatomic and neurophysiologic mechanisms.[63] The

circadian contribution to cognition also may arise from the synchronized activities of an integrated network of clocks in the brain under the direction of the SCN pacemaker.[64] It is also possible that circadian activity rhythms are biomarkers of advanced physiologic aging that provide additional risk over and beyond that of traditional covariates, but which may have no direct causal association with dementia or MCI. Sleep and rhythm disturbances are common in many neurodegenerative diseases including Alzheimer disease,[65-67] dementia,[68] and Lewy body disease.[69] The major sleep complaints associated with neurodegenerative diseases include insomnia, hypersomnia, parasomnia, excessive nocturnal motor activity, sleep apnea, and sleep-wake rhythm disturbances.[70] Sleep is disturbed early in the neurodegenerative process, and sleep disturbances are observed in the presence of MCI.[71,72] Furthermore, it has been suggested that sleep disturbance increases with severity of the neurodegeneration.[73] Although neurodegenerative diseases are believed to be proteinopathies resulting from excessive protein misfolding and intracellular protein aggregation, the role of sleep and rhythm disturbances in the neurodegenerative process has been largely unexplored. The accumulation of amyloid-beta in the brain extracellular space is a critical factor in the pathogenesis of Alzheimer disease, and both the sleep-wake cycle and orexin have been shown to play a role in regulating amyloid-beta dynamics.[74] Orexins and their receptors are involved in a number of central[75] and peripheral[76] functions and play an important role in maintaining wakefulness by preventing unwanted transitions into sleep, as seen in narcolepsy.[77]

A second possible mechanism to consider in interpreting the evidence to date is that circadian activity rhythms are biomarkers of advanced physiologic aging that provide additional risk over and beyond that of traditional covariates, but which may have no direct causal association with mortality. In this instance, available data may provide evidence that circadian activity rhythms are markers for a greater risk for disease or death not measured by conventional markers. Although the biologic mechanisms underlying the associations between disrupted activity rhythms and increased risk of cardiovascular events are unknown, the results of the study by Paudel and associates[38] suggest that the associations are independent of age, race, instrumental activities of daily living impairments, smoking status, cognitive function, use of antidepressants, walking for exercise, and history of CVD, stroke, or peripheral vascular disease. Although it generally is perceived that circadian rhythm disruptions precede CVD-related events, it is plausible that comorbid CVD and/or other conditions such as diabetes worsen circadian rhythm disruptions through their debilitating impact on sleep-wake activity. It also is possible that circadian rhythms and conditions such as diabetes share common causative factors.

IMPLICATIONS

Circadian rhythms and sleep are influenced by circadian and homeostatic processes. The results of recent work in large population-based cohorts of elderly participants suggest that weak and shifted circadian activity rhythms are associated with increased risks of mortality, MCI-dementia, and CVD. Depression also appears to be associated with more desynchronized circadian activity rhythms. If these associations are confirmed in additional cohorts, future studies should examine whether interventions such as physical activity and bright light exposure that influence activity rhythms will reduce morbidity and mortality in the elderly population.

For example, identifying new approaches for interventions that slow the age-related decline in cognitive function will potentially benefit the health and well-being of elderly persons before the onset of cognitive impairment. Sleep interventions and light therapy in patients or rodent models have been shown to demonstrate acute beneficial effects.[78,79] Exposure to ambient light also is known to influence cognition and affective state,[49,60,80-82] and bright light has phase-shifting properties that have been shown to improve cognitive performance.[83] In addition, several studies examining the efficacy of bright light for treatment of depressive symptoms in older adults without dementia have yielded promising results,[84,85] although at least one has found no benefit.[86] In older adults, depression also has proved difficult to treat, with a high rate of refractoriness to first-line pharmacotherapies.[86] Meanwhile, several studies suggest that bright light therapy, a chronobiologic modality designed to correct circadian rhythm disruption, may be an effective treatment for depression in older adults.[84,85] Overall, further elucidation of the relationship between circadian rhythm disruption and age-related diseases is critical and may lay the groundwork for the development of better treatment strategies.

CLINICAL PEARLS

- Circadian rhythms are intrinsic 24-hour physiologic cycles that are critically involved in control of sleep-wake cycles and a variety of other physiologic processes.
- With aging, circadian rhythms tend to become weaker, and the timing of peak activity also may change.
- Twenty-four-hour patterns of activity, which can be estimated using wrist actigraphy to measure limb movements, serve as a marker for intrinsic circadian rhythms.
- Older women and men with weaker and time-shifted circadian activity rhythms are at greater risk of death, MCI and dementia, and CVD.
- Interventions that regulate circadian rhythms may improve health care outcomes in older adults.

SUMMARY

In mammals, circadian rhythms are driven by the primary circadian pacemaker, the suprachiasmatic nucleus, located in the hypothalamus. These 24-hour rhythms are entrained by light and control a variety of physiologic processes including release of certain hormones, blood pressure and heart rate, bone remodeling, and sleep and activity cycles (among others). Rhythms of rest and activity, which can be approximated by measuring limb movement using actigraphy, are correlated with intrinsic circadian rhythms. In older adults, disturbances in circadian rest-activity rhythms are associated with increased prevalence of CVD, increased risk of death, and higher frequency of cognitive impairment and dementia. Rest-activity rhythms also are disturbed in depressed older adults. It is unclear whether health is directly affected by deterioration in circadian rhythms, or whether changes in the strength and

timing of rhythms constitute a biomarker for other changes that occur with aging. Future studies are warranted to test whether interventions to improve circadian rhythms, such as physical activity or light exposure regimens, may improve health in older adults.

ACKNOWLEDGMENTS

Work on which this chapter is based was supported by the following grants: National Institute on Aging (NIA) AG05407, AR35582, AG05394, AR35584, AR35583, AR46238, AG005407, AG08415, AG027576-22, AG005394-22A1, AG027574-22A1, AG030474, and AG08415, and National Heart, Lung, and Blood Institute (NHLBI) HL071194, HL070848, HL070847, HL070842, HL070841, HL070837, HL070838, and HL070839.

Selected Readings

Luik AI, Zuurbier LA, Hofman A, et al. Stability and fragmentation of the activity rhythm across the sleep-wake cycle: the importance of age, lifestyle, and mental health. *Chronobiol Int* 2013;**30**(10):1223–30.

Maglione JE, Ancoli-Israel S, Peters KW, et al. Depressive symptoms and circadian activity rhythm disturbances in community-dwelling older women. *Am J Geriatr Psychiatry* 2014;**22**(4):349–61.

Maglione JE, Nievergelt CM, Parimi N, et al. Associations of PER3 and RORA circadian gene polymorphisms and depressive symptoms in older adults. *Am J Geriatr Psychiatry* 2015;**23**:1075–87.

Paudel ML, Taylor BC, Ancoli-Israel S, et al. Rest/activity rhythms and cardiovascular disease in older men. *Chronobiol Int* 2011;**28**(3):258–66.

Paudel ML, Taylor BC, Ancoli-Israel S, et al. Rest/activity rhythms and mortality rates in older men: MrOS Sleep Study. *Chronobiol Int* 2010;**27**(2):363–77.

Paudel ML, Taylor BC, Ancoli-Israel S, et al. Sleep disturbances and risk of depression in older men. *Sleep* 2013;**36**(7):1033–40.

Sherman SM, Mumford JA, Schnyer DM. Hippocampal activity mediates the relationship between circadian activity rhythms and memory in older adults. *Neuropsychologia* 2015;**75**:617–25.

Tranah GJ, Blackwell T, Ancoli-Israel S, et al. Circadian activity rhythms and mortality: the study of osteoporotic fractures. *J Am Geriatr Soc* 2010;**58**(2):282–91.

Tranah GJ, Blackwell T, Stone KL, et al. Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. *Ann Neurol* 2011;**70**(5):722–32.

A complete reference list can be found online at ExpertConsult.com.

Section
20

Sleep in Women

Chapter
155

Sex Differences and Menstrual-Related Changes in Sleep and Circadian Rhythms

Fiona C. Baker; Louise M. O'Brien

Chapter Highlights

- The risk for developing insomnia is greater in women than in men, starting at the onset of menses and persisting across the lifespan. Women also report more sleep difficulties, a need for more sleep, and a longer sleep duration. In contrast, objective sleep recordings with polysomnography or actigraphy show that women have a shorter sleep onset latency, better sleep efficiency, more N3 sleep, and more slow wave activity than men. The reasons for the different findings with subjective and objective measures is unclear. Possibly, polysomnography measures are appropriate measures of subjective sleep quality in men but not in women.
- Although some sex differences in sleep behavior are apparent in infants and young children, differences become more prominent during adolescence. On the basis of actigraphy data, boys sleep less and have a less efficient sleep than girls. The timing of age-related changes in sleep also differs, with later shift to eveningness as well as later adolescent decline in slow wave activity in boys than in girls, possibly related to a later onset of puberty in boys.
- In women, sleep is affected by the menstrual cycle. In the postovulatory luteal phase, amount of N3 sleep remains constant but that of rapid eye movement sleep is less, and spindle frequency activity increases. Amplitude of the body temperature rhythm, but not the melatonin rhythm, is reduced in the luteal phase compared with that in the follicular phase.
- Menstrual cycle–associated disorders affect sleep in different ways. Women with polycystic ovary syndrome are at increased risk for sleep-disordered breathing. Women with severe dysmenorrhea experience more wakefulness in association with their menstrual pain. Women with severe premenstrual syndrome exhibit differences in sleep and circadian rhythms that are evident in symptomatic and asymptomatic phases of the menstrual cycle.

SEX DIFFERENCES IN SLEEP FROM INFANCY TO ADULTHOOD

Sexual dimorphism describes morphologic differences between the sexes, although it also refers to any biologic process that differs between men and women. Extensive evidence from diverse species—fruit flies to humans—demonstrates sexual dimorphism in structure, function, and regulation of the brain. Differences between male and female brains are believed to result from the actions of gonadal secretions during a critical

period of brain development, although gonadal secretion is not the only mechanism, because gene expression in neuronal cells also plays a role in sexual dimorphism of the brain before gonadal secretion occurs. Animal studies show evidence of sex differences in sleep and circadian regulation, some of which are hormone-dependent.[1] However, research into sex differences in sleep and circadian rhythm regulation lags behind other studies of sex differences in both humans and animal models. Also, emerging evidence suggests that sex differences may vary depending on stage of reproductive development. Presented next is an overview of the current evidence regarding sex differences in sleep and circadian rhythms, from infancy to adulthood. Major findings are summarized in Table 155-1.

Table 155-1 Sex Differences in Sleep throughout the Life Cycle

Subjective Reports	Objective Measures
Infants	
Poor sleepers are more likely to be boys	*Video*: Girls have a longer sleep period and more quiet sleep than boys *Polysomnography*: Boys exhibit more infraslow activity (<0.5 Hz) during sleep than girls Boys have increased arousability compared with girls
Children	
Girls sleep longer than boys	*Actigraphy*: Girls sleep longer than boys *Polysomnography*: Girls have a higher sleep efficiency than boys Girls experience more arousals than boys
Adolescents	
Girls take longer to get to sleep and sleep longer than boys Girls are more likely to be extreme sleepers than boys (i.e., sleep periods of <6 hours and >10 hours)	*Actigraphy*: Girls have more efficient sleep and fewer awakenings than boys *Polysomnography*: Girls begin the decline in NREM delta power earlier than boys Depressed boys experience a greater degree of sleep disturbance than do depressed girls
Adults	
Women report more sleep difficulties than men Women have a longer sleep duration than men	*Actigraphy*: Women have better sleep quality and sleep more than men *Polysomnography*: Women have a shorter sleep latency, better sleep efficiency, and more sleep than men Women exhibit higher delta power activity during NREM sleep

NREM, Non–rapid eye movement.

Infancy

In humans, circadian rhythms exhibit a cyclic tendency, with a periodicity of approximately 24 hours. Individuals undergo a gradual development in rhythmicity after birth, strengthening from interaction with a 24-hour light-dark cycle, social cues, and other environmental conditions. In healthy newborns, robust circadian rhythms usually are entrained by 2 to 3 months of age. The development of sleep states and circadian rhythms in neonates has been reported in multiple studies, although few have reported on comparisons between the sexes. In those that have compared male and female infants, findings are contradictory. Some studies using nighttime video recordings of infants find no differences in sleep-wake state organization,[2] but others find that female infants have higher proportions of quiet sleep and longer sleep periods than male infants.[3,4] Of interest, preterm infant boys may have significantly less active sleep, more wake time, and more diffuse states than infant girls.[5] On electroencephalogram (EEG) recordings obtained during the first few months of life, however, no sex differences in sleep states were found,[6] although healthy newborn boys had more infraslow activity (less than 0.5 Hz) during sleep than newborn girls,[7] and boys in the first month of life had increased arousability in quiet sleep compared with girls, with this difference no longer evident by the age of 2 to 3 months.[8]

Taken together, these studies support the hypothesis that infant girls have a more mature central nervous system (CNS) than infant boys. Infant girls also have more mature respiratory systems than infant boys during the first 6 months of life—possibly related to accelerated CNS maturation.[9] In healthy infants, the ratio of nighttime sleep to total sleep becomes progressively greater with age and demonstrates a more organized pattern of sleep, particularly in female infants. The time course of rapid eye movement (REM) sleep development corresponds with brain maturation.[10] Higher proportions of REM sleep, together with a maturational lag in REM sleep, have been reported in infants at risk for sudden infant death syndrome,[11] for which a clear male predominance has been demonstrated.

Childhood

In many studies, the term *childhood* is not clearly defined, and this designation generally is used to refer to the period from the end of infancy to the onset of adolescence, with inherent variability in timing. This lack of standardization clearly contributes to inconsistent findings. Even so, most studies of sleep in children have not investigated potential sex differences, and when researchers do report results by sex, findings often are conflicting.

Survey Data

Worldwide, several cross-sectional survey-based studies of hundreds of school-aged children have found that girls report sleeping longer than boys.[12,13] Other large, survey-based longitudinal studies, however, have failed to find any sex differences in sleep parameters.[14] Lack of consistency between findings may be due to factors such as cultural or racial differences and the limitations of using questionnaires to investigate sleep. Although questionnaire-based sleep measures show few sex differences, the evidence does point to a higher risk for sleep problems in girls[15] and underscores interactions among sex, sleep problems, and hypothalamic-pituitary-adrenocortical axis responses to stress. Compared to boys

without parent-reported sleep problems, boys with sleep problems had lower diurnal levels of cortisol and lower cortisol levels in response to stress, whereas girls with sleep problems had similar diurnal levels of cortisol but higher levels of cortisol in response to stress.[16]

Actigraphy and Polysomnographic Data

Data from actigraphy recordings have been more consistent in showing sex differences in sleep of school-aged children[17,18] although data remain conflicting for preschool-aged children, with sex differences varying by age even within this narrow age range.[19-21] Several studies have found that girls sleep more and spend more time in motionless sleep than boys.[22,23]

Although many reports of sleep in children include polysomnography (PSG) data, few studies compare the sexes, and findings vary by age. A large community-based study of 542 children, 3.2 to 8.6 years of age,[24] found that younger girls (ages 3 to 5 years) had a higher sleep efficiency, less wake, and more stage 3 (N3) sleep than boys of similar age, although effect sizes were small. Among older children (6 to 8.6 years), girls had a higher proportion of stage 1 (N1) sleep than boys. Another large study of white children between 1 and 18 years of age found that girls had more arousals but no other differences in sleep compared with boys.[25] Measures of the cyclic alternating pattern, which refers to EEG activity characterized by sequences of transient episodes that differ from background basal EEG, suggest that girls have more non–rapid eye movement (NREM) instability, even when standard EEG criteria show no sex differences.[26] However, another study using similar methodology did not find these sex differences.[27]

These results suggest that young girls have a longer sleep duration but only marginal differences in sleep architecture compared with young boys. Sleep characteristics begin to diverge more between boys and girls after the onset of adolescence.

Adolescence

Survey Data

Adolescence is marked by dramatic biologic and social changes that can affect health and behavior, including sleep.[28] Using questionnaire-based methodologies, several studies have found sex differences in sleep behavior. Although sleep-time preference shifts toward eveningness across adolescence in both boys and girls, girls reach the peak in eveningness at an earlier age than boys,[29,30] and boys tend to go to bed later than girls.[31] A meta-analysis of data for more than 90,000 adolescents (9 to 18 years old) from 20 countries found that on average, girls reported a longer time in bed than boys on both school and nonschool days.[32] This finding matches reports of a longer ideal sleep duration in adolescent girls than in boys.[30] A large study in the United States, however, found that adolescent girls reported a slightly shorter—by 10 minutes—sleep duration than boys; this difference reversed after the age of 19 years, when young women reported a longer sleep duration—of more than 30 minutes—than young men.[33] Adolescent girls in this study sample also were more likely than boys to have a short sleep duration (less than 6 hours), and this pattern also reversed in emerging and early adulthood.[33]

Results may vary between surveys depending on whether participants are asked about sleep duration, time in bed, or actual clock times of sleep and wake, especially during adolescence, when a marked change in these variables is seen. Finally, a recent twin study reported a sex-age interaction effect in the contribution of nonshared environmental factors (such as playing video games or participation in evening sports) to variance in sleep duration, being a major factor in both young and older boys as well as older girls, but of lesser importance for younger adolescent girls.[34]

In investigating changes in sleep across adolescence, it is important to consider not only age but also pubertal development. The onset of menses is associated with an increased risk for insomnia in girls,[35] a sex difference that persists across the lifespan. This increased risk for insomnia in girls after menarche corresponds with an increased risk for depression,[36] which is a risk factor for insomnia; however, adjusting for depression did not reduce the association between menses onset and insomnia in girls.[35] Clearly the relationship among insomnia, depression, and pubertal development is complex, with several social and biologic changes occurring around the onset of menses that also may increase risk for both depression and insomnia.

Actigraphy and Polysomnographic Data

In studies using actigraphy, adolescent boys have been found to sleep less, have less efficient sleep, less motionless sleep, and awaken earlier than girls, probably related to greater movement activity during the night in boys.[22,37,38] Of interest, measures of total sleep time by actigraphy and PSG are in close agreement for adolescent girls, but actigraphy may underestimate total sleep time in boys, possibly because of the differential pattern of nocturnal movement between girls and boys.

In adolescence, a well-described EEG change in sleep is a steep decline in delta slow wave activity (SWA), and data from cross-sectional studies indicate that delta activity declines by approximately 50% between the ages of 10 and 20 years.[28,39,40] Longitudinal data from cohorts of 9- and 12- year-olds accumulated over a 4-year period have shown that girls begin the steep adolescent decline in SWA approximately 1.2 ± 0.25 years earlier than boys, which has been hypothesized to reflect an earlier onset of adolescent synaptic pruning in girls.[41,42] It has long been speculated that sexual maturation and declining SWA may be causally related, because the decline is significantly correlated with increasing Tanner stage and is independent of age.[43]

Recent longitudinal data for adolescents spanning ages 9 to 18 years indicated that the timing of the delta decline is linked to timing of pubertal maturation, even with introduction of controls for sex differences.[42] Indeed, sex differences and the relation to the timing of puberty jointly explained two thirds of the between-subject variance in the timing of the delta decline. The study investigators suggested that these data show a temporal relationship between puberty and an electrophysiologic marker of adolescent brain development. Additional evidence includes sex-specific differences in SWA that vary according to EEG scalp topography in adolescence, which could reflect underlying sex differences in cortical structure and function. Adolescent girls exhibit higher SWA over bilateral cortical areas related to language function compared with age-matched boys, and adolescent boys demonstrate increased SWA over the right prefrontal cortex, which is associated with spatial abilities.[44] Feinberg[45] suggests

that the delta decline during adolescence is a component of widespread brain reorganization, of which other manifestations include the emergence of adult cognitive capacity. It is possible that abnormalities in these maturational brain processes may give rise to abnormal circuits that cause psychiatric illnesses.

Major depressive disorder (MDD) is associated with altered sleep architecture in adolescents, with effects dependent on sex.[46] Depressed adolescent boys experience the greatest degree of sleep disturbance (more Stage 1 sleep, less SWS), whereas sleep in depressed girls does not differ significantly from that in healthy girls. However, temporal coherence of sleep EEG rhythms (a measure of synchronization of EEG frequencies of the same periodicity) is significantly lower in adolescents with MDD, with the lowest coherence in affected girls, even in those younger than 13 years of age.[47] In addition, children and adolescents with MDD, as well as those at high risk for MDD, exhibit lower spindle frequency activity during sleep, with a more prominent deficit in girls.[48] A striking finding is that almost one half of adolescents with the most abnormal coherence values showed either symptoms of depression at the time or met diagnostic criteria within 2 years.[49] Low temporal coherence represents a more chaotic organization of sleep EEG rhythms, suggesting that early-onset depression is associated with a reduction in synchronization of sleep EEG rhythms that shows a differential maturational course in boys and girls.[47] Girls with MDD demonstrate dampened amplitude of circadian rest-activity cycles, even when they have not yet reached developmental maturity.[50] By contrast, dampened circadian amplitude is not evident in depressed boys until adolescence. Together, these findings point to a differential developmental influence on sleep in early depression that is sex-dependent.[46]

Adulthood

Survey Data

In comparison with men, women report needing more sleep, spending more time in bed, and a longer sleep duration.[51,52] Women prefer to go to bed earlier[53] and are more likely to be morning types.[54] Women also report more sleep difficulties[52] and are more frequent users of sedative-hypnotic drugs[55] than men across adulthood. Women are at increased risk for insomnia (female-to-male ratio of 1.41:1) that emerges after menarche[35] and is maximal in the elderly age group.[56] Affected women are more likely to report two or more insomnia symptoms, whereas men typically report one symptom.[57] The sex difference in insomnia persists even after underlying psychiatric disorders are taken into account.[56] In a large study of primarily retired adults followed for 19 years, increased mortality was associated with nighttime sleep duration of more than 9 hours in women, but with daytime napping of more than 30 minutes in men.[58]

Actigraphy and Polysomnographic Data

Actigraph assessment of sleep has shown that women exhibit better objective sleep quality with higher sleep efficiency index and lower frequency of transitions between sleep and wakefulness than men.[59,60] Women tend to sleep more than men and have shorter sleep onset latency. Yet compared with older women, older men have a greater tendency for fragmented sleep.[61] Regardless of sex, age is a factor in sleep fragmentation because adults have a strong decline in actual sleep time and sleep efficiency, as well as increased sleep latency as they age.

In parallel with actigraphy, PSG data demonstrate better objective sleep quality, with shorter sleep onset latency, better sleep efficiency, less N1 sleep, more N3 sleep, and less time awake in women than in men.[62-64,65] Men also exhibit more SWA during NREM sleep than women,[66-68] although men and women show similar rates of decline in power density across the night.[68] Finally, men show a more rapid decline in N3 with age,[63,64,69] although no change in the amount of SWA across midlife years is seen for either men or women.[67] The discordance between perceived and objective sleep measures may possibly reflect a greater need for sleep in women that is not being met,[1,60] or objective measures of sleep quality such as total sleep time or sleep fragmentation may be appropriate measures of good subjective sleep quality in men but not in women.[70] Of note, subjective and objective assessments of sleep tap different constructs and should not necessarily be expected to align.

Sex differences in circadian rhythmicity also are apparent, with women having a shorter intrinsic circadian period, correlating with an earlier timing of core body temperature nadir and melatonin peak relative to sleep.[71] Accordingly, women tend to sleep at a later biologic time than men, which may contribute to a higher level of sleep-maintenance insomnia in women.

Under challenge conditions, further sex differences in sleep emerge.[72] For example, compared with young men, young women have a greater delta response to the stress of sleep deprivation,[73] and sex differences in sleep normally seen in healthy adults become even more magnified in comparing depressed men with depressed women. Men with MDD show less SWA than women with MDD or men and women without MDD, whereas depressed women show lower temporal coherence than control subjects.[72] Women, but not men, with MDD demonstrate significantly greater amounts of SWA in multiple cortical areas relative to control subjects.[74] These disturbances in sleep microarchitecture that are more apparent in women with MDD may be relevant to understanding the increased risk of MDD observed in women.

Women exhibit a greater increase in sleep spindle activity after zolpidem and a larger increase in SWA after gaboxadol administration compared with men.[75] It is hypothesized that these sex-dependent effects of hypnotics on the sleep EEG may be related to sex differences in pharmacokinetics, $GABA_A$ receptor subtypes and activation, or interactions between neurosteroids and $GABA_A$ receptors.[75] Women also have a more sustained response to zolpidem, requiring a lower dose for equivalent efficacy than that in men; the U.S. Food and Drug Administration stipulates that the recommended dose for zolpidem in women should be one-half the maximum dosage approved for men.[76]

Whereas women are at greater risk for developing insomnia and restless legs syndrome than men,[52,77] they have a lower risk for obstructive sleep apnea (OSA), with a male-to-female ratio of between 2:1 and 4:1, although up to 2% of women overall may have OSA.[78] OSA is underdiagnosed in women, possibly because they may not present with the classic symptoms (snoring, gasping, or apneic events) and may have additional nonspecific symptoms that make a diagnosis of OSA more difficult.[79] Sex differences in OSA can be explained by several mechanisms, including differences in obesity and fat

distribution, upper airway anatomy and function, and hormone status.[79] The prevalence of sleep-disordered breathing (SDB) in women varies across their reproductive life cycle, with higher levels seen in pregnant and postmenopausal women, as discussed in subsequent chapters (see Chapters 156 to 159).

Observed morphologic variation between men and women in the suprachiasmatic nucleus (SCN) and other hypothalamic structures, along with laboratory and clinical data on sleep and circadian rhythms, provides a strong neurophysiologic and neuroanatomic basis for sex differences in sleep regulation.[1] Although the evidence is strongest for adulthood, emerging data suggest a differential maturational time course in boys and girls that may influence circadian rhythms and sleep-wake cycles. Indeed, the roots of such maturational influences may be evident very early in development. Further longitudinal studies are necessary to determine the time period at which these sex differences emerge and how they are shaped by neuroendocrine function, gene transcription factors, environment, and maturation of sex steroids with menstrual cycle hormonal fluctuations.

In considering potential sex differences in sleep, it is important to be aware of the potential impact on sleep of the menstrual cycle, as described next, as well as other stages of a woman's life cycle, as discussed in Chapters 156 to 159.

THE MENSTRUAL CYCLE AND EFFECTS OF OVARIAN HORMONES ON SLEEP AND CIRCADIAN RHYTHMS

Female reproductive hormones, specifically estrogen and progesterone, not only regulate reproductive tissue function during the menstrual cycle but also have secondary actions in the CNS to influence other physiologic processes, such as sleep and circadian rhythms.

Conventionally, in a normal menstrual cycle of 28 days, day 1 is identified as the first day of bleeding (menses). Ovulation occurs around day 14, dividing the cycle into two phases: a preovulatory follicular phase and a postovulatory luteal phase[80] (Figure 155-1). Under the control of the hypothalamic pituitary ovarian axis, ovarian follicles grow during the follicular phase, with an associated rise in plasma follicle-stimulating hormone (FSH) and estradiol. Circulating estradiol levels peak just before ovulation, triggering a surge in luteinizing hormone (LH) secretion from the anterior pituitary. Ovulation occurs 12 to 16 hours later, around day 14. In the luteal phase, progesterone dominates, being secreted from the ruptured dominant follicle (corpus luteum), together with estradiol. Approximately 14 days after ovulation, if implantation of a fertilized ovum does not occur, hormone levels drop precipitously, heralding the onset of menstruation. Ovulatory cycles typically are between 25 and 35 days long, although menstrual cycle length tends to shorten as women enter their forties. Most menstrual symptoms are experienced by women during the last few days of the cycle, as progesterone and estrogen levels decline, and during the first few days of menstruation.[81]

The reproductive endocrine and sleep systems have the potential to profoundly influence each other. Sleep has a stimulatory effect on LH pulsatility during puberty but an inhibitory effect on LH pulsatility in the early follicular phase, raising the possibility that disturbed sleep patterns could influence reproductive function in women.[82] Receptors for

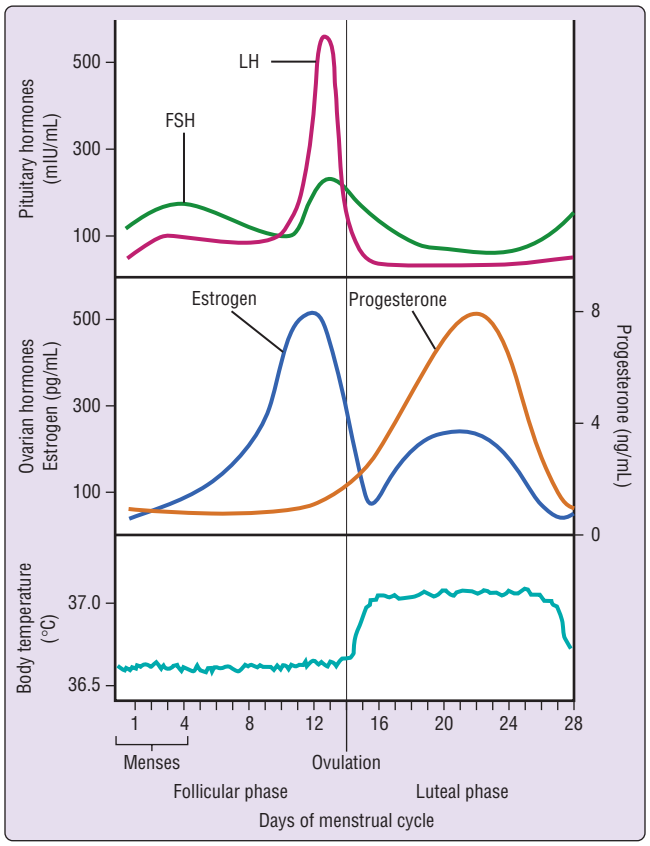


Figure 155-1 Mean daily plasma concentrations of estradiol, progesterone, follicle-stimulating hormone (FSH), and and luteinizing hormone (LH) and basal body temperature throughout a "typical" 28-day ovulatory menstrual cycle. (Modified from Pocock G, Richards CD. *Human physiology: the basis of medicine.* New York: Oxford University Press; 1999, p. 450.)

estrogen and progesterone are present in many sleep-wake–regulatory nuclei in the CNS, and estradiol induces increased arousal in rodents, possibly through its inhibitory effect on sleep-promoting neurons in the ventrolateral preoptic nucleus, an important sleep-active nucleus.[1] Estradiol also consolidates sleep-wake rhythms, possibly through its actions on the SCN.[83] The mechanistic framework is therefore in place for menstrual cycle–related changes in reproductive hormones to influence sleep and circadian rhythms.

Sleep across the Menstrual Cycle

Surveys research based on subjective reports have found that women across a wide age range (18 to 50 years) report more sleep disturbances during the premenstrual week and during the first few days of menstruation than at other times.[52] The Study of Women's Health Across the Nation, which included women in their late reproductive stage or just entering the menopausal transition, reported that trouble sleeping was more likely during the early follicular and late luteal phases of the menstrual cycle.[84] Young women (21 ± 3 years of age) without significant menstrual cycle–associated complaints also report poorer sleep quality around the time of menstruation.[85] As discussed later, women who experience severe menstrual cramps or premenstrual symptoms may experience more significant declines in sleep quality.

Studies that have investigated sleep objectively with PSG measures have found variable menstrual cycle phase effects on sleep that do not always corroborate findings from subjective sleep studies.[52,81,86] One reason for the inconclusive findings from studies of sleep in women across the menstrual cycle is the methodologic challenges for researchers, including variability in cycle length, presence and timing of ovulation, changes with age, and sampling times during the menstrual cycle.[81] Individual studies generally have been performed using small samples of women (fewer than 10) and at selected phases of the menstrual cycle.

Most studies have found that sleep onset latency and sleep efficiency remain stable at different phases of the menstrual cycle in young women. Percentages of N3 activity and SWA in NREM sleep, averaged across the night, do not change,[87,88] suggesting that sleep homeostasis is maintained across the menstrual cycle. The menstrual cycle has some influence on REM sleep. Women may have an earlier onset of REM sleep,[89] shorter REM episodes,[81,90] and less REM sleep in the luteal phase than in the follicular phase, with the amount of REM sleep negatively correlated with levels of progesterone and estradiol.[90] Women with ovulatory cycles have a shorter REM onset latency in the luteal phase compared with women with anovulatory cycles.[91] In the absence of ovulation, women do not secrete sufficient ovarian hormones or increase their core body temperature and thus do not experience a true luteal phase in the 2 weeks before menses. Variation in the timing and amount of REM sleep during the menstrual cycle may be related to altered circadian processes or the raised body temperature.[88,89]

Although not a finding in all studies, increased intermittent awakening or wake time has been documented during the late luteal phase compared with the follicular phase,[92,93] which corresponds with women's premenstrual reports of poorer subjective sleep quality. A recent actigraphy study of 163 women of late reproductive age reported a moderate decline in sleep efficiency (5%) and a decrease in total sleep time (by 25 minutes) in the premenstrual week.[94] With a larger sample that can be studied with actigraphy, subtle changes in sleep efficiency may be more apparent. Alternatively, these results raise the possibility that the menstrual cycle may have a more pronounced impact on sleep with advancing age. Hormone dynamics across the menstrual cycle also are relevant; a steeper rate of change in progesterone from the follicular to the luteal phase is associated with a greater amount of wakefulness after sleep onset in the luteal phase.[95]

The most dramatic change in sleep in association with the menstrual cycle is evident from spectral analysis of the sleep EEG. EEG activity in the 14.25- to 15.0-Hz band, which corresponds to the upper frequency range of sleep spindles, is significantly increased in the luteal phase over that in the follicular phase.[87,96] An increase in visually scored stage 2 (N2) sleep in the luteal phase also may be apparent.[87,88] The increased spindle frequency activity is reminiscent of the effects of the progesterone metabolite allopregnanolone on the EEG in rats[97] and may represent an interaction between endogenous progesterone metabolites and $GABA_A$ membrane receptors in the luteal phase.[87] Benzodiazepines and barbiturates also exert their sedative effects by binding to the $GABA_A$ receptor, but probably at a site that differs from that for progesterone metabolites.[98] Alternatively, the increased spindle frequency activity may be related to the increased body temperature in the luteal phase[81] and may help maintain sleep quality in the presence of substantial physiologic and hormonal changes in the luteal phase.[86]

Sleep is thus remarkably stable across the normal menstrual cycle, despite the large changes in the hormonal milieu. There are, however, some changes in sleep, most notably an increase in upper spindle frequency activity and a decrease in REM sleep in the luteal phase compared with the follicular phase.

Circadian Rhythms across the Menstrual Cycle

In women, circadian rhythms for hormone secretion, body temperature, and sleep-wake activity are superimposed on the menstrual cycle rhythm. As shown in Figure 155-2, men and women have the most similar temperatures when women are in their follicular phase. If ovulation occurs at the end of the follicular phase, the luteal phase is marked by an increase in body temperature of approximately 0.4° C, a consequence of the thermogenic action of progesterone secreted from the corpus luteum.[99] Most studies have found that the normal circadian nocturnal nadir in body temperature is blunted, reducing the amplitude of the temperature rhythm in the luteal phase compared with the follicular phase.[96] These findings have been confirmed in women monitored under controlled conditions of an ultrashort sleep-wake cycle procedure.[88] Under controlled conditions, no difference is seen in timing of the circadian temperature rhythm peaks and nadirs between the follicular and luteal phases.[88,100] The blunted amplitude of the temperature rhythm may be mediated by progesterone acting either directly on the SCN or downstream of the SCN; the amplitude of the body temperature rhythm is negatively related to progesterone concentration.[88]

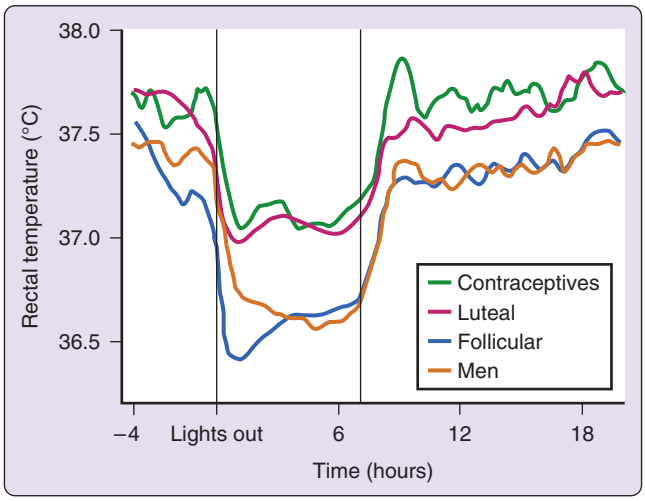


Figure 155-2 Mean diurnal rhythms in rectal temperature, plotted for 4 hours before "lights out" and 20 hours thereafter in 8 young men (*orange line*), 8 young women taking monophasic oral contraceptives (active pill) (*green line*), and 15 young women in the midfollicular and midluteal phases of their ovulatory menstrual cycles (*pink* and *blue lines,* respectively). *Vertical lines* indicate average time in bed. Subjects followed their usual daytime schedules and spent the night in a sleep laboratory. (With permission from Baker FC, Driver HS. Circadian rhythms, sleep, and the menstrual cycle. *Sleep Med* 2007;8:613–22.)

In studies with appropriate behavior controls, no difference was found in timing of onset or offset of melatonin secretion, duration, or acrophase of the melatonin rhythm between the follicular and luteal phases of the menstrual cycle in young women.[88,100,101] Two of the three studies also found no difference in amplitude of the melatonin rhythm across menstrual phases.[88,100] Taken together, findings show an interaction between the menstrual cycle and sleep/circadian factors. When the temperature rhythm is raised and blunted, REM sleep is decreased, and spindle frequency activity is increased in the absence of a change in sleep homeostasis or circadian phase in the luteal phase compared with the follicular phase.

Shift Work and Menstrual Rhythms

Women who work on schedules outside of a regular workday have disrupted circadian rhythms, which could lead to the development of medical problems, including altered reproductive function.[102,103] Female shift workers have more menstrual cycle irregularities, more painful menstruation (dysmenorrhea), and longer menstrual cycles compared with non–shift workers.[102,103] Shift-working women also are at increased risk for breast cancer, possibly because of exposure to artificial light and subsequent melatonin suppression.[103] Shift work nurses who report changes in menstrual function, compared with those who do not, report significantly more sleep disturbances, symptoms of shift work intolerance, and longer sleep onset latencies, suggesting an association between sleep disturbances and menstrual irregularities.[104]

Reproductive health problems associated with shift work could be due to increased stress as well as disrupted rhythms in hormone secretion, particularly LH and FSH, that directly influence ovarian hormone secretion.[104] Some evidence suggests that menstrual cycle irregularities and extended shift work may lead to fertility problems and increased risk for preterm birth.[105] However, meta-analyses of the literature investigating the association between various working conditions and fetal and maternal health concluded that shift work poses minimal risk to the female reproductive system,[106] and that the available evidence is insufficient for clinicians to advise restricting shift work in reproductive-age women.[107]

Sleep in Women with Menstrual Cycle Disorders

Menstrual cycle–related complaints, including mood disorders, pain associated with menstruation, and endocrine problems such as polycystic ovary syndrome (PCOS), are common among women of reproductive age. Women who suffer from a menstrual cycle–associated disorder may experience concomitant changes in their sleep. Women who have menstrual cycle–related problems are between two and three times more likely than other women to report insomnia and excessive sleepiness during the day.[108]

Polycystic Ovary Syndrome

PCOS affects 4% to 12% of women of reproductive age.[109] Women with PCOS typically present with irregular or absent cycles, androgen excess (evident as hirsutism), and bilateral polycystic ovaries. Insulin resistance also is an important component of PCOS,[110] and obesity is a factor in approximately 50% of cases.[111] Medical management of PCOS involves control of irregular menses, typically with oral contraceptives, management of infertility, and long-term management of insulin resistance.[109] The combination of obesity and excess androgen production places women with PCOS at increased risk for SDB, as well as for hypertension and cardiovascular disease. Indeed, women with PCOS have been found to be 30 times more likely to suffer from SDB than control subjects.[112] Obesity is an important risk factor for SDB, but even when compared with healthy age- and weight-matched control subjects, women with PCOS are more likely to have SDB during their reproductive years (Figure 155-3) and are more likely to report excessive daytime sleepiness (80%) than healthy women (27%), even after introduction of controls for body weight.[113]

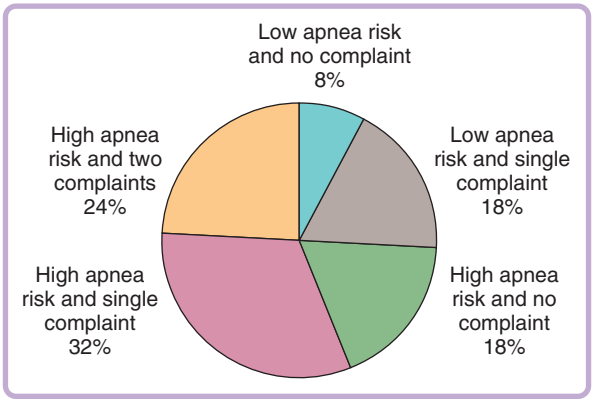


Figure 155-3 Frequency distribution of sleep apnea risk and sleep complaints in 40 women with polycystic ovary syndrome (PCOS). (With permission from Tasali E, Van Cauter E, Ehrmann DA. Relationships between sleep disordered breathing and glucose metabolism in polycystic ovary syndrome. *J Clin Endocrinol Metab* 2006;91:36–42.)

Increased central obesity, common in women with PCOS, may be more closely associated with increased risk for SDB than is elevated body mass index.[114] Increased severity of SDB is associated with glucose intolerance and insulin resistance in women with PCOS, suggesting that SDB may contribute to the metabolic abnormalities in these women.[113] Although larger trials are needed, preliminary evidence indicates that successful treatment with continuous positive airway pressure (CPAP) therapy in extremely obese women with PCOS leads to a modest improvement in insulin sensitivity and a reduction in diastolic blood pressure.[110]

Premenstrual Syndrome

Premenstrual syndrome (PMS) is characterized by emotional, behavioral, and physical symptoms that occur in the late luteal, premenstrual phase of the menstrual cycle, with resolution at the onset of menses or shortly thereafter. Many women of reproductive age experience some symptoms, but up to 18% of women have clinically relevant premenstrual symptoms that they perceive as distressing and impacting on daily function.[115] Premenstrual dysphoric disorder (PMDD) is a severe form of PMS that occurs in 3% to 8% of women.[115] PMDD is classified as a depressive disorder in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition*. A diagnosis of PMDD requires the occurrence of five specified symptoms, of which at least one must be a mood-related symptom, during the late-luteal phase, for at least two consecutive cycles. One of these symptoms is sleep disturbance (insomnia or hypersomnia).

Pharmacologic management of severe PMS/PMDD includes use of selective serotonin reuptake inhibitors (the drugs of choice recommended by the American College of Obstetricians and Gynecologists), anxiolytics, and agents that suppress ovulation.[116] Nonpharmacologic agents and approaches such as calcium supplements, L-tryptophan, and cognitive-behavioral therapy also have been shown to be effective.[116]

Women with severe PMS typically report sleep-related complaints such as insomnia, sleep perturbation by body movements and awakenings, disturbing dreams, and poor sleep quality associated with their PMS symptoms in the late-luteal phase.[117] They also report sleepiness, fatigue, less alertness, and an inability to concentrate during the premenstrual phase.[117] Laboratory studies, however, have found little evidence of PSG-defined disturbed sleep in the late-luteal phase. A recent study found that women with severe PMS/PMDD experienced a poorer subjective sleep quality in the absence of PSG measures of poor sleep quality, including measures of sleep efficiency, arousals, sleep onset latency, and the sleep EEG in the late-luteal phase relative to the follicular phase,[90] confirming findings from most earlier studies.[86,117] The perception of a poorer sleep quality correlated with anxiety levels, suggesting that the mood state of women with severe PMS affects their sleep assessments in the late-luteal phase.[90] Of interest, some evidence points to differences in sleep between control subjects and women with PMS/PMDD across the menstrual cycle (i.e., trait-like), although findings have varied between studies.[86,117] Two recent studies found that N3 sleep was increased in both the follicular and luteal phases in women with severe PMS or PMDD compared with control subjects.[90,118] This finding leads to the hypothesis that N3 sleep may be functionally linked with decreased melatonin secretion, which also was evident in the women with PMDD.[90,118,119]

Additional evidence shows altered circadian rhythms in women with PMDD. High mean nocturnal temperatures and disturbances in melatonin rhythms and in the timing of the rhythms of cortisol and thyroid-stimulating hormone have been reported in women with PMDD compared with asymptomatic control subjects.[86,96,120] Under controlled conditions, a small group of women with PMDD were found to have lower nocturnal melatonin levels in both menstrual phases and decreased melatonin amplitude in the symptomatic luteal phase compared with control subjects.[119] In view of these disturbances in circadian rhythmicity in women with PMDD, investigators have explored the possibility of treating PMDD by manipulating the timing of sleep, with sleep deprivation protocols or with light therapy. Appropriately timed light therapy has shown some promise as a treatment strategy for PMDD, possibly by altering nocturnal melatonin secretion.[121] A meta-analysis of clinical trials of bright light therapy, however, concluded that larger trials are needed to define the role of such therapy in the management of PMDD.[122]

For women with PMDD, sleep deprivation during the symptomatic luteal menstrual cycle phase may be therapeutic. Eight of 10 women responded to sleep deprivation in one study and maintained improved mood after a night of recovery sleep.[123] In a follow-up study, partial sleep deprivation had similar positive effects on mood in 60% to 67% of patients, but these effects were significant only after recovery sleep.[124] Changes in REM latency and REM density in the first REM sleep period from baseline to recovery nights were significantly correlated with improvement in mood in responders to sleep deprivation.[125] The effect of sleep deprivation on mood in women with PMDD differs from that found in patients with MDD, who have improved mood the day after total sleep deprivation but tend to relapse after recovery sleep.[126]

Dysmenorrhea

Dysmenorrhea, defined as painful menstrual cramps of uterine origin, is the most common gynecologic condition among women of reproductive age, and is very severe in approximately 10% to 25% of women.[127] *Primary* dysmenorrhea is menstrual pain without organic disease, and *secondary* dysmenorrhea is associated with conditions such as endometriosis and pelvic inflammatory disease. Dysmenorrhea is effectively treated with analgesics and nonsteroidal anti-inflammatory drugs in most women. Alternative therapies such as heat and dietary supplements of thiamine or magnesium also have shown some effectiveness.[128] Menstrual cramps experienced by these women every month significantly affect quality of life and are associated with a restriction of activity.[127] PSG studies in women with primary dysmenorrhea found that the menstrual cramps were associated with disturbed sleep, specifically poorer subjective sleep quality, lower sleep efficiency, more awake time and stage 1 (N1) light sleep, and less REM sleep than in pain-free phases of the menstrual cycle, and in comparison with asymptomatic women.[129,130] Disturbed sleep, in turn, may exacerbate pain, because sleep deprivation is associated with a decreased pain threshold.[131] Treatment of nocturnal pain with a nonsteroidal antiinflammatory drug alleviates nocturnal pain and restores both subjective and objective sleep quality in women with primary dysmenorrhea.[130]

Sleep Disorders and the Menstrual Cycle

In the most recent (2014) edition of the *International Classification of Sleep Disorders* (*ICSD3*), the only gynecologic disorder listed is menstrual cycle–related Kleine-Levin syndrome (or menstrual cycle–related hypersomnia). This exceedingly rare condition is characterized by hypersomnolence during the week before or during menses. The patient has no complaints of persistent, excessive sleepiness at other times in the menstrual cycle.[117]

Women with premenstrual hypersomnia have been successfully treated with either estrogen or combined oral contraceptives.[117] Little research has been conducted to investigate the variation in severity of sleep disorders such as insomnia and OSA according to menstrual cycle phase. The severity of OSA may be greatest in the follicular phase; upper airway resistance is lowest in the luteal phase in healthy women.[132]

Oral Contraceptives, Body Temperature, and Sleep

Oral contraceptives are combined formulations of a low-dose progestin and a synthetic estrogen and are taken by healthy women for long periods of time. Oral contraceptives suppress endogenous reproductive hormones and therefore prevent ovulation so that women taking these preparations do not have ovulatory cycles.

Women taking oral contraceptives, containing estrogen and a progestin, have raised 24-hour body temperature profiles, with the timing of temperature minima similar to that in ovulating women in the luteal phase[96] (Figure 155-2). This

body temperature elevation probably is caused by the thermogenic action of progestins contained in the contraceptive pill. During the 7-day placebo period of the oral contraceptive pack, 24-hour body temperatures remain elevated,[133] suggesting that continuous use of synthetic reproductive steroids may have long-term influences on thermoregulation. By contrast, body temperature rapidly decreases as progesterone levels decline before menstruation in women with ovulatory cycles.

Oral contraceptives also may influence melatonin levels, although findings are inconsistent.[96] In one study using a modified constant routine procedure, researchers found no significant differences in melatonin levels between naturally cycling women and women taking oral contraceptives, although a trend was observed in this small sample for increased melatonin levels in the latter part of the night in women taking oral contraceptives.[100]

Sleep architecture is altered by oral contraceptives. Women taking oral contraceptives have less N3 sleep than women with natural menstrual cycles.[81] N2 sleep is significantly increased in women taking oral contraceptive pills over that during the week of placebo pills, and compared with that in naturally cycling women in both menstrual cycle phases.[133] Also, use of a synthetic progestin (medroxyprogesterone) is associated with a specific increase in upper spindle frequency activity in women.[134] Other sleep effects reported in hormonal contraceptive users include a shorter REM onset latency and more REM sleep than in women with natural ovulatory menstrual cycles.[81,135] Exogenous steroid hormones therefore appear to influence sleep differently from the variations in endogenous hormones across the menstrual cycle. Consideration of oral contraceptive use also may be of importance in comparing healthy women and women with clinical disorders such as depression. The sleep effects of oral contraceptives may compromise sleep differences between healthy and depressed women.[135]

CLINICAL PEARLS

- As reviewed in this chapter, strong evidence supports neurophysiologic and neuroanatomic sex differences in sleep, particularly with maturation.
- Sex differences in sleep and rhythms may influence risk level for neurologic, immunologic, and psychiatric disease.
- The relationship between sleep and psychological health and behavior appears to be strongly influenced by sex-specific differences across the lifespan.
- Within groups of women, the menstrual cycle also affects sleep. REM sleep is reduced and spindle frequency activity is increased in the absence of a change in N3 amount in the luteal compared with the follicular phase. Women, especially those who experience severe premenstrual mood changes or menstrual pain, are likely to experience a decrease in sleep quality before and during menstruation compared with other phases of their menstrual cycle. Assessment of sleep complaints in women should include an investigation of any association between symptoms and menstrual cycle phase or menstrual cycle–related disorders.
- In view of the high incidence of SDB and its association with glucose intolerance in women with PCOS, screening for sleep apnea and appropriate treatment with continuous positive airway pressure may be beneficial in these women.

SUMMARY

A significant and still growing body of literature demonstrates that sex differences in sleep emerge at a very early age. Girls report longer sleeping periods and begin the adolescent decline in SWA (delta sleep) earlier than boys. Women report a poorer sleep quality and are at increased risk for insomnia compared with men across the adult lifespan. However, women have a better sleep efficiency and higher SWA during NREM sleep than men. Premenopausal women are at lower risk for development of SDB, mediated in part by a protective effect of sex steroids on the upper airway. Sex also is an important factor in considering sleep disturbances associated with disorders such as MDD; depressed men have lower SWA, whereas depressed women have lower synchronization between sleep EEG frequency bands compared with healthy control subjects. These observations highlight the importance of sex differences in sleep and circadian rhythm research studies and have broader implications for women's health issues relating to these topics. Within specific groups of women, sleep may be affected by variation in reproductive hormones, such as occurs over the menstrual cycle. The menstrual cycle is associated with changes in circadian rhythms and sleep architecture, most notably a blunted amplitude of the body temperature rhythm, increased spindle frequency activity during sleep, and reduced REM sleep in the luteal phase compared with the follicular phase. Women with PCOS are at increased risk for developing SDB, which may contribute to insulin resistance and other metabolic abnormalities. Women with severe PMS or dysmenorrhea (painful menstrual cramps) may experience transient sleep disturbances or insomnia coupled with their other mood and/or physical symptoms before and during menstruation. Assessment of sleep complaints in women should include an investigation of any association between symptoms and menstrual cycle phase or menstrual cycle–related disorders.

Selected Readings

Armitage R, Hoffmann RF, Sleep EEG. depression and gender. *Sleep Med Rev* 2001;**5**:237–46.

Baker FC, Driver HS. Circadian rhythms, sleep, and the menstrual cycle. *Sleep Med* 2007;**8**:613–22.

Campbell IG, Grimm KJ, de Bie E, Feinberg I. Sex, puberty, and the timing of sleep EEG measured adolescent brain maturation. *Proc Natl Acad Sci U S A* 2012;**109**:5740–3.

de Zambotti M, Willoughby AR, Sassoon SA, et al. Menstrual cycle-related variation in physiological sleep in women in the early menopausal transition. *J Clin Endocrinol Metab* 2015;**100**:2918–26.

Driver HS, Dijk DJ, Werth E, et al. Sleep and the sleep electroencephalogram across the menstrual cycle in young healthy women. *J Clin Endocrinol Metab* 1996;**81**:728–35.

Mong JA, Baker FC, Mahoney MM, et al. Sleep, rhythms, and the endocrine brain: influence of sex and gonadal hormones. *J Neurosci* 2011;**31**:16107–16.

Moran LJ, March WA, Whitrow MJ, et al. Sleep disturbances in a community-based sample of women with polycystic ovary syndrome. *Hum Reprod* 2015;**30**(2):466–72.

Redline S, Kirchner HL, Quan SF, et al. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch Intern Med* 2004;**164**:406–18.

Shechter A, Boivin DB. Sleep, hormones, and circadian rhythms throughout the menstrual cycle in healthy women and women with premenstrual dysphoric disorder. *Int J Endocrinol* 2010;**2010**:259345.

Tasali E, Van Cauter E, Ehrmann DA. Polycystic ovary syndrome and obstructuve sleep apnea. *Sleep Med Clin* 2008;**3**:37–46.

A complete reference list can be found online at ExpertConsult.com.

Sleep and Sleep Disorders Associated with Pregnancy

Bilgay Izci Balserak; Kathryn Aldrich Lee

Chapter 156

Chapter Highlights

- Sleep disturbances often begin with the onset of pregnancy and become quite common in the third trimester. Pregnancy-related physiologic and anatomic changes can influence sleep quality and duration.
- In the first trimester, hormonal fluctuations and associated physiologic changes may disturb sleep. By the end of the second trimester, discomfort associated with the enlarging uterus and anxiety regarding labor and delivery, as well as impending motherhood, are the main reasons for sleep disturbance, resulting in considerable daytime sleepiness and fatigue.
- Sleep disorders such as sleep-disordered breathing or restless legs syndrome may occur or worsen during pregnancy. Sleep-disordered breathing can result from added abdominal weight gain, nasal congestion, and changes in the respiratory system. Pregnancy-related restless legs syndrome is a secondary form of the disorder attributed to iron deficiency and hormonal changes.
- Complaints of nocturnal esophageal reflux and sleep-related leg cramps also are common and increase in frequency as the pregnancy advances.
- Emerging evidence suggests an association between sleep disturbance and pregnancy complications such as gestational diabetes, preeclampsia or pregnancy-induced hypertension, depression, and prolonged labor or cesarean birth.
- Sleep disturbances may represent novel and modifiable risk factors for adverse pregnancy outcomes. Early identification and management of prenatal sleep disorders may minimize adverse maternal and fetal outcomes.

OVERVIEW

Sleep disturbance during pregnancy is surprisingly common and multifaceted, beginning with conception and early hormonal changes. With increasing gestation and growth of the uterus and fetus, sleep is significantly disturbed by the third trimester of pregnancy.[1-4] In this chapter, *sleep disturbances* refer to internal and external factors that change the duration or structure of a normal sleep pattern and cause poor sleep quality or daytime sleepiness. Specific disturbances include disorders of initiating and maintaining sleep, disorders of sleep-wake schedule, and dysfunctions associated with sleep, sleep stages, or partial arousals in the absence of physical or mental disorders, prescribed medications, or substance abuse.[5] Pregnancy-related anatomic, physiologic, hormonal, and psychological factors, such as anxiety associated with labor or forthcoming life changes, can emerge in different stages of pregnancy and affect the degree and severity of sleep disturbances. Some of these pregnancy changes also can cause sleep disorders or exacerbate existing sleep disorders, which can threaten maternal and fetal health. Accordingly, clinicians need to be able to identify pathologic sleep disturbances in pregnancy. Early recognition and management of sleep problems during pregnancy can minimize or even prevent adverse maternal and fetal outcomes. Sleep-disordered breathing in pregnancy is covered in detail in the following chapter.

PREGNANCY-INDUCED HORMONAL CHANGES AND SLEEP EFFECTS

Pregnancy causes dramatic changes in melatonin and cortisol as well as gonadal steroids (estrogen and progesterone) and pituitary hormones (gonadatropins, prolactin, growth hormone). These hormonal changes not only directly influence the sleep-wake cycles and sleep structure but also cause physiologic changes that increase risk of sleep disorders (Figure 156-1). Sleep and circadian rhythms also have a role in regulating the secretion of these hormones.

Progesterone

The level of progesterone released from the placenta is 10 to 500 times higher at term than in the nonpregnant state.[6] Serum progesterone exhibits a diurnal change, with higher concentrations in the evening.[7] Progesterone, intracellular progesterone receptors, and progesterone metabolites acting on brain gamma-aminobutyric acid (GABA) receptors produce a soporific effect, with a significant increase in non–rapid eye movement (NREM) sleep.[8-10] This effect may partly explain daytime sleepiness and fatigue in the first trimester, when progesterone is steadily rising. Animal and human studies have demonstrated that exogenous progesterone administration shortens latency to sleep onset, lengthens latency to REM sleep onset, and reduces total

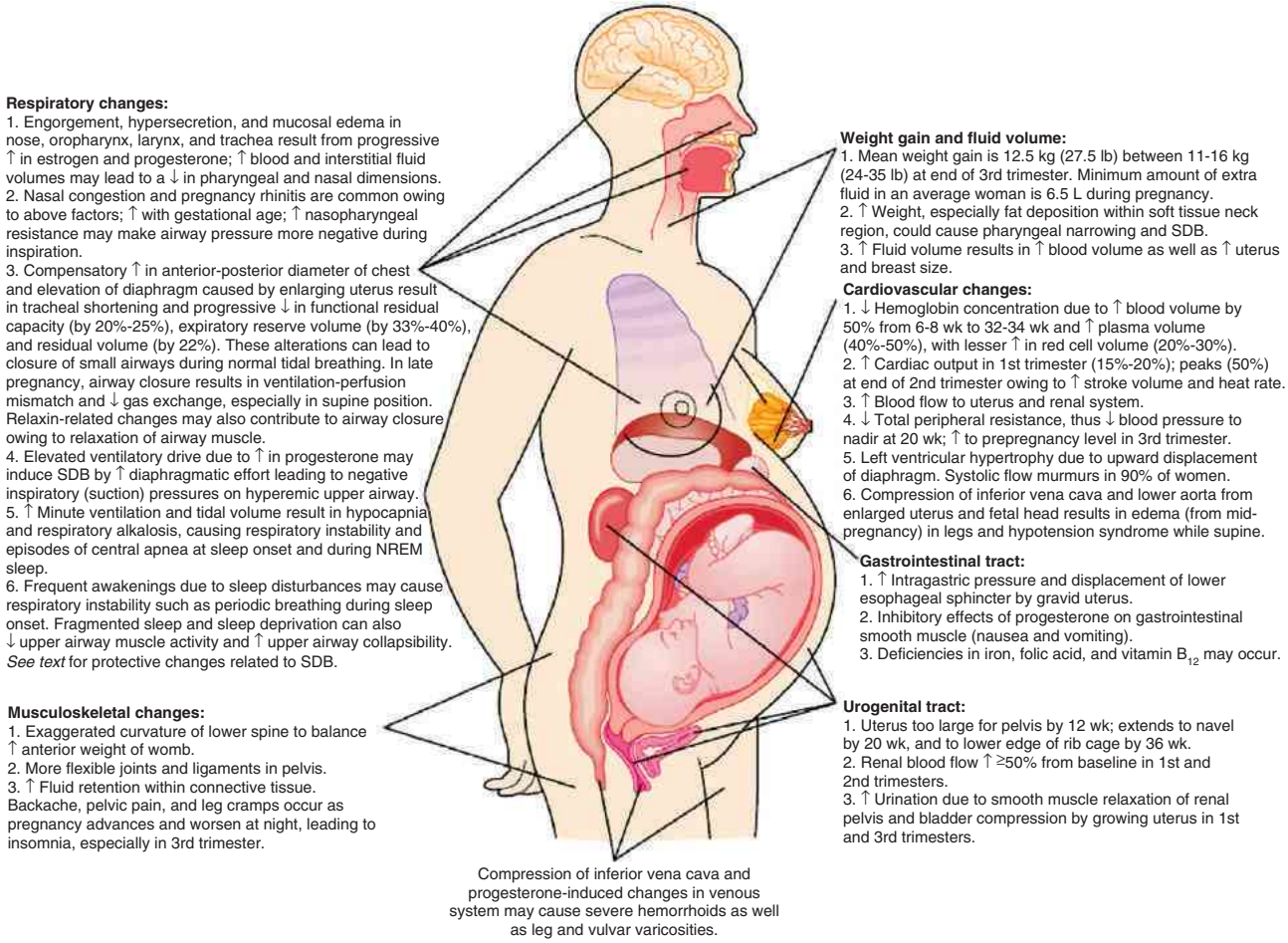


Figure 156-1 Pregnancy-related changes that can affect sleep. NREM, Non–rapid eye movement; SDB, sleep-disordered breathing.

amount of REM sleep.[9-11] Progesterone's soporific effect, its thermogenic effect that elevates core body temperature, and its inhibitory effect on smooth muscle (including gastrointestinal tract, ureters, and bladder) also indirectly influence sleep.[1,12-14] Furthermore, the increased respiratory rate caused by progesterone may protect the airway from occlusion, but complaints of feeling short of breath are common.[14]

Estrogen

Estrogen secreted by the placenta increases significantly during pregnancy, reaching peak levels before birth and declining thereafter.[6] Estrogen has excitatory effects on the nervous system and selectively decreases REM sleep activation of sleep-active neurons in the ventrolateral preoptic area.[15] Maternal plasma estriol has a 24-hour rhythm at 35 weeks of gestation, which occurs in the opposite direction of the cortisol rhythm.[7] The higher estrogen concentration during pregnancy, however, causes vasodilation, and with the extra fluid that accumulates during pregnancy, women typically experience nasal congestion and ankle edema.[6,14] Estrogen also stimulates prolactin production and suppresses dopamine release into the blood circulation, which may contribute to RLS.[16,17]

Cortisol

Cortisol starts to increase from the 25th week of pregnancy, with a two-fold increase seen in late pregnancy, followed by rapid return to normal concentrations after birth of the infant.[6,7] This elevation is mostly due to placental secretion of corticotropin-releasing hormone and adrenocorticotropic hormone (ACTH), and to increased synthesis of cortisol-binding globulin by the liver.[18] Progesterone and cortisol also share binding sites on cortisol-binding globulin.[7] Consequently, an increase in the level of progesterone during pregnancy leads to higher levels of free cortisol. The normal diurnal rhythm in cortisol includes a nadir level around midnight and marked elevation during early morning hours.[7,19] In pregnant women, the morning peak is not obvious, probably owing to the blunting effect of placental ACTH on maternal cortisol concentrations.[6,7] Sleep loss has been found to result in elevated cortisol levels the next evening.[20] Experimental studies in humans showed that cortisol infusions reduce rapid eye movement (REM) sleep but increase slow wave sleep.[18] Pregnant women with poor sleep in the third trimester have lower cortisol-melatonin ratios compared with good sleepers in the same trimester, as a result of a lower early-morning peak in their cortisol levels and a relatively higher concentration of melatonin.[7,19]

Melatonin

Melatonin is secreted by the pineal gland. Secretion is activated by darkness and suppressed by light. Melatonin is involved in the regulation of sleep patterns and circadian rhythms of reproductive hormones.[21] The circadian rhythm of melatonin secretion, peaking in the middle of the night and gradually decreasing by morning, continues during pregnancy.[21] The diurnal rhythm observed during the first and second trimesters is similar to that in a nonpregnant state, but levels increase in the third trimester.[19] In twin pregnancies, nocturnal melatonin levels are significantly higher after 28 weeks of gestation than in normal singleton pregnancies.[21]

Melatonin synergizes with oxytocin to promote the birth process.[21] The concentration of melatonin does not change in either labor induction or operative cesarean delivery.[22] Altered rhythm or low levels of melatonin secretion may potentially result in some pregnancy complications.[21] Research indicates that melatonin, of both pineal and placental origin, plays a role in fetal maturation and placental-uterine homeostasis[21] and is involved in correcting the pathophysiology of complications such as preeclampsia and fetal brain damage.[22] Plasma melatonin concentrations, particularly morning levels, are reported to be lower in depressed pregnant women but higher in depressed women in the postpartum period than in healthy control subjects.[23]

Prolactin

Prolactin concentrations are 10 times higher at term than those in the nonpregnant state.[6] Prolactin is involved in immunoregulation, lactogenesis, and mammary tissue growth. However, prolactin secretion rhythms do not differ between pregnant good sleepers and poor sleepers.[19] Studies with a small number of pregnant women show episodic prolactin secretion and elevated levels during nocturnal sleep.[7] During a normal vaginal delivery, prolactin levels peak for 4 to 6 hours and then descend back to a normal circadian pattern.[24] By contrast, prolactin concentrations are significantly lower in women who elect to undergo cesarean delivery.[24] Prolactin secretion may enhance slow wave sleep, as suggested by evidence that slow wave sleep is increased in patients with prolactinomas.[25]

Oxytocin

Oxytocin promotes uterine contractions and lactation.[11] In late pregnancy, it reaches the highest level at night, together with peaks in uterine activity rhythm.[26] In animals, oxytocin has been found to promote sleep at basal levels under stress-free conditions, but it may induce wakefulness in high concentrations.[11] Oxytocin and melatonin act synergistically to assist in the process of labor and birth. This interaction between oxytocin and melatonin during darkness, with increased oxytocin concentrations during the night, could explain the higher incidence of nocturnal births.[21,26]

Growth Hormone

The secretion of pituitary growth hormone (GH), crucial for tissue growth and protein anabolism, is regulated primarily by growth hormone–releasing hormone and neurons in the arcuate nucleus of the hypothalamus.[27] Levels of placental GH in maternal circulation increase throughout pregnancy starting as early as week 8, with a peak around week 35 of gestation.[28] Pituitary GH is released in a pulsatile manner, but the release of placental GH is continuous.[28] Ghrelin also is a GH stimulator produced by placental tissue.[6] These hormones are closely associated with slow wave sleep and play a significant role in sleep regulation,[25,27] suggesting that GH and maintenance of slow wave sleep are important for fetal development and maternal health.

Relaxin

The corpus luteum is the primary source of circulating relaxin during pregnancy. Serum relaxin peaks at the end of the first trimester.[29] Thereafter, levels fall to an intermediate value until birth. It is associated with connective tissue remodeling and relaxing pelvic ligaments to prepare for birth.[6,30] Relaxin can contribute to discomfort and sleep disturbances through relaxation of the airway, development of carpal tunnel syndrome from fluid retention in the connective tissue, and onset of low back pain associated with relaxation of supporting ligaments.[6,30] Precise mechanisms for these changes, however, remain to be clarified.

Leptin

Serum leptin levels increase during prgnancy, peaking during the second trimester and remaining two to four times those typical for the nonpregnant state. A significant amount of leptin is released by the placenta and the increasing number of fat cells associated with pregnancy weight gain.[6] Leptin has a key role in regulating body fat, energy expenditure, and fetal growth.[6,12] Extremes in either short or long sleep duration in early pregnancy are associated with altered leptin levels.[31]

The relationship between hormonal changes and altered sleep during pregnancy is thus bidirectional. Although these hormones have significant effects on sleep and circadian rhythm during pregnancy, altered sleep during pregnancy also can influence hormone levels. Further research is required, particularly to characterize sleep-induced hormonal fluctuations and implications for adverse obstetric outcomes.

PHYSIOLOGIC CHANGES IN PREGNANCY AND POTENTIAL FOR SLEEP PROBLEMS

In addition to hormonal changes, pregnancy leads to a multitude of anatomic and physiologic changes including changes in respiratory and cardiovascular systems, which are detailed in Figure 156-1.[6] These changes are essential to maintain a healthy pregnancy, but some changes contribute to expected and common sleep problems, whereas other changes have potential for adverse effects on either the fetus or the mother.

Sleep in Normal Pregnancy

With the physical and hormonal adaptations seen in pregnancy, changes in sleep[32-36] are common, with a majority of women (66% to 97%) affected. They report nocturnal awakenings that become more frequent in the third trimester.[2,37,38] Unlike other sleep parameters, both objective and subjective sleep measures show that sleep efficiency (SE) progressively decreases, not because of problems falling asleep but because of increased and longer episodes of nocturnal awakenings after sleep onset.[33,35,38,39] However, night-to-night and individual variability does exist.[34,40] For example, in several recent studies, women reported having no sleep disturbances or noted “improved” or “worsening” sleep, as well as short (6

hours or less), intermediate (7 to 8 hours), and long (10 hours or longer) sleep duration even in early pregnancy.[3,38,41]

Only a small number of studies have comprehensively evaluated sleep across pregnancy with validated tools to record sleep on multiple nights and at different weeks of gestation. Although numerous studies have attempted to describe sleep characteristics during pregnancy, findings are not completely in agreement. The discrepancy could be related to different research designs, settings (laboratory or home), assessment methods (PSG or actigraphy), small sample sizes, collecting data at different gestational periods, or comparing pregnant women at various gestational ages with nonpregnant control subjects in different menstrual cycle phases. Most studies also failed to record sleep during daytime naps, which complicates any sleep assessment. Furthermore, recent studies comparing objective (polysomnography or actigraphy) and subjective (questionnaire and sleep diaries) measures find discrepancies in sleep time estimates.[37,42,43] These discrepancies may lead to spurious associations between sleep duration and adverse pregnancy outcomes, particularly if a study relies on self-reported sleep parameters during pregnancy.

Sleep in First Trimester

Women experience daytime sleepiness and fatigue and report more frequent naps as early as the tenth week of pregnancy.[33,36,38,2,44-47] Increases in total sleep time (TST), longer sleep onset latency, and more wake time after sleep onset (WASO) also are reported. Increased TST may be ascribed to daytime naps.[33,34,39,48,44,45] Overall sleep quality and slow wave sleep decrease relative to pre-pregnancy assessments or the nonpregnant state.[2,3,33,35,36,38] Sleep disturbances occur mainly as a result of dramatic alterations in hormone levels in the first trimester.[6,9] These hormones are responsible for fatigue and daytime sleepiness, morning sickness, waking with nausea, increased urinary frequency, physical discomforts (tender breasts or back pain), and mood changes.[2,44-47] Psychosocial stressors, especially in the case of first-time or unplanned pregnancies, including absence of psychosocial support in first-time pregnancies, also have been reported.[49]

Sleep in the Second Trimester

In most cases, subjective and objective sleep parameters improve compared to the first trimester.[2,5,20,33,38] Most women report less fatigue and more energy[2,3,44-47] likely due to the stabilization of hormone levels. In a multicenter prospective cohort study of 369 first-time mothers monitored with actigraphy for 7 days, sleep averaged more than 7 hours per night in the second trimester.[50] It also has been noted that women in their second trimester have better SE and less WASO compared with those in the other trimesters[38,39,48] (Table 156-1). By the end of the second trimester, however, the number of awakenings increases.[5] Accordingly, women may experience disturbed sleep as a result of the onset of snoring, heartburn, irregular uterine contractions (Braxton-Hicks), fetal movements, leg cramps, or RLS.[2,3,44,51,52] Vivid dreams and pain in the back, neck, and joints were additional reasons for sleep disruption during the second trimester, according to a National Sleep Foundation survey.[52]

Objective measures of slow wave sleep and REM sleep during the second trimester are inconsistent.[33,34,53] A longitudinal PSG study in the home setting, however, showed that slow wave sleep (%) slightly decreased but REM sleep (%) did not change relative to these measures in the first trimester.[33]

Table 156-1 Sleep Pattern, Nocturnal Features, and Daytime Symptoms in Each Trimester and during Labor/Delivery

Characteristic	First Trimester	Second Trimester	Third Trimester	Labor/Delivery
Pattern	↑ TST ↑ Number of naps ↑ WASO ↓ SE ↓ SWS	↓ TST ↑ SE ↓ SWS ↓ WASO	↑ TST ↑ Number of naps ↑ WASO ↑ Stage 1 sleep ↓ SE ↓ SWS ↓ REM	↓ TST ↓ SE ↓ NREM ↓ REM
Nocturnal features	Urinary frequency Physical discomforts—tender breasts/back pain	Onset of snoring, restless legs, irregular uterine contractions Dreams Back, neck, and joint pain	Urinary frequency Physical discomfort Irregular uterine contractions Muscle/leg cramps Shortness of breath Heartburn Dreams/nightmares Snoring Restless legs	Anxiety Forceful uterin contractions
Daytime symptoms	Fatigue Drowsiness Waking with nausea Mood changes	Nasal congestion	Fatigue Drowsiness Impaired vigilance Nasal congestion	Fatigue Anxiety Pain

NREM, Non–rapid eye movement; REM, rapid eye movement; SE, sleep efficiency; SWS, slow wave sleep; TST, total sleep time; WASO, wake [time] after sleep onset.

Sleep in the Third Trimester

The overwhelming majority (75% to 98%) of women report sleep disturbances resulting in multiple nocturnal awakenings as they approach the 40th week of pregnancy.[1,2,47] Sleep patterns in the third trimester feature longer sleep latency, poor SE, more WASO, and increased stage 1 and 2 sleep, with some reduction in slow wave sleep.[2,3,20,32-38,47,52] Nocturnal sleep time is lower than the first two trimesters but TST may approach prepregnancy values.[2,33,35-37] This is likely because shortened night-time sleep may be compensated by more frequent and longer daytime naps. In fact more than 75% of the women reported at least one nap each week in the third trimester.[33,47,52] Despite inconsistent findings, most PSG studies show less slow wave sleep and REM sleep when compared with previous trimesters or data for nonpregnant control subjects.[33-35,38,39,54] Secondary analysis of a prospective study of 127 women showed that delta power in NREM sleep significantly decreased from the first to the third trimester after controlling for apnea-hypopnea index, age, TST, and other covariates.[55] In exchange for this decrease, women spent more time in light sleep stage 1 or 2.[33,35] An additional covariate is employment schedules, in that working women were less likely to nap and reported sleeping 1 hour less in the last month of pregnancy.[56]

Physical changes associated with a rapidly growing uterus, in addition to hormonal fluctuations, are the main causes of sleep disturbances in third trimester.[3,4,37,47] Most women complain of urinary frequency, general physical discomfort (e.g. backache), heartburn, leg cramps, spontaneous awakenings, and fatigue.[3,4,35,47,52] Fetal movements, difficulty maintaining sleep, shortness of breath, and other physical discomforts (breast tenderness, joint pain/carpal tunnel symptoms, and itching) often are reported.[1-4,44] Women also attribute their sleep loss to internal factors (vivid dreams/nightmares associated with anxiety about labor/delivery, the health of the fetus, and pregnancy complications) and external factors such as the needs of their other children or environmental noise.[1,4,44,49,52] Additionally, the risk of a primary sleep disorder increases during the third trimester, as discussed further on.[17,32]

By the third trimester, women also often complain about difficulties with attention, concentration, and memory. A meta-analysis of 14 studies comparing pregnant women with matched control subjects found that some memory measures, particularly for executive cognitive control, were significantly impaired, and these difficulties were attributed to sleep disturbance and fatigue.[57]

Sleep During Childbirth

Pain, anxiety, uterine contractions, and administration of medications all affect sleep and result in sleep loss and low sleep quality during labor and immediately after delivery.[34,58-60] The rhythm of nocturnal uterine activity, presumably a result of peaks in oxytocin secretion patterns, may contribute to nocturnal awakenings and most women experience spontaneous labor onset with forceful contractions at night.[21,26]

In a longitudinal study of 35 women, sleep quality deteriorated progressively over the last 5 days of pregnancy and was lowest the night before contractions started and admission to the hospital for delivery.[59] In one study, 20 women were asked about their experience of sleep during pregnancy and labor, and all reported they were unable to sleep once contractions started. As the latent phase of labor becomes prolonged, sleep is not possible even when sleep aids are used.[60] Significant associations between sleep duration in late pregnancy and adverse obstetric outcomes are reported, as discussed later on.[59,61,62]

Parity Differences in Sleep

Sleep during pregnancy may be influenced by parity or by the presence of other children living in the home. Studies using objective sleep measures found that TST, slow wave sleep, and REM sleep during pregnancy were not significantly affected by parity,[33,35,40] but nulliparas (no prior birthing experience), especially employed women, were at risk for poor sleep quality,[33,35,37,40] possibly due to challenges in adjusting to a new role. Additional studies indicate that parity influences SE, with nulliparas having significantly worse SE from third trimester to first month post partum.[46,63] Multiparas, who already have prior birthing experience, may be awakened during the night by children,[1,44] but when controlling for other children, multiparas have higher SE than nulliparas during pregnancy.[40,46]

During third trimester, younger women (<30 yrs), who are also more likely to be nulliparas, had more TST than older women (>30 yrs)[2,64] whose sleep may be affected by other children in the home[1,2,44] or by a primary sleep disorder. For example, increasing age is recognized as a risk factor for RLS, SDB, and insomnia.[3,17,47] Thus parity as well as age should be considered in sleep research with pregnant women.

In summary, most women experience sleep disturbance during pregnancy. Each trimester brings its own reasons for disrupting sleep, and the degree of sleep disturbance experienced throughout pregnancy is highly variable. If opportunities for napping are available, TST can be higher in each trimester compared with TST for nonpregnant women. Yet nocturnal TST and SE decrease progressively during pregnancy while sleep stage characteristics remain generally constant. Pregnant women have more light sleep and less deep sleep (i.e., slow wave sleep), owing to nocturnal awakenings. Nulliparity, age older than 30 years, and being employed each are associated with different aspects of poor sleep quality.

SLEEP DISORDERS DURING PREGNANCY

Primary sleep disorders can occur or worsen during pregnancy because of the pregnancy hormones and physiology as just discussed. In view of the potentially serious consequences for both maternal and fetal health, clinicians should consider the possibility of sleep disorders, particularly insomnia, SDB (see Chapter 157) and RLS, when symptoms associated with these disorders are observed.

Pregnancy-Associated Insomnia

Insomnia is defined as difficulty initiating or maintaining sleep, early-morning awakening or nonrestorative sleep associated with daytime consequences such as fatigue, irritability, and lack of concentration (see Chapter 83).[65,66] Pregnant women may experience some or all of these symptoms within the range of acute to chronic insomnia.[3,63,67,68] In some women, the acute insomnia[66] persists into a chronic state lasting a month or longer, even after pregnancy. A genetic disposition to chronic insomnia may have a pivotal role in this transition.[66,67]

Epidemiology

The prevalence of insomnia in the general and pregnant populations is difficult to estimate owing to various definitions of insomnia and the subjective nature of the symptoms. Symptoms can occur independent of primary insomnia or can be associated with a comorbid condition. Furthermore, the frequency, severity, and pattern of insomnia can vary during pregnancy. Many studies have used self-reported insomnia in assessing its prevalence and response to treatment.[12] As determined using the Women's Health Initiative Insomnia Rating Scale, Bergen Insomnia Scale and Insomnia Severity Index, a majority (52%) of women have some degree of insomnia at any point in pregnancy,[68] with 54% to 74% reporting insomnia during their third trimester.[3,63,67] In a national telephone survey of women in the United States, 84% reported one or more symptoms of insomnia at least a few nights each week during pregnancy.[52] Approximately 40% of women worry about their sleep during the first and second trimesters, and this rate increases to 57% by the third trimester.[44] The highest prevalence of insomnia and most wake episodes occur in the third trimester.[3,63,67,68]

Risk Factors

In adults with insomnia, the available evidence implicates physiologic hyperactivity and hyperarousal, commonly seen in pregnancy as a higher body temperature and metabolic rate, with increased secretion of ACTH and cortisol.[6,69] Increased pregnancy concentrations of estrogen and progesterone may result in insomnia due to changing GABA.[8-10] In pregnancy, however, this is a more complex phenomenon.

Insomnia in pregnancy is considered to be secondary to the many pregnancy-related discomforts described earlier. Although poor sleep quality and sleep loss among pregnant women can be the result of hormonal and physical factors, voluntary sleep restriction also is common owing to employment and household responsibilities.[3,56,64] Maladaptive sleep behaviors and coping strategies such as napping, spending more time in bed, or increasing caffeine intake[3,64,67] also can perpetuate insomnia. Weight gain and obesity contribute to sleep-onset or maintenance insomnia and SDB with multiple cortical arousals.[3,5,42] A recent study confirmed that arousals are associated with respiratory events and limb movements during late pregnancy.[35] Cortical arousals also can be associated with emotional distress, especially in a woman at risk for insomnia because of anxious thoughts about impending lifestyle changes with a new child, balancing employment and motherhood, and approaching labor and delivery, including concerns for the health of the infant.[35,49,60-63,66] Insomnia has not been adequately studied in pregnant women and cannot be treated effectively until the primary condition is better understood.

Other risk factors for insomnia in pregnancy include age older than 30 years,[2,3] nulliparity,[63] single motherhood,[20] preeclampsia or pregnancy-induced hypertension,[3,41] prepregnancy affective disorders,[48,63,70] perinatal depression,[58,63,70] smoking,[67] negative body image,[71] and environmental factors such as noise from other children, bed partners, or pets.[4,52] Exposure to light during the night also can reduce melatonin levels and exacerbate sleep disturbances.[72,73]

Management

The diagnosis of insomnia in a pregnant woman is challenging. Diagnosis requires adequate opportunity to sleep, with an inadequate ability to fall and stay asleep.[12,65] These criteria may be confused with multiple awakenings associated with pregnancy discomforts. Pregnant women also may have insomnia as a comorbid condition associated with SDB or depression. Current evidence suggests that untreated comorbid insomnia may reoccur even if the originating cause is treated.[66] In addition, insomnia in pregnancy can significantly affect physical and cognitive function and has potential implications for pregnancy outcomes.[74] Timely assessment and appropriate management (see Chapters 85 to 88) are essential to prevent potential adverse pregnancy outcomes and reoccurrence of chronic insomnia.

After considering the potential adverse effects of medication on the fetus during pregnancy, most women and clinicians opt for nonpharmacologic therapies to treat insomnia. Behavioral and cognitive therapies should be the initial treatment for women with pregnancy-related insomnia, after excluding primary sleep disorders. These nonpharmacologic therapies include cognitive-behavioral therapy for insomnia,[75] improving sleep hygiene, using relaxation techniques, and implementation of lifestyle modifications such as regular exercise and avoidance of smoking and alcohol. In an intervention study of 15 women attending mindful yoga classes, starting the classes in the second trimester improved SE and reduced nocturnal awakenings better than starting classes in the third trimester.[76] Further research is needed, however, to determine timing and optimal amount of exercise during pregnancy.

Of note, many pregnant women do not seek treatment for insomnia, because they either think it will naturally resolve after birth or wish to avoid medication owing to concerns about adverse effects on the fetus. In the United States, 11% of pregnant women used a sleep aid, including over-the counter medications, at least a few nights a week and 1% used alcohol at some point in pregnancy to help them sleep.[52] If nonpharmacologic therapies do not help ease the problem in cases of persistent and severe insomnia, medications should be prescribed at the lowest effective dose for the shortest possible duration after discussion of potential risks and benefits with the patient (Table 156-2). Currently, two hypnotic agents, diphenhydramine and doxylamine (histamine H_1 receptor antagonists with sedative) for insomnia during pregnancy are categorized as possible but unlikely to harm the fetus (former Category B drugs).[77,78]

In a recent randomized control study comparing trazodone (an antidepressant medication) or diphenhydramine with placebo, sleep profiles improved after 6 weeks of treatment in the third trimester, and depressive symptoms were reduced for 2 to 6 weeks after delivery.[77] Zolpidem and zaleplon, despite possible adverse fetal effects (former Category C drug), also may be an option owing to their short duration of action.[12] Most other sedative-hypnotics (Table 156-2) should be avoided or used with extreme caution in pregnancy. Drugs with human data showing fetal risks (formerly Category D) should only be used if the potential benefits outweigh the risk to the fetus. Drugs with clear evidence of abnormalities and high risk to the fetus are contraindicated (formerly Category X designation).

Table 156-2 Pregnancy Classification* of Medications Used for Sleep-Related Disorders

Disorder	Harm to Fetus Unlikely, but Possible (Formerly A Category B Medication)	Potential Benefit Outweighs Risk (Formerly A Category C Medication)	Evidence of Fetal Risk (Formerly A Category D Medication)	Contraindicated (Formerly A Category X Medication)
Insomnia	Doxylamine (7) Diphenhydramine (7)	Zaleplon (1) Eszopiclone (1) Ramelteon (1) Zolpidem (1) Gabapentin (2) Amitriptyline (5) Doxepin (5) Trazodone (5)	Alprazolam(1) Diazepam (1) Lorazepam (1) Midazolam(1) Secobarbital(1)	Estazolam (1) Flurazepam (1) Quazepam (1) Temazepam (1) Triazolam (1)
Restless legs syndrome (RLS)	Cabergoline (3) Pergolide (4) Oxycodone (4)	Zolpidem (1) Gabapentin (2) Oxcarbazepine (2) Pregabalin (2) Carbidopa/levodopa (3) Pramipexole (3) Ropinirole (3) Rotigotine (3) Methadone (4) Codeine-C/D* (4) Oxymorphone (4) Propoxyphene-C/D* (4) Hydrocodone-C/D* (4) Tramadol (4)	Clonazepam (1) Carbamazepine (2)	Temazepam (1) Triazolam (1)
Narcolepsy		Sodium oxybate (5) Fluoxetine (5) Venlafaxine (5) Selegiline (5) Clomipramine (5) Protriptyline (5) Dextroamphetamine (6) Mazindol (6) Methamphetamine (6) Methylphenidate (6) Modafinil/armodafinil (6) Protriptyline (5)		
Parasomnias			Clonazepam (1) Topiramate (2) Sertraline (5)	
Gastroesophageal reflux disease (GERD)	Sucralfate (8) Cimetidine (9) Ranitidine (9) Famotidine (9) Nizatidine (9) Metoclopramide (10) Lansoprazole (11) Rabeprazole (11) Pantoprazole (11) Esomeprazole (11)	Cisapride (10) Omeprazole (11)		

*Numbers in parentheses refer to medication classes: (1) sedative-hypnotics, (2) anticonvulsants, (3) dopaminergics agents, (4) opioids, (5) antidepressants and depressants, (6) stimulants, (7) antihistamines, (8) antacids, (9) histamine type 2 receptor antagonists, (10) promotility agents, and (11) proton pump inhibitors. Former U.S. Food and Drug Administration (FDA) pregnancy categories:

Category A: There is no fetal risk in first trimester in controlled human studies (no evidence in later trimesters). Possibility of fetal harm is remote.

Category B: Animal studies have not indicated a fetal risk but no controlled studies in pregnant women exist, or animal studies have shown an adverse fetal effect that is not confirmed in controlled studies in women in the first trimester (no evidence in later trimesters).

Category C: Animal studies have shown adverse fetal events, but no controlled studies in pregnant women exist, or studies in humans and animals are not available; thus, give if potential benefit outweighs the risk.

Category D: Risk to fetus is present, but benefits may outweigh risk if life-threatening or serious disease exists.

Category X: Studies in animals or humans show fetal abnormalities; drug contraindicated for pregnant women.

C/D: With prolonged use or in high doses, its Category C level of risk becomes a higher risk (Category D).

Studies provide some evidence that insomnia symptoms during pregnancy may be alleviated by mindful yoga, acupuncture, massage, or exercise.[74,76] The safety of acupuncture during pregnancy, however, needs further investigation. Alternative therapies (herbal or dietary supplements such as chamomile tea or lavender pillows) also are used as sleep aids,[12] but controlled studies are needed to assess the benefits and risks to fetal and maternal health.

Pregnancy-Related Restless Legs Syndrome (Willis-Ekbom Disease)

Restless legs syndrome (RLS) is a common neuromotor disorder characterized by an irresistible urge to move the legs to stop the uncomfortable sensations. Symptoms are distinctly worse at night while at rest, relieved by movement, and not solely accounted for by another medical or behavioral condition such as leg cramps or positional discomfort.[79,80] The diagnosis depends on five cardinal criteria[17,80] (see Chapter 95). Pregnancy-related RLS, or Willis-Ekbom disease, is classified as a secondary form of RLS because symptoms typically resolve at the time of delivery.[16,17,79] Pregnancy-related RLS is a risk factor for development of RLS that may occur several years after pregnancy.[79,81]

RLS may coexist with periodic leg movements in sleep (PLMS).[80] Such movements are characterized by rhythmic extension of the big toe and ankle dorsiflexion, with occasional flexion at the knee and hip during sleep. The resulting repeated brief arousals can contribute significantly to disturbed sleep during pregnancy.[80,82] Researchers using leg actigraphy monitoring found that the frequency of PLMS decreased by 50% to 100% after delivery in women with RLS.[83] RLS symptoms are associated with shorter TST, more difficulty initiating and maintaining sleep, and more daytime sleepiness compared with these measures in pregnant women without RLS.[16,17,79,84] In extreme cases, symptoms are so disturbing that evening relaxation and falling asleep are almost impossible, creating a high risk for depression.[79,85]

Epidemiology

Although the cause of RLS in pregnancy is uncertain, the association between RLS and pregnancy is not new. In 1940 Mussio-Fournier and Rawak made the first observation that RLS was exacerbated during pregnancy in German women.[17] Since 1940, numerous reports from around the world have estimated prevalence at 3% to 34% in pregnant populations,[16,17] and all agree that RLS is two to three times more common in pregnant women than among nonpregnant women,[16] even in ethnic groups in which RLS is a rare condition. Most prospective studies report that prevalence rates of RLS increase as pregnancy progresses, peaking in the third trimester and resolving a few days before delivery.[16,17,83,84] Variations in pregnancy-related RLS estimates can be attributed to use of differing diagnostic criteria (earlier studies were conducted before establishment of diagnostic criteria by the International Restless Legs Syndrome Study Group), self-report questionnaires based on retrospective symptoms, and different risk factors including maternal age, parity, ethnicity, regional variations, and genetics. In Western countries where studies used established diagnostic criteria for RLS, reported prevalence rates range from 11% to 34%.[16,17,79,83,84] In Asian countries, prevalence rates range from 3% to 20%.[16] In a recent epidemiologic study of 461 Taiwanese women interviewed using established criteria, the overall prevalence of RLS was 10.4%.[86]

One of the first prevalence studies using international criteria and a structured clinical interview was performed in Italy with 642 pregnant women.[79] The RLS prevalence rate was 26.6%, and the mean time of onset of RLS symptoms was around the sixth month of pregnancy. Although 16.7% had never experienced RLS symptoms before pregnancy, approximately 10% had preexisting symptoms, and 15% noted symptoms at least three times per week.[79] A recent prospective study in Switzerland, however, found a lower prevalence (12%) and earlier onset (before the fifth month).[83] These differences could be explained by ethnicity, because both studies used similar methodology and the same diagnostic criteria.[79,83]

Pathophysiology

The pathophysiology of pregnancy-induced RLS has yet to be established. Several factors are thought to be involved, including hormonal mechanisms, iron and folate metabolism, family history, depression, and multiparity.[16,17,84]

Iron and folate deficiencies, often seen during pregnancy, are associated with RLS.[16,79,84] Earlier studies reported that women in whom RLS developed by their third trimester had lower folate and ferritin levels at preconception and lower folate, plasma iron, hemoglobin values, and mean corpuscular volume throughout pregnancy, in comparison with healthy control subjects.[16,79,84] However, a later study of 19 pregnant women did not support associations between RLS and reduced iron, ferritin, or hemoglobin,[82] although levels were lower in the women with RLS. This conflict may be due to small sample sizes. Two recent epidemiologic studies also reported that ferritin, hemoglobin, and hematocrit values did not differ between women with and without RLS,[83,87] but iron intake was higher among women with RLS in one of these studies.[87]

Iron deficiency in the central nervous system, which results in the dysfunction of dopaminergic systems, may be responsible for RLS symptoms among women who have normal ferritin, an indicator of systemic iron storage. In this regard, cerebrospinal fluid (CSF) ferritin levels are informative because patients with RLS, even those with normal peripheral iron stores, have low CSF ferritin and high CSF transferrin.[17] Studies on central nervous system iron metabolism in pregnancy-related RLS are needed. However, the hypothesis of iron deficiency in pregnancy-related RLS cannot be supported for two reasons. First, all pregnant women in the second half of pregnancy have similar hemoglobin and ferritin reductions, attributed to fetal growth and hemodilution, regardless of iron and vitamin supplementation.[79,84] Second, RLS symptoms resolve just before labor onset, even though the largest loss of blood and iron is at delivery and it takes at least 3 months for repletion of storage iron.[6]

Hormones such as prolactin and estrogens may contribute to development of RLS. Prolactin may be involved in development of RLS during pregnancy as a consequence of its antidopaminergic activity.[17,82] However, no strong evidence for a causative role of prolactin in RLS is lacking,[82] because levels rise throughout pregnancy and continue to be elevated in breast-feeding women, yet RLS resolves with labor and delivery.[79,83] In addition, RLS is not associated with diurnal rhythms of prolactin in the general population.[16,82] Elevated

estradiol levels during pregnancy, and sudden decline with delivery of the placenta, were linked with RLS symptoms,[82] yet larger studies failed to confirm this association.[16,83,88] Progesterone levels do not differ between women with and without RLS.[16,82] Thyroid-stimulating hormone also has been implicated in the pathophysiology of RLS,[16] but levels did not differ between pregnant women with and without RLS in one recent study.[87] With these conflicting findings, more research on the relationships between hormonal factors and pregnancy-related RLS is needed.

Family history of RLS and multiparity (especially closely spaced pregnancies) may be independent risk factors for RLS.[16,79,81,83] RLS was reported by 75% of multiparas who experienced RLS in a previous pregnancy[83] but was not experienced between pregnancies.[2] The risk of RLS increases with the number of children in a "dose-dependent" way.[12,81] Other factors associated with RLS include older age, heavier maternal weight, smoking, peptic ulcer disease, varicosities, and concomitant SDB or depression.[16,79,83,86,87] Some medications such as SSRIs, antihistamines, and antiemetics also may trigger or exacerbate RLS symptoms.[12,16,17] Sleep deprivation, anxiety and stress, insomnia, and fatigue also are documented contributors to RLS in pregnancy.[16,17,79,81] In a prospective study of 1428 women, pre-pregnancy RLS was an increased risk factor for both antenatal and postnatal depression, whereas no added risk was seen in pregnant women with new-onset RLS during pregnancy.[85]

Management

RLS that develops during pregnancy is likely to be short-term and highly symptomatic until delivery. Pregnant women should consider nonpharmacologic comfort strategies (Figure 156-2), rather than medication that may affect the fetus. If sleep is severely disrupted by RLS, iron therapy may be warranted,[89] or the lowest dose of pharmacologic therapy (see Figure 156-2) can be considered (also see treatment options in Chapter 95).[81,90]

Sleep-Related Leg Cramps

Leg cramps are painful muscle contractions in the foot or leg. When the cramping occurs during sleep, a sudden awakening is typical, and the pain intensity can prevent return to

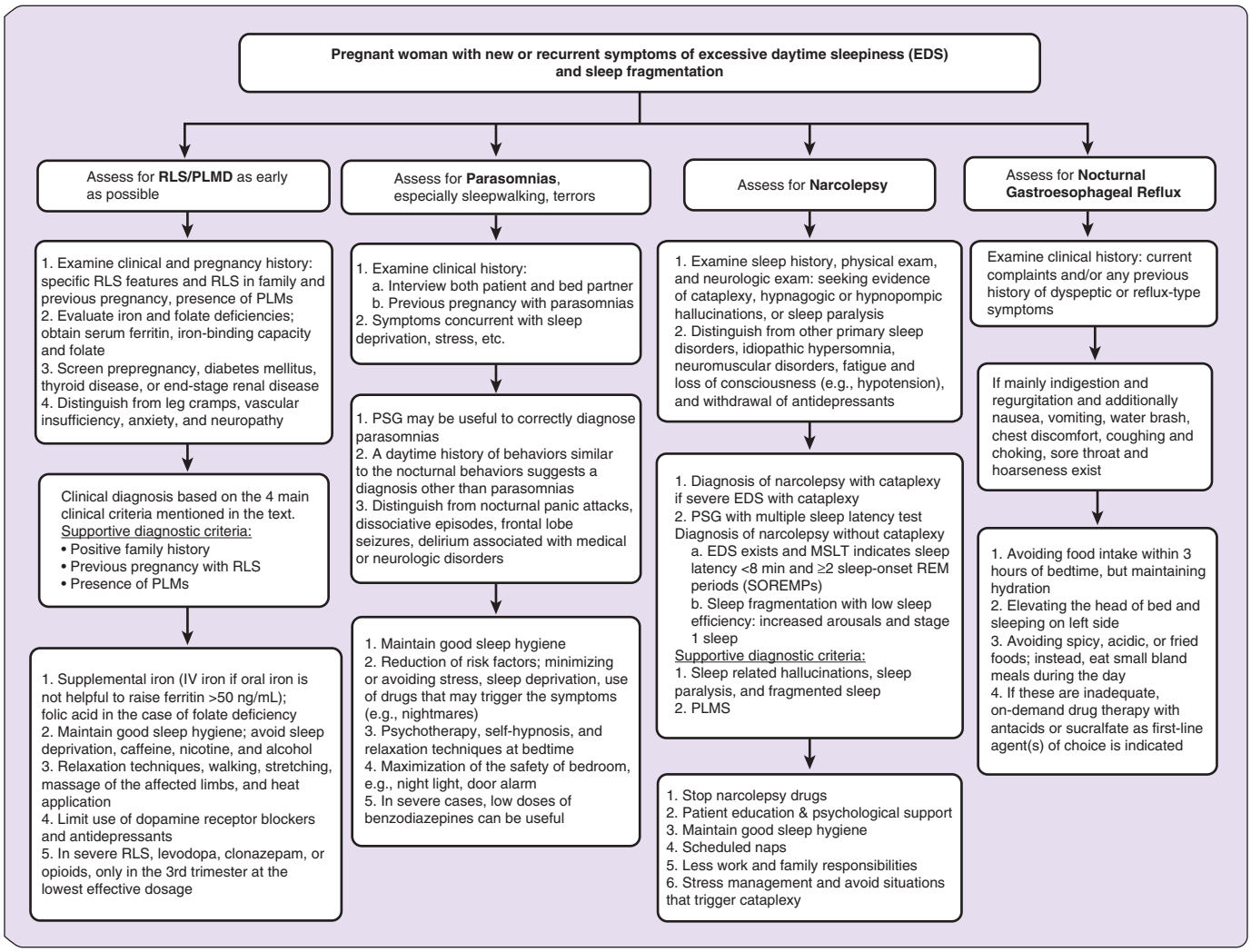


Figure 156-2 Management and treatment of sleep-related disorders. IV, Intravenous; MSLT, Multiple Sleep Latency Test; PLMs, periodic leg movements; PLMD, periodic leg movement disorder; PSG, polysomnography; REM, rapid eye movement; RLS, restless legs syndrome; SOREMP, sleep-onset rapid eye movement sleep period.

sleep for hours. Leg cramps during the night are one of the most common reasons for pregnancy-related sleep interruption.[91]

Epidemiology

The prevalence of nocturnal leg cramps increases from 10% before pregnancy to 21% in the first trimester, 57% in the second trimester, and up to 75% in the third trimester.[1,44,84]

Pathophysiology

It is unclear why these intense muscle spasms occur more often in pregnancy. One mechanism to consider is the altered metabolism of calcium, magnesium, and phosphate while fetal bones are developing, but muscle spasms increase in prevalence toward the end of pregnancy, when there is less demand for fetal bone formation. Other potential mechanisms include slowed venous return secondary to increased intraabdominal pressure and inhibitory effects of progesterone on the venous smooth muscle.[92]

Management

Abundant anecdotal evidence is available, but few studies have been performed to investigate the utility of treatment modalities to relieve or prevent nocturnal cramp attacks. These modalities include hyperextension to counter the contraction during the cramp, leg massage, and reducing intake of phosphorus-containing substances such as milk and meat.[44,92]

A systematic review of placebo-controlled trials involving 352 women showed no benefit from calcium supplements but suggested some relief with magnesium supplementation.[92-94] More recent research indicates that vitamin B is beneficial for eliminating leg cramps during pregnancy. In a 2-week clinical trial, 84 pregnant women were randomly assigned to one of four groups, receiving supplementation with either calcium, magnesium, or a combination of vitamins B_1 (thiamine) and B_6 (pyridoxine), with the fourth group consisting of control (no supplementation) subjects. Compared with that for the no supplementation group, B vitamin supplementation was most beneficial in eliminating the cramps over a 1-month time frame (72%) or in reducing their frequency and intensity (19%); only 2 (9%) of the 21 women in the vitamin B group reported no change in leg cramps.[94] By contrast, all 21 women in the magnesium supplementation group improved, but only 29% experienced absolute elimination of cramps. Although an adequate control group was in place and the sample size for each group was adequate, the researchers were not specific about any baseline pregnancy characteristics (parity, weeks of gestation, or age) and had no sleep parameters to report.[94] Vitamin B supplementation is found in most perinatal vitamins and should be discussed with the health care provider if leg cramps are disrupting sleep during the night.

Nocturnal Gastroesophageal Reflux Disease

Gastroesophageal reflux disease (GERD) is a chronic disorder in which gastric contents enter the esophagus, particularly after large meals, with ingestion of certain types of foods, or on reclining. This becomes problematic when the reflux leads to fragmentation of sleep while lying down at night. It is rare for younger healthy women to be diagnosed with GERD as a chronic condition before pregnancy. If a woman is already diagnosed with GERD and becomes pregnant, the condition can worsen because of the additional abdominal weight, the anatomic and hormonal changes that occur during pregnancy (see Figure 156-1),[95] and the different food items that may elicit symptoms. If reflux begins in pregnancy, the incidence typically peaks in the third trimester and resolves after childbirth, when anatomy returns to normal.

Indigestion or heartburn, acid taste, and regurgitation are common symptoms of GERD. However, other symptoms include nausea, vomiting, chest discomfort, coughing, choking, sore throat, and hoarseness that can be present on awakening in the morning. Nighttime symptoms include not only fragmented sleep but discomfort from injury to esophageal tissue caused by the acidic pH.[96,97] GERD impairs quality of life and also has the potential to interfere with healthy weight gain during pregnancy. Fortunately, GERD typically is confined to pregnancy, and long-term complications are rare.

Epidemiology

The incidence of GERD at any point in pregnancy ranges from 30% to 80%.[96] Overall, aproximately 25% of women experience reflux in each trimester.[97-99] In a recent prospective longitudinal study of 408 pregnant women, the prevalence of GERD symptoms doubled from 26% in the first trimester to 51% in the third trimester.[99] What remains unknown are the specific rates of reflux associated with nighttime awakenings while lying down, with all meals across the day, and with particular cultures and types of food choices.

Pathophysiology

The pathophysiology of GER in pregnancy is multifactorial (see Figure 156-1). One key factor is elevated progesterone hormone,[96,97,99] with its inhibitory effect on smooth muscle. In addition, as pregnancy progresses, repositioning of the stomach and intestinal tract occurs. Along with this repositioning, the mass effect of the growing fetus and abdominal weight gain increase intraabdominal pressure and reduce esophageal sphincter pressure, making GERD more common.[95]

Management

For the diagnosis of esophageal reflux during pregnancy, or the diagnosis of GERD as a longer-term disorder, a careful history should be obtained, focusing on current symptoms and pattern of weight gain in pregnancy.[96] Mild symptoms can be managed with temporary modifications to diet and lifestyle. Food intake should not occur within 3 hours of bedtime, but continued water intake into the evening hours is important to maintain hydration. Women with GERD symptoms are encouraged to eat smaller meals more frequently, and to sleep upright with more pillows to support the upper body. Sleeping on the left side with the head elevated can help counter the pressure from the repositioned esophageal sphincter, stomach, and intestines. The head of the bed also can be elevated with wedges, or the patient may feel more comfortable sleeping in a recliner chair, especially during the third trimester.[79,97]

When these lifestyle and dietary modifications are inadequate, sodium bicarbonate should not be used, because of high potential for electrolyte imbalance. Severe symptoms of GERD often are responsive to antacids, which have a low risk of harm to the fetus (see Table 156-2) and have no documented adverse effects in animal studies.[95] Antacid

regimens are useful to prevent or minimize GERD symptoms, reduce discomfort, improve sleep, and help maximize quality of life.[96]

In a European prospective study of 553 cases of first-trimester exposure to histamine H_2 receptor antagonists (e.g., cimetidine) to treat GERD, a higher relative risk of preterm birth (8.9%) compared with control cases (5.6%) (odds ratio [OR], 1.67; 95% CI, 1.18 to 2.35) was noted, and two cases of neural tube defect in the offspring of women taking famotidine were reported.[100] H_2 receptor antagonists have low risk to the fetus (formerly Category B; see Table 156-2), and although effective, may be best to use after any danger of preterm birth (before 36 weeks) has passed. With so few studies available, more data are needed from both animal studies and human research.

Parasomnias During Pregnancy

Parasomnias are undesirable physical events occurring during entry into sleep or during arousals from sleep, and the affected person usually is unaware that the event took place.[65] Parasomnias specific to NREM (e.g., confusional arousals, sleepwalking, sleep terrors) and to REM (e.g., sleep paralysis, nightmare disorder, REM sleep behavior disorder) have been well documented.[65] See Chapters 101 to 106 for more details on parasomnias in general.

Epidemiology

No strong evidence supports any increase in the incidence of parasomnias during pregnancy. In a longitudinal survey study of 325 women, sleep paralysis significantly increased in the second half of pregnancy (from 5.8% to 13.2%), but other findings included decreases in sleepwalking, sleeptalking, hypnagogic hallucinations, and sleep bruxism during pregnancy from values for the 3-month prepregnancy period, especially in nulliparas.[101] Data on dreams, nightmares, and their content during pregnancy are inconsistent, probably owing to different data collection methods and time frames. Although nightmares decreased significantly in the same sample overall in the survey study,[101] earlier reports have indicated increased frequency of dreams with gestational age.[44,93] At least one frightening dream about the infant or pregnancy has been reported by 25% of women during pregnancy.[93] Kennedy and associates[60] found that 80% of new mothers reported particularly vivid, peculiar, detailed and disturbing dreams during pregnancy. However, Nielsen and coworkers[102] noted that dream recall was equally prevalent among pregnant, postpartum, and nonpregnant women (88% to 91%). Pregnant women's dreams involved anxiety about the infant and birth outcomes, and the dreams often were accompanied by dream-associated behavior and confusional arousals. Nulliparas and multiparas also differed in dream recall, with more anxious dream content about the infant in nulliparas.[102] In a study of women with a previous pregnancy loss, 80% reported dreams associated with anxiety about the infant and pregnancy complications.[103]

A recent cross-sectional prospective study found that women in their third trimester and nonpregnant women have similar dream recall, but the pregnant women had more disturbing dreams.[104] The incidence of nightmares that exceeded criteria for moderately severe pathology (more than 1 nightmare/week) was three times higher in the same night among pregnant women (21%) than among nonpregnant women (7%).[104] Further research using polysomnography is required to clarify associations between sleep and pregnancy-related dream experiences.

Pathophysiology

Hormonal changes, fragmented sleep, and intense emotional stress may predispose pregnant women to frightening dreams, night terrors, or sleepwalking, especially in those with a genetic predisposition or previous history of parasomnia.[49,60,102,104] Cortisol levels may have a pivotal role in causation, with peaks occurring during the second half of the night, when REM sleep dominates and dream imagery and emotions are most intense.[104] Furthermore, pregnancy parasomnias may be associated with an undiagnosed sleep disorder (e.g., SDB) that fragments sleep. Finally, dreams and nightmares may be important indicators of the woman's psychological state. Pregnant women experiencing an active parasomnia with complex, vigorous, or violent behavior are at risk for sleep-related injury and injury to the fetus.

Management

Data on treating parasomnias during pregnancy are scarce, and more research is required. Informing women about the likelihood of parasomnias, such as disturbing dreams, may reduce anxiety. In the general population, clonazepam is prescribed for some parasomnias, but due to its high risk for fetal harm (formerly Category D), it is often discontinued during pregnancy.[105] In most cases, no special treatment is necessary, but effective measures include maintaining good sleep hygiene, reducing or avoiding stress and sleep deprivation, and making the bedroom safe (see Figure 156-2 and Chapters 101 to 106 for more detail).

Narcolepsy

Narcolepsy is a neurologic sleep disorder characterized by excessive daytime sleepiness, cataplexy (loss of voluntary muscle tone following emotion), hypnagogic hallucinations, and sleep paralysis.[65,106,107] Diagnostic criteria for narcolepsy are detailed in Chapter 90 and Figure 156-2. Diagnosis of the condition during pregnancy, however, may require special attention owing to pregnancy-related daytime sleepiness or vivid dreams. The weight gain that often accompanies pregnancy also could be falsely attributed to onset of narcolepsy.[107]

Epidemiology

The prevalence of narcolepsy in the general population is low (20 to 160 cases/100,000 population) in western Europe and the United States,[107-109] and the disorder usually emerges during young adulthood, thereby affecting women in the reproductive phase of life. The prevalence among pregnant women probably is similar to that in the nonpregnant population. The effects of pregnancy on narcolepsy, and how narcolepsy may affect the pregnancy, are not well understood. Symptoms of narcolepsy may be either exacerbated or attenuated during pregnancy.

Pathophysiology

Hypocretin (orexin) has a critical role in the pathophysiology of narcolepsy,[109,110] along with the loss of hypothalamic hypocretin-producing cells[65] in genetically susceptible persons who carry human leukocyte antigen (HLA)-DQB1*0602 and

HLA-DR2.[109] The significance of these gene abnormalities, especially in pregnancy, is not clear, but the relationship between narcolepsy and HLA type suggests an autoimmune disease that destroys hypocretin-producing cells.[109] Environmental factors also may play a causal role. Studies in northern Europe and China indicate that H1N1 virus infection may trigger narcolepsy.[109] Regarding pregnancy, however, symptoms of narcolepsy initially appeared in 16.6% of affected women up to 1 year after the birth of their first child,[107] so pregnancy is not a likely contributing factor in the pathogenesis of narcolepsy.

Consequences of Narcolepsy in Pregnancy

In retrospective European cohort studies, women with narcolepsy symptoms experienced before or during pregnancy tended to have a higher incidence of impaired glucose metabolism and anemia compared with asymptomatic women. However, these studies lacked control groups without narcolepsy.[106,107] Reported rates of cesarean births were higher in women with narcolepsy-cataplexy than in women with narcolepsy and no cataplexy.[106] Cataplexy itself can be dangerous and stressful during pregnancy owing to its uncontrollable nature. Mothers may worry about their children's inheriting narcolepsy. In addition, women with narcolepsy are inclined to develop obesity, which may increase the risk of SDB and pregnancy complications.[107] In a recent animal study, an association between pregnancy and increased risk of maternal death was evident in the narcolepsy-cataplexy mice with complete hypocretin deficiency.[110]

Management

The confirmation of narcolepsy diagnosis is based on PSG and a multiple sleep latency test (see Chapter 90 and Figure 156-2). Treatment is challenging in pregnant women. Detailed practice parameters for treatment of narcolepsy during pregnancy are not currently available. A standard narcolepsy treatment regimen includes scheduled daytime naps and may require multiple pharmacologic agents, including a stimulant to control excessive daytime sleepiness and an anticataplectic (sodium oxybate) or a selective serotonin reuptake inhibitor (SSRI) for cataplexy. However, most of these medications should be used with caution (formerly Category C drugs), as human data are lacking to determine pregnancy risks (Table 156-2).[111] Spontaneous miscarriages have been documented in women using sodium oxybate[111] and prematurity, low birth weight, and withdrawal symptoms have been reported in infants of women taking amphetamines.[12,111] A recent survey of 75 clinicians who treated narcolepsy during pregnancy, however, found no evidence of teratogenicity with modafinil, sodium oxybate, methylphenidate, amphetamines, or SSRIs.[111]

In a retrospective study of 249 pregnant women with narcolepsy, symptoms did not change in 40%, worsened in 40%, and improved in 18% who withdrew from medication during the first trimester.[106,111] These types of studies are limited by recall bias using retrospective data and small numbers of pregnancies that occur in women taking antinarcolepsy medication.[100,105] On the basis of current evidence, medication for narcolepsy should be discontinued for conception and during pregnancy and prescribed only if potential benefit outweighs potential risk. If a woman requires medication during pregnancy or decides to continue medication while attempting to conceive, however, she can be reassured that evidence for harm is largely lacking. With narcolepsy associated with cataplexy, an elective cesarean should be considered owing to the potential dangers for mother and newborn if a cataplexy episode occurs during labor.[111]

Patient education and psychological support are important for the woman and her family.[109] Good sleep hygiene and a supportive network are imperative to ensure adequate nocturnal sleep, adherence to scheduled daytime naps, and fewer role responsibilities. Behavioral interventions including scheduled naps, stress management, and avoiding situations that trigger cataplexy could decrease risk of physical injury as well as unpredictable sleepiness.

In summary, narcolepsy is a rare but disabling sleep disorder. Little is known about how pregnancy affects narcoleptic symptoms or how narcolepsy and pharmacologic treatments affect pregnancy. An autoimmune disorder in genetically predisposed women, possibly precipitated by infection, can alter hypothalamic hypocretin production. Based on animal studies, narcoleptic medications may be harmful to the fetus (formerly Category C medication), but human data are lacking for maternal-fetal risk.

SLEEP DURING PREGNANCY WITH PREEXISTING MEDICAL CONDITIONS

Sleep in pregnant women with medical conditions such as preeclampsia, affective disorders, asthma, and migraine headaches has not been widely studied. These conditions may have a further impact on sleep during pregnancy. Studies using objective sleep measures report that women with gestational hypertension have impaired sleep quality and markedly altered sleep architecture, with significantly more slow wave sleep and less REM sleep.[112-114] These women also napped more frequently than healthy pregnant women.[114] In a large cohort study (n = 1332) during early pregnancy, women with a psychiatric diagnosis reported shorter sleep duration, increased frequency of exhaustion, and higher perceived stress levels compared with women without such diagnoses.[115] A systematic review of earlier studies also noted that sleep disturbance, fatigue, and stress were more frequent during pregnancy among women with anxiety disorders than among other pregnant women.[58] In another cross-sectional study of 1335 pregnant women, women with a history of asthma experienced more sleep disturbances, including snoring, than did nonasthmatic women.[116]

MATERNAL AND FETAL CONSEQUENCES OF SLEEP DISTURBANCE

Sleep disorders, especially untreated apnea or extremes of sleep duration (both short and long), are potential contributors to adverse health conditions such as cardiovascular disease and type 2 diabetes. An analogous association in pregnancy also is recognized, with women at increased risk for development of adverse outcomes such as gestational diabetes or preeclampsia.[5] Adverse outcomes associated with SDB are covered in Chapter 157. Although study results in pregnant women should be interpreted cautiously owing to small sample sizes or other limitations such as recall bias, evidence supports the concept that sleep disturbances may be novel modifiable risk factors for preventing and managing pregnancy complications.

Obesity or Excessive Weight Gain

Experimental and observational studies reveal that both short and long sleep duration (generally defined as less than 6 hours and more than 9 hours, respectively), sleep restriction, and sleep disturbance can predispose affected persons to weight gain and obesity-related cardiometabolic disorders.[5,31] Although the evidence for a relationship between sleep duration and weight gain in pregnancy is limited, a few mechanisms have been proposed to explain the risk of excessive weight gain. Sleep deprivation associated with pregnancy discomforts may result in fatigue that leads to comfort eating, physical inactivity, and eventual obesity. Short sleep duration also allows additional time for food consumption. In a population-based cohort study of 710 pregnant women recruited before the 17th week of pregnancy and followed until delivery, sleeping less than 8 hours/day during the second or third trimester was a risk factor for inadequate gestational weight gain.[117] Sleep disturbance may also lead to altered leptin levels and contribute to obesity. In a study of 830 women during early pregnancy, Qiu and colleagues[31] noted that short sleep (5 hours or less) and, to a lesser extent, long sleep (9 hours or longer) was associated with elevated leptin levels if women also were overweight or obese.

Sleep loss, sleepiness, and fatigue, common conditions in pregnancy, are independently associated with insulin resistance,[118] resulting in increased risk of obesity.[5] Thus, avoiding obesity before and during pregnancy may minimize the risk of developing sleep disorders and adverse pregnancy outcomes.

Sleep-Disordered Breathing

Decreased REM sleep may prevent apneic events during pregnancy, but sleep onset and arousals from sleep can destabilize respiration (see Chapter 157). Studies and case reports in pregnancy revealed that sleep-disordered breathing (SDB) is associated with gestational hypertension and preeclampsia, gestational diabetes, pulmonary hypertension, and other cardiovascular diseases, as detailed in Chapter 157.

Pregnancy-Induced Hypertension and Preeclampsia

Pregnancy-induced hypertension, or gestational hypertension, is defined as repeated blood pressure recordings higher than 140/90 mm Hg, first diagnosed during pregnancy in a previously normotensive woman. If proteinuria also is present, the condition is called preeclampsia, but overt proteinuria is not an indication of severity and may not even be evident in some cases.[6] Gestational hypertension is a major contributor to maternal and fetal morbidity and mortality worldwide.[6,38,119]

The mechanisms linking sleep disturbance with gestational hypertension and preeclampsia are complex and involve several pathways detailed in Chapter 157. In nonpregnant populations, acute sleep deprivation is associated with intravascular inflammation and endothelial cell dysfunction, which also is a common pathway for gestational hypertension and preeclampsia.[119] Longitudinal studies of pregnant women indicate that poor sleep quality, fragmented sleep, and short sleep duration are associated with higher circulating inflammatory cytokines (interleukins IL-6 and IL-8, and tumor necrosis factor) and C-reactive protein.[36,120,121] Nevertheless, in view of the relatively small sample sizes and other study limitations, these results should be interpreted with caution.

In a recent study of 161 pregnant women, sleep diaries and actigraphy were used to evaluate sleep parameters, blood pressure, and BMI. Two measures of sleep continuity (latency to sleep onset and wake after sleep onset) were associated with higher blood pressures, even after controlling for covariates.[38] In a prospective cohort study of 1272 women, Williams and colleagues[41] found associations between self-reported nocturnal sleep duration and blood pressure. Both short (less than 6 hours) and long (more than 9 hours) sleep durations in early pregnancy were associated with elevated mean blood pressure in the third trimester. After controlling for confounding factors, they also observed the highest risk of preeclampsia in women who slept less than 5 hours per night.[41] Studies using objective measures confirm that women with preeclampsia have a significant increase in slow wave sleep (probably owing to cerebral edema and release of cytokines), less REM sleep, and more daytime naps compared with healthy pregnant women in the same trimester.[112-114]

Hyperglycemia and Gestational Diabetes Mellitus

Hyperglycemia (1-hour oral glucose tolerance test result of 140 mg/dL or higher) can be detected at any time during pregnancy. Severity ranges from glucose concentrations slightly above normal to gestational diabetes mellitus (GDM).[5] Accumulating evidence demonstrates an association between sleep duration and risk of hyperglycemia,[121a] but a causal relationship has not been established. Even mild hyperglycemia is associated with adverse maternal and fetal outcomes (i.e., preeclampsia and fetal growth retardation) and future development of type 2 diabetes, obesity, and cardiovascular disease in mother and child.[5] In a prospective study of 189 nulliparas with a singleton gestation, women who reported short sleep (less than 7 hours sleep/night) were 2.6 (95% CI, 1.3 to 5.7) times more likely to have a 1-hour oral glucose test value of 130 or higher and 10 times more likely to have GDM (95% CI, 1.3 to 85.5).[122] In a large prospective cohort study of 1290 women before 20 weeks of gestation, both short and long sleep durations were associated with hyperglycemia. Women sleeping ≤4 hours/night had a 5-fold increase (95% CI, 1.31 to 23.69) in risk of GDM compared with women sleeping 9 hours/night after adjusting for age and race. However, this association was no longer significant after adjustment for prepregnancy BMI.[123]

In a third study of 169 women who completed several sleep questionnaires, a significant inverse relationship was noted between sleep duration and 1-hour glucose values; for every 1 hour less sleep time, there was a 4% increase in glucose levels. Controls for risk factors for GDM, however, were lacking.[124] In a prospective study of 260 pregnant women, self-report sleep parameters were not significantly different between pregnancies with and without adverse outcomes including GDM.[20] Studies with small sample sizes ($n = 52$ to 104) are less likely to find statistically significant associations between sleep duration and hyperglycemia.[125,126] However, self-reported nap duration was associated with high glucose challenge test values in adjusted models. Bourjeily and associates[118] also have demonstrated that severe daytime sleepiness increases the risk of GDM in nonsnoring pregnant women, independent of confounding factors. These studies suggest that sleep, regardless of time of day, has modulatory influences

on glucose regulation. Further studies are required to elucidate whether napping may be therapeutic for preventing or treating GDM.

Adverse Obstetric Outcomes

Short sleep duration in pregnancy can be considered a physiologic stressor, leading to longer labor and cesarean delivery.[58] In a prospective observational study of 131 nulliparas in their last month of pregnancy, those who slept less than 6 hours/night had a significantly longer labor duration and higher rate (4.5-fold) of cesarean birth (with controls for infant birth weight) than nulliparas sleeping 7 hours or more by actigraphy measures.[61] A population-based study in 10,662 pregnant women confirmed these findings, indicating that short sleep in the last trimester is an independent risk factor for emergency cesarean delivery (adjusted OR 1.57; 95% CI, 1.14 to 2.16).[127] These correlations between sleep and length of labor or type of delivery also were confirmed by a cross-sectional study of 457 pregnant women after exclusion of hypertension, gestational diabetes, or emergency cesarean, noting that poor sleep quality and sleeping less than 7 hours were associated with longer labor and a higher rate of cesarean delivery.[128] The association between perception of sleep quality and duration of labor and type of delivery also was confirmed in subsequent studies. Using the Pittsburgh Sleep Quality Index (PSQI), poor sleepers (with a PSQI score above 5) were 20% more likely to have longer labor duration and to undergo cesarean delivery.[129] Despite these findings suggesting adverse effects of nocturnal sleep on labor and delivery, daytime sleep was not considered. In another study comprising 120 women, longer daytime sleep duration shortened the labor duration among women with vaginal delivery, after institution of controls for maternal age, but did not influence the need for cesarean delivery or prolong labor in women who subsequently underwent cesarean section.[62] Additional prospective studies are necessary to investigate whether daytime naps can offset the increased risk of adverse obstetric outcomes associated with poor sleep quality and short nocturnal sleep duration.

One mechanism for these adverse outcomes may be an increased perception of pain during labor. Beebe and associates[59] reported a significant relationship between pain perception and sleep duration recorded by wrist actigraphy the night before hospitalization in women with spontaneous labor onset. This finding suggests that women with less total sleep time and more wake time on the last night before labor starts may perceive labor as more painful. Furthermore, higher fatigue levels during pregnancy may increase risk of cesarean birth. Sleep loss before hospitalization may affect motivation to comply with planned strategies for labor progression.

Fetal Complications

Maternal sleep is important for fetal growth and well-being because the secretion of growth hormone and uteroplacental blood flow are at their peak during sleep.[6,7,113] Thus persistent poor sleep may alter the uterine environment and adversely affect fetal development.[130] A recent animal study found that fragmented sleep during late gestation induced epigenetic modifications in adiponectin in offspring visceral white adipose tissue adipocytes, along with a metabolic syndrome–like phenotype. The study investigators concluded that changes elicited by disrupted sleep can impose adverse long-lasting metabolic outcomes on the next generation.[130]

Sleep disturbance associated with preeclampsia and GDM may increase risk for fetal complications. The sleep disturbance associated with shift work carried a very low risk of preterm birth in a meta-analysis of epidemiologic studies on the effects of shift work.[131] Unfortunately, the available data on preeclampsia or GDM were very limited. Thus further research is warranted to better understand potential adverse obstetric outcomes related to different types of shiftwork schedules.

Sleeping less than 5 hours/day was a risk factor for preterm birth (RR, 1.7; 95% CI, 1.1 to 2.8) after adjustment for confounding factors in a prospective cohort of more than 10,000 nulliparas.[132] In a smaller study of 166 pregnant women who completed the PSQI each trimester, poor sleep quality was associated with higher risk for preterm birth. Researchers found that each 1-point increase in PSQI score increased the odds of preterm birth by 25% in early pregnancy and by 18% in later pregnancy.[45] Similarly, Reutrakul and coworkers[124] reported that short sleep and poor sleep quality were correlated with preterm birth in a study of 169 women. A cross-sectional study also noted that sleeping longer than 8 hours was associated with higher Apgar scores (> 9) compared with sleeping less than 8 hours, but no significant difference in birth weight was noted.[128] In a case-control study of 734 pregnant women, women sleeping less than 8 hours/day were at high risk for first-trimester miscarriage (OR, 3.80; 95% CI, 1.01 to 14.3) or second-trimester miscarriage (OR, 2.04; 95% CI, 1.24 to 3.37).[133] Some studies linking sleep parameters with preterm birth or stillbirth were limited by cross-sectional data and failure to adjust for confounding variables such as preeclampsia.[58] Accordingly, findings should be interpreted with caution. In a secondary analysis of pregnant women monitored with home polysomnography, 82.4% spent some time sleeping on the back, in the supine position.[134] The supine position for any length of time is a particularly strong risk factor for stillbirth or low birth weight, because the heavy uterus compresses the mother's vena cava, thereby decreasing placental perfusion to the fetus.[134]

Maternal Psychosocial Consequences

Accumulated sleep loss, regardless of the cause during pregnancy, affects level of energy, mood, interpersonal functioning, concentration, and memory. As mentioned earlier, insufficient sleep may even trigger depression.[1,34,58,85] In a population-based survey of more than 2800 women, depressive symptoms were strongly associated with insomnia during late pregnancy, especially when sleep duration was less than 5 or more than 10 hours, when sleep efficiency was below 75%, or when sleep onset latency was prolonged.[70]

These studies cannot prove causation owing to limitations in design. Nevertheless, these reported associations can be a helpful clinical guide. With accumulated data presented in a systematic review of the published evidence on associations between sleep and pregnancy outcomes, the investigators concluded that sleep loss may be a risk factor for the development of prenatal depression or for onset of preterm labor.[60] A study comparing depressed and nondepressed pregnant women suggested that giving birth to smaller infants was associated with short sleep duration or symptoms of insomnia (difficulty

initiating or maintaining sleep, or unrefreshing sleep) among the depressed women in the sample.[121]

Persistent sleep loss also is likely to interfere with the pregnant woman's ability to work efficiently, as a result of feeling very tired, especially by the third trimester.[60] Consequently, rates of absenteeism from the workplace may increase, and risk for accidents is higher at work and on the highway.

As indicated by the available evidence, maternal-fetal health, labor duration and type of delivery, mood status, pain sensitivity, work performance, and quality of life all may be affected by sleep disturbance. Good sleep is essential to optimal maternal physical functioning and mental health. Early intervention to promote sleep may improve health outcomes for both mother and infant.

CLINICAL PEARLS

- Good sleep quality is necessary for optimal maternal health and fetal development, and for the family's smooth transition from pregnancy to early parenting. Sleep disturbance is common during pregnancy, and emerging evidence indicates a link between sleep disturbance and adverse pregnancy outcomes. Nevertheless, a clinical sleep assessment has not been part of routine prenatal care. Associations between sleep disturbance and adverse pregnancy outcomes, the diagnosis and management of sleep disorders during pregnancy, and cost-benefit analysis and long-term benefits of treating sleep disorders during pregnancy have not been well studied, suggesting that well-designed, large prospective observational studies as well as intervention trials are required to clarify these issues.
- Recognition, management, and treatment of sleep disorders are of confirmed benefit for improving short- and long-term maternal and fetal health outcomes. Preferably, the obstetric clinician and the sleep specialist should communicate and share their clinical assessments and test results in an effort to prepare a successful management plan for pregnant women at risk for adverse outcomes or events. Clinical management of sleep disturbances during pregnancy should aim to inform women about the importance of sleep, the value of sleep hygiene and nonpharmacologic options as first-line therapeutic interventions, and the likely temporary nature of the disturbance during pregnancy. Women with a preexisting sleep disorder should be advised to take added precautions during pregnancy to maintain their own health and well-being as well as those of the fetus.

SUMMARY

Physical and hormonal changes during pregnancy can cause sleep disturbances and predispose women to development of a temporary sleep disorder (such as RLS, GERD, or SDB) or may worsen a preexisting disorder. Such disorders are associated with adverse maternal and fetal outcomes. Proper diagnosis and treatment of sleep disorders are necessary to decrease adverse pregnancy outcomes, to improve maternal health during and after pregnancy, and to promote both short- and long-term optimal health consequences for the child. Thus treatment of sleep disorders during pregnancy has implications for long-term benefits for public health.[74,135] Large longitudinal studies are needed to examine prevalence and to prospectively identify the gestational time frame for onset of a sleep disorder, to examine the dose-response relationship between severity of a sleep disorder and subsequent pregnancy complications, and to test interventions for these sleep disorders. Of most importance, validated screening tools to identify sleep characteristics and disturbances specific to pregnancy are needed.

Selected Readings

Bonzini M, Palmer KT, Coggon D, et al. Shift work and pregnancy outcomes: a systematic review with meta-analysis of currently available epidemiological studies. *BJOG* 2011;**118**(12):1429–37.

Briggs GG, Freeman RK, Yaffe SJ. *Drugs in pregnancy and lactation: a refe rence guide to fetal and neonatal risk*, vol. 9. Philadelphia (Pa.): Lippincott Williams & Wilkins; 2011.

Howe LD, Signal TL, Paine SJ, et al. Self-reported sleep in late pregnancy in relation to birth size and fetal distress: the E Moe, Māmā prospective cohort study. *BMJ Open* 2015;**5**(10):e008910.

Izci-Balserak B, Pien GW. The relationship and potential mechanistic pathways between sleep disturbances and maternal hyperglycemia. *Curr Diab Rep* 2014;**14**:459–69.

Kizilirmak A, Timur S, Kartal B. Insomnia in pregnancy and factors related to insomnia. *Scientificworldjournal* 2012;**2012**:1–8.

Lee KA, Gay CL. Sleep in late pregnancy predicts length of labor and type of delivery. *Am J Obstet Gynecol* 2004;**191**(6):2041–6.

Lee KA, Zaffke ME, McEnany G. Parity and sleep patterns during and after pregnancy. *Obstet Gynecol* 2000;**95**(1):14–18.

Neyal A, Senel GB, Aslan R, et al. A prospective study of Willis-Ekbom disease/restless legs syndrome during and after pregnancy. *Sleep Med* 2015;**16**(9):1036–40.

Picchietti DL, Hensley JG, Bainbridge JL, et al. Consensus clinical practice guidelines for the diagnosis and treatment of restless legs syndrome/Willis-Ekbom disease during pregnancy and lactation. *Sleep Med Rev* 2015;**22**:64–77.

Qiu C, Frederick IO, Sorensen TK, et al. Sleep duration and plasma leptin concentrations in early pregnancy among lean and overweight/obese women: a cross sectional study. *BMC Res Notes* 2014;**7**:20.

Srivanitchapoom P, Pandey S, Hallett M. Restless legs syndrome and pregnancy: a review. *Parkinsonism Relat Disord* 2014;**20**(7):716–22.

Thorpy M, Zhao CG, Dauvilliers Y. Management of narcolepsy during pregnancy. *Sleep Med* 2013;**14**(4):367–76.

Tsai SY, Lin JW, Kuo LT, Thomas KA. Daily sleep and fatigue characteristics in nulliparous women during the third trimester of pregnancy. *Sleep* 2012;**35**(2):257–62.

Twedt R, Bradley M, Deiseroth D, et al. Sleep duration and blood glucose control in women with gestational diabetes mellitus. *Obstet Gynecol* 2015;**126**(2):326–31.

Williams MA, Miller RS, Qiu C, et al. Associations of early pregnancy sleep duration with trimester-specific blood pressures and hypertensive disorders in pregnancy. *Sleep* 2010;**33**(10):1363–71.

Wilson DL, Fung A, Walker SP, Barnes M. Subjective reports versus objective measurement of sleep latency and sleep duration in pregnancy. *Behav Sleep Med* 2013;**11**(3):207–21.

A complete reference list can be found online at ExpertConsult.com.

Chapter
157

Sleep-Disordered Breathing in Pregnancy

Francesca Facco; Judette Louis; Melissa Pauline Knavert; Bilgay Izci Balserak

Chapter Highlights

- Pregnant women may be particularly predisposed to obstructive sleep apnea and other major sleep-related breathing disorders as a consequence of the physiologic changes associated with the gravid state.
- Sleep-disordered breathing (SDB) symptoms are common during pregnancy and worsen as the pregnancy progresses.
- Outcomes that have been linked to SDB in the nonpregnant population, such as hypertension and insulin-resistant diabetes, have correlates in pregnancy (e.g., gestational hypertension, preeclampsia, gestational diabetes).
- This chapter reviews the physiology that may influence SDB prevalence and severity in pregnancy, the epidemiology of SDB in pregnancy, the possible link between SDB and adverse pregnancy outcomes, and special considerations in screening for and treating SDB in pregnancy.

PREGNANCY PHYSIOLOGY AND SLEEP-DISORDERED BREATHING

As reviewed in some detail in Chapter 156, pregancy is accompanied by certain alterations in physiology secondary to hormonal, mechanical, and circulatory changes characteristic of the gravid state. A Clinician who examines a pregnant woman for potential sleep-disordered breathing (SDB) needs to be familiar with these pregnancy adaptations; while some changes predispose to SDB, others may protect from it.

Respiratory System Changes that Predispose Pregnant Women to Sleep-Disordered Breathing

Multiple mechanisms lead to anatomic narrowing and increased resistance within the respiratory system during pregnancy. Increased levels of estrogen and progesterone induce capillary engorgement, hypersecretion, and mucosal edema of the upper airway.[1-3] These changes begin early in the first trimester and increase progressively throughout pregnancy. They may lead to a reduction in dimensions of the nasopharynx, oropharynx, and larynx, with consequent increased airflow resistance and an increase in the Mallampati score as pregnancy progresses.[4-6] Furthermore, pregnancy rhinitis is nasal congestion of pregnancy without other signs of respiratory tract infection and with no known allergic cause. This rhinitis results in difficulty breathing and resolves within 2 weeks after delivery. It occurs in up to 42% of women by the third trimester of pregnancy.[1,3]

Increased nasal congestion also may cause increased nasopharyngeal resistance and produce more intrapharyngeal pressure during inspiration. Elevated intrapharyngeal pressure during inspiration contributes to pharyngeal airway narrowing during sleep.[7] The narrowed airway results in snoring and obstructed breathing during sleep. Thus pregnant women are more likely to snore than nonpregnant women, and the prevalence of habitual snoring (on 3 or more nights/week) increases from the first to the third trimester.[5,8,9] Increased fat deposition within the soft tissue regions of the neck with weight gain of pregnancy also could cause pharyngeal narrowing and predispose affected women to snoring and SDB.[4,5,10,11] Pregnant women with a larger neck circumference and higher baseline body mass index (BMI) report more symptoms of SDB than other women.[4,5,10,12]

In addition, maternal blood volume peaks at 40% to 50% greater over baseline by third trimester. The combination of increased blood volume, interstitial fluid, and recumbent position during sleep displaces fluid, which could adversely affect upper airway patency.[2] Evidence regarding nocturnal displacement of fluid is conflicting. Whereas some studies indicated that nocturnal fluid shifting from the legs into the neck increases susceptibility to or severity of pharyngeal obstruction,[13,14] others reported that such rostral fluid shifts do not increase the frequency of obstructed breathing events.[15]

A compensatory increase in the anterior-posterior diameter of the chest and elevated diaphragm caused by the enlarging uterus result in tracheal shortening and progressive functional residual capacity reductions by 20% to 25%, expiratory reserve volume by 15% to 20%, and residual volume by 22%.[2,16] These alterations can lead to the closure of small airways during normal tidal breathing.[4,5,17] In late pregnancy, airway closure results in ventilation-perfusion mismatch and reduced gas exchange,[18,19] especially in the supine position, owing to gravity, increased intraabdominal pressure, and loss of muscle tone during sleep.[2,4,17,20]

Oxygen consumption and minute ventilation progressively increase during pregnancy by 20% and 30% to 50%, respectively.[2] The increased ventilatory drive may induce obstructive respiratory events by increasing diaphragmatic effort that creates negative inspiratory (suction) pressure on the hyperemic upper airway.[7] Furthermore, higher ventilatory drive, along with resultant respiratory alkalosis, may cause instability in respiratory control pathways, potentially increasing the

likelihood of central apnea episodes at sleep onset and during sleep.[2,20,21] However, findings from one recent study suggest that pregnancy does not increase risk for central apnea.[22]

Finally, frequent awakenings due to pregnancy-related discomfort may cause respiratory instability and periodic breathing at sleep onset.[21] The resulting sleep deprivation also can increase arousal threshold, impair upper airway muscle activity, and increase upper airway collapsibility.

Respiratory and Circulatory System Changes that May Protect against Sleep-Disordered Breathing

Several mechanisms that influence respiratory and cardiovascular changes in pregnancy may also lessen risk of apnea or hypopnea episodes. High circulating progesterone during pregnancy may protect the upper airway from obstruction by increasing upper airway dilator muscle (genioglossal) activity and its responsiveness to chemical stimuli (carbon dioxide) during sleep.[23] Pregnancy-related increases in heart rate, stroke volume, and cardiac output with reductions in peripheral vascular may diminish the impact of apneic episodes.[23] Furthermore, as pregnancy advances, women tend to spend less time in the supine position during sleep.[10,24] This may decrease the rate of adverse respiratory events during sleep, as the supine position is frequently associated with increased event rates.[9] However, a recent study reported that 82% of women spend some time sleeping in the supine position in the second and third trimesters of pregnancy.[25]

EPIDEMIOLOGY OF SLEEP-DISORDERED BREATHING IN PREGNANCY

A majority of studies evaluating the prevalence of SDB have been carried out in middle-aged nonpregnant women populations. General population studies have estimated that obstructive sleep apnea (OSA), occurs in 2% to 25% of middle-aged adults in the community, but certain populations are at greater risk,[26,27] particularly the obese and morbidly obese.[28-30] Among reproductive-age women, epidemiologic studies suggest a 2% to 13% prevalence of SDB, with higher rates in certain populations of women.[30,31] For example, among 420 premenopausal women with sleep complaints who were referred for polysomnography (PSG) sleep studies, SDB was present in 70% of those younger than 30 years and in 83% of those older than 30 years of age. Their apnea-hypopnea index (AHI) indicated that younger women had less severe SDB (AHI of 15.5 ± 22) than women older than 30 years of age (AHI of 22.4 ± 34.6).[32] Findings also suggested that SDB is common in pregnancy and worsens as pregnancy progresses.[9,33-35]

Frequent snoring during pregnancy has been well characterized, and studies have consistently demonstrated that SDB symptoms, including snoring, increase as pregnancy progresses.[8,36,37] The prevalence of pre-pregnancy snoring has been reported to be 7% to 11%.[8,36,37] By the third trimester, frequent snoring ranges from 16% to 25%.[8,36-38] Using the apnea symptom score from the Multivariable Apnea Prediction Index, one study found that SDB symptoms increased significantly from the first trimester to the month of delivery, and the increase in symptoms was not limited to snoring but included gasping, choking, difficulty breathing, and apneic events.[9] In this study, 11.4% of the participants reported an apnea symptom score increase of 2 units or more, consistent with a clinically significant increase in symptoms; these women also experienced a significant increase in subjective sleepiness compared with other women.[9]

Epidemiologic data on the prevalence of objectively assessed SDB are more limited. Olivarez and colleagues[39] performed sleep studies in 100 hospitalized pregnant women, admitted for a variety of obstetric and nonobstetric complications. The mean gestational age at the time of the sleep study was 32 weeks, and these investigators reported a 20% incidence of SDB (AHI of 5 or greater).[39] Louis and associates[40] used ambulatory sleep monitoring to assess SDB in 175 obese women at an average of 21 weeks' gestation. The prevalence of SDB was 15.4%, and most cases were mild (AHI of 5.0 to 14.9).[40]

Two studies have reported serial assessments of SDB across pregnancy.[12,41] Pien and colleagues[12] studied 105 women (28% normal weight, 24% overweight, and 50% obese). SDB was present in 10.5% of their subjects during the first trimester (median, 12 weeks). By the third trimester (median, 33.6 weeks), 26.7% had SDB.[12] Facco and associates[41] studied 128 high-risk pregnant women with one or more of the following risk factors: obesity, chronic hypertension, presentational diabetes, previous preeclampsia, and twin pregnancies.[41] Early SDB assessments were performed at 6 to 20 weeks of gestation (mean, 17 weeks) and late-pregnancy SDB assessments were performed at 28 to 37 weeks of gestation (mean, 33 weeks). In early pregnancy the frequency of mild, moderate, and severe SDB was 12%, 6%, and 3%, respectively, and these rates increased to 35%, 7%, and 5% in late pregnancy (Figure 157-1). Of the 128 women, 34 (27%) experienced a worsening of SDB during pregnancy; of the women so affected, 26 had new-onset SDB, and the other 8 had SDB in early pregnancy that became more severe. The incidence of new-onset SDB was 20%, and most of these new-onset cases were mild.[41]

Of note, all of the studies objectively assessing SDB in early or late pregnancy reported that most cases are mild

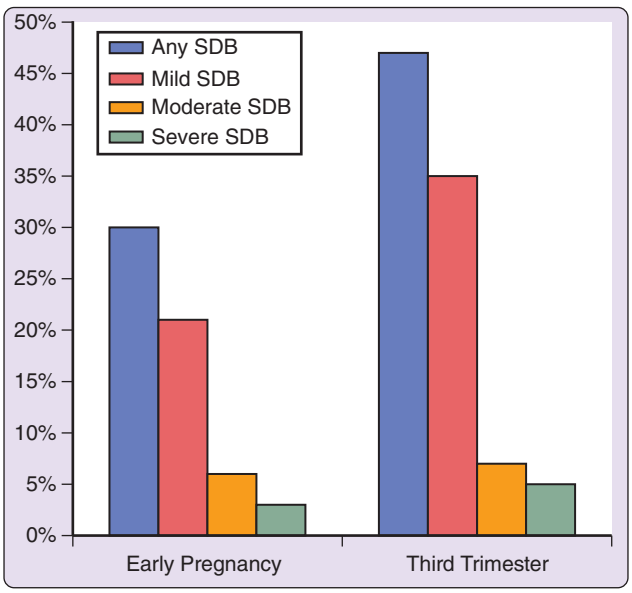


Figure 157-1 Trends in sleep-disordered breathing (SDB) across pregnancy in a high-risk cohort: early pregnancy (6 to 20 weeks), third trimester (32 to 37 weeks). (From Facco FL, Ouyang DW, Zee PC, Grobman WA. Sleep disordered breathing in a high-risk cohort: prevalence and severity across pregnancy. *Am J Perinatol* 2014;31:899–904.)

forms of SDB. Generalizability from these studies is limited, however, because they all studied high-risk populations. Pien et al attempted to correct for this limitation by using BMI distribution for all women screened for their study to estimate SDB among the general obstetrical population from which their subjects were recruited.[12] These investigators then estimated SDB prevalence in their sample at 8.4% (95% CI, 5.6% to 11.9%) in early pregnancy and 19.7% (95% CI, 15.6% to 24.4%) in the third trimester.

Risk factors for SDB in pregnancy have not been well characterized. Pre-pregnancy BMI, age, and chronic hypertension are known risk factors for SDB outside of pregnancy and are also associated with SDB in early pregnancy.[12,41] Excessive gestational weight gain is theorized to be a risk factor for developing new-onset SDB in pregnancy, but epidemiologic data supporting this theory are lacking. Facco and colleagues[41] reported that twin gestation is associated with increased risk of developing new-onset SDB in pregnancy. Although women with twin pregnancies in their cohort, as expected, had greater weight gain, weight gain did not significantly differ between women who developed new-onset SDB and women who did not. Similarly, Pien and colleagues[12] also reported that gestational weight gain was not associated with third-trimester SDB. Maternal weight gain and weight distribution may play a role in incident SDB in pregnancy, but assessing this variable by merely measuring total weight gain may be inadequate. Other measures such as changes in neck circumference, body fat composition, and trajectory of weight gain in relation to baseline BMI warrant further study as risk factors for incident SDB in pregnancy.

SLEEP-DISORDERED BREATHING AND ADVERSE MATERNAL OUTCOMES

Potential Mechanisms for Adverse Maternal Outcomes

Any underlying mechanistic pathway linking sleep disturbances and adverse outcomes for the mother is likely to be multifactorial.[42] Dysregulation of pregnancy adaptations to the cardiovascular, metabolic, and immune systems can make women vulnerable to complications.[43,44] Even small changes in sleep parameters and subtle obstructive respiratory events could exacerbate these adaptations and increase risk for adverse outcomes. SDB causes oxidative stress, autonomic dysfunction, inflammation, and altered hormonal regulation of energy expenditure.[45] These same pathways are associated with adverse pregnancy outcomes.[46,47] Figure 157-2 is a conceptual model depicting potential pathways linking SDB and pregnancy complications.

Oxidative stress, a consequence of intermittent hypoxia-reoxygenation cycles in SDB, plays a pivotal role in development of hypertensive disorders of pregnancy, inducing proinflammatory cytokines that trigger further oxidative stress, sympathetic activation, and endothelial dysfunction.[48] Increased oxidative stress also is linked to development of gestational diabetes.[49] In an animal model of SDB, gestational exposure to hypoxia was associated with increased pancreatic beta cell proliferation, cell death, impaired fetal growth, and hyperlipidemia.[50,51]

SDB leads to sympathetic nervous system and hypothalamic-pituitary axis activation.[52-54] Disproportionate sympathetic

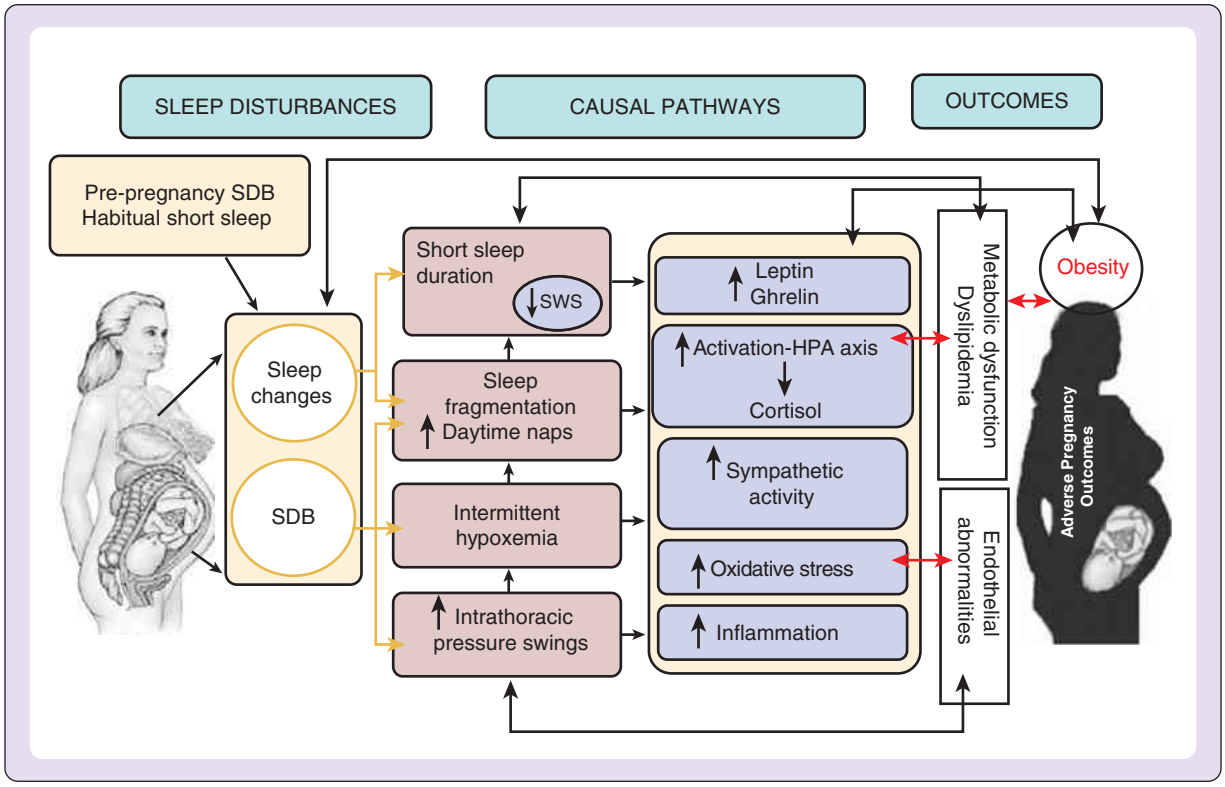


Figure 157-2 Schematic illustration of potential causal pathways linking sleep disturbances during pregnancy with adverse pregnancy outcomes. HPA, Hypothalamic-pituitary-adrenal; SDB, sleep-disordered breathing; SWS, slow wave sleep.

activation persists into the daytime, leading to increased peripheral vascular reactivity and catecholamine production, blunted baroreflex sensitivity, hindered pancreatic insulin secretion, and altered hepatic glucose release.[55] All of these downstream SDB effects have been linked to processes seen in preeclampsia: endothelial dysfunction, elevated systemic arterial blood pressure, and decreased cardiac output.[56-58] SDB also has been strongly linked to systemic inflammation, as evidenced by markers including elevated interleukin-6 (IL-6), tumor necrosis factor alpha (TNF-α), and C-reactive protein levels and leukocyte counts.[59-61] Increased inflammation in early pregnancy is associated with adverse outcomes, particularly preeclampsia and preterm birth.[46,62,63]

Slow wave delta sleep (N3 sleep) is disrupted as a consequence of SDB, and this effect may be the mechanism for adverse pregnancy outcomes. Experimental studies in healthy subjects have demonstrated that disruption of N3 sleep can adversely alter insulin and glucose metabolism, sympathovagal balance, and cortisol levels.[64-66] Sleep loss and intermittent hypoxia also have been reported to induce changes in leptin and ghrelin hormones that regulate appetite, satiety, and energy metabolism.[67-69] Recent data suggest that dysregulation of leptin and ghrelin may contribute to the pathophysiology of gestational diabetes and preeclampsia.[70,71]

Impact of Sleep-Disordered Breathing on Pregnancy

Considerable heterogeneity in the definition of SDB is apparent among the studies examining pregnancy outcomes. In some of the largest cohorts, SDB was defined on the basis of symptoms such as self-reported habitual snoring. The method of ascertainment of those symptoms varied, with some using interviews and others using questionnaires. Among studies using objective testing, both in-laboratory and portable PSG, there is variation types of devices used, AHI criteria for establishing a SDB diagnosis, and timing of assessment (prospective-longitudinal versus cross-sectional). These differences lead to difficulty in summarizing the research data and in determining significance for clinical practice.

Sleep-Disordered Breathing and Hypertensive Disorders of Pregnancy

Hypertensive disease complicates 5% to 10% of all pregnancies[72] and can be classified into subtypes according to clinical features: gestational hypertension, preeclampsia-eclampsia, and preeclampsia superimposed on chronic hypertension.[73] The risk factors include obesity and increased maternal age, which overlap with risk factors for SDB. This overlap makes it difficult to investigate potential links between the diseases. Substantial evidence, however, suggests a link between SDB and pregnancy-related hypertension, with most studies demonstrating a two-fold increase in odds of gestational hypertension and preeclampsia.[36,40,74,75]

In two of the largest epidemiologic studies to date, OSA diagnosis was associated with preeclampsia (adjusted odds ratio [OR], 1.60 and 1.89, respectively).[74,76] Although the two studies have been criticized for limitations in the quality of assessment for SDB and pregnancy outcomes, smaller studies using symptoms based on PSG-diagnosed OSA show similar findings of a two-fold increase in the likelihood of preeclampsia in pregnancies complicated by SDB.[77,78] Other studies, however, have failed to confirm this increased risk.[12,79]

Sleep-Disordered Breathing and Diabetes in Pregnancy

Gestational diabetes, a condition characterized by carbohydrate intolerance with onset or recognition during pregnancy, is a common complication, affecting 5% to 8% of pregnant women.[72,80] The incidence of gestational diabetes has increased in recent years, paralleling the increase in maternal obesity.[81] Recognized complications of diabetes include gestational hypertension, preterm birth, malformations, fetal growth restriction or macrosomia, and stillbirth.[79] Studies of gestational diabetes have included SDB as an outcome and as a predictor. The relationship between SDB and diabetes in the general population is well established. Although a causal link has not been demonstrated, SDB has been noted to precede the onset of diabetes, and initiating treatment of SDB with continuous positive airway pressure (CPAP) improves glucose control in nonpregnant populations.[82,83]

Independent of other risk factors, all patients with SDB are at increased risk for development of type 2 diabetes (see Chapter 118). Emerging data indicate that this relationship also may also occur in pregnant women specifically as it relates to gestational diabetes. A systematic review of five observational studies noted that pregnant women with SDB were at increased risk for development of gestational diabetes (adjusted OR, 1.86; 95% CI, 1.30 to 2.42).[78] The presence of SDB was ascertained with questionnaires in four studies and with PSG in one study. A prospective study using longitudinal, objective assessment of SDB demonstrated a dose-response relationship between SDB and gestational diabetes. The prevalence rates of gestational diabetes with no SDB, mild SDB, and moderate to severe SDB were 25%, 43%, and 63%, respectively.[79] Conclusions drawn from these data must consider that not all studies were able to demonstrate a relationship between SDB and gestational diabetes that was independent of BMI.

Sleep-Disordered Breathing and Severe Maternal Morbidity

Severe maternal morbidity refers to conditions or events that if uninterrupted are proximal causes of maternal death. After significant gains in reducing maternal mortality, recent years have been plagued by plateaus and slight increases in maternal death rates in even developed countries.[84] Although most studies regarding SDB and pregnancy have been underpowered to detect severe morbidity, data from a large database of delivery-related hospital discharges are compelling: Among 55,781,965 women, OSA was associated with increased risk of hospital death (adjusted OR, 5.28), pulmonary embolism (adjusted OR, 4.47), and cardiomyopathy (adjusted OR, 9.01).[74] These relationships persisted and were exacerbated by obesity.

Implications of Sleep-Disordered Breathing for Cesarean Delivery

In a large cohort study, pregnant women who snored were more likely to undergo elective cesarean (OR, 2.25; 95% CI, 1.22 to 4.18) or emergency cesarean delivery (OR, 1.68; 95% CI, 1.22 to 2.30).[85] This relationship also has been reported by other smaller, observational studies that have used both symptom-based and PSG-diagnosed SDB.[38,40,75] Although the reason for this association is not immediately clear, it is postulated to be secondary to the high prevalence of obesity,

hypertension, and diabetes among women with SDB. These conditions increase the rate of pregnancy complications and promote induction of labor, which in turn can result in a higher rate of cesarean births.

Sleep-Disordered Breathing and Adverse Fetal Outcomes

To date, the potential effects of SDB on the fetus have received limited attention. Effects may be exerted directly or indirectly by exacerbation of the underlying comorbid conditions that often track with SDB. Initial reports of fetal complications came from case reports of fetal growth restriction and fetal heart rate decelerations.[77] In many of those cases, however, confounding comorbid conditions such as hypertensive disease and diabetes coexisted and associations could not be assumed. More recent studies have attempted to examine that relationship more closely.

Stillbirth

Stillbirth is a fetal death at or beyond 20 weeks of gestation.[72] No large-scale studies have been conducted to examine the role of SDB in stillbirth. Risk factors for stillbirth overlap with SDB risk factors and include advancing age, African American ancestry, smoking, maternal conditions such obesity, diabetes or hypertension, fetal growth restriction, and previous adverse pregnancy outcomes.[86] These correlations provide biologic plausibility for an association between stillbirth and maternal SDB, but such an association has not yet been demonstrated.

Miscarriage

Miscarriage or spontaneous abortion involves the loss of a pregnancy, usually within the first 3 months of conception. The estimated frequency of spontaneous abortion is between 12% and 24% of all clinically identified pregnancies. The risk factors for miscarriage include extremes of age, smoking history, obesity, previous miscarriage, hypertension, and diabetes. All of these are also overlapping risk factors for SDB. Data linking SDB and miscarriage are limited, and any discussions are mostly theoretical in nature. In a retrospective review of sequential clinic charts of 147 premenopausal women who had been referred to a sleep disorders clinic for an evaluation of sleep complaints, an association between SDB and number of miscarriages was demonstrated. In that review, overweight or obese women with SDB, especially those with moderate to severe SDB, were more likely to have had a miscarriage than women without SDB.[87]

Preterm Delivery

Preterm births are births that occur before 37 weeks of gestation. Although preterm birth occurs in 11.6% of births, it is responsible for a significant amount of neonatal morbidity and mortality. Preterm birth may be classified as spontaneous or medically indicated if it was precipitated by obstetric intervention for maternal or fetal benefit.[72] Data on preterm birth and SDB have been inconsistent. Large cross-sectional studies of SDB in pregnancy have reported higher risk of preterm delivery.[74,88] However, these studies were unable to differentiate spontaneous from medically indicated preterm birth. One smaller retrospective study noted an increase in medically indicated preterm delivery associated with preeclampsia among women with SDB[75] and a higher rate of preterm birth among pregnant women with SDB (29.8%) compared with control subjects of normal weight (12.3%).

Fetal Heart Rate Abnormalities

The fetal nonstress test was introduced in 1975 to describe fetal heart rate acceleration in response to fetal movement as a sign of fetal health.[72] It currently is the most common form of fetal assessment in obstetrics. A few studies have attempted to examine fetal well-being in response to nocturnal desaturations and results are conflicting. In a small study, 3 of the 4 women with snoring had fetal heart decelerations that accompanied maternal desaturation, but types of decelerations were not characterized.[89] In larger prospective studies with sample sizes of 20 women with PSG evidence of SDB, findings have not been replicated.[90,91] In these studies, despite episodes of oxygen desaturation, no fetal heart rate decelerations were noted during apneic events. In summary, it is unclear whether fetal hypoxia during maternal apnea occurs and, moreover, whether it is a primary contributor to adverse pregnancy outcomes associated with SDB.

Fetal Growth Abnormalities

Intrauterine growth restriction (IUGR) is defined as retardation of fetal development resulting in small size in relation to gestational age, with most studies using less than the tenth percentile as a cutoff growth criterion.[72] Findings from retrospective and observational studies of SDB and IUGR are mixed, with some failing to show any association. However, studies point to an increased likelihood of fetal growth restriction among pregnant women with moderate to severe SDB (OR, 1.44; 95% CI, 1.22 to 1.71).[92]

Low birth weight is an important cause of neonatal morbidity and is defined as birth weight less than 2500 g. Two potential mechanisms for low birth weight are well documented: preterm delivery and growth restriction. Regardless of the etiology, low birth weight is associated with higher rates of short- and long-term morbidity[72] and with maternal SDB among pooled studies (unadjusted OR, 1.39; 95% CI, 1.14 to 1.65).[78] Most studies have not found a difference between mothers with and without SDB when birth weight is assessed as a continuous value.[78] Two large studies found significant differences in birth weight, with maternal SDB–exposed neonates weighing 100 g less than unexposed neonates. However, this difference is of questionable clinical significance.[75,93]

Most of the focus on growth abnormalities and SDB has been on growth restriction. However, it is also important to consider that there may also be an association with large for gestational age (LGA) neonates due to the increased prevalence of obesity and diabetes among women with SDB. LGA neonates have higher rates of birth trauma, respiratory complications, and short- and long-term morbidity.[72] One study found more LGA infants born to women with SDB than to obese and normal-weight control subjects (17% vs. 8% and 2.6%, respectively).[75]

SCREENING FOR SLEEP-DISORDERED BREATHING IN PREGNANCY

Despite the growing prevalence of SDB and risk factors for SDB in pregnancy, obstetric providers generally remain unaware of this important sleep disorder. In a survey of

practitioners and patients regarding prenatal sleep assessment, less than 3% of clinicians reported routinely asking patients about snoring. Yet 32% of women reported that they snored, and only 5% reported being asked about snoring during a prenatal visit.[94]

Investigators have struggled to identify the most appropriate screening tools for SDB in pregnancy. As with all screening tools, questionnaires need to be easily implemented, inexpensive, and useful in the clinical setting for identifying patients requiring referral for more definitive diagnostic testing. Canonical symptoms of SDB include excessive daytime sleepiness, snoring, and witnessed pauses during sleep. Excessive daytime sleepiness most typically has been assessed with the Epworth Sleepiness Scale (ESS). However, because daytime sleepiness is so prevalent in pregnancy, the ESS does not seem to inform risk assessment of pregnant patients for OSA.[95,96]

Pre-pregnancy BMI and maternal age are inconsistent predictors of SDB. Older women and those entering pregnancy with higher baseline BMI, however, are at high risk for pregnancy-onset SDB.[36] In a study using PSG to assess for SDB in the first and third trimesters, maternal weight before pregnancy and maternal age were major predictors of SDB risk.[97]

Snoring may be one of the most useful single symptoms to identify at initial clinical assessment. In nonpregnant patients, habitual snoring has good correlation with PSG.[98,99] For example, for women who say they "often" snore, the PSG odds ratio is 3.8 higher than for nonsnoring peers. Similarly, for women who say they usually (always or almost always) snore, the PSG odds ratio is 16.3 higher than that for nonsnoring peers.[98] Thus, in the first trimester, habitual pre-pregnancy snoring may be a useful indicator of *preexisting* OSA. The Berlin questionnaire is widely used for SDB screening in nonobstetric populations, and a high-risk Berlin score has a sensitivity ranging from 68% to 86% and a specificity ranging from 46% to 95% for OSA.[100] In pregnancy, however, the Berlin score has poor predictive capacity (sensitivity of 35% to 39%, specificity of 64% to 68%).[39,96] By contrast, a pregnancy-specific tool that includes frequent snoring ("yes or no"), chronic hypertension ("yes or no"), and maternal age and baseline BMI as continuous variables performed well in predicting SDB in early pregnancy.[96]

In summary, habitual snoring, chronic hypertension, baseline BMI greater than 25 to 30 kg/m^2, and older maternal age are easily ascertained assessment elements that can effectively indicate the risk for either preexisting SDB or pregnancy-onset SDB.

DIAGNOSING SLEEP-DISORDERED BREATHING IN PREGNANCY

The available evidence is insufficient to guide diagnostic evaluation of SDB specific to pregnant patients. Accordingly, pregnant women in whom SDB is suspected should be evaluated in line with standard guidelines that recommend evaluation by a sleep specialist for a sleep-directed history and physical examination and sleep testing.[101] Sleep testing can be reasonably accomplished by laboratory PSG,[97] or home testing can be performed using type 3 portable monitors.[40] Validation of home monitors for pregnant women is limited, but the convenience makes it a viable option for women who are unable to spend a night away from home for a PSG study. The usual caveats regarding decreased negative predictive value associated with home monitoring apply to pregnant patients, and the need for laboratory PSG for sleep testing in patients with comorbid cardiac, pulmonary, psychiatric, or neurologic disease also applies to pregnant women.[102]

TREATMENT OF SLEEP-DISORDERED BREATHING IN PREGNANCY

Women with a preexisting SDB diagnosis and an established treatment regimen should continue that treatment during pregnancy. CPAP is considered to be safe and effective during pregnancy.[103,104] AHI increases appear to be relatively modest for most women,[9] and a pressure setting increase of 1 to 3 cm H_2O during the second trimester usually is needed.[103,104] CPAP settings are now more easily monitored because devices function in an autoset mode that adjusts pressures within a designated range and reports compliance and residual AHI data. Pregnancy-induced nasal congestion and increased BMI may necessitate adjustments in mask fit and humidification.

Use of a mandibular advancement device (MAD) with previously documented efficacy could be continued during early pregnancy. One study, however, demonstrated autoset CPAP superiority over MAD plus nasal strip in treating SDB in pregnant women.[105] At the very least, effectiveness of MAD should be monitored and a sleep study should be done early in the third trimester, when women are likely to have a higher AHI and require increased CPAP support.[103,104] Positional therapy can be recommended as an adjuvant, in the absence of extenuating circumstances, because most women have a positional component to their SDB, and nonsupine sleep generally is preferred during pregnancy.[106] Postpartum AHI values can be expected to return to pre-pregnancy levels.[107]

In women with known preexisting SDB not already undergoing treatment, immediate initiation of a CPAP therapy regimen is indicated. Autoset CPAP with data download is appropriate for these women and allows for rapid treatment, tracking, and adjustment as pregnancy progresses. MAD treatment is not recommended, because it generally takes several weeks to months to fit, adjust, and test the device. In the absence of a preexisting SDB diagnosis, data are lacking to support a strategy of universal screening for SDB in pregnancy. It is important to recognize, however, that as the general population is becoming more obese, clinicians are likely to encounter more women with symptomatic SDB in pregnancy. Obstetric care providers should refer any patient with suspected SDB to a sleep specialist for diagnosis and possible treatment.

To date, studies examining the effect of CPAP treatment on pregnancy end points have been insufficiently powered or limited in the scope of the end points.[103,104,108-110] The largest of these trials followed women already diagnosed with preeclampsia who were treated with CPAP and used improvements in fetal movement and cardiac output as primary clinical end points.[108,109]

In caring for obstetric patients with SDB, it is important to consider increased risk of perioperative complications associated with SDB.[111,112] An analgesic strategy that minimizes the need for systemic opioids should be considered. If used, opioids should be prescribed as single doses rather than by standing order. Monitoring maternal oxygen saturation should

be considered if systemic opioids are administered, and women should wear their CPAP device while in the hospital recovering from labor and delivery, as well as after discharge to home. Predelivery consultation with an anesthesiologist to plan intrapartum and postpartum pain management also should be considered.

CLINICAL PEARLS

- SDB prevalence and severity increase from the first to the third trimester of pregnancy.
- Risk factors for SDB in pregnancy have not been well characterized. Habitual snoring, chronic hypertension, maternal baseline BMI greater than 25 to 30 kg/m^2, and older maternal age are easily obtained information that may effectively indicate risk of either preexisting SDB or pregnancy-onset SDB.
- Data suggest that SDB during pregnancy may increase the incidence of adverse pregnancy outcomes such as gestational hypertension, preeclampsia, and gestational diabetes. Many studies, however, did not control for obesity, a strong risk factor for both SDB and adverse outcomes, nor did they clearly define a temporal relationship between SDB and the subsequent development of adverse outcomes.
- In pregnant women with a preexisting SDB diagnosis and an established treatment protocol, treatment should be continued during the pregnancy and the AHI evaluated for an increase in the early third trimester.
- In pregnant women without preexisting SDB, data are currently lacking to recommend a strategy of universal screening for SDB in pregnancy.

SUMMARY

SDB in pregnancy is an ongoing area of research. It is clear that SDB prevalence and severity increase as pregnancy progresses, especially among high-risk obese women. Data also suggest that SDB during pregnancy may increase the incidence of adverse pregnancy outcomes such as gestational hypertension, preeclampsia, and gestational diabetes. Many of the reported studies, however, did not control for obesity, a strong risk factor for both SDB and adverse pregnancy outcomes, nor did they clearly define a temporal relationship between SDB and subsequent development of adverse outcomes. The optimal way to screen for SDB during pregnancy has yet to be determined, but data suggest that instruments used in nonpregnant populations (e.g., Berlin questionnaire, ESS) perform poorly in pregnancy. Pregnant women in whom SDB is suspected should be evaluated and treated using standard guidelines; however, the available evidence currently is too limited to suggest that treatment of SDB during pregnancy can alter pregnancy outcomes.

Selected Readings

Chen YH, Kang JH, Lin CC, et al. Obstructive sleep apnea and the risk of adverse pregnancy outcomes. *Am J Obstet Gynecol* 2012;**206**(2):136 e131–5.

Facco FL, Ouyang DW, Zee PC, Grobman WA. Development of a pregnancy-specific screening tool for sleep apnea. *J Clin Sleep Med* 2012;**8**(4):389–94.

Izci-Balserak B, Pien GW. Sleep-disordered breathing and pregnancy: potential mechanisms and evidence for maternal and fetal morbidity. *Curr Opin Pulm Med* 2010;**16**(6):574–82.

Lockhart EM, Ben Abdallah A, Tuuli MG, Leighton BL. Obstructive sleep apnea in pregnancy: assessment of current screening tools. *Obstet Gynecol* 2015;**126**(1):93–102.

Louis J, Auckley D, Bolden N. Management of obstructive sleep apnea in pregnant women. *Obstet Gynecol* 2012;**119**(4):864–8.

Louis J, Auckley D, Miladinovic B, et al. Perinatal outcomes associated with obstructive sleep apnea in obese pregnant women. *Obstet Gynecol* 2012;**120**(5):1085–92.

O'Brien LM, Bullough AS, Owusu JT, et al. Pregnancy-onset habitual snoring, gestational hypertension, and preeclampsia: prospective cohort study. *Am J Obstet Gynecol* 2012;**207**(6):487 e481–9.

Pamidi S, Pinto LM, Marc I, et al. Maternal sleep-disordered breathing and adverse pregnancy outcomes: a systematic review and metaanalysis. *Am J Obstet Gynecol* 2014;**210**(1):52.e1–14.

Rice JR, Larrabure-Torrealva GT, Luque Fernandez MA, et al. High risk for obstructive sleep apnea and other sleep disorders among overweight and obese pregnant women. *BMC Pregnancy Childbirth* 2015;**15**:198.

Romero R, Badr MS. A role for sleep disorders in pregnancy complications: challenges and opportunities. *Am J Obstet Gynecol* 2014;**210**(1):3–11.

A complete reference list can be found online at ExpertConsult.com.

Postpartum Period and Early Motherhood

Robyn Stremler; Katherine M. Sharkey; Amy R. Wolfson

Chapter 158

Chapter Highlights

- Sleep patterns are greatly disturbed during the postpartum period, particularly in the first few months, resulting in considerable daytime sleepiness and fatigue in the new mother. Infant care demands contribute to the poor sleep and altered circadian rhythms experienced by postpartum women.
- Multiple influences on maternal sleep are recognized, including the presence of other children in the home, return to work outside the home, infant feeding method, and infant sleep location. Women who report significant sleep loss and infant sleep problems are more likely to experience postpartum depression. Few evaluations of interventions to improve maternal sleep in the postpartum period have been performed.
- Reviewed in this chapter are the sleep changes that occur during the postpartum period and the impact of sleep patterns on maternal health during the early months of motherhood. Conclusions are presented within the context of sleep and circadian rhythm disturbances for childbearing women, and the implications for maternal role functioning and maternal–infant interaction are discussed.

OVERVIEW

It is likely that as long as there have been mothers, the experience of sleep disturbance and fatigue in the postpartum period has existed. However, sleep is difficult to study during postpartum recovery and the first few months after birth for several reasons: As might be expected, new mothers find it difficult to attend sleep laboratories for study; home polysomnography (PSG) may be intrusive; and sleep data collection, even by actigraphy or questionnaire, may represent a significant burden in the face of the demands of infant care.

Sleep disturbance during the postpartum period and its effects on maternal role functioning and maternal-infant interactions are not well understood. Confounding factors such as maternal age, type of delivery, type of infant feeding, infant temperament, return-to-work issues, previous birth experience, number of other children in the home, and availability of nighttime support from the partner or others may have an impact on quality and quantity of sleep in new mothers, especially until the infant begins to sleep through the night. Interventions to improve maternal sleep and fatigue in the postpartum period are limited, perhaps because of the universal nature of the experience and the belief that disturbed sleep is an unavoidable part of motherhood. Furthermore, consensus is lacking on what intervention modalities are most effective or regarding which women are most likely to benefit from treatment, although one intervention study showed greater improvements in sleep in women at higher risk for sleep disruption, including those with worse baseline sleep and greater social disadvantage[1] (see later under Interventions to Improve Maternal Sleep in the Postpartum Period).

PHYSIOLOGIC CHANGES DURING THE POSTPARTUM PERIOD

First-time mothers experience more disrupted sleep patterns than experienced mothers[2] (Box 158-1). All new mothers are at risk for development of anemia from iron demands of pregnancy, blood loss during delivery, common infections (endometritis, urinary tract, mastitis), and thyroid dysfunction. Any woman experiencing high levels of fatigue in the first few months after childbirth should be assessed for these conditions and treated.[3]

CIRCADIAN RHYTHMS AND THE POSTPARTUM PERIOD

Although growth hormone, prolactin, melatonin, cortisol, and thyroid-stimulating hormone each have a distinct circadian pattern of secretion, little is known about changes that may occur in the postpartum period and effects on sleep and daytime sleepiness. Delivery of the placenta reduces estriol and progesterone levels,[4] with accompanying increases in oxytocin and prolactin.

Melatonin levels and patterns of secretion may be altered in postpartum women by increased nighttime light exposure during provision of infant care, and by decreased daylight exposure secondary to limited activity outside the home. Changes in circadian rhythms, including melatonin secretion patterns, may influence subsequent sleep quality. Yet few studies have examined the influence of circadian rhythm disruption on sleep disturbance or fatigue in the postpartum period. In fact, the most commonly dispensed advice

Box 158-1 TYPICAL HORMONE, SLEEP, AND BEHAVIOR CHANGES DURING THE POSTPARTUM PERIOD

Hormones

Abrupt drop in progesterone levels from those in pregancy
Abrupt drop in estrogen levels from those in pregnancy
Prolactin levels fluctuate with lactation
Melatonin amplitude decreased

Sleep Patterns and Architecture

Sleep disrupted owing to infant care and feeding
Sleep more disrupted for first-time mothers than for experienced mothers
Tendency to sleep later into the morning hours
Increased slow wave sleep (delta sleep [i.e., N3 sleep stage])

Sleepiness and Sleep Challenges

Increased daytime sleepiness
Difficulty with sleeping; new mothers may nap if not employed outside the home
Issues of bed-sharing emerge

Mental Health Concerns

Risk of postpartum depression

regarding maternal sleep, the admonishment to "sleep when the baby sleeps," may not promote circadian rhythm entrainment. Sleep at night is most restorative when it occurs during usual sleep hours; entraining the mother's sleep to the infant's is less ideal than finding ways to more quickly facilitate the newborn's sleeping through the middle of the night.

Preliminary evidence in 22 infants studied between 2 and 10 weeks post partum indicates that greater daytime light exposure is associated with higher circadian amplitude, suggesting that manipulating ambient light may promote development of more consolidated sleeping patterns in infants.[5] Women at 4 and 10 weeks post partum exhibited elevated daytime melatonin excretion as well as blunted circadian rhythm amplitude relative to those nonpregnant nulliparous women,[6] and a recent study of circadian phase across the perinatal period showed differences in the time of dim-light melatonin onset of up to 2.5 hours in 12 women from the third trimester to 6 weeks post partum.[7] In addition, most women in this small sample showed a shortening of phase angle between melatonin onset and self-selected bedtime, indicating that they may go to bed "earlier" in their biologic night during the postpartum period. These differences suggest that altered circadian rhythms or failure of circadian entrainment may contribute to postpartum sleep disruption.[6,7]

SLEEP DURING POSTPARTUM RECOVERY

For the most part, maternal sleep efficiency is poor on the first night after delivery but begins to improve over the following days.[8] Being in labor during the night may be related to depressed mood during the first week after delivery[9] (see later under Postpartum Depression and Disturbed Sleep). Postpartum recovery generally is defined as that occurring during the first 6 months after delivery, although most women would say that the postpartum period continues from birth until the infant is sleeping through the night with predictable day and night sleeping patterns.[10] Average maternity leave for employed women varies by region; it is 12 months in Canada and around 10 weeks in the United States.[11] These differences in return to employment undoubtedly affect the mother and her family's sleep experience in the first year, but such comparisons have not been studied.

In a qualitative study of women during both pregnancy and the postpartum period,[12] naps were used as a way of coping with sleep loss at night. However, paid work demands and partner availability for child care determined if naps could occur. In a study of Israeli mothers still on maternity leave at 4 to 5 months post partum, findings included more infant wake time at night, shorter infant daytime naps, more maternal night awakenings, and greater parenting stress when compared with the mothers who returned to work.[13] The investigators suggested that women on leave from paid employment may not find opportunities to nap, and that perhaps women who return to work experience less stress because they have more control over their own schedule, with breaks from continuous caregiving. An alternate explanation is that returning to work was more feasible for a woman if her infant was sleeping more at night and if her family was experiencing less stress.

In the early postpartum period, a rapid decline in all placental hormones is seen, and 35% to 80% of women experience postpartum distress or "blues" approximately 3 to 5 days after childbirth. For most women, theses blues are confined to the first 2 weeks, but a major episode of postpartum depression can manifest at any time between 4 weeks and 12 months after childbirth.[14,15] Approximately 10% to 15% of women experience a major depressive episode in the postpartum period, a prevalence that is similar to that for women at other stages of life (see later under Postpartum Depression and Disturbed Sleep).

Both self-report and actigraphy studies have demonstrated that mothers have disturbed sleep after the birth of their baby and that they experience more nighttime awakenings from disruptions during the first 4 weeks in comparison with the end of pregnancy and later postpartum months.[16-19] Although total sleep time may change little over this period, a mother's sleep is far more disrupted and less efficient during these early postpartum weeks than at other points in her life. Women seem to compensate for their sleep disruptions by spending more time napping and sleeping later in the morning during the first month.[16-20] Compared with those during early postpartum recovery, average wake and rise times occur earlier at 3 to 4 months and at 12 to 15 months,[19] presumably because some women have returned to paid employment by those time points.

Although women often are advised to "sleep when the baby sleeps," investigations of napping in the postpartum period are few. A study of 51 mothers of 3-month-old infants examined factors related to mothers' decisions to take naps.[21] Nap taking was predicted by maternal perceptions of sleep disturbance, rather than by actual amount of sleep or infant wake time at night. Mothers who were engaged in more hours of employment or who had more children at home reported less napping. Almost one half of the women took no naps at all, and among those who did nap, the average length was 30 minutes. Similarly, actigraphic evidence determined that after postpartum week 2, fewer than 50% napped at least once per week, and among women who did nap, it was an average of only twice

a week.[22] A small study designed to assess the relationships between maternal sleep disturbance and fatigue and mother-infant interactions found that maternal napping was associated with improved interactions related to fostering cognitive growth.[23] Further research is needed to determine optimal timing and duration of napping and its effects on daytime alertness and cognitive function in the postpartum period, as well as to investigate any deleterious effects of napping, such as circadian disruption or insomnia.

Researchers have examined changes in fatigue from pregnancy into the postpartum period and through the year after birth, using self-report and physiologic measures such as progesterone levels, thyroid functioning, and iron or other nutrient deficiencies.[3,24-26] On the basis of these studies, primiparas and multiparas do not differ significantly in their perception of fatigue. The 3 to 6 weeks after delivery is especially fatiguing for first-time mothers, however, and fatigue remains high during the first 3 months post partum compared with pre-pregnancy reports.[24,26] Postpartum fatigue is associated with reduced sleep time, low hemoglobin, and low ferritin levels.[24] When these measures are taken together, researchers have concluded that the level of progesterone is unrelated to fatigue and that social factors, such as employment and family responsibilities, influence perception of fatigue during the postpartum period.[24,26] Work is emerging to suggest that the chronic partial sleep loss and postpartum fatigue experienced by women in the early months of caring for a new infant are associated with impaired cognitive performance.[27] The time course over which women return to their baseline levels of sleep quality and daytime functioning after delivery is poorly understood.

A small number of studies have described sleep architecture in the postpartum period. It appears that mothers have less stage 1 (N1) and stage 2 (N2) sleep and a significant increase in N3 slow wave (delta) sleep at 1 month post partum.[2,16,24,28-30] These sleep pattern alterations may be caused in part by prolactin-mediated increases in N3 (see later under Breastfeeding versus Formula Feeding).

INTERVENTIONS TO IMPROVE MATERNAL SLEEP

Interventions to prevent infant sleep problems have focused on infant sleep outcomes without measuring parental sleep outcomes.[31-33] These interventions typically provided parents with information related to infant sleep and strategies to limit development of unwanted sleep associations, to increase the infant's self-soothing ability, and to facilitate day-night entrainment. Although none of these studies evaluated the effects of the intervention on parental sleep outcomes, Wolfson and colleagues[33] examined effects on parents' stress and found that parents who received training on infant sleep reported significantly fewer life "hassles" than parents in the control group when the infant was 6 to 9 weeks of age. Similarly, several randomized clinical trials (RCTs) conducted with older infants[34-37] focused on providing parents with information on infant sleep and strategies to decrease night feeding and waking, but with no advice related to improving parents' sleep. Three of the studies[34,35,37] included maternal sleep outcomes and found reductions in problematic infant sleep behaviors and a concomitant improvement in self-reported sleep outcomes using the Pittsburgh Sleep Quality Index (PSQI).

Inasmuch as infant and parent sleep are presumed to affect one another, it is surprising that few studies have examined strategies to improve parents' sleep. Two samples of expectant parents were enrolled in an RCT of a modified sleep hygiene intervention to reduce parental nighttime sleep disruption.[1] Subjects in the experimental groups were seen in the last month of pregnancy and instructed to have the infant in a bassinet close to the maternal bedside, to use a white noise machine, and to use a nightlight to ensure low lighting. The two samples of first-time expectant parents consisted of those in stable relationships (n = 118) from fee-based childbirth classes and women (n = 122) recruited from free prenatal classes and clinics serving low-income women. In the socioeconomically advantaged sample, no group differences were seen in any sleep outcomes measured by actigraphy. In the low-income sample (in which the subjects reported worse sleep at baseline than socially advantaged participants), the intervention group attained significantly more nocturnal sleep (7.1 hours vs. 6.5 hours), better sleep efficiency (80% vs. 75%), and less wake time after sleep onset (19% vs. 23%) than control subjects at 3 months post partum.

A pilot RCT tested a behavioral-educational intervention designed to increase nighttime sleep and sleep continuity for 30 mother-infant dyads in the early postpartum period.[38] First-time mothers in the intervention group received a 45-minute meeting with a nurse in the immediate postpartum period in hospital, a written booklet, and weekly phone contact to reinforce information and problem-solve. On the basis of actigraphy data, women in the intervention group achieved almost an hour more sleep than the control group. In the large-scale RCT (n = 246) that followed this pilot study, no group differences were found on any outcome at 6 or 12 weeks post partum.[39] Although women in the larger trial were ethnically diverse, high socioeconomic status and support may explain lack of effect of the intervention.

The few published intervention studies to improve parent sleep provide little support for such strategies. Socially disadvantaged families may benefit most from advice regarding infant and maternal sleep, and future interventions should be tested in this population. An exploration of parents' beliefs around infant and parent sleep, and their experience of barriers, facilitators, and motivations related to implementing sleep strategies, would provide insight for development and testing interventions at points farther along in the first postpartum year.

Cognitive-behavioral therapy for insomnia (CBT-I) may be a feasible and effective treatment for women with postpartum sleep difficulties. In an open pilot study of CBT-I, Swanson and colleagues[40] treated 12 new mothers with comorbid postpartum depression and insomnia. They adapted core CBT-I strategies to accommodate infant care and also included components on facilitating infant sleep and requesting assistance with infant care. Participants demonstrated significant improvements in diary-reported sleep efficiency, sleep latency, and total sleep time, as well as a reduction in depressive symptoms. Future studies could examine this intervention in larger, randomized samples and test its efficacy as a preventive strategy.

Hall and colleagues[41] tested an intervention to help parents resolve their infant's sleep problems. In this pre- and post-test study of 39 couples, a statistically significant improvement in subjective sleep quality (2.2 point decrease in PSQI scores

from baseline) was achieved for both mothers and fathers at 6 weeks after the intervention. Without a control group, it is difficult to ascertain whether the improvement was due to the intervention or to better infant sleep with development over the 6-week period of the study.

Breastfeeding versus Formula Feeding

During the first 2 weeks after birth, rapid eye movement (REM) sleep is stable for women who breastfeed, but it gradually decreases for women who are not lactating.[42] Some parents choose to supplement infant feeding with formula in the hope that the infant would be better satiated, settle faster, and sleep longer at night, so that they could achieve more nighttime sleep.[43] Studies examining the relationships between infant feeding method and parental sleep with objective measures of sleep have found either no difference in total nighttime sleep[20,44] or preservation of sleep[43] for parents whose infants are breastfed. Although the available evidence related to characteristics of parent and infant sleep is conflicting, promotion of breastfeeding is recommended to optimize many other important infant and maternal health outcomes. Because supplementation of breastfeeding with formula often is perceived as a strategy to increase infant sleep at night, parents should be made aware that research findings do not support a recommendation of formula feeding as a means to improve sleep. For women planning to breastfeed, it may be helpful to know that no study asking parents to report on the quality of their sleep has reported any difference between parents with breastfed and those with formula-fed infants.

For lactating women, basal levels of prolactin are high, and a burst of prolactin secretion occurs at the onset of each breastfeeding event, regardless of when sleep occurs, although these bursts seem to be of a higher magnitude in the evening than in the morning.[45] Both basal levels and bursts of prolactin diminish to pre-pregnancy levels by approximately 3 months post partum. Blyton and colleagues[46] studied 31 women in their home environment with portable PSG. Lactating women had less light sleep (N1 and N2), fewer arousals, and more deep sleep (N3), especially in the second half of the night, compared with nonlactating postpartum women. No difference was observed in the amount of REM sleep or total sleep time between the two groups.[46] Within 24 hours of weaning, prolactin levels return to low basal levels and to the circadian sleep–associated patterns found in healthy adults.[47]

Bed-Sharing and Room-Sharing

Bed-sharing is defined as an infant sleeping in the same bed with the caregiver; previously, this practice was commonly referred to as cosleeping. Room-sharing refers to sleep situations in which the infant sleeps in the same room as, and in close proximity to, a parent or caregiver, but does not occupy the same bed. Room-sharing is the infant sleep arrangement recommended by the American Academy of Pediatrics and the Canadian Paediatric Society. Although few studies have examined the relationship between bed-sharing or room sharing and mothers' sleep-wake patterns, bed-sharing has become increasingly common during the postpartum period.[48-50] The proportion of usually bed-sharing infants in the United States rose from 5.5% in the period 1993-1994 to 12.8% in 1999-2000[50] and to 13.5% in 2010.[51] Bed-sharing has been associated with longer duration of breastfeeding.[52,53] Some studies suggest that bed-sharing reduces the risk of sudden infant death syndrome (SIDS), because a bed-sharing infant has more arousals from the mother's body heat, sounds, oxygen and carbon dioxide exchange, smells, movement, and touch.[49] It has been postulated that infants are at risk for SIDS because of an immature neurologic system and difficulty arousing from sleep to breathe.[49] However, greater risk of SIDS or asphyxiation for bed-sharing infants may be directly related to overheating, the presence of pillows and soft surfaces, and exposure to adults who are smokers, extremely fatigued, intoxicated, or obese.[54,55] In contrast with risks imposed by bed-sharing, room-sharing may decrease risk of SIDS by providing infant proximity to parental cues for arousal.[54,55]

Bed-sharing mothers report disrupted sleep that also is objectively verifiable. Mosko and colleagues[48] assigned 20 usually bed-sharing and 15 usually separate-sleeping mother-infant dyads to bed-share for one night and then to sleep separately the following night. In this PSG study, bed-sharing had no effect on REM sleep but increased the number of arousals and modestly reduced N3 deep sleep while increasing light sleep (N1 and N2) in the group of breastfeeding women. Women who routinely had their infant sleep in a separate room rated their sleep quality as lower on the night when they were assigned to bed-share, indicating that occasional bed-sharing may be more disruptive to self-reported sleep quality than routine bed-sharing.

In another study, first-time mothers (n = 246) provided information regarding infant sleep location at 6 and 12 weeks post partum along with actigraphy evaluation of sleep and subjective maternal sleep disturbance ratings.[56] Room-sharing was the most typical sleeping arrangement (for 46% of infants at 6 weeks and 39% at 12 weeks). At 6 weeks, 17% of families were usually bed-sharing; rates decreased to 12% at 12 weeks. At 6 weeks, usually bed-sharing mothers had shorter stretches of sleep than mothers with infants sleeping alone (130 vs. 156 minutes; P = .03) and more awakenings than mothers who were usually room-sharing or whose baby slept alone (11 vs. 9 vs. 8; P = .001). At 12 weeks, usually room-sharing mothers had shorter stretches of uninterrupted sleep than solitary sleepers (164 vs. 192 minutes; P = .04). No significant difference in maternal subjective sleep disturbance was found between infant sleep location groups. Because the trend for infant bed-sharing is on the rise,[51] the impact of bed-sharing or room-sharing on sleep for new mothers and fathers, as well as the infant's sleep, requires further investigation.

Sociocultural Factors

International and cross-cultural studies have demonstrated that mother-infant sleep behaviors differ in the postpartum period. The international literature highlights significant cultural variation in the sleep environment, infant care practices, and, of course, the meanings attributed to them. Ethnographic data suggest that sleep is construed as a form of social behavior in many societies.[57] Whereas a solitary-sleeping infant is a common goal for families in Western postindustrial settings, close maternal-infant contact at night is the cultural norm in many other cultural groups.[58] For example, bed-sharing is a common practice in Malaysia, Brazil, Thailand, and Japan, and among Maori and Pacific Islanders.[59-63] Outside of bed-sharing at night, infant sleep may occur in other locations, such as outdoors, as practiced in Finland and other Northern European nations,[64] or while the infant is carried in a sling, as in African hunter-gatherer cultures while the mother works during the day.[65,66]

Common elements in the social structuring of the postpartum period within a culture, such as mandated rest or sleep and assistance in tasks from relatives or midwives,[67] are likely to have effects on maternal sleep and well-being. In keeping with evidence of variation in risk factors and rates of postpartum depression across cultures, new mothers' experiences vary greatly; accordingly, it is imperative that their sleep patterns continue to be studied across cultures. These types of studies will add to a more nuanced understanding of sleep changes for mothers as well as to development of new interventions and strategies to maximize the family's well-being.

POSTPARTUM DEPRESSION AND DISTURBED SLEEP

Postpartum depression is the most common complication of childbirth. Up to 50% to 60% of all new mothers experience postpartum blues during the first 2 postpartum weeks. The blues manifest as excessive and unpredictable crying episodes, labile mood, and sadness during a time that is expected to be joyful. At some point 3 to 6 months after delivery, between 10% and 15% of new mothers experience diagnosable postpartum major depression.[68,69] Symptoms are similar to those of depression experienced at other times of life (see Chapter 137) but also include difficulty sleeping when the baby is sleeping at night and worrying about hurting the baby.[68] Various mechanisms for development of postpartum depression have been proposed, including the dramatic drop in progesterone and estrogen levels during the first few postpartum days,[69] poor supportive relationships, overly high expectations regarding the baby and motherhood, unexpected difficulties with labor and delivery or postpartum infant care, and a sudden shift in attention from the mother to the baby. These contextual variables, however, have not been directly related to postpartum depression.[68,69]

One additional contextual variable that seems obvious but has previously been overlooked is sleep deprivation and disrupted sleep-wake cycles. Sleep deprivation is likely to contribute to postpartum mood changes.[29,30,42,70] Being in labor during the night and a previous history of sleep disruption at the end of pregnancy may result in a higher incidence of postpartum blues.[9,19] Specifically, one study reported that new mothers had increased dysphoric mood during the first postpartum week, but this significant effect was eliminated when "time awake" was controlled in the analysis.[18] The frequency of night awakenings, rather than a change in hormone levels, was related to negative mood at 1 month post partum, and REM sleep onset latency was significantly shorter (less than 60 minutes) for postpartum women who reported higher levels of depressed mood.[70]

Moreover, mothers in whom depressive symptoms developed at 2 to 4 weeks post partum had significantly different sleep schedules at the end of their pregnancies compared with nondepressed mothers, with less total sleep from the end of pregnancy to early in the postpartum period than that for nondepressed women.[19] On average, mothers with subsequent postpartum depressive symptoms reported later rise times, longer naps, and more total sleep at the end of their pregnancies. Increased time awake during the night and poor sleep quality were strongly associated with increased negative daytime mood, or "blues," particularly in the first 4 weeks after childbirth.[18,19,43] Lee and colleagues[70] compared sleep patterns for women with positive postpartum affect at 1 month post partum with those for women with negative postpartum affect. The positive mood group had stable sleep times from the last trimester to 1 month post partum, whereas the negative mood group slept 80 minutes less at 1 month post partum.[70]

Associations between depressed mood and sleep schedule changes across the perinatal period may be mediated by circadian disruption. Sharkey and colleagues showed that later circadian phase in the third trimester of pregnancy was associated with more depressive symptoms at 2 and 6 weeks post partum in a small sample of women with a history of major depression who were not depressed during their third trimester.[7] Studies are needed to further examine the effect of altered sleep patterns from pregnancy to weeks or months post partum on development of postpartum depression.

The effects of infant sleep patterns on maternal sleep and development of postpartum depression also are essential to consider. Women with no major depressive symptoms at 1 week post partum were evaluated at 4 and 8 weeks to examine relationships among infant sleep patterns, maternal sleep, and depressive symptoms.[71] Mothers exhibiting major depressive symptoms at 4 and 8 weeks post partum were more likely to report getting less than 6 hours of sleep in a 24-hour period during the past week and being awakened by their baby three or more times between 10 PM and 6 AM. These findings suggest that infant sleep patterns and maternal sleep deprivation and fragmentation are significant contributors to the pathogenesis of postpartum depression. Report of infant sleep problems at 6 to 12 months of age also is associated with maternal major depressive symptomatology, even after adjustment for known risk factors such as history of depressive illness.[72]

Because infant sleep problems in these studies were based on maternal report rather than on objective evaluations of infant sleep, the possibility exists that depressed mothers may perceive their infant's sleep more negatively than nondepressed mothers, or may be more likely to report sleep problems whether they exist or not. Some evidence, however, suggests that treatment of infant sleep problems results in decreased incidence of maternal depression. Two randomized controlled trials of brief behavioral interventions to reduce infant sleep problems significantly decreased maternal reports of infant sleep problems and maternal symptoms of depression.[34,35] Similarly, Wolfson and colleagues[33] showed that parent training focused on developing healthy infant sleep patterns improves both parental competence and marital satisfaction and also decreases parental stress.

Postpartum depression can have an impact on maternal-infant bonding, as well as on marital satisfaction and infant development.[33] In a few women, the disorder culminates in postpartum psychosis, of which insomnia is one of the most frequent symptoms.[68,69,73,74] New mothers with a history of bipolar disorder may be at particular risk for development of postpartum psychosis consequent to sleep deprivation and require close monitoring for inability to sleep or decreased need for sleep.

Women with major postpartum depression had more REM sleep than women without a pregnancy-related depressive episode,[75] suggesting a greater sleep-related risk for depression. By contrast, among women with a history of mood disorder, less REM sleep was noted early in pregnancy and persisted up to 8 months postpartum in one study,[29] and REM sleep was not different during pregnancy or the postpartum period from that in matched control subjects.[76]

Women with depression during the perinatal period, and clinicians who treat them, must balance the risk of medication side effects in the breastfed infant with the risk of untreated depression for the mother's health and the infant's well-being. For these women, it is recommended that any drug therapy start with a low dose, to be increased slowly in small increments.[68] Cognitive-behavioral therapy, total or partial sleep deprivation, and light therapy may be useful behavioral interventions (see Chapters 85 and 137). Women with postpartum depression who were allowed to sleep only from 9 PM to 1 AM responded more favorably to this schedule of sleep deprivation therapy than women allowed to sleep from 3 to 7 AM.[75,77] Light treatment for postpartum depression may be more effective if administered over weeks rather than days.[78]

OTHER POSTPARTUM HEALTH ISSUES ASSOCIATED WITH DISTURBED SLEEP

Sleep disorders that may arise during pregnancy, such as sleep-disordered breathing, esophageal reflux, or restless legs syndrome, typically resolve after labor and birth of the infant.[79,80] Pharmacologic treatment for insomnia during lactation is of concern because of potential effects on the newborn's growth and development, as well as safety concerns regarding the mother's ability to awaken to respond to the baby at night. One observational study in 124 women, however, showed minimal sedation among infants breastfed by new mothers taking benzodiazepines, suggesting that breastfeeding need not be contraindicated when benzodiazepines are being used post partum.[81] (See Chapter 56 for a summary of common sleep-related medications and their ratings for potential harm to the newborn.)

Severe sleep deprivation in the postpartum period contributes to substantial postpartum weight retention. Women reporting short sleep duration (5 hours or less in a 24-hour period) at 6 months post partum are more than twice as likely to retain at least 5 kg of the weight gained during pregnancy.[82,83] Strategies to decrease postpartum obesity should include interventions to increase maternal sleep duration.

CLINICAL PEARLS

- If a specific sleep disorder such as restless legs syndrome, sleep apnea, or insomnia developed during pregnancy, evaluations during the postpartum period should determine if the difficulties have resolved.
- Complaints of excessive sleepiness, fatigue, or sleep loss should be pursued and evaluated by health care providers because of increased morbidity and mortality in the mother, potential harm to the newborn, and relationship to anemia, infection, thyroid dysfunction, and postpartum depression or psychosis.
- Evaluation for sleep disturbance and depression should be included in routine postpartum assessments; advice related to optimizing sleep in the postpartum period should be provided, taking into account cultural sleep practices.
- Severe sleep restriction in the postpartum period may contribute to retention of much of the pregnancy weight gain, development of depressive symptoms, and poor maternal-infant interactions, with potential risk to the health and safety of the newborn.

SUMMARY

The postpartum period is a time in a woman's life when sleep patterns are greatly disturbed. The unpredictable sleep patterns of the newborn and intense infant care requirements contribute to the poor sleep and alteration of circadian rhythms common among new mothers. These women attempt to counteract sleep loss at night through changes in their sleep schedules, including more frequent naps and sleeping later in the morning. Nevertheless, many women experience considerable daytime sleepiness and fatigue through the first few months. Additionally, the prevalence of postpartum depression is approximately 13%, and women who report significant sleep loss and infant sleep problems are more likely to experience depression. Complaints of excessive sleepiness, fatigue, or sleep loss should be evaluated by health care providers because of increased maternal morbidity and mortality, potential harm to the newborn, and high risk of postpartum depression. Evaluation for sleep disturbance and depression should be a part of routine postpartum assessments, and advice related to optimizing sleep should be provided. Only a few studies have investigated the feasibility and efficacy of improving maternal sleep in the postpartum period, and although some results are promising, more work is needed in this area.

Selected Readings

Colson ER, Willinger M, Rybin D, et al. Trends and factors associated with infant bed sharing, 1993-2010: the National Infant Sleep Position Study. *JAMA Pediatr* 2013;**167**(11):1032–7.

Clout D, Brown R. Sociodemographic, pregnancy, obstetric, and postnatal predictors of postpartum stress, anxiety and depression in new mothers. *J Affect Disord* 2015;**188**:60–7.

Hiscock H, Wake M. Randomised controlled trial of behavioural infant sleep intervention to improve infant sleep and maternal mood. *BMJ* 2002;**324**(7345):1062–5.

Kempler L, Sharpe L, Miller CB, Bartlett DJ. Do psychosocial sleep interventions improve infant sleep or maternal mood in the postnatal period? A systematic review and meta-analysis of randomised controlled trials. *Sleep Med Rev* 2015;**29**:15–22.

Lee KA, Gay CL. Can modifications to the bedroom environment improve the sleep of new parents? Two randomized controlled trials. *Res Nurs Health* 2011;**34**(1):7–19.

Lee KA, McEnany G, Zaffke ME. REM sleep and mood state in childbearing women: sleepy or weepy? *Sleep* 2000;**23**(7):877–84.

Lee KA, Zaffke ME, McEnany G. Parity and sleep patterns during and after pregnancy. *Obstet Gynecol* 2000;**95**(1):14–18.

Montgomery-Downs HE, Insana SP, Clegg-Kraynok MM, Mancini LM. Normative longitudinal maternal sleep: the first 4 postpartum months. *Am J Obstet Gynecol* 2010;**203**(5):465.e1–7.

Sharkey KM, Pearlstein TB, Carskadon MA. Circadian phase shifts and mood across the perinatal period in women with a history of major depressive disorder: a preliminary communication. *J Affect Disord* 2013;**150**(3):1103–8.

Stremler R, Hodnett E, Kenton L, et al. Effect of a behavioural-educational intervention on sleep for primiparous women and their infants in the early postpartum: multisite randomised controlled trial. *BMJ* 2013;**346**:f1164.

Swanson LM, Flynn H, Adams-Mundy JD, et al. An open pilot of cognitive-behavioral therapy for insomnia in women with postpartum depression. *Behav Sleep Med* 2013;**11**(4):297–307.

Volkovich E, Ben-Zion H, Karny D, et al. Sleep patterns of co-sleeping and solitary sleeping infants and mothers: a longitudinal study. *Sleep Med* 2015;**16**(11):1305–12.

Wolfson AR, Crowley S, Anwer U, Bassett JL. Changes in sleep patterns and depressive symptoms in first-time mothers: last trimester to one-year postpartum. *Behav Sleep Med* 2003;**1**(1):54–67.

A complete reference list can be found online at ExpertConsult.com.

Chapter 159

Sleep and Menopause

Fiona C. Baker; Hadine Joffe; Kathryn Aldrich Lee

Chapter Highlights

- The prevalence of sleep disturbances, especially difficulty maintaining sleep, escalates as women undergo the transition through menopause. Hot flashes are an important contributor to the increase in sleep disturbances.
- Depression, which also increases in prevalence as women approach menopause, contributes to poor sleep and also may develop secondarily to disrupted sleep, especially sleep disruption associated with hot flashes.
- When hot flashes are a primary source of sleep problems in peri- and postmenopausal women, hormone therapy can be an effective treatment. Nonhormonal therapies for vasomotor symptoms, such as selective serotonin reuptake inhibitors, also have been shown to improve sleep quality. More trials are needed to determine efficacy of nonpharmacologic approaches, especially cognitive-behavioral therapy, for alleviating sleep disruption associated with hot flashes in perimenopausal women.
- The prevalence of sleep-disordered breathing increases with the transition to menopause, attributed to weight gain and/or changes in adipose tissue disposition, as well as declines in reproductive hormones that can have an adverse impact on the upper airway. The preferred treatment is continuous positive airway pressure along with weight loss and exercise programs.
- In women, aging and the transition through menopause are accompanied by an increased prevalence of clinical conditions such as breast cancer, arthritis, fibromyalgia, and hypothyroidism. These clinical conditions or their treatments can adversely affect sleep. Sleep disturbances associated with these conditions may be further exacerbated by menopausal symptoms such as hot flashes.

PHASES OF THE MENOPAUSAL TRANSITION

Menopause is the anchor point of a woman's transition during midlife to nonreproductive status and is confirmed to be present after 12 months of amenorrhea. The median age at onset of the menopause transition is 47 years, and the median age for the final menstrual period (FMP) is 51.4 years,[1] but can range between 40 and 58 years of age. For many years, menopause was simply considered to be a consequence of depleted ovarian follicles (the primary source of estrogen and progesterone). Today, however, it is well recognized to be a transitional process consisting of a series of complex and interactive changes in the central nervous and endocrine systems, beginning several years before complete cessation of menstruation and continuing for several years thereafter.[2]

A menopause staging system was developed in 2001 and updated in 2011 by the Stages of Reproductive Aging Workshop (STRAW) to provide a consistent way to describe reproductive aging through menopause[3,4] (Figure 159-1). The *menopause transition* is divided into early and late stages. The early transition is marked by increased variability in menstrual cycle length (variation by 7 days or more in the length of consecutive cycles), and the late menopause transition is marked by the occurrence of amenorrhea for 60 days or longer, usually accompanied by intermittent increases in follicle-stimulating hormone (FSH) level to greater than 25 IU/L.[3] During the menopause transition, breast tenderness, vasomotor symptoms (hot flashes and night sweats), and sleep difficulties within ovulatory cycles are maximal in the late-luteal phase,[5,6] and hot flashes become more prominent, with increasing frequency of anovulatory cycles, during the late menopause transition.

Postmenopause is divided into an early stage, lasting 5 to 8 years until levels of estradiol and FSH stabilize, and a late stage, when reproductive hormone changes are limited.[3] During the first 2 years of postmenopause, FSH levels continue to increase, estradiol levels continue to decrease, and vasomotor symptoms are most likely to occur. The term *perimenopause* describes the menopause transition and the first year of postmenopause.

Menopause is a universal phenomenon, but the timing and duration of these transitional stages and associated signs and symptoms vary considerably from woman to woman and cannot be easily predicted. Midlife (middle age) is considered to be between 40 and 60 years of age, but there is no "generic" midlife woman.[7] In addition to the complexity of menopausal stages and hormonal fluctuations, midlife women may have young children at home, have grown children leaving home, and be caring for elderly parents or spouses. Other common challenges and new-onset problems in this age group include

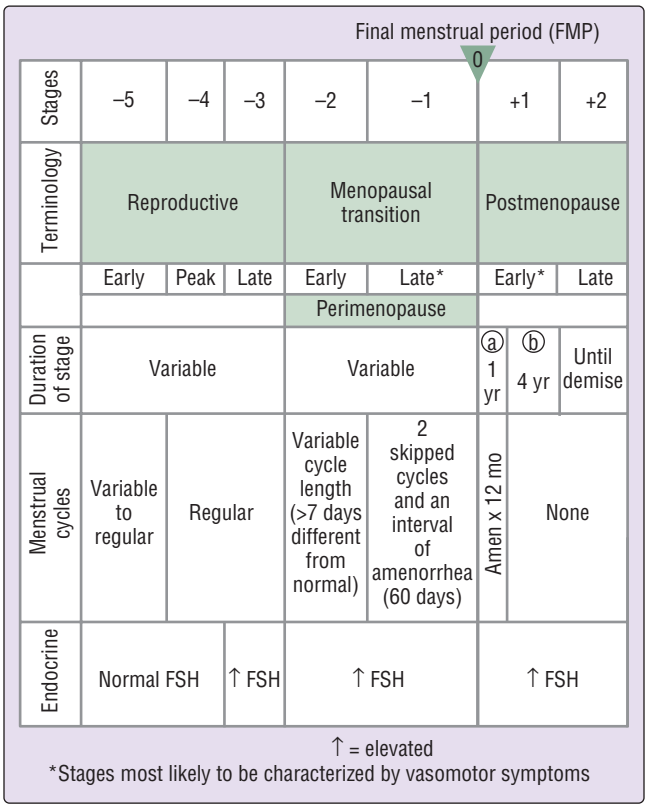


Figure 159-1 Stages of normal reproductive aging in women. The final menstrual period (FMP) is the time in a woman's life when she has missed 12 consecutive menstrual periods (amenorrhea) (*a*) and is considered early postmenopausal for the next 4 years (*b*) before becoming late postmenopausal for the duration of her life. Before permanent cessation of menstrual cycles, women vary in the duration of their cycles during the stage known as perimenopause. Follicle-stimulating hormone (FSH) is elevated throughout early and late perimenopause because of a lack of adequate levels of ovarian hormones to inhibit secretion of FSH from the hypothalamus. Note that the reproductive stages most likely to include hot flashes and night sweats are late perimenopause and early postmenopause, with wide and unpredictable variations in duration of these stages of normal reproductive aging. (Modified from Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop [STRAW]. *Fertil Steril* 2001;76:874-8.)

increasing career demands, changes in lifestyle, weight gain, and chronic health conditions. All of these scenarios can adversely affect a woman's sleep during menopause.

Surgical Menopause

Menopause can be induced by hysterectomy or bilateral oophorectomy (removal of both ovaries). After cesarean births, hysterectomy is the second most common surgery for women in the United States.[8] The average age at hysterectomy is 40 to 45 years; approximately 37% of American women have had a hysterectomy by the age of 60 years. Historically, at least one half of these women also underwent concurrent bilateral oophorectomy, although that percentage has declined in recent years.[9] In women who are premenopausal or perimenopausal, a bilateral oophorectomy induces an abrupt cessation of any ovarian hormone secretion, leading to a decline in estradiol and increased likelihood of menopausal symptoms. Although risk status for vasomotor symptoms may differ between women who undergo hysterectomy with ovarian conservation and those who have had a bilateral oophorectomy,[10] most studies have combined these two groups of women into a "surgical menopause" group and compared them with women who experience natural menopause.

NORMAL SLEEP PATTERNS DURING MENOPAUSE

Surveys and Self-Reported Sleep Measures

Population survey studies show that sleep problems are more common in midlife women transitioning to menopause and during postmenopause compared with premenopause (as confirmed in several reviews[11-13]). Intermittent awakenings are the most common sleep complaint and also are reported to be one of the most bothersome symptoms.[14,15] Being perimenopausal was associated with greater likelihood of trouble sleeping relative to being premenopausal, even after adjustments for age and ethnicity in the multicenter Study of Women's Health Across the Nation (SWAN).[6,16] These results were confirmed in the SWAN longitudinal analysis of women undergoing the menopausal transition (see Figure 159-2).[15] Similarly, the Australian Longitudinal Study on Women's Health found that difficulty sleeping was associated with menopause status, but not age, after adjusting for several confounders including mental health score and night sweats.[17] Another prospective study of midlife women in the United Kingdom reported increased trouble sleeping as the women progressed through menopause, although the relationship was most evident for women with severe as opposed to moderate sleep difficulties, especially after adjustment for factors such as vasomotor and psychological symptoms.[18] Longitudinal data from the Seattle Midlife Women's Health study also showed an increase in sleep problems, especially intermittent awakenings, across the menopause transition and early postmenopause.[19]

Sleep difficulties across the menopause transition and early postmenopause have been linked to several factors, including changing reproductive hormone levels (decrease in estradiol and increase in FSH),[15] hot flashes, depression, perceived poor health, and stress[19] (Table 159-1). Personality traits such as neuroticism have also been linked with midlife sleep difficulties.[20]

Sleep and Surgical Menopause

Surgical menopause (hysterectomy with or without bilateral oophorectomy) has been linked with poor sleep as well as other menopausal symptoms.[15,18,21] In a cross-sectional analysis of the SWAN cohort, a sample of women who had a bilateral oophorectomy and who were not using hormone therapy showed the highest prevalence of sleep difficulty, independent of age or years since surgery.[16] This effect was related to vasomotor symptoms. Women who undergo bilateral oophorectomy are at increased risk for more severe hot flashes than women in natural menopause,[10] and these severe hot flashes could have a greater impact on sleep. Disturbed sleep and fatigue are common postoperative symptoms after hysterectomy[22]; however, women who had a hysterectomy (with or without oophorectomy) were still more likely than other midlife women to complain of trouble sleeping even years after surgery,[23] suggesting greater vulnerability to sleep difficulties in these women. Increased sleep disturbances after surgical menopause is not necessarily only due to abrupt changes in reproductive hormone levels but could also be related to a worse health profile before menopause or worse health after hysterectomy compared to women with

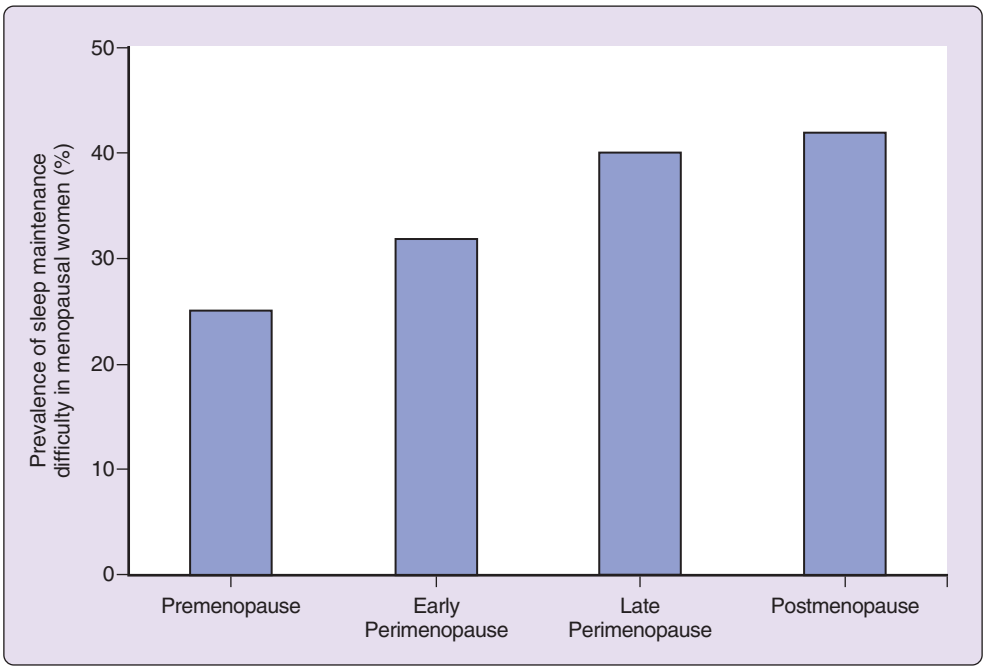


Figure 159-2 Percentage of women participating in the Study of Women's Health Across the Nation (SWAN) ($n = 3045$) reporting difficulty maintaining sleep at least three times per week in the past 2 weeks as they progress through menopause transition. Women in transition from pre- to early menopause were more likely to have difficulty maintaining sleep than women who stayed premenopausal. Women in transition from early to late menopause also were more likely to have difficulty maintaining sleep. (Data from Kravitz HM, Zhao X, Bromberger JT, et al. Sleep disturbance during the menopausal transition in a multi-ethnic community sample of women. *Sleep* 2008;31:979-90.)

Table 159-1 Common Sources of Sleep Disturbance in Peri- and Postmenopause

Menopause-Specific	General	Sleep Disorders	Mental Health Issues	Comorbid Illnesses
Hot flashes ↓ Estradiol ↑ FSH ↓ Inhibin B ↑ Testosterone	Stress Age-related Caffeine	Insomnia disorder Obstructive sleep apnea Periodic limb movement disorder Restless legs syndrome	Depression Anxiety	Chronic pain, fibromyalgia Obesity Gastroesophageal reflux Cancer Thyroid disease

FSH, Follicle-stimulating hormone.
Data from Joffe H, Massler A, Sharkey KM. Evaluation and management of sleep disturbance during the menopause transition. *Semin Reprod Med* 2010;28:404-21; and from Hall M, Buysse DJ, Nofzinger EA, et al. Financial strain is a significant correlate of sleep continuity disturbances in late-life. *Biol Psychol* 2008;77:217-22.

natural menopause.[18] Significant depression or anxiety may occur after hysterectomy and contribute to sleep problems in some women,[22] although psychological symptoms are not more common overall after hysterectomy with or without oophorectomy.[24] As discussed later, use of hormone therapy (HT) effectively improves sleep in surgically menopausal women.

Race and Ethnic Factors

Limited evidence suggests that race and ethnicity influence the extent of sleep disturbances across the menopause transition. SWAN researchers have shown that prevalence rates for difficulty sleeping were lowest in midlife Japanese women (28.2%) and highest in midlife white women (40.3%).[25] In their longitudinal analysis, few interaction effects between race/ethnicity and menopausal status were found.[15] However, a recent meta-analysis of 24 studies investigating the relationship between menopausal status and subjective sleep disturbance reported that perimenopausal, postmenopausal, and surgical-menopausal white and Asian women, but not Hispanic women, are more likely to experience sleep disturbance than premenopausal women, suggesting that culture and ethnicity may influence the extent of sleep disturbance associated with menopause transition.[26]

Insomnia Disorder

In addition to an increase in symptoms of insomnia and dissatisfaction with sleep typical for the menopause transition, an increase in the prevalence of insomnia disorder also has been documented. Chronic difficulty initiating sleep, nonrestorative sleep, global sleep dissatisfaction, and a diagnosis of insomnia (as defined in the *Diagnostic and Statistical Manual of Mental Disorders,* 4th Edition) all are more frequent among perimenopausal women than among

premenopausal and postmenopausal women.[27] A major factor in the increased prevalence of insomnia disorder in midlife women appears to be hot flashes; experiencing hot flashes (mild, moderate, or severe) was strongly associated with a diagnosis of insomnia, along with chronic pain, poor health, and white race.[27] In this situation, nocturnal hot flashes and night sweats may be precipitating factors for the development of insomnia in vulnerable women. For women who have preexisting insomnia, the approach to menopause may exacerbate their symptoms, with hot flashes precipitating further sleep disruption and possibly worsening their insomnia.

Objective Sleep Measures

Although strong evidence from epidemiologic studies shows an increase in trouble sleeping during the menopause transition and early postmenopause, findings from objective polysomnography (PSG) studies have been mixed. Cross-sectional studies comparing sleep architecture in pre-, peri-, and postmenopausal women have shown few differences in sleep,[28] or show better sleep after menopause. The Wisconsin Sleep Cohort Study of 589 midlife women showed that postmenopausal women had more slow wave sleep and a lower proportion of time spent awake despite reporting less sleep satisfaction compared with premenopausal women.[29] Perimenopausal women in this study also had a better objective sleep quality with more slow wave sleep and less stage 1 sleep than premenopausal women. Kalleinen and colleagues showed that age had a much greater effect on PSG-recorded sleep than menopause status, and although women in late-reproductive stage (45 to 51 years of age) and postmenopausal women had similar PSG measures, their objective sleep quality was much poorer than that in a group of young women.[30]

It should be kept in mind that subjective and objective sleep variables measure different constructs, such that they may deviate in their estimates of sleep quality. It also is possible that the menopause transition may be associated with disturbances in the sleep electroencephalogram (EEG) or processes related to sleep that are not evident from PSG scoring, but that could adversely affect subjective sleep quality. Recently, Campbell and colleagues applied quantitative analysis to the sleep EEG in women participating in the SWAN study.[31] As reported by other investigators, no differences were found in PSG measures by menopause status, despite a subjective worsening of sleep with the transition to menopause. However, quantitative analysis of EEG activity revealed elevated beta EEG activity during sleep in late perimenopausal and postmenopausal women compared with early perimenopausal and premenopausal women; an effect partly explained by the presence of frequent self-reported hot flashes. Elevated beta activity during sleep suggests a greater level of hyperarousal in these women, which could contribute to their perception of a poorer sleep quality.

A theoretical basis for the menopausal hormone changes with potential impact on sleep comes from studies in animals showing that reproductive hormones, specifically progesterone and estradiol, affect sleep-wake regulation, although the precise mechanisms remain unclear.[32] In women participating in SWAN, an increase in FSH over time was associated with reports of difficulty staying asleep, and a declining level of estradiol over time was associated with trouble falling asleep and staying asleep.[15] Other studies also have reported associations between low or faster rates of change in estradiol levels and poor subjective sleep.[33,34] Changes in PSG sleep parameters also have been linked with hormonal changes. SWAN data showed that a more rapid rate of change in FSH over a 5- to 7-year period was associated with higher percentage of slow wave sleep and a longer total sleep time.[35] At the same time, a greater rate of change in FSH was associated with a poorer subjective sleep quality.[35] Change in estradiol was unrelated to any PSG measures, although a lower estradiol/testosterone ratio (sampled 3 to 6 months before the sleep study) was associated with less wakefulness after sleep onset.[35] Although these analyses have been limited by infrequent, annual assessments of hormonal profiles during a reproductive stage with frequent fluctuations in hormone levels, further longitudinal data can refine the current understanding of how objective and subjective sleep quality is impacted by the hormonal changes that characterize menopause transition.

Circadian Rhythm Influence during Menopause

One prominent theory of reproductive aging is that menopause results from the aging of multiple pacemakers in the brain and ovaries that control and coordinate a variety of circadian and other rhythms.[2] The suprachiasmatic nuclei of the hypothalamus are a primary source of these endogenous rhythms and their synchrony. The sleep-wake cycle is the most visible human circadian rhythm, and it is profoundly influenced by the suprachiasmatic nuclei and by other circadian rhythms, particularly rhythms of body temperature and melatonin secretion.

Numerous studies have shown estrogen's impact on circadian rhythms in female mammals.[32] These data suggest that circadian control of sleep might be perturbed by menopause, and preliminary data from studies in women suggest a change in the circadian system. For example, under constant routine conditions, postmenopausal women have an advanced melatonin acrophase time and tend to have an earlier time of melatonin onset compared with late-reproductive stage women, which could be a consequence of menopausal hormone changes, although an effect of chronologic aging cannot be completely excluded.[36] An advanced circadian phase could contribute to more fragmented sleep or early-morning awakening in postmenopausal women. Of interest, after 6 months of HT, postmenopausal women showed a delay in melatonin acrophase, with no change in absolute melatonin levels.[37]

SLEEP DISTURBANCE ASSOCIATED WITH VASOMOTOR SYMPTOMS

It has long been suspected that at least some component of the self-reported sleep disturbance associated with menopause is secondary to vasomotor symptoms (hot flashes and night sweats). A hot flash (often called a hot flush or "night sweat" when it occurs during the night) is a sudden, transient, and recurrent sensation of moderate to intense heat that usually begins in the upper body. It is primarily a thermoregulatory phenomenon,[38] with all the characteristics of a heat dissipation response: (1) peripheral vasodilation, which causes increased heat loss, and (2) increased sweating, which causes evaporative cooling.

Tremendous individual variability in the frequency of hot flashes between women and within women over the course

of menopause transition and postmenopausal years is characteristic, probably as a result of individual differences in hypothalamic thermoregulatory activity and higher levels of cortical activity in the insula.[39] Recent evidence suggests that inputs to the gonadotropin pulse regulator from KNDy neurons (kisspeptin, neurokinin B, and dynorphin) in the hypothalamus also may contribute to vasomotor symptoms during menopause transition.[40] A hot flash typically lasts no more than a few minutes, but such episodes can reoccur frequently throughout the day and night in some women. The perceived intensity of the flash also can vary widely, from mild to severely disruptive. Some women experience 20 or more hot flashes each day, whereas others report only 1 or 2 per week. Data from large population studies suggest that some women continue to have hot flashes for many years, even decades, after menopause, highlighting the potential importance of hot flashes as a source of sleep disturbance well after midlife.[41]

Although hot flashes most commonly are associated with menopause, they also occur after bilateral oophorectomy, with use of aromatase inhibitors and selective estrogen receptor modulators such as tamoxifen for treatment of estrogen-sensitive breast cancer, and during gonadotropin-releasing hormone agonist (GnRHa) therapy. These therapies are used in women to treat breast cancer, uterine fibroids, and endometriosis. These same therapies also are used in men to treat prostate cancer. Surgical menopause, smoking, obesity, depression, anxiety, and heightened somatic attunement increase the likelihood of hot flashes.[42] The prevalence of hot flashes also varies by racial/ethnic groups, with SWAN data showing that African American women are most likely, and Chinese American and Japanese Americans least likely, to report experiencing hot flashes.[42] It is not clear whether differences in the experience of hot flashes are the result of differences in lifestyle stressors, diet, cultural factors, or unidentified biologic factors,[43] although genetic variability plays a role.[44] Self-reported hot flashes and night sweats are consistently associated with reduced self-reported sleep quality and chronic insomnia.[6,15,16,27] However, results from several cross-sectional studies examining the association between hot flashes (either self-reported or objectively recorded with skin conductance measures) and objectively measured sleep quality have been mixed.[11] In a recent controlled model of new-onset hot flashes, nocturnal hot flashes were definitively linked with more frequent PSG awakenings, more wake time after sleep onset (WASO), and more stage 1 sleep in young premenopausal women treated with a GnRHa.[45] The results were the same when hot flashes were measured by both self-report in the morning and by physiologic changes in skin conductance during the night. These experimental data validate the subjective complaint of poor sleep quality associated with menopausal hot flashes. The magnitude of sleep fragmentation appears to be proportional to the number of nocturnal hot flashes, whereas daytime hot flashes show no association with objective or subjective sleep quality.[45] This finding suggests that interindividual variability in the link between hot flashes and sleep disruption may be explained in part by variability in the proportion of hot flashes experienced at night, as well as the absolute number of nocturnal hot flashes. Of interest, the sleep fragmentation occurring in women with nocturnal hot flashes induced after GnRHa treatment appears to exceed the number of hot flashes measured objectively or reported,[45] suggesting that factors other than individual hot flash events contribute to sleep disruption in these women. Another recent study quantified the impact of hot flashes on objective sleep in perimenopausal women by totaling the intervals of wakefulness associated with any hot flashes during the night. A majority of hot flashes (69%) were associated with awakenings across the night, and hot flash–associated wake time was responsible for more than 25% of objective WASO. Hot flash–associated wake time correlated with self-reported estimates of wakefulness.[46] These findings confirm that hot flashes are a significant—but not the only—contributor to subjective and objective sleep disturbance in perimenopausal women.

Recent studies have shown the importance of examining the impact of hot flashes on aspects of sleep other than traditional PSG measures and on non-EEG correlates of sleep. Campbell and colleagues[31] reported that beta EEG activity during sleep was related to menopausal status, a relationship that was partly explained by self-reported hot flash frequency. Other investigators have shown that hot flashes are associated with changes in autonomic nervous system activity,[47,48] with a reduction in parasympathetic (vagal) activity and increased heart rate at the onset of a nocturnal hot flash even in the absence of an EEG arousal.[49] When a hot flash is associated with an arousal or awakening, autonomic regulation is further affected, because even brief arousals are associated with a shift to sympathetic dominance.[50] Multiple hot flashes throughout the night could therefore affect nocturnal autonomic nervous system balance and, consequently, the restorative aspects of sleep.

Although nocturnal hot flashes constitute an important component of sleep disturbance during midlife, not all women who have menopause-related sleep problems complain of hot flashes.[11] As discussed later, sleep in midlife women may potentially be affected by the presence of a sleep-disordered breathing (SDB) problem, mood disturbance, or medical condition. It is therefore critical to evaluate other causes of sleep disruption in midlife women, many of which may co-occur with hot flashes.

PSYCHOLOGICAL SYMPTOMS AND SLEEP DISTURBANCE IN MENOPAUSE

Depression

Several longitudinal studies have found that depressive symptoms increase during menopause transition.[51] In addition to an increase in the proportion of women experiencing mild depressive symptoms during the transition, a smaller number of women experience a major depressive episode during this period of hormonal fluctuation.[52] The first onset of depression can occur in menopause[53-55]; however, the vast majority of women who experience depression during menopause transition have a previous history of depression, and their illness episode represents a recurrent depression episode.[52]

Depressive symptoms in menopausal women are strongly linked to vasomotor symptoms.[56,57,11] In some women, this association is explained at least in part by the fragmented sleep that occurs with a hot flash[45] and independently contributes to depressive symptoms.[57] Although fragmented sleep could contribute to depression, depression also has a significant negative impact on sleep in general (see Chapter 137), and this is equally the case for depression during menopause.

Peri- and postmenopausal women with depression report worse sleep quality[58] and have more objectively measured sleep disturbance than women without depression.[59] Compared with nondepressed women with hot flashes, women with concurrent hot flashes and depression spend less time in bed and have longer sleep onset latency, shorter total sleep time, and lower sleep efficiency, rather than more frequent awakenings.[59] Peri- and postmenopausal women with major depression also have increased levels of nocturnal melatonin, as well as a phase delay in their melatonin secretion.[60]

Anxiety and Life Stressors

Whereas the association between depression and sleep disturbance in perimenopausal women has been well studied, anxiety has received less attention, and it remains unclear whether the frequency of anxiety symptoms or anxiety disorder is increased in women undergoing the menopausal transition.[61] However, higher levels of anxiety[62,63] and perceived stress[16] have been linked with complaints of poor sleep quality in midlife women, and anxiety also is associated with PSG-measured longer sleep latency and lower sleep efficiency.[64] Financial strain, a chronic stressor associated with lower socioeconomic status, was found to be an independent correlate (along with race) of sleep complaints and lower sleep efficiency in SWAN study participants; financial strain, as well as other stressors, could interfere with sleep by way of stress pathways including negative affect and autonomic-endocrine dysregulation.[65] Multiple stressors for midlife women are recognized, including jobs, family responsibilities, relationships, and caring for sick or elderly relatives,[66] and these stressors may impact sleep. Midlife women may consider sleep a low priority that competes with many other demands of motherhood, career, marriage, and caring for aging parents.[67] Conversely, disrupted or poor sleep quality may affect a woman's ability to cope with life stressors. It has been suggested that being symptomatic during menopause transition is itself a unique stressor and may compound preexisting stressors.[66]

TREATMENT OF SLEEP DIFFICULTIES IN MENOPAUSE

Estrogen and progesterone levels fluctuate dramatically during menopause transition, ultimately declining to very low levels in the postmenopausal period. Consequently, estrogen therapy (ET), typically in combination with a progestin (together called hormone therapy [HT]), was commonly prescribed for midlife and older women on a long-term basis in an effort to counter hormone deficiency and to protect against osteoporosis, heart disease, and Alzheimer dementia. In 2003, however, the Women's Health Initiative clinical trial results abruptly reversed this practice by showing that use of a common HT regimen over an average of 7 years significantly increased risk of breast cancer, stroke, heart disease, and vascular dementia.[68]

Women are now being advised to avoid long-term exposure to HT, and to use HT for only a short time to provide relief from hot flashes and improve quality of life during menopause transition. As a result, although historical data are available on the benefit of ET or HT for specific sleep conditions, hormonal treatments are rarely used as firstline therapy for sleep conditions unrelated to hot flashes. Included in this section is information about the effects of HT, when available, on all sleep problems associated with menopause, because these conditions may be concurrent and complicate hot flash–related sleep disturbance.

Hormone Therapy and Sleep

Multiple studies in midlife women without sleep complaints show that ET, with or without a progestin, and progestin therapy (PT) alone, improve perceived and, to a lesser extent, PSG indices of sleep quality.[11,13] When PSG measures were obtained in several small studies of midlife women without sleep complaints, some showed inconsistent and small benefits of ET or PT for sleep fragmentation measured with PSG, and others showed no benefit or a negative effect. A limited number of investigations have examined the effectiveness of HT for insomnia in midlife women, and findings have been mixed. Some show improvement[69,70] and others show no effect[71] on perceived sleep quality and PSG measures for women with an insomnia diagnosis.

Inconsistent findings among studies may be attributed to differences in treatment duration, timing of treatment in relation to menopause transition, or even differences in HT formulations. Preliminary studies suggest benefit of natural progesterone over the synthetic progestin medroxyprogesterone on perceived sleep quality and selected PSG parameters.[11] One reason for these potential differences is that unlike medroxyprogesterone acetate, micronized progesterone is metabolized to potent neurosteroids such as allopregnanolone and pregnanolone. These neurosteroids interact with the same brain gamma-aminobutyric acid (GABA) type A receptors as for sedative-hypnotic medications, and they are soporific.[72] When taken in the morning, natural progesterone can result in drowsiness,[73] particularly at higher doses. Accordingly, women on HT may be advised to take progesterone at night.

The mechanisms through which ET or HT may improve sleep are poorly understood. Animal models suggest that estrogen may increase homeostatic drive for sleep[74] and reduce prostaglandin synthesis in the ventrolateral preoptic nucleus of the hypothalamus,[75] whereas the hypnotic effect of progestins is mediated through an effect on GABA-active metabolites.[76] Because the greatest benefits of ET for sleep have been observed in women with co-occurring hot flashes, ET presumably may improve sleep as an indirect consequence of its salutary effects on nocturnal hot flashes.[77]

In summary, ET and PT independently and together have positive effects on sleep quality in midlife and older women, independent of hot flashes. Data are more extensive and stronger for self-reported sleep problems than for PSG sleep measures; several small studies on the latter report mixed results. Data supporting the efficacy of HT for primary sleep disorders in midlife women are limited and do not support use of these interventions for underlying sleep disorders. Progesterone also may contribute potential adverse sedating effects as well. As a result, HT is not typically recommended for sleep problems unless hot flashes are thought to be the primary source of the sleep disruption.

Hormone Therapy for Sleep Disturbance Associated with Hot Flashes

Sleep disruption associated with hot flashes may be treated in several ways. ET has historically been the standard treatment.

Numerous epidemiologic studies and small clinical trials have shown that ET/HT reduces hot flashes and concurrently improves self-reported sleep quality.[11,13] The effect varies, ranging from a significant but clinically small benefit to alleviation of hot flashes as highly predictive of improved sleep. PSG studies generally have replicated these findings. In women who experienced frequent hot flashes, ET decreased the number and duration of nighttime awakenings, increased REM sleep, and shortened sleep onset latency.[13] However, in several small studies of women with only mild or infrequent hot flashes, neither ET nor HT had a measurable effect on PSG measures of sleep stages.

Recent data also show efficacy of a novel treatment that combines estrogen therapy and a selective estrogen receptor modulator for perceived sleep quality in women with hot flashes, particularly at lower doses of the selective estrogen receptor modulator.[78] Overall, HT can be expected to have a therapeutic effect on perceived sleep quality among women with hot flash–related sleep disruption.

Hormone Withdrawal

When HT is discontinued, it commonly is stopped abruptly. The sleep effects of withdrawal from HT are unknown, although abrupt estrogen withdrawal is shown to result in significant hot flashes,[79,80] which points to probable sleep consequences with abrupt HT discontinuation. Among women who stop HT, sleep disturbance is an importance predictor of HT reinitiation.[81]

Selective Serotonin Reuptake Inhibitors, Serotonin-Norepinephrine Reuptake Inhibitors, and Gabapentin

In the past decade, nonhormonal therapies using selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), or GABAergic agents have become established treatments for vasomotor symptoms, both in patients with breast cancer and in healthy midlife women.[82] The vast majority of women enrolled in these hot flash clinical trials had comorbid insomnia symptoms, permitting investigation of the effects of these interventions on insomnia symptoms that commonly co-occur with hot flashes. In recent trials, the serotoninergic antidepressants escitalopram,[83] venlafaxine,[84] and paroxetine[85] all have been shown to be more effective than placebo in reducing insomnia symptoms and improving sleep quality. Although the magnitude of the effect on sleep symptoms is modest in some studies, it was unexpected that sleep concerns would diminish, inasmuch as these agents can induce or exacerbate sleep problems in other populations receiving such treatments for mental illness.[86] Similar to serotoninergic agents, the synthetic GABA-type neurotransmitter gabapentin is another nonhormonal agent used to treat hot flashes, and preliminary evidence also shows a benefit for sleep complaints.[87]

Hypnotics

Selective GABAergic agents such as zolpidem and eszopiclone have been shown to improve sleep onset[88,89] and sleep maintenance[89,90] in women with vasomotor symptom–associated insomnia.[88-90] Randomized controlled trials show efficacy of zolpidem 10 mg[90] and eszopiclone 3 mg[88] for ameliorating poor sleep quality and for managing sleep-onset and sleep-maintenance insomnia, resulting in improved quality of life. For those with hot flashes, eszopiclone also reduced the number of hot flashes reported at night, but not during the day,[89] suggesting that it helps women sleep through their hot flashes, although a direct benefit for alleviation of hot flashes may be obtained when drug levels are high. Zolpidem also has been shown to augment the therapeutic effect of SSRIs/SNRIs on sleep disruption in women with hot flashes, resulting in improved quality of life.[91]

Although these selective GABAergic therapies are effective, dosages may need to be adjusted in midlife women. Recent data show that women are more likely than men to have detectable zolpidem levels on the morning after ingestion, resulting in next-day slowed reaction time.[92]

Cognitive Behavior and Alternative Therapies

Although complementary and behavioral therapies may reduce hot flash frequency or severity in some women,[93,94] data from randomized clinical trials do not support the efficacy of treatments such as soy, black cohosh, or omega-3 regimens, or of yoga or exercise.[95-98] Exercise did have a modest salutory effect on insomnia symptoms in women assigned to an exercise regimen to treat hot flashes.[98] Studies assessing the efficacy of cognitive-behavioral therapy (CBT) for insomnia, specifically in women with hot flashes, are ongoing, and a initial trial of a menopause symptom–focused CBT trial in patients with breast cancer showed some benefit for sleep quality among women with hot flashes.[99]

PRIMARY SLEEP DISORDERS

As women proceed through the menopausal process and beyond, they are at increased risk for development of a primary sleep disorder such as SDB, attributed to aging as well as menopause-related factors. In a sample of midlife women with sleep complaints, 53% had sleep apnea, periodic limb movement disorder, or both. The major predictors of poor sleep efficiency were apnea-hypopnea index (AHI), periodic limb movement index, arousals associated with these disorders, and total number of arousals, whereas the presence of hot flashes was not a significant predictor of objective sleep efficiency in the model.[62] These data highlight the importance of recognizing and treating primary sleep disorders in midlife women.

Sleep-Disordered Breathing

The well-documented gender gap in prevalence of SDB[100] begins to narrow with menopause, and menopause has long been described as a risk factor for SDB.[101] Partial upper airway obstruction, characterized by hypoventilation and carbon dioxide retention, appears to be more common than sleep apnea in postmenopausal women.[102] In studies with large samples of women that have controlled for important confounders influencing SDB severity, strong support has emerged for the hypothesis that menopause increases the risk of SDB.[103-105] Bixler and colleagues[103] found a ratio of 1 woman for every 3.3 men with apnea in their sample of 1741 men and women between 20 and 100 years of age; this ratio fell to 1 postmenopausal women for every 1.44 men once women were matched with men by age and body mass index (BMI). In an analysis of the women in the sample, the prevalence rates for mild SDB (defined as AHI between 0 and 15, together with a self-report of moderate or severe snoring) and

more severe SDB (AHI of at least 15) were higher in postmenopausal women not using HT than in premenopausal women, even after adjusting for age and BMI. In another large study, the prevalence of sleep apnea was higher in women older than 55 years of age, presumed to be postmenopausal (47%), than in women younger than 45 years, presumed to be premenopausal (21%), even after controlling for BMI and neck circumference and confounding effect of age.[104] Postmenopausal women also had a higher mean AHI than premenopausal women.

Young and colleagues[105] examined cross-sectional and longitudinal PSG data collected on 589 midlife women participating in the longitudinal Wisconsin Sleep Cohort Study. The detailed attention paid to menopause staging in this study allowed the investigators to determine SDB prevalence and odds ratios for premenopause as well as perimenopause and postmenopause. Postmenopausal women were 2.6 times more likely than premenopausal women to have SDB (defined as AHI of at least 5) and 3.5 times more likely to have more severe SDB (AHI of at least 15). The likelihood of having SDB was not significantly higher for perimenopausal women than for midlife premenopausal women. However, the findings did suggest that the risk for SDB increases throughout the menopause transition. With stratification of data on the years since the subjects' last menstrual period, a significant linear trend emerged toward increased risk for AHI of at least 5, with increasing postmenopause duration up to 5 years (see Figure 159-3). Strengths of this study were the combined cross-sectional and longitudinal data analysis and the careful use of multivariate models to adjust for several known risk factors for SDB, particularly age, BMI, smoking, and alcohol use. Despite the inclusion of these potent risk factors in the analyses, menopause status remained a strong independent risk factor for SDB.

Several factors may contribute to an increased risk of SDB after menopause. An important factor is weight gain or a change in the distribution of adipose tissue, which progressively accumulates in the upper part of the body after menopause. Excess weight is a problem for midlife and older women, with more than 35% in the United States considered obese.[106] The prevalence of SDB is highly correlated with excess weight.[107] After menopause, there is also a preferential increase in intraabdominal or visceral deposition of fat relative to other areas of the body,[108] which may be due in part to menopause-related hormonal changes, as supported by the finding that HT decreases the shift to visceral adiposity and can lower serum lipid levels.[109,110] The relationship between visceral fat and SDB is especially strong and believed by some to be the principal culprit leading to SDB.[111] Thus the increased adiposity and visceral fat associated with menopause places women at increased risk for SDB. It would be worthwhile for future studies to consider these factors, in addition to BMI, when investigating predictors of SDB in midlife women. Another factor that could contribute to the increased incidence of SDB after menopause is the decline in levels of endogenous estrogen or progesterone.[101] Studies show that progesterone increases ventilatory drive and increases activity of upper airway dilatory muscles.[112] Popovic and White[113] found that genioglossus muscle activity was highest in the luteal phase of the menstrual cycle (when progesterone levels are high) in young women and lowest in a group of postmenopausal women, although upper airway resistance did not differ between the two groups of women. Muscle activity increased in the postmenopausal women after they received combination HT. Progesterone's stimulatory effect on respiration is believed to be mediated through estrogen-dependent receptors,[114] so the menopausal decline in both of these hormones presumably may affect respiratory function.

On the basis of the evidence for a protective role of progesterone on ventilation, together with epidemiologic evidence of the association between menopause and increased

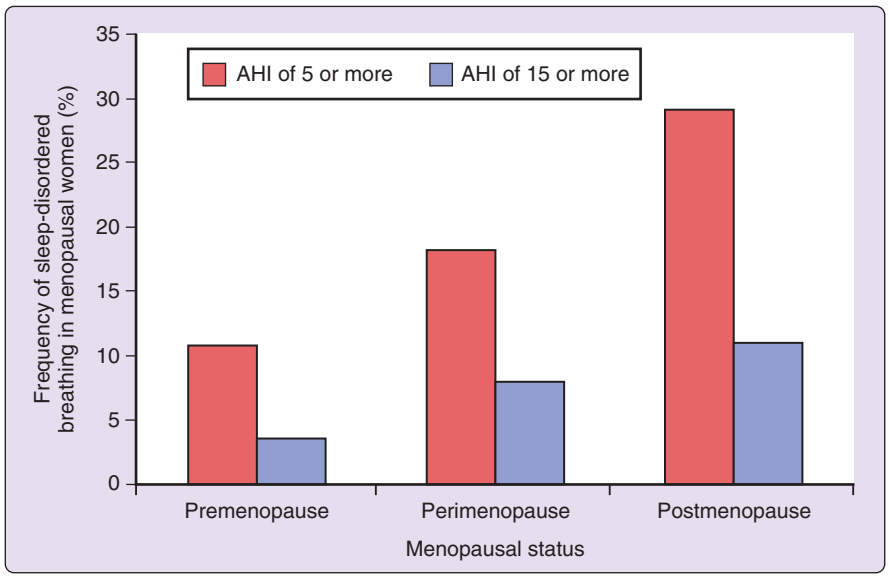


Figure 159-3 Prevalence of sleep-disordered breathing (SDB), indicated by apnea-hypopnea index (AHI) cut-off levels of 5 and 15 events/hour, in premenopausal ($n = 498$), perimenopausal ($n = 125$), and postmenopausal ($n = 375$) women who participated in the Wisconsin Sleep Cohort Study. Prevalence increased across menopausal groups. (Data from Young T, Finn L, Austin D, Peterson A. Menopausal status and sleep-disordered breathing in the Wisconsin Sleep Cohort Study. *Am J Respir Crit Care Med* 2003;167:1181-5.)

risk of SDB, it might be expected that HT would be effective at preventing or treating SDB. Indeed, in epidemiologic population studies, HT has been associated with a lower prevalence of sleep apnea in postmenopausal women.[103,105] This relationship was confirmed in the Sleep Heart Health Study[115] even after introduction of controls for well-documented differences (e.g., education level, body weight, health awareness) between women who used HT and those who did not. However, clinical trials that have evaluated the effects of estrogen, progesterone, or both on SDB in postmenopausal women have yielded conflicting results.[116] Although exogenous progesterone administration in postmenopausal women with SDB was associated with improved nocturnal ventilation in a number of studies, no change was seen in the number of apneas or hypopneas.[117-119] The wide variability in responses to HT among women suggests that if these hormones affect SDB, they do so through a specific mechanism that is not common to all cases of SDB. In view of the health risks associated with using HT, continuous positive airway pressure remains the treatment of choice for SDB in perimenopausal and postmenopausal women. Also, weight loss and exercise (specifically to reduce adiposity) should be strongly considered in the SDB treatment plan for any midlife or postmenopausal woman.

Restless Legs Syndrome and Periodic Limb Movement Disorder

The prevalence of restless legs syndrome (RLS) and periodic limb movement disorder (PLMD) increases with age, and women are 37% more likely than men to report RLS symptoms.[120] Women also are more likely than men to report the early-onset form of RLS (symptoms experienced before the age of 45 years). Whereas RLS and PLMD have been strongly linked with female reproductive life events such as pregnancy (see Chapters 95 and 156), an association between RLS or PLMD and the hormonal changes of menopause is less clear.

On the basis of survey data, most female patients with RLS (69%) retrospectively reported a worsening of their symptoms after menopause.[121] However, the prevalence of RLS among women increases with age,[122] which may confound the relationship between RLS and menopause per se. In a population study of Swedish women, a strong association was found between vasomotor symptoms and RLS; however, no relationship was noted between use of HT and RLS.[122] The incidence of PLMD is high in postmenopausal women,[123] and this disorder contributes to poor objective sleep quality in midlife women.[62] Evidence from current studies, however, does not support a strong link between PLMD and menopausal hormone changes. In a group of asymptomatic postmenopausal women, the incidence of periodic limb movements was unrelated to estradiol or FSH levels.[123] Furthermore, short-term ET did not alter the incidence or intensity of limb movements.[123] The increase in prevalence of RLS and PLMD after menopause may be related more to aging than to the menopause transition.

OTHER CLINICAL CONDITIONS WITH POTENTIAL EFFECTS ON SLEEP

As women age, coincident with the onset of menopause, health conditions are likely to develop, which can have an adverse impact on sleep. Cancer, neurologic disorders, cardiovascular or pulmonary disease, diabetes, hypothyroidism, gastroesophageal reflux disease, and musculoskeletal disease all are associated with sleep disruption.[13] Poor sleep could increase the risk of developing some of these conditions and/or exacerbate their severity.

Cancer

With age, women are more likely to develop cancer of the breast, lung, colon, ovary, gallbladder, or thyroid gland, and sleep disturbance is a common correlate. As discussed in Chapter 130, the etiology of cancer-related sleep disruption often is complex, with multiple factors likely to precipitate sleep problems. These include pain and discomfort, physical effects of the cancer itself, depression and anxiety, and side effects of chemotherapy or radiation treatment such as nausea, vomiting, diarrhea, or urinary frequency. In addition to the sleep-disrupting factors commonly experienced by most patients with cancer, women undergoing treatment for breast cancer are likely to experience hot flashes.[124] Hot flashes are a side effect of chemotherapy-induced ovarian disruption, and they also occur with the use of adjuvant hormone therapy.

Women with estrogen receptor–positive tumors are treated with the antiestrogen tamoxifen, aromatase inhibitors (e.g., anastrozole, letrozole, exemestane), and a GnRHa (e.g., leuprolide), or they undergo bilateral oophorectomy to prevent endogenous estrogens from stimulating the growth of residual tumors or micrometastases. More than 50% of tamoxifen users experience hot flashes, which usually are more frequent and more severe than hot flashes associated with natural menopause.[125]

Hot flashes could be a precipitating factor in the development of insomnia in breast cancer survivors.[126] As discussed in detail earlier, a strong association has been documented between hot flashes and sleep disruption in women going through menopause transition, and the same is true for women with breast cancer. Hot flashes in breast cancer survivors are associated with a less efficient, more disrupted sleep.[127] HT is not indicated for treatment of hot flashes in women with a history of breast cancer,[128] making alternative therapies a necessity. SSRIs and SNRIs, clonidine, and gabapentin have been shown to reduce the number and severity of hot flashes in women with a history of breast cancer.[129] Coadministration of a hypnotic, such as zolpidem, with an SSRI/SNRI improves sleep and quality of life more than use of an SSRI/SNRI alone in women with breast cancer who also have hot flash–associated sleep disturbance.[91] CBT also is beneficial, leading to improvements in subjective sleep as well as mood and quality of life that are maintained beyond 1 year in women with insomnia secondary to breast cancer.[130] Other interventions such as acupuncture, exercise, or melatonin also have shown some benefit for alleviating hot flashes and improving sleep quality, although larger clinical trials are needed.

Thyroid Dysfunction

The prevalence of thyroid disease, particularly hypothyroidism, increases with age and is far higher in women than in men. For midlife women living in iodine-replete areas, the prevalence of impaired thyroid function (i.e., thyroid-stimulating hormone [TSH] values outside the euthyroid range) is 9.6%.[131] In these cases, a majority (6.2%) of affected

women have elevated TSH, indicating clinical or subclinical hypothyroidism. Because hypothyroidism typically is characterized by tiredness and fatigue, and not sleepiness, such complaints in perimenopausal and postmenopausal women should be clinically evaluated in light of a TSH level.

Women with hypothyroidism may be more likely than euthyroid women to have SDB (see Chapters 20 and 132), suggesting that hypothyroidism also may be a risk factor for SDB.[132] Midlife and older women with hypothyroidism also should be screened for clinical signs and symptoms indicative of SDB.

Hypertension

The prevalence of hypertension rises sharply with onset of menopause. The etiopathogenic mechanism of this phenomenon is complex and still under investigation, but two factors strongly associated with hypertension are obesity and SDB, conditions common in perimenopausal and postmenopausal women.[133,134] The National Health and Nutrition Examination Survey III (NHANES III) documented a strong association between hypertension and BMI in women.[133] For women in midlife, the prevalence of hypertension was approximately 10% when BMI was less than 25 but rose to 39% when BMI was 30 or higher. For women older than 60 years of age, hypertension occurred in 52% with a BMI less than 25, but the prevalence was greater than 72% if BMI was 30 or higher. Midlife women are at increased risk for development of hypertension and SDB by virtue of the higher prevalence of obesity in this age group, rather than hormonal or menopausal factors per se. Sleep duration and efficiency were unrelated to hypertension with introduction of controls for several confounders in midlife women who participated in the SWAN study.[135]

Fibromyalgia

Fibromyalgia is a clinical disorder characterized by widespread pain and the presence of specific tender points. Sleep disturbance is a core symptom. The prevalence of fibromyalgia is higher among women (3.4%) than among men (0.5%), and the disorder is most common in women older than 50 years of age.[136] A sex difference in clinical presentation also has been noted, with symptoms of sleep disturbance and fatigue being three times more common in women.[136] A link between reproductive hormone changes and fibromyalgia has been hypothesized,[137] and some evidence suggests that at least for some women, fibromyalgia symptoms start after the onset of menopause,[138] when estradiol levels decline. However, estrogen therapy had no effect on pain thresholds or tolerance in postmenopausal women with fibromyalgia, but the potential benefit for relief of symptoms such as sleep disturbance and depression was not investigated.[139]

Neurodegenerative Disorders

Sleep disruption can be an early symptom of neurodegenerative conditions such as Parkinson or Alzheimer disease (see Chapters 92 and 96). These disorders occur more often in postmenopausal women than in age-matched men. Genetic factors and changes in biologic hormone mileu related to menopause may play an important role.[140] However, more research is needed to investigate possible associations among sleep disruption, estradiol decline, and development of neurodegenerative disorders in women.

CLINICAL PEARLS

- Sleep difficulties are more common in midlife women transitioning to menopause compared with premenopause, with intermittent awakenings being the most common and bothersome complaint. Nocturnal hot flashes are an important component of sleep disturbance in midlife women, and both hormonal and non-hormonal therapies that alleviate hot flashes are associated with improvements in sleep quality.
- Sleep-disordered breathing is more common in women after menopause, which may in part be attributed to a change in the distribution of adipose tissue, with an increase in intraabdominal fat. Reports of fatigue or sleep complaints in postmenopausal women, possibly combined with hypertension and excessive weight, should prompt consideration of snoring and SDB in a clinical evaluation regardless of whether excessive daytime sleepiness is reported.

SUMMARY

Menopause is a normal event in a woman's life. It is the physiologic centerpiece of a major developmental stage in the normal aging process, marking the transition from a reproductive to a nonreproductive stage of life. Most women now live long enough to become menopausal and can expect to live at least another 30 years beyond their FMP. The number of postmenopausal women in developed countries is increasing as a consequence of longer life expectancy. In the United States, approximately 1.5 million women reach menopause each year, and it is predicted that by the year 2020, 45.9 million women will be older than 55 years of age.[141] The transition to menopause and postmenopause is associated with an increased prevalence of sleep problems, which may be attributed partly to menopausal hormone changes and associated vasomotor (hot flashes and night sweats) and mood symptoms but also to concomitant primary sleep disorders, chronic health conditions, and midlife stressors.

Insomnia and fatigue are among the most frequent health complaints of perimenopausal women, including those who are not seeking treatment for menopausal symptoms. SDB is more prevalent in midlife women, which may be related to hormonal changes as well as increased adiposity and centripetal weight gain. An association between menopausal symptoms and sleep disruption, as well as the impact of SDB and other clinical conditions on sleep in menopause, has been well documented and is the subject of ongoing research. Hormonal and nonhormonal treatments for menopause-related sleep disruption of variable efficacy have been described.

Selected Readings

Baker FC, Willoughby AR, Sassoon SA, et al. Insomnia in women approaching menopause: beyond perception. *Psychoneuroendocrinol* 2015;**60**: 96–104.

de Zambotti M, Colrain IM, Javitz HS, Baker FC. Magnitude of the impact of hot flashes on sleep in perimenopausal women. *Fertil Steril* 2014;**102**: 1708–15.

Freedman RR. Hot flashes: behavioral treatments, mechanisms, and relation to sleep. *Am J Med* 2005;**118**(Suppl. 12B):124–30.

Joffe H, Massler A, Sharkey KM. Evaluation and management of sleep disturbance during the menopause transition. *Semin Reprod Med* 2010; **28**:404–21.

Kapsimalis F, Kryger MH. Gender and obstructive sleep apnea syndrome, part 2: mechanisms. *Sleep* 2002;**25**:499–506.

Kravitz HM, Joffe H. Sleep during the perimenopause: a SWAN story. *Obstet Gynecol Clin North Am* 2011;**38**:567–86.

Ohayon MM. Severe hot flashes are associated with chronic insomnia. *Arch Intern Med* 2006;**166**:1262–8.

Polo-Kantola P. Sleep problems in midlife and beyond. *Maturitas* 2011; **68**:224–32.

Polo-Kantola P, Erkkola R, Helenius H, et al. When does estrogen replacement therapy improve sleep quality? *Am J Obstet Gynecol* 1998;**178**: 1002–9.

Young T, Finn L, Austin D, Peterson A. Menopausal status and sleep-disordered breathing in the Wisconsin Sleep Cohort Study. *Am J Respir Crit Care Med* 2003;**167**:1181–5.

A complete reference list can be found online at ExpertConsult.com.

Section
21 Instrumentation and Methodology

Chapter
160 Polysomnography and Beyond

Max Hirshkowitz

Chapter Highlights

- Polysomnography began as a research method and provided fundamental descriptions of the sleep process. Sleep staging resulted. PSG evolved into a medical diagnostic procedure applied largely for diagnosing and treating sleep-related breathing disorders but also for identifying parasomnias and nocturnal seizures.
- Advances in biosensor and digital technologies improved polysomnography by eliminating paper, reducing data storage requirements, and allowing for additional signal analysis. A subset of polysomnographic cardiopulmonary recording channels were adapted for sleep apnea diagnosis and are now often used in home sleep testing.
- Actigraphy emerged as a research method for studying insomnia and circadian rhythm disorders. It too evolved into a medical technology.
- With advances in miniaturized sensor technology, wearable and other devices have become popular as consumer products for self-monitoring fitness and sleep. Whether these devices will advance the current understanding of sleep in large populations and/or improve sleep health remains to be seen.

Polysomnography (PSG) began as a research tool informing early investigators about brain activity during sleep. It provided an objective, quantifiable method and enabled scientific discovery. From PSG studies it became clear that sleep was not uniform throughout a sleep episode. Distinctly different processes emerged and took prominence in a fairly orderly manner. However, each polysomnogram produced massive amounts of data. Consequently, researchers needed a data reduction scheme to summarize and make generalizations about sleep; thus sleep staging was invented. Armed with sleep stage metrics, scientists could describe progressive changes across a night and across the lifespan, as well as differences between men and women. Concurrent activity of organs, tissues, muscles, and gland became focus areas. Sleep alterations occurring in response to interventions (e.g., sleep deprivation, drugs, stressors) were subjected to laboratory scrutiny.

When Eugene Aserinsky discovered rapid eye movement (REM) sleep,[1] another natural avenue for comparison opened. Because dreaming occurred during REM sleep, the REM

process captured the imagination of many researchers. Sleep science migrated to a place under the umbrella "psychophysiology." A research society formed with the moniker "The Association for the Psychophysiological Study of Sleep" (APSS). Animal studies revealed neuronal innervations arising from the pons, ascending to the lateral geniculate nuclei, and then continuing to the occipital regions (pons-geniculate-occipital [PGO] waves) during REM sleep. REM sleep was conceptualized as a different state of human consciousness (or, for more reductionist thinkers, as a unique state of central nervous system organization). REM's discovery reinforced conceptualization of sleep as being composed of unique states, rather than as an interplay of different processes. Immediately, any measurable phenomenon occurring during sleep became fair game for comparing activity during REM sleep versus all sleep states other than REM. Consequently, a major component of the sleep state became known as non-REM (NREM) sleep, even though REM sleep represented only 20% to 25% of sleep time.

With sleep divided into REM and four different NREM categories (stages 1, 2, 3, and 4) investigators applied PSG to characterize underlying and overt sleep processes.[2] The most obvious characteristic of the temporal relationship between NREM and REM sleep was a consistent alternation, with an approximate 90-minute cycle length. Intrigued also by REM sleep's seemingly independent homeostatic response to deprivation, neurologists, psychiatrists, and psychologists seized the opportunity to test hypotheses concerning the role of REM and NREM sleep in physical and mental functions.

Methodologic issues spawned innovation. Long-duration recording necessitated improved recording equipment. Sensor technology advanced to permit study of concurrent physiologic activities. Analytic techniques borrowed from other disciplines were applied to summarize finer-grain activity over time.

Although still the standard technique for scientifically studying sleep physiology, in the 1970s, PSG evolved into a medical procedure used to diagnose specific sleep disorders. Within a decade, the importance and prevalence of sleep-related breathing disorder became very apparent to sleep clinician-scientists. Before long, diagnosing obstructive sleep apnea developed as the predominant application for PSG (in Silicon Valley parlance, PSG's "killer app"). Even today, the vast majority of PSGs performed on any given night will be for diagnosing or treating sleep apnea.

As PSG's momentum increased clinically, recording techniques evolved dramatically. The digital age arrived and analogue amplifiers and paper chart drives rapidly became extinct. Computerization rendered PSG data storage problems (a huge issue when PSGs were recorded on paper) moot. Warehouses were replaced by filing cabinets and eventually by high-density disk drives or cloud storage. Quality, compatibility, and other technological issues emerged during this transition; over time, however, things steadily improved.[3] Ultimately, standards were adopted by clinical societies, and they continue to evolve.[4]

Of interest, PSG created "big data" long before tools were available to handle it all. Right out of the gate, sleep staging brought PSG summary down to less than 1000, 30-second pages. These stages could be summarized and indices calculated. Computerization made calculation easier, but very little changed with respect to the way sleep was analyzed. To a large extent, the computer became a recorder-reviewing station with basic parameter calculation capability. For decades, and even now, the signal processing and computational power of digital PSG are vastly underutilized. Computers can count specific waveforms (e.g., sleep spindles) and analyze wave patterns (using Fourier transforms, period-amplitude analysis, or complex demodulation). They can detect movements and respiratory events, calculate their periodicity, and determine their association with central nervous system changes. Initially, or at least up to the turn of this century, available computing resources were arguably stretched to accomplish such analysis in real time. Researchers faced memory, storage, and processor speed limitations. However, today's wireless (cloud) technology and inexpensive massive storage devices provide access to nearly unlimited computing resources.

The success of PSG for diagnosing sleep-disordered breathing may have ultimately rung its own death knell. Diagnostic procedures' goals involve disease verification and/or severity determination. Mainstream procedures are subject to close economic scrutiny by payers at all levels. Expensive procedures attract the attention of gatekeepers. Consequently, cardiopulmonary home sleep testing has arisen as a more economical alternative for confirming sleep apnea when clinical suspicion and pretest probability are high. PSG's indication for diagnosing insomnia had already been jettisoned by third party payers, except in cases in which all therapeutics remained notably unsuccessful. Nonetheless, PSG continues to be used as the objective technique for measuring sleep in scientific studies and clinical drug trials.

On a different technologic front, actigraphy arose as an alternate method for assessing activity-inactivity patterns over prolonged time periods. Harking back to Kleitman's basic rest-activity cycle (BRAC)[5] and creating a human analog to caged animals' wheel running, actigraphy provides information about sleep-wake patterns and circadian rhythm. Finer-grain analysis in connection with use of a light sensor potentially informs about insomnia, nocturnal awakenings, and circadian disorders. Integrating additional sensors (e.g., pulse, temperature, and skin conductance) stand to improve concordance between actigraphic and PSG-derived measures.

As often happens, after science's application to medical technology, consumer products begin emerging. The recent renaissance of interest in personal fitness using actigraphy opened the doors to sleep trackers. Wrist-worn consumer actigraphs quickly achieved market penetration. Uplink technologies to transfer data for cloud storage, analysis, and retrieval make devices affordable and promising. On any given night, more consumer product actigraphic data probably are uplinked than all of the research actigraphy ever recorded. Current understanding and knowledge garnered from actigraphy research are therefore ripe for application on a wider scale. A basic axiom in actigraphy research is that each device needs validation, and it is well recognized that reliability varies widely from device to device (see Chapter 171). Although this maxim regarding validation is undeniable, it represents a principle characterizing every new technology during its infancy. First-wave devices often perform erratically or marginally.[6-8] However, whether the main topic of discussion is wristwatches, automobiles, or airplanes, successive refinement ultimately improves performance and homogenizes the available products until a revolutionary new approach starts the process anew.

A research group in Munich has launched a human sleep project to collect actigraphic uplinked and downloaded data to elucidate sleep-wake patterns in the overall population.[9] In a completely separate initiative, the Consumer Electronics Association launched a project to standardize sleep terminology and performance criteria for wearable devices. These developments are already advancing the current understanding of different chronotypes, health consequences of sleep fasting and binging, and the medical costs of social jetlag.

The juggernaut of advancing sensor technology continues unabated.[10] Temperature, heart rate, blood pressure, blood sugar level, electrodermal skin conductance, heart sounds, breathing sounds, heart rate variability, cardiopulmonary coupling, pulse wave analysis—all are fair game for inclusion. Bedside monitors using static-charged strips, standing wave patterns in the room, and thermal sensors can detect breathing, snoring, movement, and heart rhythm noninvasively and unobtrusively. Another trend is use of sensors built into the sleep surface (i.e., mattress). Embedded transducers not only can detect high and low pressure points on the sleep surface but also can direct an air coil mattress to alter inflation to optimize comfort. These same sensors can detect movement, breathing, and heartbeat. Integrating these technologies and finding a desired or practical use for them consiititue the current challenge.

Are these consumer devices here to stay or do they represent a fad? Recall that more than 100 million Hula Hoops were sold between 1958 and 1960—whereas they are a rarity today. If self-monitoring for fitness persists, the data thus acquired undoubtedly will contribute much to the body of knowledge on sleep health worldwide. In the same manner in which satellite weather monitoring informed meteorologists about ocean currents, polar melt, and storm tracking, big-data analytic techniques applied to consumer uplinked actigraphs can be expected to answer many questions about sleep. These answers in turn will raise even more questions. For example, what population sleep effects correlate with seasonal variations, weather, daylight saving time changes, and latitude locations? What percentage of the population sample experiences disrupted sleep, exhibits different circadian chronotypes, and shows extreme long or short sleep periods?

Beyond these questions, the road ahead holds additional possibilities for us to further understand both sleep health and sleep disorders. Centralized electronic medical record systems are increasingly used and improving by leaps and bounds. With proper deidentification, such records meshed with uploaded sleep information potentially may uncover hitherto-unknown associations between sleep and health. Major issues must be resolved for this to happen, including techniques to ensure privacy, improved recording, validated analysis, and useful data summarization. The first hurdle—privacy—has social, political, and legal aspects. Important questions must be answered. For example, are wearable device data discoverable as evidence in legal proceedings? Technologic hurdles are easier to surmount.

Technologic challenges usually are met, given enough time and continued interest. Improvements proceed in a stepwise manner, much in the way mobile phones evolved from cumbersome, heavy, bulky, unreliable contraptions to the pocket-size devices of today (on which their users may thoroughly depend). Hopefully, self-monitoring fitness enthusiasts will begin to recognize sleep's value by its reflection on performance and daily mental outlook. Researchers, clinicians, and patients alike may be perched on the doorstep of sleep health with evolution from research laboratory to medical clinic and from medical clinic to personal device worn on the body or situated at the bedside. Interfacing self-monitoring devices and patient health care records remains wide open for innovation. Sleep represents a physiologic process critical to health; optimizing this process will undoubtedly improve health.

CLINICAL PEARLS

- PSG represents both a tool of discovery and the most sophisticated technique available for diagnosing sleep disorders.
- The cardiopulmonary channels can be used to verify sleep-disordered breathing in clear-cut cases.
- Actigraphy provides rest-activity patterns useful for interpreting sleep-wake and circadian rhythm patterns.
- The recent popularity of consumer self-monitoring fitness devices holds promise for future translational understanding of sleep health.

SUMMARY

PSG began as a research tool for objectively characterizing the sleep process. PSG studies create massive amounts of data, and sleep staging was a logical approach. Sleep studies provided a physiologic tool for investigating sleep correlates and functions. PSG's usefulness for differentially diagnosing sleep disorders soon became apparent, especially for evaluating sleep-related breathing disorders. Recently, the subset of cardiopulmonary recording channels is supplanting full PSG for apnea diagnostics. In a parallel development, actigraphy became popular for assessing insomnia and circadian rhythm disorders. New sensor technology and advanced analytical techniques increase the information extractable for biologic signals. The latest development involves entry of actigraphy into the consumer product space for fitness self-monitoring. Actigraphy devices can potentially provide information about sleep. If and when they meet performance standards, current understanding about sleep in large populations and overall sleep health can be expected to advance dramatically.

Selected Readings

Ferguson T, et al. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act* 2015;**12**(1):42.

Hirshkowitz M. Polysomnography: understanding this technology's past might guide future developments. *IEEE Pulse* 2014;**5**(5):26–8.

Hirshkowitz M, Sharafkhaneh A. Comparison of portable monitoring with laboratory polysomnography for diagnosing sleep-related breathing disorders: scoring and interpretation. *Sleep Med Clin* 2011;**6**(3):283–92.

Roenneberg T, et al. Social jetlag and obesity. *Curr Biol* 2012;**22**(10):939–43.

Russo K, et al. Consumer sleep monitors: is there a baby in the bathwater? *Nat Sci Sleep* 2015;**7**:147–57.

A complete reference list can be found online at ExpertConsult.com.

Chapter 161

Sleep Stage Scoring

Sharon Keenan; Max Hirshkowitz

Chapter Highlights

- The electroencephalogram (EEG) remains the most objective measure for determining level of consciousness and allowing determination of sleep stages. Fundamental to discoveries in sleep research and sleep medicine is our ability to appreciate subtle changes in brain activity.
- Human visual pattern recognition of the EEG in the context of changes in the electrooculogram and electromyogram remains the primary skill used to analyze these data to perform sleep stage scoring. The results of this analysis are used to inform interpretation of the data and recommendation for clinical care. Close collaboration between technical and medical members of the clinical team will facilitate these ends.
- In research, sleep stage scoring and the recognition of other patterns emerging from data remain active areas of intense study. More sophisticated analysis of EEG data using advances in technology holds great promise for expanding the current understanding of the central nervous system. Sleep provides a unique window for studying physiology, pathophysiology, and consciousness.
- This chapter reviews the basics of sleep stage scoring.

HISTORY

From a behavioral perspective, immobility and reduced environmental responsiveness characterize human sleep. This state stands in contrast to purposeful (presumably) activities and provides the basis for dichotomizing observable living existence as either sleep or wakefulness. Furthermore, sleep and wakefulness cycle in a lawful, orderly fashion. Some rhythms are seasonal, some are daily (circadian), and some occur more than once a day (ultradian). In addition to the underlying rhythms, the sleep cycle responds to reduced sleep time. This response testifies to sleep-wake cycle autoregulation, with a dynamic tension providing overall system homeostasis. Once techniques were developed to augment mere observation, electroencephalography revealed a complex array of brain activities clustered in a manner strongly suggestive of multiple sleep processes.

All scientific inquiry begins with observation and description. From there it proceeds to classification based ultimately upon measurement. Accordingly, when Loomis and colleagues[1] electroencephalographically reported their first studies, in 1937, they faced the daunting task of devising a system to describe sleep patterns in normal healthy human subjects. Thus sleep staging was born. In the original studies, amplified activity derived from electrodes that were placed on the scalp's surface at several loci produced ink tracings on paper wrapped around a slowly rotating cylinder. An enormous 8-foot "drum polygraph" enabled all-night sleep recording. One electrode was located near the eye and undoubtedly detected eye movement. However, rapid eye movement (REM) sleep remained unrecognized until 16 years later, when Aserinsky published, in part, his University of Chicago doctoral study results.[2] Aserinsky actually christened these movements "jerky eye movements" (JEMs) and in the first paper referred to the phenomenon as *periodic ocular motility*.

Perhaps it was the quirkiness of the original commercially available polygraph systems (e.g., Ofner, Beckman, Grass), with their tendency to polarize electrodes, problematic rechargeable car battery–like systems, and aperiodic (and difficult to predict) recording interference artifacts, or perhaps it was Loomis's silence on the matter of eye movements during sleep. In either case, Aserinsky's pilot work reportedly met with considerable skepticism. Ultimately, however, REM sleep's discovery, and particularly its correlation with dreaming by Dement,[3] altered the course of sleep research for decades. The near-exclusive focus on REM sleep, to the point that all other sleep states were considered simply non-REM (NREM), overshadowed substantial findings (and probably impeded progress) in other sleep research arenas (such as neuroendocrinology, physiology, and medicine). The spotlight on REM sleep made electrooculographic recording critical in performing sleep studies.

Meanwhile, in Lyon, France, Michel Jouvet noted postural differences during sleep in cats.[4] These differences correlated with sleep state and reduced skeletal electromyographic activity. REM sleep (and, by association, dreaming) coincided with marked hypotonia in descending alpha and gamma motor neurons. This hypotonia induced functional paralysis that was quickly ascribed the purpose of keeping the sleeper from enacting dreamed activities. This sleep state–related electromyographic alteration added the final compulsory recording component to the procedure now known as polysomnography (PSG).

Clinical PSG, in addition to brain wave, eye movement, and muscle tone recording, also assesses respiratory, cardiac, and limb movement activity (discussed in detail in other

chapters and elsewhere in the literature[5]). PSG in its simplest form, however—consisting of an electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG)—provides the basic information requisite for classifying sleep state and examining sleep processes.

ELECTRODE PLACEMENT AND APPLICATION

To record the EEG, EOG, and EMG, electrodes are placed on the scalp and skin surfaces. The site must be cleansed and properly prepared to ensure good contact and to maintain electrical impedance at or below 5000 Ω. Scalp electrodes can be affixed with collodion or electrode paste. Facial electrodes can be applied with double-sided adhesive electrode collars and paper tape. Although prescribed sites for electrode application have changed over the years, the system used to identify location remains the American Electroencephalographic society's international 10-20 system. In this system, the intersection of lines drawn from the left to right preauricular point, with the midpoint along the scalp between the nasion and inion, serves to landmark the vertex, designated Cz. Other loci can be found by measuring 10% and 20% along longitudinal and lateral surfaces. Specific locations are designated with a letter indicating the brain area below the electrode (e.g., C for central lobe, O for occipital lobe, F for frontal lobe) and a number ascribing specific points (odd numbers for the left side, even numbers for the right, and z for midline). EEG electrode placements should be precise; consequently, appropriate measurement techniques must be are essential to ensure accuracy. Additionally, EEG amplifiers require calibration at the beginning and end of PSG recording to document proper functioning. This calibration-recalibration provides verification that amplitude changes of the recorded signal accurately reflect oscillating voltages from brain activity.

The classic and amazingly long-lived standardized technique (i.e., the manual produced by the ad hoc committee chaired by Rechtschaffen and Kales) requires a single monopolar central-lobe scalp EEG electrode referenced to a contralateral mastoid electrode (either C3-M2 or C4-M1). This single-channel brain wave recording, when paired with right and left eye EOGs and submentalis EMG, sufficiently reveals brain, eye, and muscle activity for classifying sleep stages.[6] With evolution of PSG from a psychophysiologic research method to a clinical procedure, addition of an occipital lead supplemented centrally derived EEG for improved visualization of waveforms needed to differentiate sleep from wakefulness and to detect central nervous system (CNS) arousals.[7,8]

EOG recording capitalizes on the eyes' cornea-retina potential difference. Strong positive corneal potential fields affect electrodes placed near the right and left outer canthi of the eyes. The recording traces the response to this positive charge moving toward or away from the recording site. Each electrode is referenced to a neutral site, typically over the mastoid behind the ear. Thus lateral eye movements produce out-of-phase tracings for right and left EOG tracings as the cornea moves toward one electrode and away from the other (provided that two channels are dedicated to tracing eye movements). This arrangement makes eye movements easily differentiable from in-phase frontal lobe EEG activity that also is present on recording from these sites. To discern vertical eye movements, the right-side EOG electrode can be placed 1 cm above the outer canthus and the left-side electrode is positioned 1 cm below (or vice versa). An alternative recording montage devised to enhance vertical eye movement detection entails lowering both recording sites to 1 cm below the outer canthi and referencing each to the middle of the forehead (Fpz).

Skeletal muscle activity level is estimated from a pair of electrodes arranged to record submentalis EMG activity. An electrode placed approximately midline but 1 cm above the mandible's inferior edge is referenced to another placed 2 cm below and 2 cm to the right (or left). As a precaution, a backup electrode also is attached at a site at least 0.5 cm away from either of the previously described locations and on the same side (to decrease ECG artifact). Of note, this physical separation is critical to maintain the integrity of the electrode. If electrodes come in contact with one another, they merge to become one recording site. The resulting submentalis EMG recording serves qualitatively to provide an overall estimate for muscle activity level.

The American Academy of Sleep Medicine (AASM) published a standardized manual for conducting clinical polysomnography in their accredited sleep disorders centers.[7,9] This AASM standards manual makes recommendations for recording, scoring, and summarizing sleep stages, CNS arousals, breathing, various kinds of movement, and electrocardiographic activity. By bringing instructional guidelines for a range of techniques into a single volume, the AASM manual has strongly influenced practice, particularly in North America. Researchers, however, should not feel constrained by these clinical guidelines. New discoveries and future techniques need to continue unshackled by even a de facto standard clinical practice cookbook.

AASM specifies recording frontal, central, and occipital monopolar EEG from F4, C4, and O2. The contralateral mastoid (M1) serves as the ostensively neutral reference. Electrodes placed at F3, C3, and O1 sites (and referenced to M2) provide redundancy and serve as backup, if needed. The AASM manual sanctions the use of midline bipolar recordings for frontal and occipital EEG; however, the AASM "frequently asked questions" (FAQ) section states that frontal bipolar derivations are not appropriate for measuring frontal EEG activity. The FAQ also states that EEG amplitudes can be measured from the C4-M1 derivation. The AASM manual recommends using mastoid-referenced EOG with separate channels for E2 and E1, but it also approves a forehead-referenced alternative montage. Submentalis EMG is recorded as a bipolar derivation—that is, one of the surface EMG electrodes is referred to one of the other surface EMG electrodes on the chin.

DIGITAL RECORDING REQUIREMENTS

The first time a polysomnographic signal was digitized, whether it originated from analog or digital amplifying circuits, an entirely new set of factors required consideration. The two most important issues involved specifying amplitude and temporal resolution. Selection of voltage per digital unit (bit) and sampling rate probably had more to do with computer hardware limitations than with conceptual considerations. Amazingly, no standard was established for digital PSG until publication of the AASM standards manual.

The AASM standards manual specifies minimum 12-bit representation for amplitude, providing 4096 units to

represent a 2.5-volt regulated current (IREG) range, or its equivalent (Video 161-1). In this manner, even the smallest signals, exceeding the level of electrical noise, can be detected. Temporal resolution during recording depends on sampling rate and ultimately must allow accurate waveform reconstruction, provide enough data to potentially overcome frequency aliasing, and be appropriate for high- and low-pass digital filter settings. One size does not fit all: The minimum temporal resolution needed during data acquisition to meet these requirements varies for different bioelectrical signals (Table 161-1).

Additional digital specifications involve data selection, display pagination, and calibration. Recorded channels must be selectable, and the channel calibration must be available for display and documentation. The viewable data should provide user-selectable time frame compression and expansion (ranging from 5 seconds to an entire night shown on a page). Display screens definition should be at least 1600 × 1200 pixels. Digital polysomnographs should provide the capability to view data as they appear on the initial recording and also as they appear after sleep staging and sleep-related events have been marked and classified manually. Accompanying video at a minimum of one frame per second should be synchronized with the polysomnographic display.

ELECTROENCEPHALOGRAM BANDWIDTHS, WAVEFORMS, AND OTHER ACTIVITY

Bandwidths

One approach to differentiating EEG involves separating activity into dominant frequency bandwidths. *Delta activity* includes brain waves with a frequency less than 4 Hz. Sleep-related delta waves occurring at the low end of the frequency spectrum are called *slow waves.* Slow waves have high amplitude (greater than 75 microvolts) and low frequency (less than 2 Hz). *Theta activity* includes 5- to 7-Hz waves prominent in central and temporal leads. *Alpha activity* is characterized by an occipitally prominent 8- to 13-Hz rhythm, and *beta waves* include the low-amplitude waves at even higher frequencies (up to approximately 25 Hz for clinical purposes).

Waveforms

In addition to ongoing background EEG activity oscillating predominantly within one or another of the specific bandwidths, distinct transient waveform events occur. These include vertex sharp waves, K-complexes, sleep spindles, and sawtooth theta waves. *Vertex sharp waves* are sharply contoured, negative ("negative" meaning upward deflection of the signal, as per EEG polarity convention) wave forms that stand out from the background EEG activity. As the name implies, they appear prominently in EEG derived from electrodes placed near the midline or vertex region (Cz).

The *K-complex* begins much like a vertex sharp wave (i.e., it begins with a sharp negative waveform but is immediately followed by a large, usually much slower positive component). Overall, the K-complex usually is clearest in central and frontal regions and has a duration criterion of 0.5 second or more. A *sleep spindle* is a readily apparent 0.5-second (or longer) burst of 12- to 14-Hz activity generated by the thalamus and sent along thalamocortical pathways to the cortex. The name derives from its spindle-like shape. A *sawtooth wave* is a variant of theta activity, with each wave also containing a notch, making it sawtooth-shaped.

Table 161-1 Recording Recommendations for Digital Polysomnography

	Sampling Rate (Hz)*		Filter Setting (Hz)	
Recording Channel	Desirable	Minimal	Low *f*	High *f*
Central EEG (C4-M1)	500	200	0.3	35
Occipital EEG (O4-M1 or Cz-Oz)	500	200	0.3	35
Frontal EEG (F4-M1 or Fz-Cz)	500	200	0.3	35
Left EOG (E1-M2 or E1-Fpz)	500	200	0.3	35
Right EOG (E2-M2 or E2-Fpz)	500	200	0.3	35
Muscle tone (submental EMG)	500	200	10	100
ECG (lead II, modified)	500	200	0.3	70
Airflow sensors at nares and mouth	100	25	0.1	15
Oximetry (ear lobe or finger)	25	10	0.1	15
Nasal pressure	100	25	0.1	15
Esophageal pressure	100	25	0.1	15
Body position	1	1		
Respiratory effort				
Snoring sounds	500	200	10	100
Rib cage and abdominal movement	100	25	0.1	15
Intercostal EMG	500	200	10	100

E1, Left eye; E2, right eye; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram; *f*, frequency; Fpz, frontal pole; M, mastoid.

*Higher sampling rates increase file storage requirements but provide increased temporal resolution. The tradeoff between fidelity and practicality is a matter of debate.

Other nonpathologic sleep-related waveforms exist (e.g., benign epileptiform transients of sleep [BETS], sensory motor rhythm [SMR], wicket rhythm [mu rhythm], and positive occipital sharp transients of sleep [POSTS]. These normal variants do not occur consistently during polysomnography.

Activity Patterns

Sleep EEG also contains dynamic activity patterns not captured by sleep staging schema or identification of individual waveforms. The *cyclic alternating pattern* (CAP) includes waveform bursts (usually high-amplitude slow, sharp, or polymorphic waves) separated by quiescent periods.[10] The pattern's burst component sometimes includes transient alpha-bandwidth components meeting CNS arousal scoring criteria and thus can index sleep disturbance. However, a CAP occurring without frank arousals is thought to signify more subtle sleep instability.

SLEEP STAGING RULES AND CENTRAL NERVOUS SYSTEM AROUSALS

More than 40 years ago, the standardized technique described in *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*[6] provided a unifying methodology for human sleep research. This standardized manual combined elements from various systems that had evolved over time, and it provided adequate detail to achieve general use. To a large extent, however, its enormous success stems from the consensus it attained from the multinational, multidiscipline stakeholders that made up its development committee. That is, when the committee members returned to their respective laboratories, they used the techniques and taught them to scientists and clinicians in training.

Staging, as a summarizing technique, necessarily must define a period over which the summary applies. The standardized manual endorsed 20- and 30-second time domains (epochs). This flexibility deferred to extant technology—that is, generally available polygraph machine paper chart drive speeds. Over time, the 30-second epoch won out because it provided enough detail to see waveforms (EEG standards dictate minimal paper speed of 10 mm/second to ensure ability to discern individual EEG waveforms); at 10 mm/second, one epoch fits on a standard 30-cm wide paper fanfold polygraph page; and one 1000-page box of polygraph paper would hold a complete recording (or two if one also is recorded on the back) (Box 161-1).

Sleep Staging Rules

Wakefulness (stage W) in a relaxed subject with eyes closed is differentiated from sleep by the presence of alpha EEG activity in 50% or more of the epoch (Figure 161-1). Poorly defined alpha EEG activity complicates differentiation of sleep onset from wakefulness. The observation of the reactivity (disappearance) of the alpha rhythm to eyes open versus eyes closed provides a helpful contrast to facilitate detection of the background alpha rhythm in wakefulness.

Stage 1 typically is defined by exclusion; that is, it appears as a low-voltage, mixed-frequency background EEG signal devoid of sleep spindles and K-complexes, minimal slow wave activity, cessation of blinking, absence of saccadic eye movements, and alpha activity for less than 50% of the epoch duration (Figure 161-2). Stage 1 sleep may, but does not necessarily, include vertex sharp waves, background activity slowing, and slow eye movements.

Stage 2 characteristics include sleep spindles and K-complexes (Figure 161-3) occurring on a low-voltage, mixed-frequency background EEG and minimal (less than 20% of the epoch) slow wave (0.5 to 2 Hz, 75 μV) activity.

Box 161-1 CLINICAL TIPS FOR SLEEP STAGE SCORING

If the goal is to score an entire sleep study, a helpful approach is to familiarize oneself with the entire recording before beginning the epoch-by-epoch task of assigning a sleep stage to each 30-second epoch. A useful first step is a general review of the study recording to discern overall patterns—for example, the alpha background, the shape of the K-complexes, presentation of sleep spindles, and the appearance or lack of slow waves. Each person has an individual "signature of sleep," and familiarity with the data beforehand from each study makes scoring easier. The scorer "shakes hands," as it were, with the recording and asks general questions. For example, did the subject sleep or not? Does the person have robust, easy-to-discern eye movements? What do the K-complexes look like? Is the breathing regular or not? Do the patient's legs move or not? Are there any ECG rhythm abnormalities? This brief review creates a context for the scoring process. Also, it is common for the scorer to have to go back to a point earlier in the data, or to look ahead, before making a decision about a specific epoch, especially when determining the exact epoch on which REM sleep began or to make the decision about the precise moment of sleep onset. All individuals engaged in sleep stage scoring for a laboratory should engage in systematic interrater reliability checks to ensure that everyone is scoring the data in similar fashion.

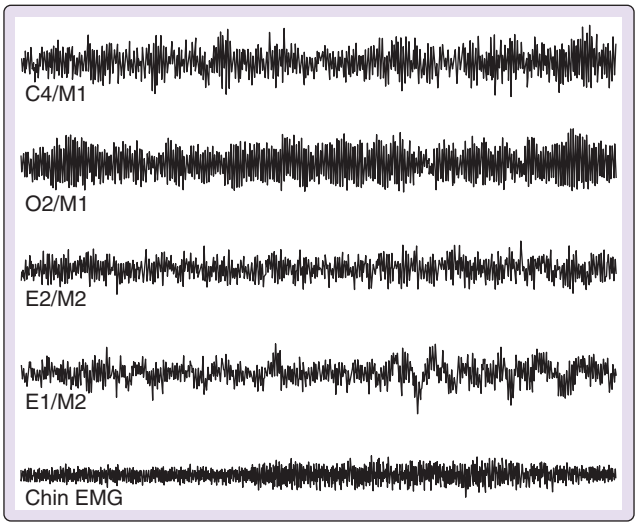


Figure 161-1 Stage wake (W), eyes closed. This example demonstrates a classic wake pattern, with alpha rhythm in the EEG and EOG. Alpha activity is most prominent in the occipital channel. The chin EMG displays normal muscle tone associated with relaxed wakefulness. C4/M1, Right central lobe EEG referenced to left mastoid; E2/M1, right eye (outer canthus) referenced to left mastoid; E1/M2, left eye (outer canthus) referenced to right mastoid; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram; O2/M1, right occipital lobe EEG referenced to left mastoid. (From Butkov N. *Atlas of clinical polysomnography.* 2nd ed. Medford [Ore.]: Synapse Media; 2010.)

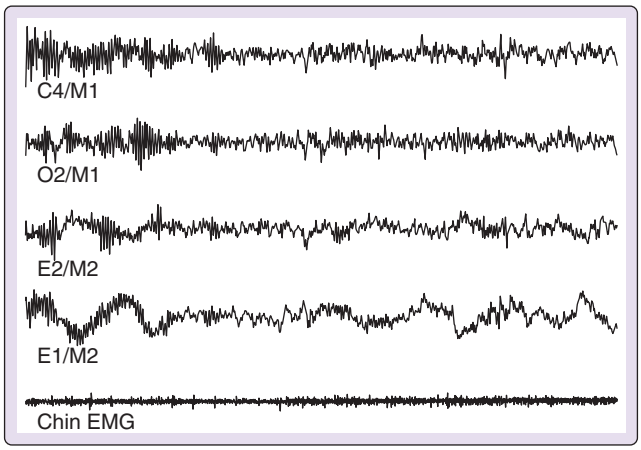


Figure 161-2 Stage 1 sleep (N1). The onset of N1 is identified by the disappearance of alpha rhythm, replaced by relatively low voltage mixed-frequency EEG with a prominence of theta activity in the range of 5 to 7 Hz. The chin EMG remains tonic, although it can attenuate slightly with sleep onset. C4/M1, Right central lobe EEG referenced to left mastoid; E2/M1, right eye (outer canthus) referenced to left mastoid; E1/M2, left eye (outer canthus) referenced to right mastoid; EEG, electroencephalogram; EMG, electromyogram; O2/M1, right occipital lobe EEG referenced to left mastoid. (From Butkov N. *Atlas of clinical polysomnography.* 2nd ed. Medford [Ore.]: Synapse Media; 2010.)

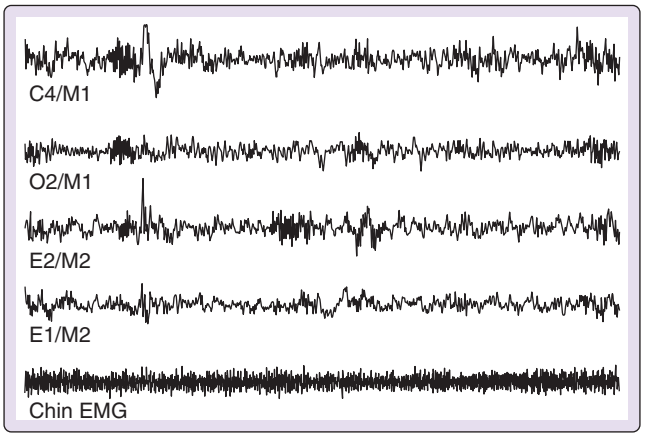


Figure 161-3 Stage 2 sleep (N2). Stage N2 is identified by the presence of K-complexes and/or sleep spindles against a background of mixed-frequency EEG. The chin EMG displays normal muscle tone, as expected during NREM sleep. C4/M1, Right central lobe EEG referenced to left mastoid; E2/M1, right eye (outer canthus) referenced to left mastoid; E1/M2, left eye (outer canthus) referenced to right mastoid; EEG, electroencephalogram; EMG, electromyogram; O2/M1, right occipital lobe EEG referenced to left mastoid. (From Butkov N. *Atlas of clinical polysomnography.* 2nd ed. Medford [Ore.]: Synapse Media; 2010.)

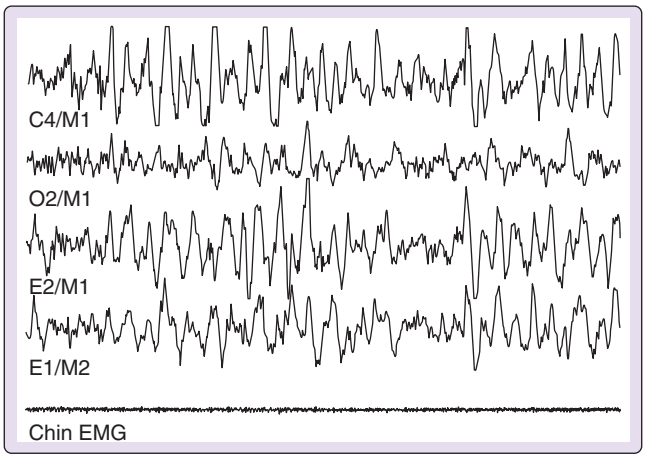


Figure 161-4 Slow wave sleep (N3). In this example, high-amplitude slow waves occupy greater than 50% of the epoch. By the Rechtschaffen and Kales (R&K) criteria, this epoch is scored as stage 4. By the revised American Association of Sleep Medicine (AASM) criteria, this epoch is scored as N3. (From Butkov N. *Atlas of clinical polysomnography.* 2nd ed. Medford [Ore.]: Synapse Media; 2010.)

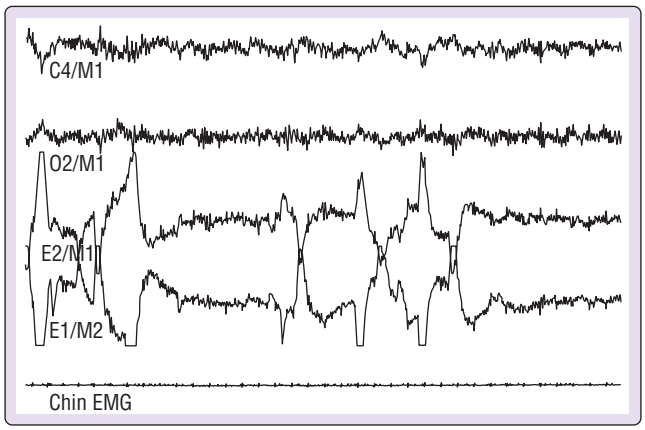


Figure 161-5 Rapid eye movement (REM) sleep. During REM sleep, chin muscle tone drops to the lowest level of the recording. REM sleep is identified by the presence of rapid eye movements in combination with relatively low-voltage, mixed-frequency EEG and low-chin EMG. C4/M1, Right central lobe EEG referenced to left mastoid; E2/M1, right eye (outer canthus) referenced to left mastoid; E1/M2, left eye (outer canthus) referenced to right mastoid; EEG, electroencephalogram; EMG, electromyogram; O2/M1, right occipital lobe EEG referenced to left mastoid. (From Butkov N. *Atlas of clinical polysomnography.* 2nd ed. Medford [(Ore.]: Synapse Media; 2010.)

Slow wave sleep (stages 3 and 4 sleep) contains delta EEG activity (recorded from a monopolar central derivation) with a 75-μV or greater amplitude enduring for 20% or more of an epoch (Figure 161-4). *Stage 3* is scored when the duration of slow waves constitutes 20% to 50% of the epoch, and *stage 4* is scored when duration reaches 50% or more.

REM sleep is scored when saccadic eye movements occur during epochs with low-voltage, mixed-frequency EEG activity in association with a very low level of submentalis EMG activity (Figure 161-5). Epochs with low-voltage, mixed-frequency EEG activity and continuing low-level submentalis EMG (without eye movements) falling between epochs of REM sleep (with eye movements) also are scored as REM sleep. Epochs falling before or after (and contiguous with) clear REM sleep that have comparable EEG and EMG features but lack rapid eye movements are scored as REM sleep until an arousal, EMG level increase, or resumption of K-complexes or sleep spindles occurs. These *smoothing rules* gloss over minor transitions on the supposition that REM sleep represents a persistent CNS organizational state distinct from wakefulness and NREM sleep.

In 2007 (last update 2015), the AASM standards manual provided revised criteria for scoring sleep stages. Changes are summarized in Table 161-2. Essentially, changes include standardizing epoch length at 30 seconds; combining stages 3 and 4 sleep and applying amplitude criteria for slow waves to

Table 161-2 Comparison of Traditional and AASM (2007) Sleep Stage Scoring Systems

Parameter	R&K Classification Criteria	AASM Classification Criteria
Epoch length	20 or 30 seconds, user's choice	30 seconds, mandated
Stage nomenclature	Wakefulness, stage 1 sleep, stage 2 sleep, stage 3 sleep, stage 4 sleep, REM sleep, movement time	Stages W, N1, N2, N3, and R
Wakefulness	EEG alpha activity for ≥50% of an epoch	Same
Slow wave sleep	EEG slow wave activity for ≥50% of the epoch for stage 4 sleep or ≥20% of the epoch for stage 3 sleep	Same, except that stages 3 and 4 are combined into N3
Stage 2 sleep	Sleep spindles or K-complexes; EEG slow wave activity for <20% of the epoch	Same
Stage 1 sleep	Low-voltage, mixed-frequency activity; possibly vertex sharp waves; possibly slow eye movements; no sleep spindles or K-complexes; EEG alpha activity for <50% of the epoch	Same
REM sleep	Low-voltage, mixed-frequency EEG activity; very low submental EMG activity; possibly saw tooth EEG theta activity; at least one unequivocal rapid eye movement	Same
Movement time	Polysomnographic activity obscured to the point of not being readable for more than 50% of the epoch; the preceding epoch is scored as stage 1, 2, 3, 4, or REM sleep	This epoch classification is eliminated
Smoothing rules	When an epoch is classified as a particular stage but is surrounded by epochs lacking unique features (e.g., a sleep spindle, slow-rolling eye movements, or CNS arousal) and would otherwise have been scored as stage 1 sleep, the *classified epoch* scoring is generalized to the surrounding epochs (but only for 3 minutes). These *smoothing rules* apply to stage 2 and REM sleep	Same, except that there is no 3-minute limit to the generalization

AASM, American Academy of Sleep Medicine; CNS, central nervous system; EEG, electroencephalogram; EMG, electromyogram; R&K, Rechtschaffen and Kales; REM, rapid eye movement.

frontal EEG activity; revising terminology (R for REM sleep, N1 for NREM stage 1, N2 for NREM stage 2, N3 for NREM stages 3 and 4, and W for wakefulness); and simplifying smoothing rules. Some changes are controversial.[11-16]

Central Nervous System Arousal Scoring

Sleep staging fails to represent brief CNS arousals because it summarizes EEG, EOG, and EMG activity over a 30-second time domain. Increasing clinical application of polysomnographic technique heightened the need to appreciate sleep fragmentation; consequently, a scoring system for arousals was developed[8] under the auspices of the American Sleep Disorders Association (later to become the AASM). Abrupt 3-second (or longer) EEG frequency increases to theta, alpha, or beta activity (but not to sleep spindles) are considered biomarkers for CNS activation. The arousals most often entail emergent occipital EEG alpha activity. To qualify as an arousal, 10 seconds of sleep must precede the event. In REM sleep, activity must also increase in submentalis EMG leads for at least 1 second (Figure 161-6). The 3-second duration represents the minimum duration that could be reliably scored by visual inspection (among the task force members). Events of shorter duration likely also have clinical significance. The AASM standards manual endorsed this scoring technique and simplified the original 11 rules to a single statement with 2 explanatory notes.

SUMMARIZING NORMAL SLEEP

The sleep stage pattern across the night can be represented diagrammatically (Figure 161-7). In healthy young adults, stage R (REM sleep) accounts for approximately 20% to 25% of total sleep time, stage N2 accounts for 50%, N3 accounts for 12.5% to 20%, and N1 accounts for the remainder. In normal sleepers, wakefulness characteristically accounts for 5% to 10% of the time in bed. Stage R typically does not appear until approximately 90 minutes after sleep onset, after which it reoccurs every 90 to 120 minutes in distinct episodes. These episodes increase in duration as sleep progresses; accordingly, the first half of the sleep session contains less REM sleep than the second half. By contrast, slow wave activity (stage N3) predominates in the first third of the night. Age-matched sleep architecture bears great similarity in men and in women; however, women may have slightly better preserved stage N3 with advancing age. Sleep can be quantitatively summarized, and Table 161-3 provides definitions for commonly used parameters.

AMBIGUOUS SLEEP STAGES AND SLEEP QUALITY

Sleep stage scoring was developed to summarize EEG, EOG, and EMG correlates of *normal sleep*. Under normal

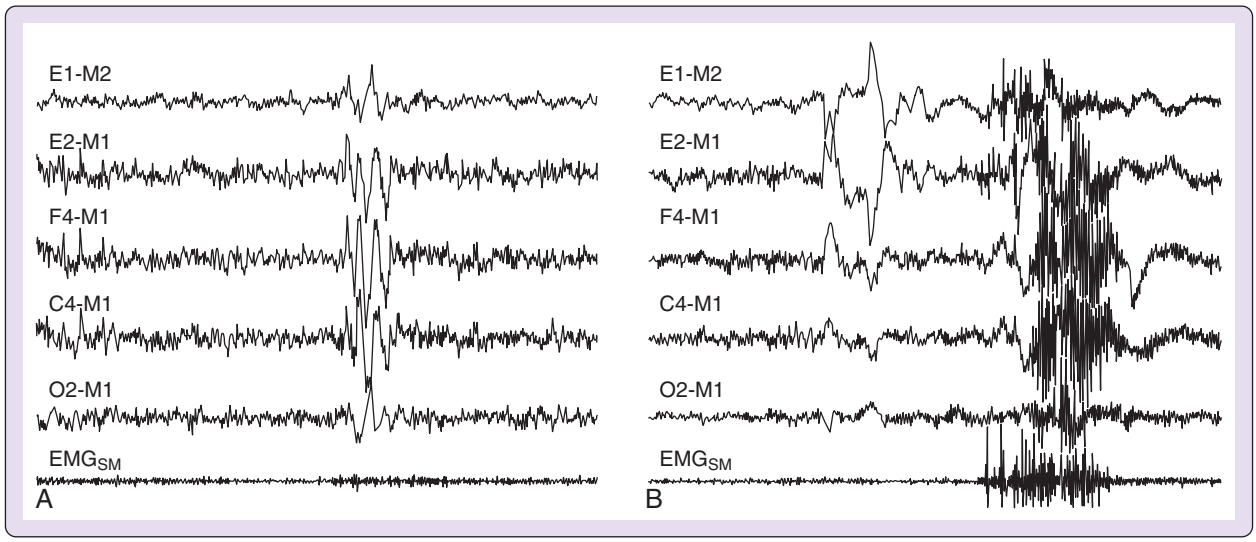


Figure 161-6 Arousals from NREM **(A)** and REM **(B)** sleep. **A,** A paroxysmal burst of high-amplitude slow wave activity appears near the center of the epoch. The distribution of the electrical field changes associated with this event can be seen reflected in the other EEG channels but with (expected) decreased amplitude. Little or no change occurs in the EMG channel in association with this event. It is common to see K-complex activity evoked by auditory stimuli. **B,** An increase in EMG activity is noted on the EMG and almost simultaneously in the E2-M1 channel. This is followed by a brief generalized presentation of EMG artifact throughout the EEG and EOG channels. Before the event, there is evidence of REM sleep: low-voltage, mixed-frequency EEG, rapid eye movements, and very low EMG. After the event, the EMG channel shows an increased tone, and the EEG background activity is low-voltage and fast. A continuation of EMG artifact is seen on the EEG channels after the short burst. These data, especially the burst of alpha activity seen in the O2-M1 channel, are consistent with a possible transition to wake from REM sleep. C4-M1, Right central lobe EEG referenced to left mastoid; E1-M2, left eye (outer canthus) referenced to right mastoid; E2-M1, right eye (outer canthus) referenced to left mastoid; EEG, electroencephalogram; EMG, electromyogram; EMGSM, submentalis EMG; EOG, electrooculogram; F4-M1, right frontal lobe EEG referenced to left mastoid; NREM, non–rapid eye movement; O2-M1, right Occipital Lobe EEG referenced to left mastoid; REM, rapid eye movement. (Courtesy Max Hirshkowitz, PhD, DABSM.)

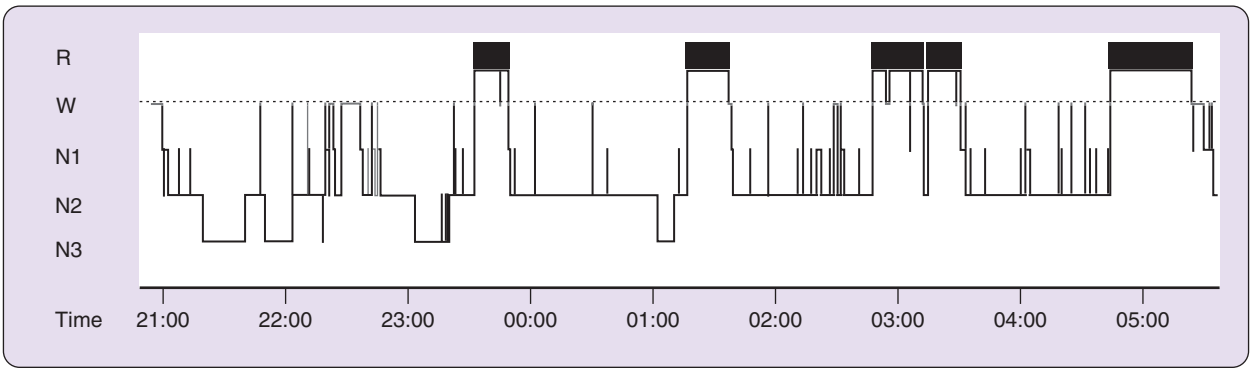


Figure 161-7 Normal sleep histogram illustrating sleep macroarchitecture (stages) for a young adult. N1, N2, and N3, NREM sleep stage 1, 2, and 3, respectively; R, REM sleep; W, wakefulness.

circumstances, particular events cluster for the vast majority of the time. By contrast, this tight coupling tends to loosen when patients rebound from sleep deprivation; sustain brain injury; are afflicted with sleep, medical, neurologic, psychiatric, or sleep disorders; or ingest psychoactive substances. The resulting intrusion, translocation, or migration of specific EEG, EOG, or EMG activity characteristic of one stage into another produces ambiguous epochs that are difficult to classify according to the usual scoring rules. This departure from normal processes can provide qualitative evidence of an underlying sleep dysfunction.

Perhaps the most common ambiguities accompany pharmacotherapy. Gamma-aminobutyric acid type A and benzodiazepine receptor agonists generally increase spindle activity in the EEG. These pharmacologically induced spindles typically are of higher frequency (16 to 18 Hz) and often of longer duration, occur more frequently (with a higher density), and can appear not only in N2 but also in other stages of sleep and in wakefulness.

Another commonly noted drug effect is serotonin agonist augmentation of eye movement activity. In some persons, rapid eye movements occur at sleep onset and in sleep stages

Table 161-3 Parameters Derived from Sleep Staging and Central Nervous System Arousal Scoring

Parameter	Notation	Explanation
AASM Recommended Parameters		
Lights out clock time	L-out	The clock time (in hh:mm) that the subject was instructed to allow himself or herself to fall asleep
Lights on clock time	L-on	The clock time (in hh:mm) that the subject was awakened
Total sleep time	TST	Minutes scored as stage N1, N2, N3, or R
Total recording time	TRT	Elapsed time from L-out to L-on (in minutes)
Sleep latency	SLAT	Elapsed time from L-out to first epoch of stage N1, N2, N3, or R (in minutes)
REM sleep latency	RLAT	Elapsed time in minutes from SLAT to first epoch of stage R
Wake after sleep onset	WASO	Minutes scored as stage W from first sleep epoch to L-on
Sleep efficiency	SEI	TST as a percentage of TRT
Time in each stage	MW, M1, M2, M3, MR	Minutes scored as W, N1, N2, N3, and R (individually)
Sleep stage percentages	P1, P2, P3, PR	Time scored as N1, N2, N3, and REM as a percentage of TST (individually)
Number of CNS arousals	NArsls	The number of CNS arousals
CNS arousal	CNS AI	The number of CNS arousals scored per hour of TST
Other Useful Parameters		
Latency to persistent sleep	LTPS	Elapsed time (in minutes) from L-out to first of 10 consecutive minutes of sleep
Latency to unequivocal sleep	LUS	Elapsed time (in minutes) from L-out to first epoch of N2, N3, or R or to three consecutive (or more) epochs of N1 If N1 is followed by an epoch of N2, N3, or R, LUS is calculated from L-out to the first epoch of N1
Sleep-period time	SPT	Minutes from first to last epoch scored as N1, N2, N3, or R
Number of REM sleep episodes	NREME	Number of stage R occurrences
Number of awakenings	NWake	Number of stage W occurrences
Wake index	WI	Number of awakenings per hour of TST
Sleep fragmentation index	SFI	Number of awakenings and CNS arousals per hour of TST
Number of stage shifts	NShifts	Number of stage transitions during TRT
Stage shift index	SSI	Number of stage transitions per hour of TRT
Latency to arising	LTA	Duration of final stage W if it was ongoing when L-on occurred

CNS, Central nervous system; REM, rapid eye movement.

N2 and N3, making the scoring of REM sleep a challenge. The phenomenon is so common that many sleep specialists refer to it as "Prozac eyes" (with reference to fluoxetine, the prototypical selective serotonin reuptake inhibitor).

Another serotonin agonist–provoked sleep alteration involves elevated muscle activity during REM sleep. In some cases, these medications produce a loss of atonia, permitting attempted dream enactments—that is, iatrogenic REM sleep behavior disorder (RBD). Individual PSG epochs during these events do not meet usual stage classification criteria. Similar REM sleep ambiguities occur in Parkinson disease–related and posttraumatic stress disorder–related RBD.

Patients suffering from neurodegenerative diseases or brain insult can manifest an overall erosion of EEG sleep events. This effect includes reduced sleep spindles, K-complexes, and slow wave activity. We also sometimes observe this effect in patients with sleep apnea, heart failure, and metabolic disorders. The resulting nearly featureless sleep EEG can be difficult to score according to normal staging rules. By contrast, another very different scoring problem can occur in persons with severely fragmented sleep produced by obstructive apnea, in whom a continual cycle of falling asleep, airway collapse, struggle to breathe, awakening, and falling asleep is observed. Thus the patient remains in a *t*ransition state that does not fit well into any sleep stage category. It was once proposed that this pattern be scored as *t-sleep*.

In some persons, copious EEG alpha activity permeates ongoing background activity. In sleep states marked by low-amplitude, mixed-frequency activity, alpha bursts meeting criteria for CNS arousal can be scored as such (alpha intrusion). However, when slow waves characterize the dominant ongoing background EEG activity and the alpha coincides with delta, arousals are not scored. This *alpha-delta* sleep sometimes accompanies pain syndromes, but it appears to lack

specificity. A related phenomenon, also ascribed to pain, consists of K-complex bursts followed by EEG alpha activity. Many sleep specialists consider this "K-alpha" activity to be a variety of the CAP.

CLINICAL PEARLS

- Sleep staging and CNS arousal scoring provide important clinical information about brain processes during sleep. Ultimately, persons who awaken sleepy or unrefreshed or who have difficulty initiating or maintaining sleep can be assayed for sleep integrity, quantity, and quality using polysomnography.
- Human sleep is a brain process. Pathophysiologic conditions such as increased airway resistance and leg movements produce CNS arousals that fragment and destroy the fabric of sleep. Disorders often alter sleep patterns and overall architecture.
- Appropriate treatments may promote return to normal. Quantitative analysis through staging and arousal scoring objectively documents sleep disruption and provides a severity index for sleep disorders.

SUMMARY

Polysomnography involves recording a wide assortment of bioparameters while a person sleeps. The only objective method for knowing if a person is sleeping is to examine brain wave activity (the electroencephalogram). Sleep stage scoring summarizes patterns in the electroencephalogram, electrooculogram, and skeletal muscle electromyogram. Well established, specific, scoring criteria exist for sleep stages N1, N2, N3, and R (previously called stage 1, 2, 3, 4, and REM, respectively). Scoring criteria depend upon EEG bandwidth activity (delta, theta, alpha, and beta), EEG events (vertex sharp waves, sleep spindles, and K-complexes), eye movement activity (slow and rapid eye movements), and the level of muscle tone. Stage N3 is characterized by high-voltage, slow wave activity. Stage N2 contains sleep spindles and K-complexes. Stage N1 has low-voltage, mixed-frequency background, possibly slow eye movements, and vertex sharp waves. If rapid eye movements accompany a low-voltage, mixed-frequency EEG and skeletal muscle tone is low, REM sleep is scored. CNS arousals also can occur from sleep, either spontaneously or resulting from pathophysiologic processes or environmental factors. Quantitative analysis of sleep stages and CNS arousals provides evidence for, contributes to the definition of, and indexes the severity of some sleep disorders. Similarly, these indices can provide objective outcome measures for assessing therapeutic interventions. This chapter summarizes recording, digital processing, and scoring techniques used for evaluating brain activity during human sleep and its disturbance by CNS arousals.

Selected Readings

Berry RB, Brooks R, Gamaldo CE, et al, for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, Version 2.2. Darien, Illinois: American Academy of Sleep Medicine; 2015.

Butkov N. *Atlas of clinical polysomnography*. 2nd ed. Medford (Ore.): Synapse Media; 2010.

Hirshkowitz M, Kryger MH. Diagnostic methods. In: Kryger MH, Avidan A, Berry R, editors. *Atlas of clinical sleep medicine*, 2nd ed. Philadelphia: Elsevier; 2014.

A complete reference list can be found online at ExpertConsult.com.

Chapter
162

Central Nervous System Arousals and Cyclic Alternating Patterns

Liborio Parrino; Mario Giovanni Terzano

Chapter Highlights

- Conventional sleep measures based on static 30-second epochs have certain limitations compared with the more dynamic interpretation of sleep based on cyclic alternating pattern (CAP) scoring criteria. To apply the rules of CAP interpretation, the concept of electroencephalogram (EEG)-confirmed arousal has been expanded to include EEG features characterized by high-voltage slow waves that have activation properties correlated with autonomic functions and muscle activity.
- CAP parameters have been clearly defined and have found clinical application in all major sleep disorders. Automated analysis to improve and expedite CAP quantification also is now available.
- As the EEG marker of sleep instability, CAP offers a practical additional tool to investigate the physiology of sleep, the pathophysiology of sleep disorders, and the mechanisms of action of medications and other therapeutic modalities.

In 1968, the standardized Rechtschaffen and Kales (R&K) criteria[1] for sleep staging became the internationally accepted system for sleep scoring. This system remained the standard for almost 40 years. An enormous body of data based on the R&K criteria has been accumulated in the scientific literature. In 2007, the American Academy of Sleep Medicine (AASM) published a new manual for scoring sleep and associated events.[2] The updated AASM manual revised the R&K sleep staging criteria and also addressed the scoring of arousals, respiratory events, sleep-related movement disorders, and cardiac abnormalities. It also considered differences for pediatric and geriatric age groups. With the AASM criteria, the visual scoring of sleep derives from at least three EEG channels, specifically, frontal, central, and occipital leads, rather than from a single central lead as with the R&K system. The classic electrooculogram (EOG) and electromyogram (EMG) derivations are retained. This provides more extensive scalp area coverage, so that significant sleep-wake rhythms and waveforms (e.g., K-complexes, delta waves) are better represented. The distinction among wakefulness ("W"), non–rapid eye movement (NREM) sleep ("N"), and REM sleep ("R") is maintained, but to simplify daily clinical practice, a reduced number of NREM stages is used : N1 (previous stage 1), N2 (previous stage 2), and N3 (previous stages 3 and 4).

The AASM system standardized the 30-second epoch as the basic time domain for summarizing activity. The dominant role of tonic background EEG activity confers a stepwise outline to the sleep histogram (*sleep macrostructure*) characterized by periods of static configuration (sleep stages) interrupted by rapid shifts (stage changes). Scoring based on 30-second sleep epochs neglects short-duration events, which therefore are not represented in the classic sleep staging reports, although such events also carry important information. Various features—K-complexes, sleep spindles, delta bursts, and others—lasting for intervals shorter than the scoring epoch (phasic events) have been described and commonly are collectively referred to as *sleep microstructure*. Analysis of the cyclic alternating pattern (CAP)[3] represents a sophisticated method for recognizing and reporting specific sleep microstructural activity.

THE CYCLIC ALTERNATING PATTERN

Spontaneous cortical activity—that is, ongoing activity in the absence of sensory stimulation—can show very complex collective features, with, in some cases, the membrane potential making spontaneous transitions between two different levels called up and down states. This alternation of down states of network quiescence and up states of generalized spiking and neuronal depolarization has been observed to occur spontaneously in a variety of systems and conditions, both in vitro and in vivo, during slow wave sleep, anesthesia, and quiet waking. The precise mechanism by which these up state transitions occur is still unclear, but network mechanisms seem to be involved.[4]

CAP is a well-defined physiologic cerebral activity occurring under conditions of reduced vigilance (sleep, coma), translating to a state of arousal instability and involving muscle, behavioral, and autonomic functions.[3] During NREM sleep, CAP is organized in sequences. A CAP sequence is composed of a succession of CAP cycles. The CAP cycle is composed of a phase A and the following phase B (Figure 162-1). All CAP sequences (each made up of a series of CAP cycles) begin with a phase A and end with a phase B. Each phase of CAP is 2 to 60 seconds in duration. This cutoff value relies on the consideration that approximately 90% of A phase events occurring during sleep are separated by an interval of less than 60 seconds.[5] The absence of CAP for more than 60 seconds is scored as non-CAP (Figure 162-2) and coincides with a condition of sustained physiologic stability.[6]

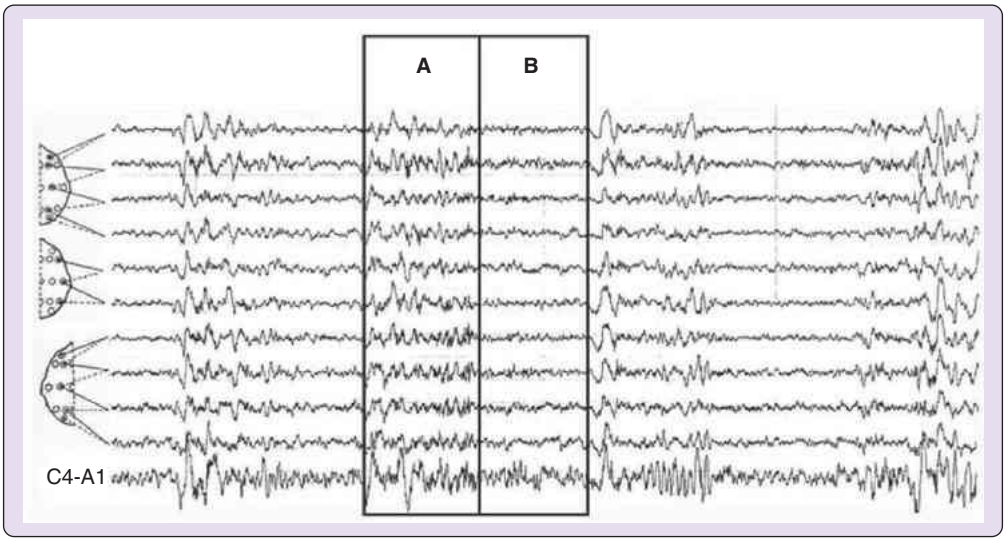


Figure 162-1 Cyclic alternating pattern (CAP) sequence. The *black outline* limits a CAP cycle composed of a phase A (cluster of K-complexes) and a phase B (interval between two successive A phases). Note that CAP is a widespread phenomenon that involves several EEG channels, with a dominance of high-amplitude waves in the anterior regions. EEG, Electroencephalogram.

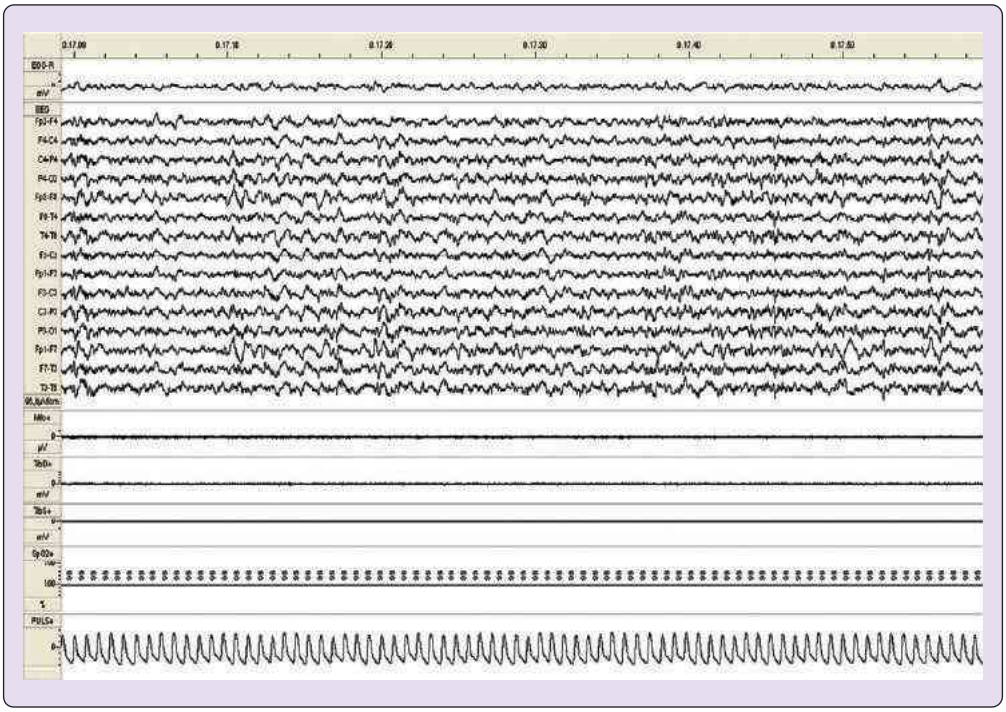


Figure 162-2 Non–cyclic alternating pattern (CAP) in stage N3 expressed by a sustained succession of slow waves, which correspond to the less-than-1-Hz oscillation of non–rapid eye movement (NREM) sleep. Arousals or other EEG phasic events are missing. Note the extreme stability of signal on the pulse and oximetry traces and the absence of muscle jerks. EEG, Electroencephalogram.

CAP and non-CAP can be manipulated by sensorial inputs. With separate application of the same arousing stimulus during the two EEG components of CAP, phase B immediately assumes the morphology of the other component, whereas the inverse transformation never occurs when the stimulus is delivered during phase A. This stereotypical reactivity persists throughout the successive phases of CAP, with no habituation. By contrast, when the same stimulus is presented during non-CAP, the EEG responses generally are brief, hypersynchronized (as slow waves), and proceed toward progressive habituation.[5] However, a robust or sustained stimulus delivered during non-CAP induces the immediate appearance of repeated CAP cycles that display the same morphology and reactive behavior of spontaneous CAP

sequences. The evoked CAP sequence heralds a lightening of sleep depth or continues as a damping oscillation before the complete recovery of non-CAP.[7]

CAP sequences have no upper limits for duration and number of CAP cycles, but at least two consecutive CAP cycles are required to define a CAP sequence. In the physiologic sleep histogram, CAP sequences occur during NREM stages, accompany most NREM stage shifts, and commonly precede the transition from NREM to REM sleep.[8] Under normal circumstances, CAP cannot be scored in REM sleep, because phase A features in REM sleep are separated by mean intervals of 3 to 4 minutes.[9] However, pathologic conditions characterized by repetitive A phase events recurring at intervals shorter than 60 seconds, such as periodic REM-related sleep apnea events,[10] can produce CAP sequences in REM sleep.

The Phase A Subtypes of Cyclic Alternating Pattern

Phase A activity can be classified into three subtypes. Subtype classification is based on the reciprocal proportion of high-voltage slow waves (EEG synchrony) and low-amplitude fast rhythms (EEG desynchrony) throughout the entire phase A duration.[6] The three phase A subtypes (Figure 162-3) are:

- *Subtype A1*: EEG synchrony (high-amplitude slow waves) is the predominant activity. If present, EEG desynchrony (low-amplitude fast waves) occupies less than 20% of the entire phase A duration. Subtype A1 specimens include delta bursts, K-complex sequences, vertex sharp transients, and polyphasic bursts with less than 20% of EEG desynchrony.
- *Subtype A2*: The EEG activity is a mixture of slow and fast rhythms, with 20% to 50% of phase A occupied by EEG desynchrony. Subtype A2 specimens include polyphasic bursts with more than 20% but less than 50% of EEG desynchrony.
- *Subtype A3*: The EEG activity is predominantly rapid low-voltage rhythms with greater than 50% of phase A occupied by EEG desynchrony. Subtype A3 specimens include K-alpha activity, EEG arousals, and polyphasic bursts with greater than 50% EEG desynchrony. A movement artifact within a CAP sequence also is classified as subtype A3.

CAP sequences include different phase A subtypes. A majority of EEG arousals occurring in NREM sleep (87%) fall within the CAP sequences and basically coincide with a phase A2 or A3.[11] In particular, 95% of subtype A3 events and 62% of subtype A2 events meet the AASM criteria for arousals.[12]

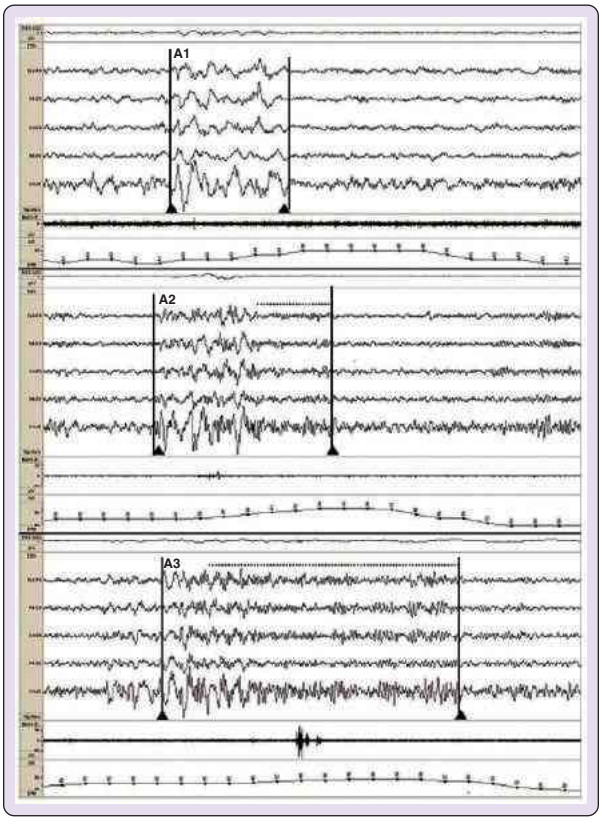


Figure 162-3 The three phase A subtypes, demarcated by *black vertical lines*. The A1 subtype is dominated by high-voltage, low-frequency waves, whereas subtypes A2 and A3 contain increasing amounts of low-voltage rapid EEG activity, identified by *dotted horizontal lines*. The A3 phase is accompanied by a mild EMG activation, whereas all three subtypes are associated with a transient heart rate acceleration. EEG, Electroencephalogram; EMG, electromyogram.

ELECTROENCEPHALOGRAPHIC AROUSALS

According to the online Psychology Dictionary, *arousal is "a state of physiological alertness and readiness for action," or "a pervasive state of cortical responsiveness believed to be associated with sensory stimulation and therefore, the activation of fibers from the reticular activating system," or "a state of excitement or energy expenditure linked to a strong emotion."*[13]

The concept of arousal has a long history that is closely connected with the development of sleep neurophysiology. Scoring criteria and measurement of arousal are controversial issues; consequently, several definitions have been proposed.[14] Behavioral and autonomic concomitants of arousal may or may not be present simultaneously. When they are present, they can be graduated in intensity, whereas presence or absence of the single components of the arousal constellation depends on the involvement of the specific cerebral compartments.

The AASM manual[2] confirms the definition of EEG arousal established in 1992 by the American Sleep Disorders Association[15] (Figure 162-4). According to the original framework, arousals are markers of sleep disruption representing a detrimental and harmful feature of sleep. For this reason, they were excluded from the conventional sleep staging procedures. Nevertheless a number of studies have established that spontaneous arousals constitute a natural concomitant of sleep and occur with increasing frequency along the life span, following the profile of maturation and aging.[16,17] Moreover, the spectral composition of arousals[18] and their ultradian distribution throughout the sleep cycle reveal that arousals are endowed within the texture of physiologic sleep[8,14,19] under the biologic control of REM-on and REM-off mechanisms.[20] In accordance with these indications, arousal scoring is now considered a fundamental process in sleep staging classification, as well as scoring of spindles and K-complexes. By contrast, in the R&K system,[1] a K-complex, with or without an arousal, was unmistakably a marker of sleep stage 2. According to the AASM manual,[2] instead, if a K-complex is associated with an arousal, the epoch is scored as N1. In other words, the R&K approach gives primacy to the role of EEG

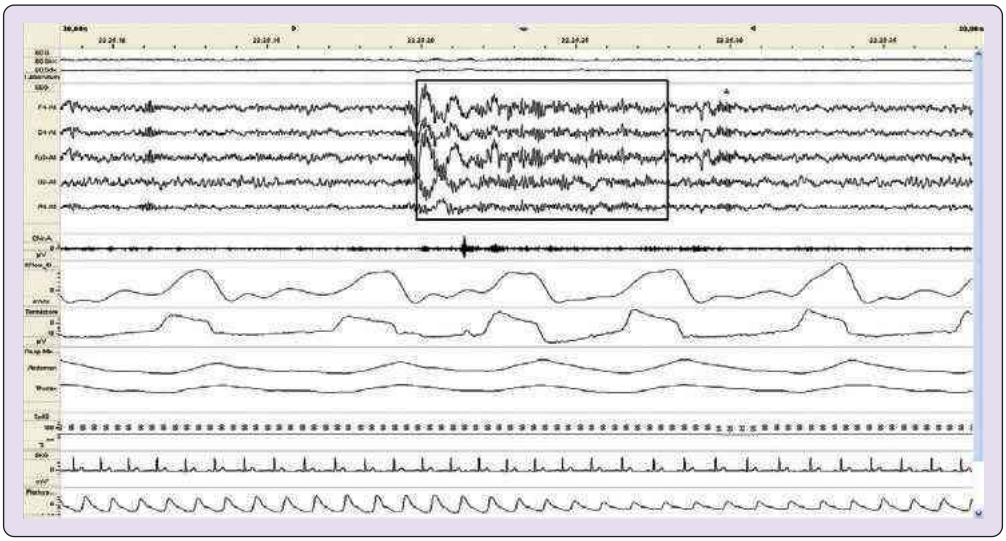


Figure 162-4 During sleep stages N1, N2, and N3 or rapid eye movement (REM) sleep, arousal is scored if there is an abrupt shift of frequency including alpha, theta, and/or frequencies greater than 16 Hz (but not spindles) that lasts at least 3 seconds, with at least 10 seconds of stable sleep preceding the change. Scoring of arousal during REM sleep also requires a concurrent increase in submental EMG activity lasting at least 1 second. Note that in non-REM (NREM) sleep, arousals often are heralded by one or more K-complexes. EMG, Electromyogram.

synchrony in the scoring process, whereas the AASM manual emphasizes the impact of EEG desynchrony.

A K-complex can be followed by other K-complexes, and in that case, stage 2 tends toward consolidation; or it can be followed by an arousal that reflects a regression toward a "shallower" neurophysiologic condition. Most arousals in NREM sleep (87%), however, are preceded by a K-complex or by a delta burst,[18] indicating that EEG synchrony and desynchrony may coexist in a natural dynamic succession as typified by CAP phases A2 and A3 (Figures 162-5 and 162-6).

Central Nervous System and Autonomic Arousals

> *Adaptive behaviour requires the selection of coherent sequences of action. What is needed for the production of organized sequential action is an internal representation of the temporal or task context, which can span periods during which the environment provides an indeterminate cue. An additional characteristic of human action is that it tends to assume a roughly hierarchic structure.*[21]

Somatosensory and auditory stimulation during sleep may result in cardiac, respiratory, and somatic modifications apparently without the EEG features of conventional central nervous system (CNS) arousal.[22] In particular, K-complexes and delta bursts are associated with activation of neurovegetative functions (heart rate, blood pressure) and can trigger muscle activity.[19] Despite their established arousal-related effect, however, K-complexes and delta bursts are neglected in scoring of arousals when they precede or occur independently of an EEG frequency speeding event.

This consideration does not imply that all arousal events are equal in terms of strength (i.e., immediate repercussions on motor and autonomic functions) and duration. It has been demonstrated that alpha EEG arousals lead to significantly different increases in systolic and diastolic blood pressure than do K-complexes, and studies carried out in patients with obstructive sleep apnea have ascertained that ASDA-defined arousals in NREM sleep are significantly longer after an apneic episode (12.6 seconds) than at the end of a hypopneic episode (9.9 seconds).[23] Moreover, quantifying arousal intensity in polysomnography (PSG) provides additional information on heart rate response.[24,25]

Taken together, these findings indicate that:

1. K-complexes and delta bursts are arousal equivalents; therefore any distinction between visible and nonvisible cortical arousals is misleading.
2. The arousal response is a continuous modulated phenomenon lying within a spectrum of EEG variability and having graduated effects on the ongoing polygraphic parameters.
3. The sleeping brain disposes of multiple arousal solutions that allow adaptive adjustments of the ongoing state to internal and external inputs.

Current knowledge about the arousal patterns has developed step by step, since the 1970s. The first step was taken by the Brussels school pioneers describing a "phase d'activation transitoire,"[9,26] the French version of AASM arousal. Subsequently, attention focused on the dual nature of arousals during NREM sleep, describing "desynchronization"- and "synchronization"-type components[14]—the latter named "antiarousal" by Hirshkowitz.[27] Finally, the whole arousal hierarchy was systematized by the Parma school in the CAP framework.[3,19] Accordingly, different expressions of arousal can be identified:

Behavioral arousal: Reported in the R&K manual[1] as movement arousal, it is characterized by any increase in electromyographic activity that is accompanied by a change in any other EEG channel.

Cortical arousal: Defined by the AASM committee as EEG arousal, it is identified by transient desynchronized EEG patterns interrupting sleep. It reflects a brief awakening of

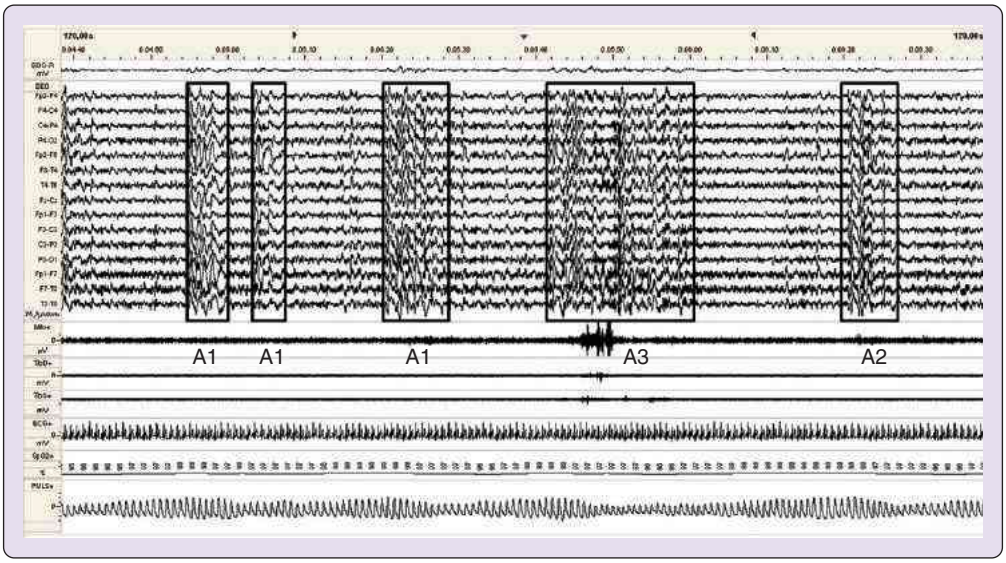


Figure 162-5 Different phase A subtypes may be scored within a common cyclic alternating pattern (CAP) sequence. The central phase A3 is combined with a strong activation of the chin EMG signal and a mild increase in muscle tone on both tibialis electrode traces. Note the waxing and waning of pulse wave amplitude (*bottom trace*), in phase with the periodicity and functional power of EEG oscillations. EEG, Electroencephalogram; EMG, electromyogram.

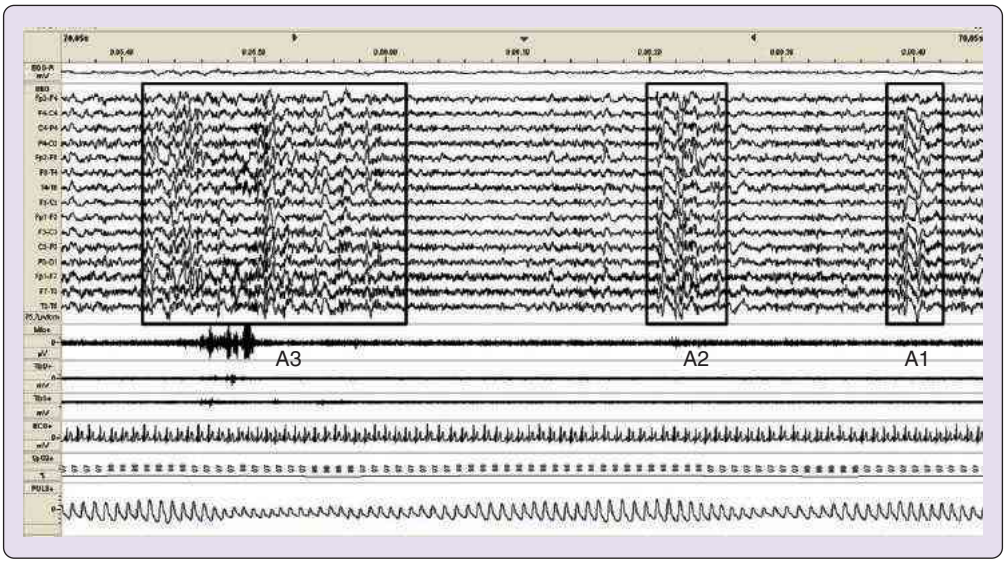


Figure 162-6 Specimen of cyclic alternating pattern (CAP) sequence with the three well-defined phase A subtypes (*black outlines*).

the cerebral cortex regardless of any concomitant participation of the autonomic system or behavioral components.

Subcortical activation: This arousal type is characterized by autonomic activation that is associated with a transient EEG pattern different from that for a conventional AASM arousal.[28-30]

Behavioral arousals and subcortical activations represent the two extremes of a gradual scale of cerebral activation. However, they are not separated by rigid boundaries. The temporal overlap between cortical, somatomotor, and autonomic events within the same arousal episode does not necessarily imply synchrony, and the order of activation of the single compartments can vary in different physiologic or pathologic circumstances.

In arousal phenomena during sleep, there is no mandatory chronologic and etiologic subordination. Such phenomena take place within interactive loops in which the cerebral cortex can be the starting or the ending point but in either case serves as a source of control. The origin of arousal is defined by the subsystem primarily activated or perturbed. The arousal can be generated directly by the brain with the physiologic

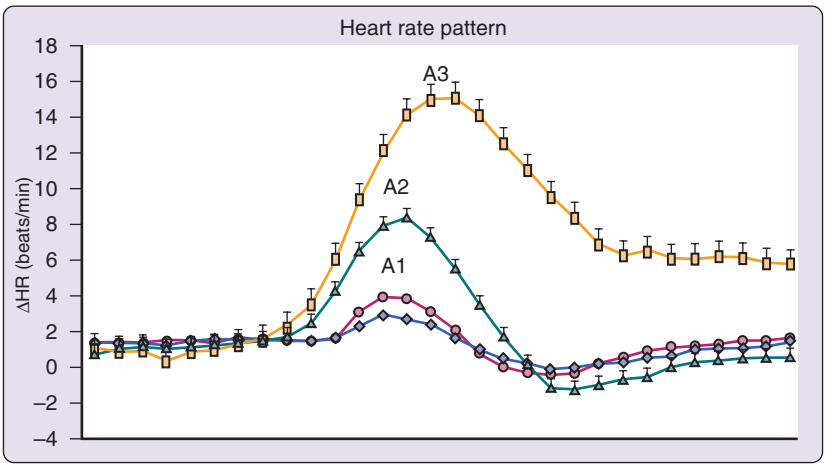


Figure 162-7 The hierarchical impact of the three phase A subtypes on heart rate (HR). The difference involves only the magnitude of reactivity, because they all induce heart rate increases. (Modified from Sforza E, Jouny C, Ibanez V. Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response. *Clin Neurophysiol* 2000;111:1611–9.)

evolution of sleep, such as with the transition from NREM to REM sleep, or in response to a sensorial perturbation, such as respiratory interruption, environmental noise, alteration of blood pressure or heart rate, or physical movements. In any case, it is the involvement of the brain that makes arousal a unitary phenomenon[31] in which activation is modulated through a hierarchy of phasic responses (Figure 162-7) ranging from slow wave high-amplitude EEG patterns (CAP subtypes A1) to fast low-voltage activity (CAP subtypes A3).

SLOW OSCILLATIONS AND THE HOMEOSTATIC NATURE OF CYCLIC ALTERNATING PATTERN

> *Any system in dynamic equilibrium tends to reach a steady state, a balance that resists outside forces of change. When such a system is disturbed, built-in regulatory devices respond to the departures to establish a new balance. All processes of integration and coordination of function are examples of homeostatic regulation.*[32]

Besides CAP, the other major EEG activity in the frequency range below 1 Hz is the so-called slow oscillation.[33] This 0.5- to 0.9-Hz EEG rhythm, which characterizes states of reduced tonic arousal, was outlined during anesthesia and NREM sleep in both animals (cats and rats) and human subjects. The slow oscillation is generated in cortical neurons and consists of phases of depolarization, characterized by intensive neural firing, followed by long-lasting hyperpolarization. Hence the two phases of the slow oscillation are characterized by opposite neural phenomena: cortical excitation, made up of synaptic potentials, and cortical inhibition, due mainly to disfacilitation in the network. The excitatory component of the slow oscillation is effective in discriminating the K-complexes and delta waves, which do not occur in isolation but are grouped into complex wave sequences. The coalescence of slow rhythms is especially visible during NREM sleep. In particular, during slow wave sleep, high-voltage slow waves rarely appear as isolated features; rather, in most cases they converge into discrete clusters, resulting in the phase A1 subtypes of CAP. The main frequency of slow oscillation occurrence during deep sleep is 0.8 Hz.[34] Accordingly, the mean interval between K-complexes and delta waves within A1 phase events hosting large portions of EEG synchronization actually ranges in frequency between 0.8 and 0.9 Hz. As NREM sleep progresses from N1 to N3, the differences in morphology and voltage between phase A (clusters of K-complexes and delta bursts) and the successive phase B (sleep stage background) become gradually less evident until the EEG is dominated by the uniform pattern of non-CAP (Figure 162-2), with the high-amplitude slow waves recurring in the frequency range of the less-than-1-Hz oscillation[35] (see Figure 162-2).

Neurophysiologic investigation reveals that slow cortical oscillation begins organizing in small territories, thereafter recruiting larger areas through coupling mechanisms, as sleep deepens. Halasz and colleagues[36] observed that sensory stimuli delivered during the descending slope of sleep cycles under high homeostatic slow wave pressure elicit a paradoxical sleep-like response (A1 phase of CAP), rather than the classic arousal response.

It is known that late components of sensory event-related potentials are sensitive to vigilance changes. The increased synchronization due to gating of the burst-firing thalamocortical working mode during slow wave sleep contributes to the enlargement of the late event-related synchronization components, taking the shape of polyphasic slow waves—the main components of K-complexes. The relationship of primary sensory-evoked potentials with K-complexes has been confirmed by Riedner and coworkers.[37] Recently, Laurino and associates[38] have shown that different modalities of sensory-evoked potentials act as traveling corticocortical excitations. When they reach the frontocentral region, depending on the proneness of the cortex at different NREM sleep levels, they elicit a biphasic slow wave response: the K-complex. In other words, during NREM sleep the impulses convey environmental information to the brain and at the same time fuel slow waves. Studies on artificial boosting of slow oscillations during NREM sleep indicate that sleep slow waves have a reactive nature. Marshall and coworkers[39] induced slow oscillation–like potential fields by transcranial application of polarizing

current through bilateral frontolateral electrodes during early nocturnal NREM sleep. The artificially induced slow oscillation subsequently increased slow wave sleep and entrained slow spindle activity over the frontal cortex. Massimini and associates[40] were able to trigger sleep slow waves, resembling spontaneous ones, by transcranial magnetic stimulation. Stimulation applied during stage 2 led to rapid transition to stage 4. Ngo and colleagues[41] enhanced slow oscillations in sleeping subjects by applying a random acoustic stimulation of 0.8 Hz. These convergent results suggest the existence of a basic slow wave–defending mechanism during NREM sleep. Its physiologic triggers are environmental or internal sensory stimuli resulting in K-complex or CAP A1 phase. It can also be artificially elicited by electrical stimulation of the brain or transcranial magnetic stimulation. This input-dependent, modality-independent stereotypical response could be considered as the elementary building block of NREM sleep.

DYNAMIC ORGANIZATION OF THE ELECTROENCEPHALOGRAM DURING CYCLIC ALTERNATING PATTERN

Interactions thus happen at all levels of organization between different objects within a level, and, internally to each level, they are not additive. Moreover, communication is both intra- and inter-level and, therefore, a change in one level may spread its effects at the same time to higher and lower ones.[42]

The dynamic properties of the EEG assessed by means of the nonlinear cross prediction test[43] indicate that sleep is a self-organized evolving process of different high- and low-dimensional states of the EEG. The CAP sequences express short periods of low-dimensional nonlinearity that interrupt a baseline EEG not distinguishable from high-dimensional noise (non-CAP). The decrease in dimensionality of the EEG structure, especially evident during the A1 CAP subtype, indicates a decreased complexity reached by means of an increased brain synchronization.[44] The different EEG frequency components of CAP are not uniformly distributed over the scalp. Two main frequency bands can be detected, characterizing, respectively, CAP subtype A1 events (0.25 to 2.5 Hz), showing a clear prevalence over the anterior frontal regions, and CAP subtype A3 events (7 to 12 Hz), located over the parietal-occipital areas. The two frequency bands coexist in the CAP A2 subtype.[45] Because the generators of the low-frequency CAP component are localized mostly over the frontal midline cortex, CAP subtype A1 events can be considered as frontal lobe arousals.[31] By contrast, those of the high-frequency band involve both midline and hemispheric areas within the parietal and occipital areas. The nonrandom association between the two different frequency components of CAP suggests the involvement of a complex pattern of EEG synchronization and resynchronization, including both longitudinal intrahemispheric pathways and transversal interhemispheric connections.[46-51]

MEASURES OF CYCLIC ALTERNATING PATTERN

But neither does time exist without change; for when the state of our own minds does not change at all, or we have not noticed its changing, we do not realize that time has elapsed, any more than those who are fabled to sleep among the heroes in Sardinia do when they are awakened; for they connect the earlier 'now' with the later and make them one, cutting out the interval because of their failure to notice it.
***Aristotle,* Physics**[52]

EEG features are highly sensitive markers of brain development. Accordingly, sleep reflects the physiologic changes accompanying different ages across the lifespan. CAP parameters also undergo dynamic changes across the maturational phases of life. The age-related values for CAP have been measured and defined to establish the physiologic ranges of normal sleep. In recent years, several studies in children revealed specific changes of CAP parameters in different sleep disorders, neuropsychological disabilities, and mental retardation.[53,54] An overview of CAP studies in childrens' sleep was published by Bruni and colleagues.[55]

A Phases

The percent of A1 subtype events for successive age groups can be plotted to show a bell-shaped curve, with reported values of 69.7% in infants, 63.2% in preschool-aged children, 84.4% in school-aged children, 85.5% in peripubertal children, and 71.3% in teenagers. This percentage tends to decrease and then slightly increase between the ages of 30 and 60 years (61.4% in young adultsand 62.0% in those of middle age) and, finally, declines after the age of 60 (46.6%). By contrast, subtypes A2 and A3 undergo a linear increase from infancy to old age, similar to the arousal pattern seen across the life span.[56]

Cyclic Alternating Pattern Rate

CAP rate is the most extensively used parameter for clinical purposes. Calculated as the percentage ratio of total CAP time during NREM sleep to that during NREM sleep time, CAP rate is the measure of sleep instability[57]; it typically increases when sleep is disturbed by internal or external factors,[58] and its variations correlate with the subjective appreciation of sleep quality, with higher values of CAP rate associated with poorer sleep quality.

In normal sleepers, CAP rate shows a low night-to-night intraindividual variability. CAP rate also shows a complex evolution during development. It is very low in the first months, then increases gradually, with a peak in adolescence, and then gradually decreases in adulthood, followed by an increase in old age.[56] As soon as NREM sleep emerges, CAP begins to develop, and the oscillating pattern of the different phasic EEG activities becomes evident. The gradual appearance of an oscillating pattern of slow EEG activities, representing the first prototype of CAP, appears at 46 to 55 weeks of conceptional age. CAP rate in infants 3 to 4 months of age increases from 6.83% ± 3.58% to 12.9% ± 2.21%, depending on the level of maturation of sleep patterns. Starting at the age of 3 years, there is a gradual increase in CAP rate that peaks at the peripubertal age and then decreases to its lowest values, in young adults (25.9% in preschool children aged 3 to 6 years old and 33.4% in school-aged children 6 to 10 years old; 62.1% in peripubertal children aged 8 to 12 years; 43.4% in teenagers; 31.9% in young adults; 37.5% in middle-aged adults, and 55.3% in elderly persons).[55,56] The age-related changes of CAP rate reflect the biologic growth processes that accompany the preparation for and onset of adolescence,

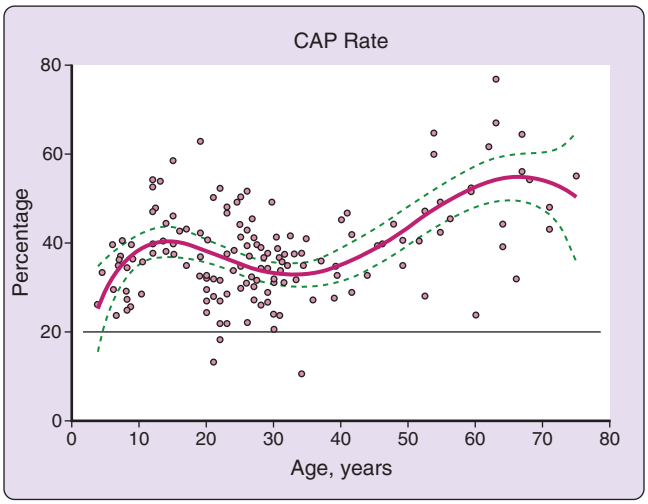


Figure 162-8 The age-related distribution of cyclic alternating pattern (CAP) rate in healthy subjects. The two peaks correspond to adolescence and senescence; the trough coincides with young adulthood. Note that almost all normal CAP rate values are higher than 20%. Narcoleptic patients and long-term benzodiazepine users and abusers usually show percentages below this threshold.

the maturational consolidation of development, and the process of senescence that increases the instability of sleep[57,58] (Figure 162-8).

CYCLIC ALTERNATING PATTERN AND SLEEP DISORDERS

Neuronal avalanches, i.e., cascade of activity with power law distribution of size and durations, are only one of the observed proprieties suggestive of criticality. Criticality is very advantageous for the brain, in terms of optimization of dynamical range, information transmission, and capacity (large repertoire of diverse activity patterns). All these intriguing results on spontaneous dynamics support the long-lasting hypothesis that brain can move in a landscape with multiple dynamical attractors, and that up states may be the result of the system falling in one of these attractors. From this point of view, the spontaneous fluctuations between up and down state may be the signature of the system posed at a non-equilibrium phase transition, where system fluctuates in the landscape, and flexibly switches from one state to another.[4]

Cyclic Alternating Pattern and Sleep-Disordered Breathing

A great body of evidence has confirmed that sleep fragmentation and punctuation of sleep with frequent, brief arousals reduce its restorative properties. Restoration is reduced even when arousals do not alter sleep macroarchitecture (as summarized using standard 30-second epoch sleep stage scoring). Correlation between the number of arousals and daytime sleepiness in patients with obstructive sleep apnea syndrome (OSAS) has been reported, but phasic delta activities during sleep (with diurnal consequences) also can exert activating effects. In subjects with upper airway resistance syndrome (UARS), airway opening may occur with a concurrent increase in delta power and not necessarily associated with a scorable EEG arousal.[59] Termination of respiratory events marked by an EEG pattern of delta waves also has been observed in patients with OSAS.[22] Involvement of either slow or fast EEG responses depends on the regulation of upper airway patency. Respiratory patterns that need correction activate the CNS. This activation varies, depending on the sensory recruitment and the adequacy of the response. A respiratory challenge is resolved by a cortical arousal only when the thalamocortical structures that control slow waves fail to modulate breathing or when ascending reticular volleys are required to restore respiration. The slow waves determine a "softer" autonomic reaction, which in certain pathologic conditions may be strong enough to overcome a disturbing factor, such as with an obstructive event. Otherwise, the slow rhythms are immediately replaced by faster EEG activities, which guarantee a more powerful activation of autonomic functions. Probably the effects on daytime function are linked not to a single phase A subtype but to the reciprocal amount and distribution of the single subtypes. In severe OSAS, CAP analysis shows increased values of CAP rate accompanied by enhanced percentages of subtypes A3 (Figure 162-9) and reduced percentages of subtypes A1.[10] When patients with OSAS are treated effectively with nasal continuous positive airway pressure (CPAP), the ventilatory-induced reduction in CAP rate, which correlates significantly with daytime sleepiness, is associated with a robust curtailment of subtypes A3 and a progressive recovery of the A1 percentage.[60,61] Different features occur in primary insomnia, in which high values of CAP rate are associated with an enhanced arousal index and increased percentages of all phase A subtypes. Compared with placebo, however, active medication significantly reduces CAP rate and percentages of subtypes A1 and A2 but exerts only marginal effects on subtypes A3 and on EEG arousals.[62]

Insomnia

In normotensive subjects with chronic insomnia, higher nighttime blood pressure and blunted blood pressure dipping are associated with brain cortical activation during sleep in the absence of arousal rate changes.[63] This correlation implies that scoring of single arousal events can be enriched by integrative sleep parameters, including the more general concept of oscillating activation. Previous studies ascertained that acoustic stimuli enhance the physiologic CAP rate and contribute to poor sleep and daytime dysfunction even without an increase in sleep fragmentation. CAP thus operates as a "double-edged sword": Whereas limited quantities of CAP mediate physiologic effects, larger quantities reflect difficulties in consolidation and preservation of sleep normally under cortical control and may therefore be associated with adverse effects. Whatever the nature of the disturbance, the outcome is the amplification of an otherwise physiologic process.[64] An investigation carried out in a large cohort of white patients with primary insomnia demonstrated that CAP parameters consistently demonstrate the reduced quality of sleep in persons with insomnia and can corroborate the efficacy of hypnotic medication. CAP rate is the PSG parameter that best reflects subjective sleep quality and is also the most sensitive sleep measure of effective drug treatment.[62] Similar findings have been reported for subjects in non-white populations. In Japanese patients with psychophysiologic insomnia, a randomized crossover comparative study with placebo showed that hypnotic treatment (with zolpidem) increased sleep stability by

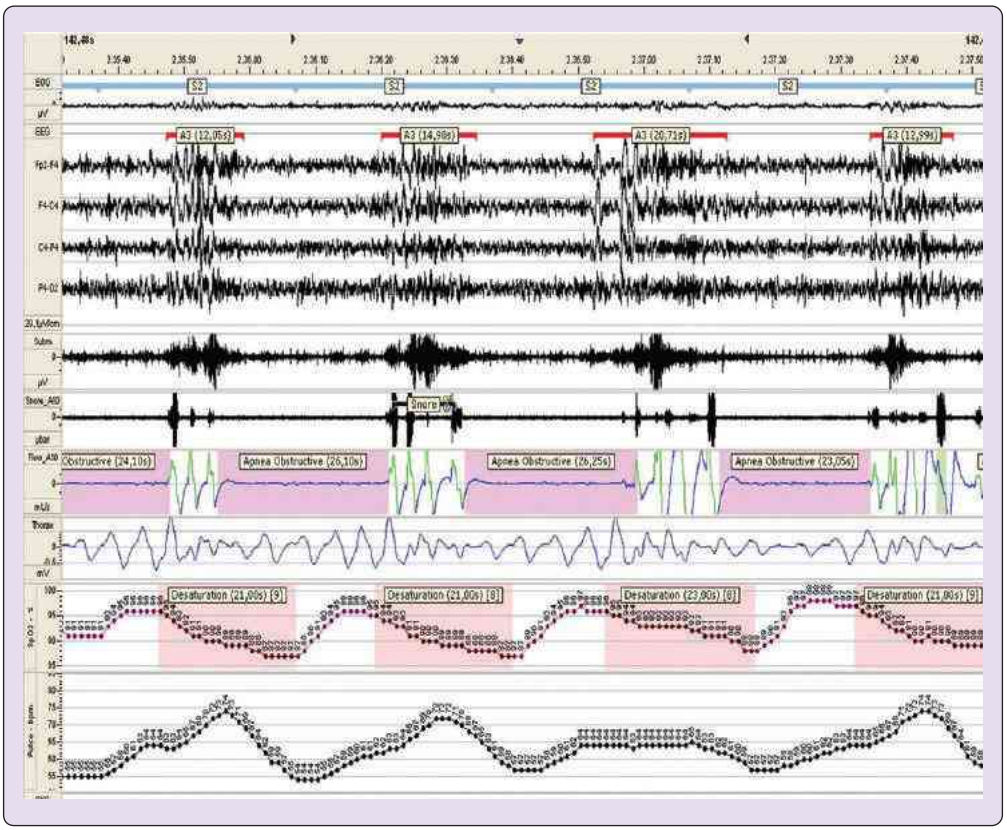


Figure 162-9 A cyclic alternating pattern (CAP) sequence of phase A3 subtype events that allow recovery of respiratory activity at the end of recurring episodes of obstructive apnea. Note the exaggeration of muscle and vegetative oscillations of what normally happens in nonpathologic conditions.

significantly reducing the overnight CAP rate as well as improving subjective sleep quality.[65]

CAP parameters also are useful tools for monitoring the effects of intermittent hypnotic treatment[66] and for investigating cases of paradoxical insomnia.[67] Compared with normal control subjects, insomniacs with so called sleep state misperception (i.e., "misperceptors") show significantly higher CAP rate in stage 1 and in stage 2 but not in slow wave sleep. The percentage of subtype A2 events, which include both sleep promoting and wake-promoting EEG features, is significantly higher ($P < .001$) in misperceptors (31%) than in control subjects (24%). Misperceptors report lower numbers but longer intervals of subjective awakenings (mean, 4) compared with objective findings (mean, 11). The mismatch could be explained in part by the large amounts of CAP between successive awakenings that were merged together in a single experience. In other words, if sleep between two successive awakenings is superficial (expressed by sleep stages 1 and 2), unstable (as reflected by increased amounts of CAP), and fragmented (increased arousal index), time separating the two events is perceived as continuous wake. These findings suggest a greater sensitivity of misperceptors to sleep problems.

Overall, assessment of sleep quality relies on a variety of PSG measures including sleep duration (quantified by total sleep time and sleep efficiency), sleep intensity (reflected by N3), sleep continuity (altered by nocturnal awakenings and arousals), and sleep stability (impaired by excessive amounts of CAP). These measures are susceptible to deterioration in various ways and in accordance with the clinical manifestations of insomnia. In the hierarchy of sleep measures, CAP variables are arguably the most sensitive to any source of internal or external perturbation during sleep.

Cyclic Alternating Pattern and Body Movements

CAP can promote or release motor activities ranging from physiologic body movements to periodic limb movements (PLMs), sleep bruxism, night terrors, and sleepwalking. Overall, the CAP A phase can be conceptualized as an "amplifier" that may facilitate the occurrence of pathophysiological events, whereas the B phase exerts a "filter" action. The magnitude and temporal variations of cerebral activity and autonomic responses before and during the motor event support the hypothesis of a continuum in arousal reactivity present not only on visual analysis but also at the spectral level. With respect to manual scoring, this continuum is translated by progression of the arousal response from delta and K-complex bursts to an EEG arousal and full awakening.[12,14] The same continuum can be detected by spectral analysis, starting with enhancement of delta activity and increase in heart and respiratory rates,[68,69] which may be followed by a transition to fast EEG activities. The hierarchy of the arousal response is related to severity provided by emergency. From a mechanistic standpoint, postapnea arousals in sleep breathing disorders are the most evident example of automatic and purposeful periodic movement during sleep.

PLMs reflect a pattern of motor phenomena and EEG changes, both recurring at intervals of 20 to 40 seconds. The periodicity of PLM parallels the recurrence of CAP cycles

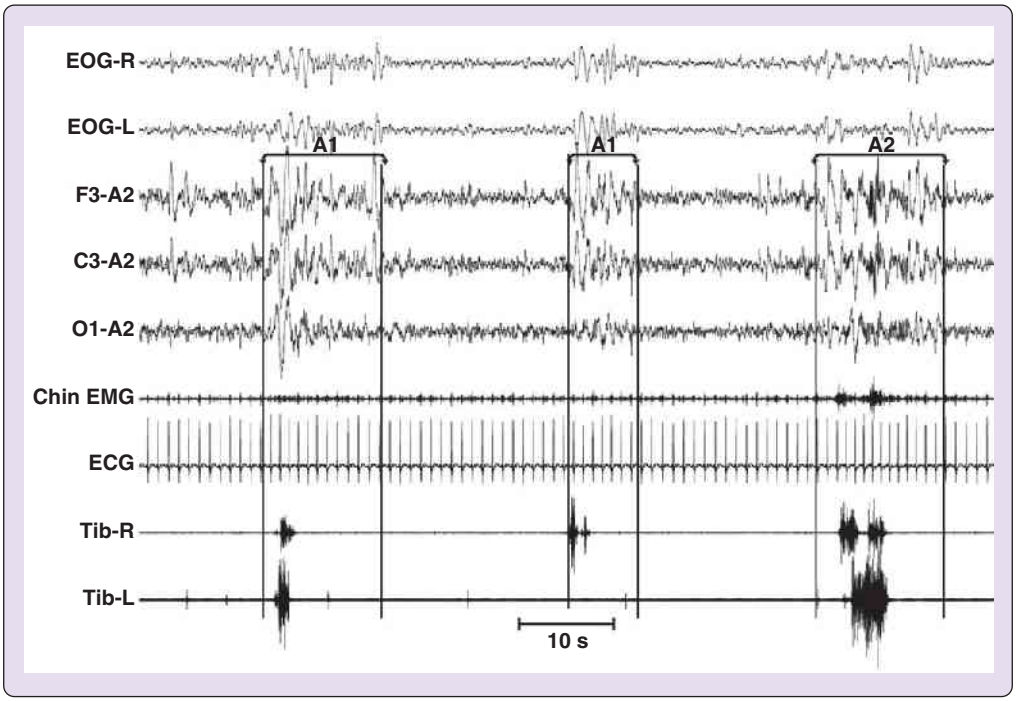


Figure 162-10 Cyclic alternating pattern (CAP) sequence coupled with periodic limb movements. Note that myoclonic jerks during non–rapid eye movement (NREM) sleep also can be associated with a phase A1 subtype event generally not scored as a conventional arousal. ECG, Electrocardiogram; EMG, electromyogram; EOG, electrooculogram; Tib, tibialis muscle.

(Figure 162-10). It is not surprising, therefore, that the two PSG manifestations (PLM and CAP) are mutually connected. This association does not imply that PLM is induced by CAP; rather, the latter represents a permissive framework that synchronizes rhythmicity. In patients with PLMs, 92% of all jerks detected in NREM sleep occurred in CAP, with the great majority of limb movements (96%) associated with a phase A, especially subtypes A2 and A3 (95%). Ninety-four per cent of the nocturnal jerks coupled with a phase A appeared in concert with onset of the phase or when it had already begun. In particular, 67% of the myoclonic events occur in the first 2.5 seconds of the A phase. PLM onset is heralded by a significant activation of delta activity power (the initial portion of the phase A2 or A3 subtype) starting approximately 3 to 4 seconds before occurrence of movement.[70-72]

Parasomnias and Disorders of Arousal

Some disorders of arousal show a typical EEG pattern defined as hypersynchronous delta activity, that corresponds to the high-voltage delta bursts of CAP A1 subtypes in slow wave sleep.[73-75] CAP recording and analysis provide a useful technique for evaluating parasomnias involving arousal. Subjects with sleepwalking or sleep terrors show an increase in CAP rate, a higher percentage of A subtype events in slow wave sleep, and a decrease in phase B and CAP cycle duration.[76-79] The shorter duration of CAP cycle and phase B determines an abnormally fast oscillatory pattern of the amplitude of EEG slow wave components (defined as hypersynchronous delta activity, slow wave sleep arousals, or CAP subtype A1), leading to recurrent arousals from slow wave sleep, with consequent sleep fragmentation, which contributes to trigger slow wave sleep parasomnias.[78] In chronic sleepwalkers, instability of NREM sleep, detectable by CAP analysis, is present even during nights without sleepwalking episodes. In children, this instability has been related to the presence of associated sleep disorders, such as upper airway resistance syndrome or PLMs.[80,81]

Cyclic Alternating Pattern and Epilepsy

Interictal Discharges

In primary generalized epilepsy, interictal discharges commonly are activated during unstable sleep, with the number of EEG paroxysms per minute of sleep significantly higher during CAP than during non-CAP. Phase A has a significant activation influence, whereas phase B exerts a powerful and prolonged inhibitory action.[82,83] In comparison with normal control subjects, the occurrence of interictal epileptiform discharges in patients with primary generalized epilepsy has no remarkable consequences on sleep macrostructure, but such events have significant effects on arousal stability, in that epileptic patients show higher CAP rate values.[84]

As occurs with primary generalized epilepsy, the presence of focal lesional interictal epileptiform discharges impairs the stability of sleep. By contrast, despite the high burst frequency during NREM sleep, interictal epileptiform discharges in benign epilepsy with rolandic spikes are not modulated by the arousal-related mechanisms of CAP.[85,86]

Ictal Events

Seizures require a process of changes in brain dynamics that start before clinical manifestation. Analysis of preictal synchronizations indicates that epileptic seizures do not occur in a behavioral vacuum, but that the functioning of the brain

before the seizure occurs is critical. Seizure foci are surrounded by pools of neurons functioning in local and large-scale interactions and are "pulled" into the seizure discharge once the seizure has started. The preictal period may reflect a state of increased susceptibility for pathologic synchronization, which acts as a route to the seizure.[87] In patients affected by focal epilepsy, the great majority of nocturnal partial motor seizures occur during NREM sleep, and almost always during a phase A.[88] These findings indicate that an NREM sleep condition of highly fluctuating vigilance constitutes a favorable substrate for the occurrence of focal epileptic seizures.

Nocturnal Frontal Lobe Epilepsy

Clinically, nocturnal frontal lobe epilepsy (NFLE) consists of a spectrum of paroxysmal motor manifestations ranging from major seizures to paroxysmal arousals and minor motor events. A common feature is the onset of all episodes during NREM sleep. Paroxysmal arousals and minor motor events may recur every night, sometimes several times per night, arising mainly from CAP in stage 2 sleep.[89] In NFLE, stereotypical features involve not only motor behaviors but also polysomnographic patterns.[90] Patients with NFLE show a significant increase in wake after sleep onset, slow wave sleep duration, and REM latency and a significant increase in CAP time, CAP rate, CAP cycles, and mean duration of CAP sequences. These findings are associated with a significant enhancement of all subtypes of the A phase of CAP (mainly subtype A1). Ninety percent of total NREM seizures occur during a CAP sequence, and CAP-related seizures occur in association with a phase A. Antiepileptic treatment effects a partial reduction in both major attacks and minor motor events. REM latency, wake after sleep onset, and sleep efficiency recover to normal values, but NREM sleep instability remains at a pathologically high level and is associated with persistence of daytime sleepiness.[91] The residual high level of NREM sleep instability probably is related to the persistence of epileptic discharges that act as internal triggers of subcontinuous arousal fluctuations during NREM sleep. Despite the reduction in nocturnal seizures, the persistence of epileptic events, high levels of NREM sleep instability, and impaired daytime vigilance under drug treatment suggest the need for new therapeutic options in this particular form of epilepsy.

AUTOMATED ANALYSIS OF CYCLIC ALTERNATING PATTERN

Consolidated evidence has demonstrated that CAP parameters provide more detailed information and are significantly more sensitive than conventional sleep measures. The visual scoring of CAP is time-consuming, however, and this requirement can limit its use in routine clinical practice. In other words, only the availability of an adequate system for automated detection of CAP can make it an easily exploitable tool. The problem of automated CAP recognition has been addressed in several studies, with interesting results.[92-94] Most proposed classification methods, however, require the preliminary measurement of sleep macrostructure, making it necessary for the human scorer to visually classify the sleep profile. Moreover, these methods rely on the extraction of spectral parameters from the EEG signal to compute descriptors on time windows and on the application of machine learning algorithms for classifying each window as belonging to a phase A of CAP or not. A preliminary study investigated a set of descriptors computed on 1-second-long windows with different techniques, and the information content of each descriptor was evaluated by means of receiver operating characteristic (ROC) curves.[95] The descriptors were used in a subsequent study training four different workers for the classification of CAP A phase subtypes, in which the linear discriminant was found to be the most accurate, with an average accuracy of 83%.[96] However, owing to the intrinsic characteristics of A phase activity, which consists of transient, nonstationary phenomena, the signal characteristics are likely to vary over time, and it is not convenient to define a fixed window length. A dataset of 16 PSG recordings from healthy subjects was used for this study, and the EEG traces first were subjected to an automated isolation of NREM sleep segments by means of an artificial neural network and then underwent a segmentation process based on the spectral error measure. The information content of the descriptors was evaluated by means of receiver operating characteristic curves and compared with that of descriptors obtained without the use of segmentation. Finally, the descriptors were used to train a discriminant function for the automated classification of CAP phase A subtypes. With respect to previous scoring methods, significant improvement was obtained in terms of both information content carried by the descriptors and accuracy of the classification. Average CAP rate obtained by automatd scoring was 44.83% ± 12.41%, versus 43.83% ± 10.13% measured with visual analysis.[96] Although the sensitivity was still not very high (69.55% ± 6.6%), the specificity (90.49% ± 2.8%) and the accuracy (87.19% ± 2.48%) showed a significant increase with respect to the previous study, indicating that EEG segmentation proves to be a useful step in the computation of descriptors for CAP scoring.

CYCLIC ALTERNATING PATTERN: THE MASTER CLOCK OF SLEEP

> *In biology, living systems dynamics are not a matter of stable or unstable equilibrium, but of far from equilibrium processes, i.e. which undergo a flow of energy or matter, yet they are "structurally stable" systems. This hard to understand simultaneous structural stability and non-conservative behavior is a blend of stability and instability due to the coexistence of opposite properties such as "order/disorder … integration/differentiation."*[42]

Objection could be raised on the usefulness of CAP in routine clinical practice.[97] Besides the multiple clinical applications of CAP,[98] attention should be focused on the underlying rationale for the CAP scoring methodology. The main assumption in CAP scoring is the fact that during certain periods of the night, the arousal level of sleep is unstable. Accordingly, CAP is decreased in narcolepsy[99,100] and multiple system atrophy,[101] with drug administration[102-104] and CPAP treatment for obstructive sleep apnea,[60,61] and during night-time recovery sleep after prolonged sleep deprivation.[105]

The concept of instability is a basic issue of all complex systems and supports the dynamics of biologic variability. Within certain ranges, instability warrants flexible and adaptive features for the complex system. To guarantee survival during prolonged unconsciousness (i.e., sleep), a strong interaction among all of the biologic subsytems is mandatory. CAP

analysis is not limited to the finding of a single event (e.g., an isolated arousal) but instead identifies a pattern (presence or absence of a CAP sequence) that translates to a physiologic state involving cerebral activities, autonomic functions, and behavioral features.

In normal sleep, CAP accompanies the stage transitions maintaining in phase both the EEG and autonomic functions through regular fluctuations. Accordingly, even without the EEG leads, the detection of an unstable cardiorespiratory pattern is likely to be associated with an oscillatory condition of cerebral activities.[106] In other words, what happens upstairs (brain) is reflected at the lower levels (autonomic and muscle parameters), and vice versa. Temporal patterns are generated within the sleeping brain, and CAP is a "master clock" that determines the framework within which temporal patterns can be generated and implemented.

With attribution of activation relevance to all A phases of CAP (including also K-complexes and delta bursts), a major advance has been to overcome the stark contrast between visible and nonvisible arousals[107] that for years froze the possibility of understanding the brain's flexible capacity to produce different EEG features in different neurophysiologic situations. The approach to arousals during sleep would be better conceived not as static single events but rather as a general dynamic process in which activating phenomena of differing intensity and morphology can be organized into sequences. Conventional EEG arousals represent the tip of an iceberg composed of other, more subtle but equally powerful activating events. This extensive interpretation of arousals and their nonrandom distribution into periodic sequences converges on CAP, a structured organization of sleep beyond the static stage setting. Perhaps the time has come to update the scoring rules of sleep and to quantify what is seen in the PSG and use CAP to evaluate brain function.[108]

> *Each of them contains all within itself, and at the same time sees all in every other, so that everywhere there is all, and all is all . . . . Each of them is great, the small is great; the sun, there, is all the stars and every star, again, is all the stars and sun. While some one manner of being is dominant in each, all are mirrored in every other.*
>
> ***Plotinus*, The Six Enneads *(ca. 250 CE)***

CLINICAL PEARLS

- The concept of instability is a basic issue of all complex systems and supports the dynamics of biologic variability. Within certain ranges, instability warrants flexible and adaptive features in the complex system.
- CAP is the EEG marker of NREM sleep instability. Conversely, non-CAP translates to a condition of stable sleep.
- CAP is the "master clock" that accompanies the stage transitions maintaining in phase both the EEG and autonomic functions through regular fluctuations.
- Conventional EEG arousals represent the tip of an iceberg composed of other, more subtle but equally powerful activating events often organized in the CAP sequences.
- K-complexes and delta bursts are arousal equivalents; therefore any distinction between visible and nonvisible cortical arousals is misleading.
- The arousal response is a continuous modulated phenomenon lying within a spectrum of EEG variability.
- The phase A subtypes of CAP allow adaptive adjustments of the ongoing state to internal and external inputs.
- CAP parameters undergo dynamic changes across the maturational phases of life and show specific alterations in the different sleep pathologic conditions.
- CAP plays a gate control role for a number of isolated and periodic events that characterize sleep disorders.
- EEG segmentation seems to be a useful tool for automated scoring of CAP.

SUMMARY

Sleep is a dynamic process with a self-regulating character. The nightly recurring sleep process is organized into consecutive cycles in which the sequence of NREM stages and the alternation between NREM and REM sleep show a quite stable tendency and a largely predictable pattern. These constraints produce the macrostructural development of sleep. However, transient EEG changes can interact with the expected development of sleep and ensure adaptation to both internal and external conditions. Arousals and CAP represent rapid adaptive adjustments of vigilance during sleep. Failure of these compensatory processes contributes to sleep disorders and nonrestorative sleep. Delineation of CAP parameters and of the role of CAP in the dynamic organization of sleep has led to multiple clinical applications of CAP, and various methods are available for the automated quantification of CAP measures.

Selected Readings

Ferri R, Bruni O, Miano S, Terzano MG. Topographic mapping of the spectral components of the cyclic alternating pattern (CAP). *Sleep Med* 2005;**6**:29–36.

Halasz P, Terzano MG, Parrino L, Bodisz R. The nature of arousal in sleep. *J Sleep Res* 2004;**13**:1–23.

Maestri M, Carnicelli L, Tognoni G, et al. Non-rapid eye movement sleep instability in mild cognitive impairment: a pilot study. *Sleep Med* 2015;**16**(9):1139–45.

Mariani S, Grassi A, Mendez MO, et al. EEG segmentation for improving automatic CAP detection. *Clin Neurophysiol* 2013;**124**:1815–23.

Parrino L. Down the rabbit-hole: new perspectives from stereo-EEG in the investigation of sleep, nocturnal epilepsy and parasomnias. *Sleep Med Rev* 2015;**25**:1–3.

Parrino L, Ferri R, Bruni O, Terzano MG. Cyclic alternating pattern (CAP): the marker of sleep instability. *Sleep Med Rev* 2012;**16**:27–45.

Parrino L, Halasz P, Tassinari CA, Terzano MG. CAP, epilepsy and motor events during sleep: the unifying role of arousal. *Sleep Med Rev* 2006;**10**:267–85.

Parrino L, Smerieri A, Rossi M, Terzano MG. Relationship of slow and rapid EEG components of CAP to ASDA arousals in normal sleep. *Sleep* 2001;**24**:881–5.

Terzano MG, Mancia D, Salati MR, et al. The cyclic alternating pattern as a physiologic component of normal NREM. *Sleep* 1985;**8**:137–45.

Terzano MG, Parrino L, Smerieri A, et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med* 2001;**2**:537–53.

Terzano MG, Parrino L, Smerieri A, et al. CAP and arousals are involved in the homeostatic and ultradian sleep processes. *J Sleep Res* 2005;**4**: 359–68.

Thomas RJ. Arousals in sleep-disordered breathing: patterns and implications. *Sleep* 2003;**26**:1042–7.

A complete reference list can be found online at ExpertConsult.com.

Chapter
163

Neurologic Monitoring Techniques

Beth A. Malow

Chapter Highlights

- The focus of this chapter is on the methodology and indications for V-EEG-PSG in the diagnosis of nocturnal events. Techniques and approaches that are described include routine electroencephalography, short-term continuous video-electroencephalogram (EEG) monitoring, long-term continuous video-EEG monitoring, and ambulatory monitoring.
- The EEG may show either a rhythmic evolving discharge characteristic of an epileptic seizure or interictal epileptiform discharges that occur apart from epileptic seizures. Lack of an EEG change, however, does not exclude some types of epilepsy, particularly those of frontal lobe origin.
- Included in this chapter are several case examples illustrating why a particular monitoring technique is selected for a specific clinical situation.

Nocturnal spells often present diagnostic problems in sleep medicine because the history alone may not provide sufficient information to differentiate among the various diagnostic possibilities. Standard polysomnography (PSG) is helpful in defining the state and stage of sleep from which such nocturnal spells emerge, but its diagnostic capability is limited because behavioral analysis often is not included, and only a certain number of channels are devoted to electroencephalography. These shortcomings are especially pertinent to the evaluation of suspected nocturnal epileptic seizures, defined by behavioral and motor manifestations in addition to electroencephalogram (EEG) criteria.[1] Both behavioral and EEG analyses are critical for characterizing epileptic seizures and for distinguishing them from parasomnias.

The behavioral and EEG manifestations of nocturnal spells caused by parasomnias, neurologic disorders, and psychiatric disorders can be characterized more precisely by combining standard PSG with video recordings and extensive (comprising 12 or more channels) EEG montages.[2] The focus of this chapter is on the methodology and indications for video-electroencephalography—PSG (V-EEG-PSG) in the diagnosis of nocturnal events. Specific techniques and approaches that are described include routine electroencephalography, short-term continuous video-EEG monitoring (STM), long-term continuous video-EEG monitoring (LTM), and ambulatory monitoring.

METHODOLOGY

Technical Aspects of Electroencephalography

The EEG measures the difference in electrical potential between pairs of electrodes placed on the scalp. These signals, reflecting synchronous postsynaptic potentials in large groups of neurons, are amplified and filtered to produce a recording.[3] The *international 10–20 system* of EEG electrode placement is customarily used, in which "10-20" refers to 10% and 20% of the distances between standard cranial landmarks[4,5] (Figure 163-1). In this system, each electrode site is identified with a letter representing the underlying region of the brain and with a number indicating a specific position above that region, with odd numbers indicating the left hemisphere and even numbers indicating the right hemisphere (e.g., T3 represents a left midtemporal electrode).

Each recording channel is derived from the signals from a pair of electrodes, and several pairs of electrodes, or derivations, are combined to form a *montage.* Montages may be either referential or bipolar. In *referential* montages, one of the electrodes in each pair is connected to a common electrode (e.g., channel 1: Fp1-A1; channel 2: F7-A1; channel 3: T3-A1; channel 4: T5-A1; channel 5: O1-A1). In *bipolar* montages, there is no common electrode. Bipolar montages usually are arranged in a chain with the same electrode in adjacent derivations (e.g., channel 1: Fp1-F7; channel 2: F7-T3; channel 3: T3-T5; channel 4: T5-01).

The EEG montages used in a combined electroencephalography-PSG study depend on the clinical indication and the number of channels available for recording (Table 163-1). The channels suggested in this chapter for inclusion in such studies are in addition to those recommended by the American Academy of Sleep Medicine for the scoring of sleep stages.[6] Physicians and technicians using EEG monitoring techniques require a solid knowledge of the principles of EEG recording and interpretation. Additional information regarding EEG methodology is available in a standard electroencephalography textbook.[3]

Computerized digital electroencephalography-PSG systems facilitate the review of large amounts of the acquired data by displaying scoring and event information in a format that allows the user to, for example, click on the stage or event of interest and bring up the corresponding EEG PSG segment.

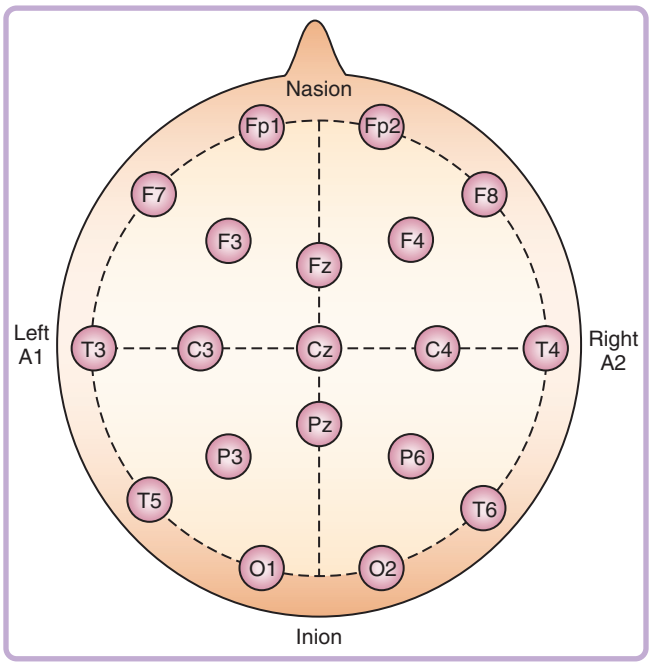


Figure 163-1 Standard international 10-20 electrode placement: Electrodes are placed at 10% or 20% of the distances between standard cranial landmarks. (From Keenan SA. Polysomnographic technique: an overview. In: Chokroverty S, editor. *Sleep disorders medicine.* Boston: Butterworth-Heinemann; 1994, p. 84.)

Table 163-1 Sample Attended Electroencephalographic Montages

Number of Available Channels	Montage
8	F7-T3, T3-T5, T5-O1, F8-T4, T4-T6, T6-O2, F3-C3, F4-C4
10	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, F3-C3, F4-C4
12	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, F3-C3, C3-P3, F4-C4, C4-P4
14	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, F3-C3, C3-P3, P3-O1, F4-C4, C4-P4, P4-O2
16	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F4, F4-C4, C4-P4, P4-O2
18	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F4, F4-C4, C4-P4, P4-O2, Fz-Cz, Cz-Pz
24	Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F4, F4-C4, C4-P4, P4-O2, Fz-Cz, Cz-Pz, T1-T3, T3-C3, C3-Cz, Cz-C4, C4-T4, T4-T2

The recording may be viewed at a variety of display settings. Filters, sensitivities, and montages may be adjusted to characterize events of interest and to help distinguish abnormalities from artifacts or normal variants. For example, certain montages can more easily identify and distinguish artifacts from interictal epileptiform discharges (IEDs), defined as epileptic-type activity occurring between seizures. Digital electroencephalography enhances the detection and review of IEDs by allowing for remontaging (displaying data in a different montage that may better reveal such activity), changing the display settings that influence temporal resolution, and isolating specific channels for review (Figure 163-2). For example, by altering the display settings, synchronous delta or theta activity characteristic of disorders of arousal from NREM sleep may be more easily distinguished from spike-wave activity or from an evolving ictal (seizure) pattern characteristic of an epileptic seizure disorder.

Several digital electroencephalography-PSG systems share a common platform with epilepsy monitoring systems, allowing data obtained during a PSG study to be analyzed by IED detection programs. Conversely, a study performed in the epilepsy-monitoring unit can be enhanced through addition of electrooculogram (EOG) and chin electromyogram (EMG) electrodes to score sleep and to determine the stage of sleep that precedes a particular event. This determination can be diagnostic in the case of distinguishing dissociative events, which rarely occur during sleep,[7] from epileptic seizures, which can occur during sleep or wakefulness.

Daytime Electroencephalography

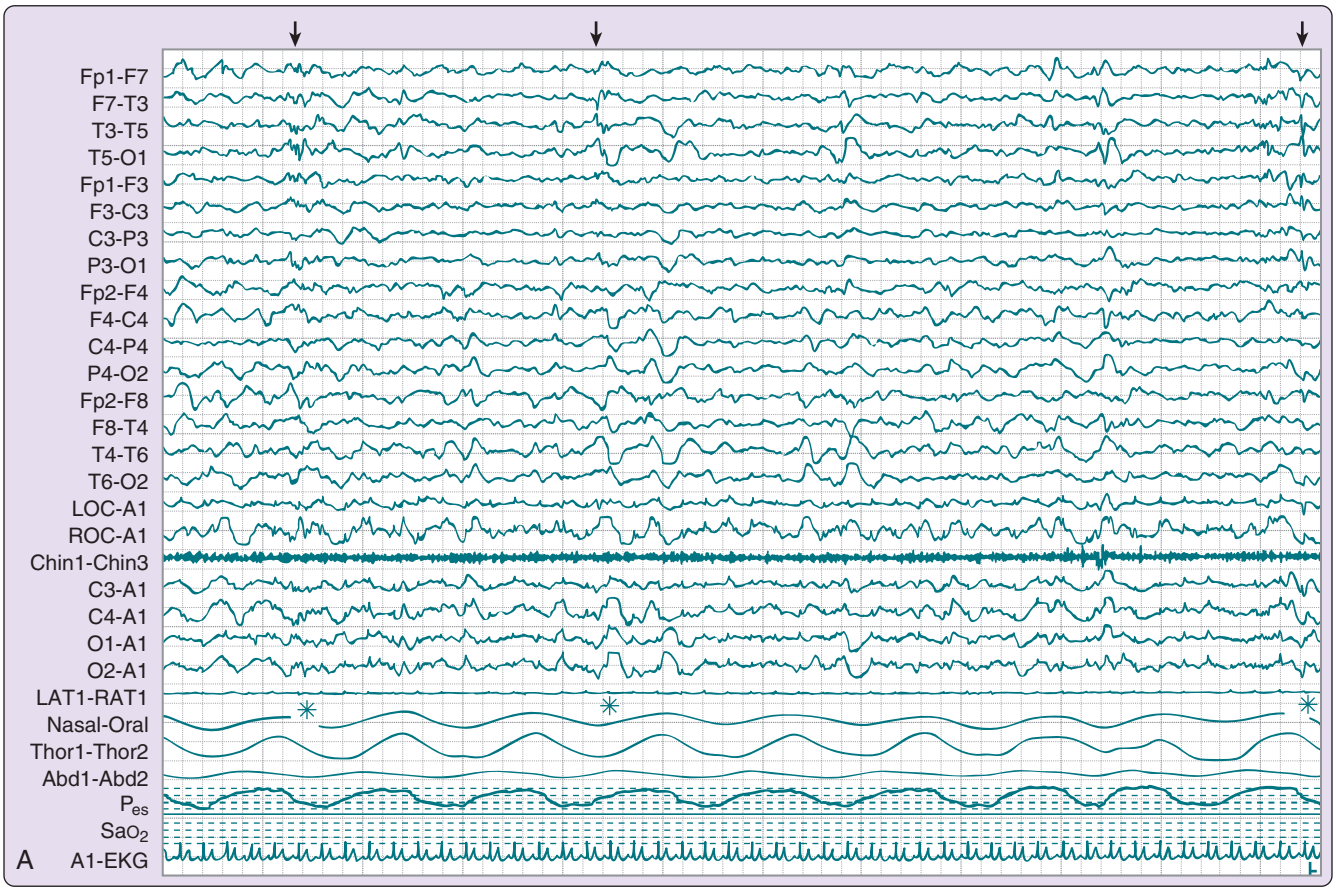
Daytime electroencephalography is used to look for IEDs, to support the diagnosis of a seizure disorder in many clinical settings.[3] In addition to the electrode placements listed earlier, central (Fz, Cz, Pz) and ear (A1, A2) electrodes are included. Nasopharyngeal electrodes, although used in the past, are not recommended, because they are uncomfortable, are prone to artifact, and rarely provide additional information. The activating techniques of hyperventilation and intermittent photic stimulation are routinely performed and can bring out focal asymmetries or epileptiform activity.

Although seizures are not uniformly recorded during a routine EEG, focal IEDs or generalized spike-and-wave discharges may be observed and can assist in the classification of an epileptic syndrome as partial or generalized. The location of IEDs, determined with the use of an extended EEG montage, can clarify the nature of the epilepsy syndrome and its prognosis.[8] For example, benign epilepsy of childhood with centrotemporal spikes carries an excellent prognosis (Figure 163-3). By contrast, some temporal lobe IEDs may be refractory to medical treatment, and affected patients can become candidates for epilepsy surgery.

Daytime studies performed while the patient is asleep for at least a portion of the recording increase the yield of finding abnormalities, because in many patients, IEDs are more common in drowsiness and NREM sleep. Stage N2 sleep is usually, but not always, recorded on the routine EEG, whereas stage N3 sleep and REM sleep are rarely recorded. When the routine EEG does not show IEDs and the clinical picture is highly suggestive of seizures, a sleep-deprived EEG improves the yield of finding epileptiform activity, at least in part because sleep is more likely to be recorded. Digital EEG recordings permit the viewing of a segment of the EEG in a variety of montages and speeds, which can help distinguish an IED from an artifact.

Video-Electroencephalography-Polysomnography

When the patient's history is not sufficient for a reasonably confident diagnosis in the case of nocturnal spells associated



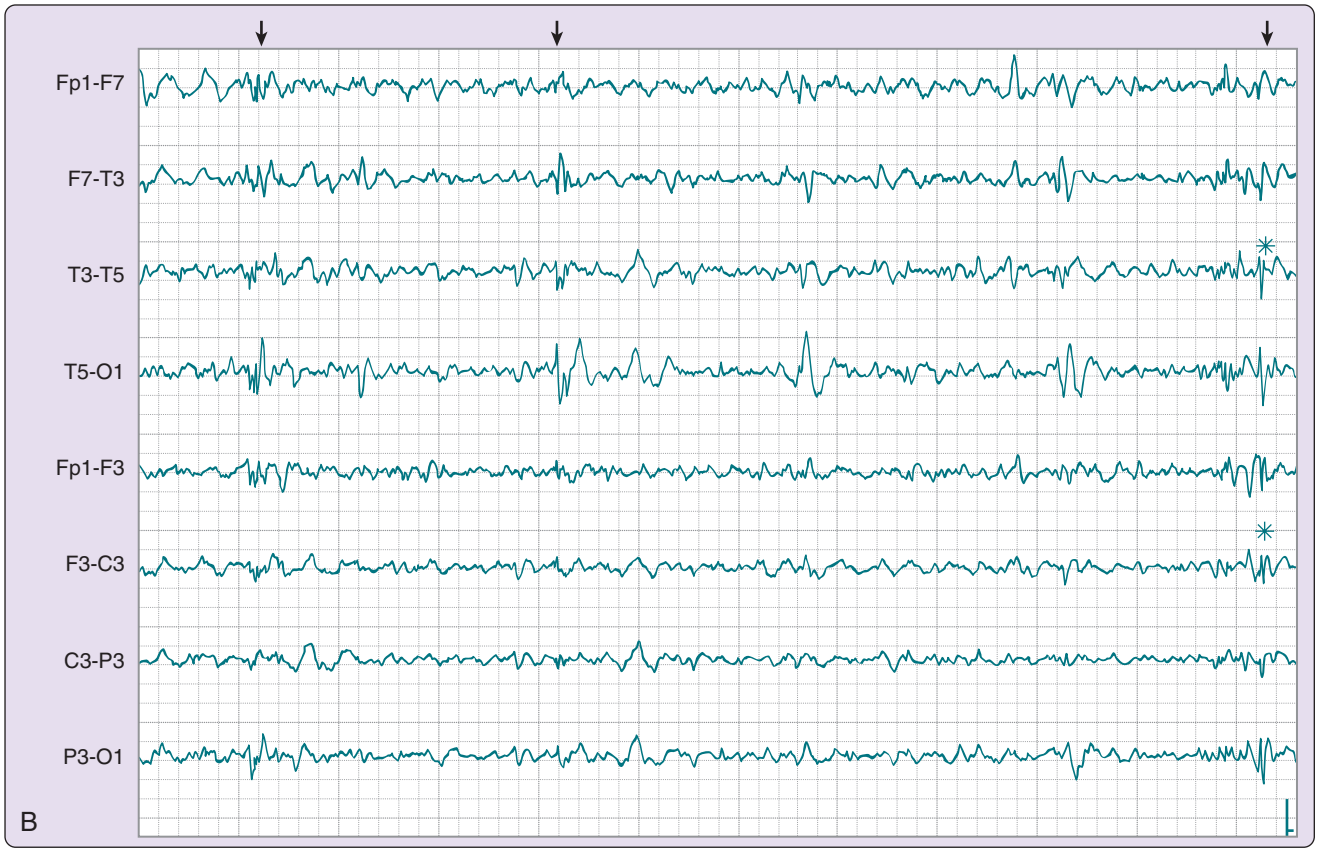


Figure 163-2 Use of different montages to enhance the review of interictal epileptiform discharges (IEDs). **A,** Left temporally dominant IEDs (*arrows*) on an extended electroencephalogram (EEG) montage that are not apparent in the central-to-ear channels (*asterisks*). **B,** Digital EEG allows the interpreter to select the relevant left temporal and parasagittal channels for review. The *arrows* highlight IEDs with phase reversals (*asterisks*) at T3 and F3, indicating maximal negativity at these electrodes, which defines the approximate location of the epileptic region. Both are 30-second epochs. Calibration symbol (*bottom right*): 100 μV.

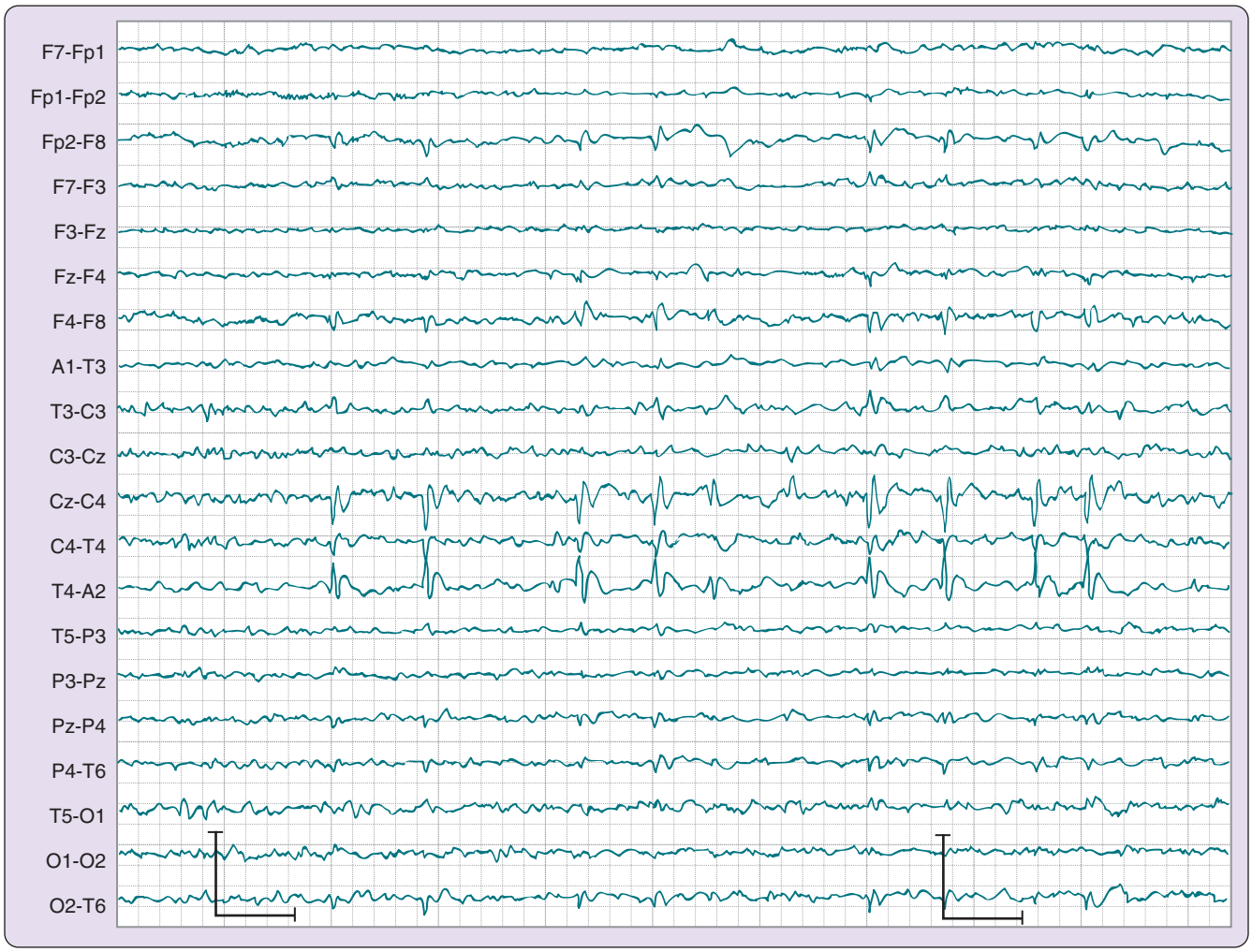


Figure 163-3 Runs of interictal epileptiform discharges with a centrotemporal dominance in benign epilepsy of childhood with centrotemporal spikes. The bipolar montage readily demonstrates peak negativity at the C4 and T4 electrodes. Such localization is not possible with electroencephalogram (EEG) montages commonly used with standard polysomnography (PSG). Calibration symbol: 500 μV, 1 second. (From Malow BA. Sleep and epilepsy. *Neurol Clin* 1996;14:774.)

with complex movements and behavior, recording of the sleep-related event in question may allow definitive confirmation. V-EEG-PSG, which combines video recording with an extended EEG montage and with other standard PSG physiologic monitoring, is useful in characterizing unusual behavior and movements during sleep. Diagnostic considerations for patients with complex actions at night can include epileptic seizures, disorders of arousal from NREM sleep, REM sleep behavior disorder (RBD), rhythmic movement disorder, or psychiatric disorders such as panic disorder or dissociative disorder. Episodes associated with these disorders have specific clinical features as discussed in Chapters 103 to 108. Events recorded with V-EEG-PSG are reviewed to characterize the motor and behavioral manifestations of the event and the EEG PSG features, including the stage of sleep preceding the event, the time of the event relative to sleep onset, and EEG patterns occurring during the event or between events. The time-locked video can be reviewed second by second along with the EEG PSG recording. Infrared cameras are useful for recording nighttime events. Movable cameras can be mounted in the patient's room and display close-ups or full-body views. Double cameras are useful for focusing on the face while simultaneously monitoring the body. During recorded events, the technologist should interact with the patient to test for level of consciousness and ability to perform commands.

The stage of sleep from which the spells emerge and the time of the spell relative to sleep onset provide useful diagnostic information. For example, the actions accompanying the disorders of arousal from NREM sleep arise from stage N3 or sometimes stage N2 sleep, usually in the first third of the sleep period (see Chapter 102), whereas those associated with RBD emerge from REM sleep, most commonly in the last third of the sleep period (see Chapter 103). Epileptic seizures are more common during NREM sleep than during REM sleep (see Chapter 97).[9] Rhythmic movements associated with rhythmic movement disorder usually occur during sleep-wake transitions, and dissociative episodes emerge from wakefulness. Nocturnal panic attacks occur from NREM sleep, usually at the transition from stage N2 to stage N3.[10]

If complex partial seizures are a diagnostic consideration, the EEG montage should emphasize the use of electrodes placed over the temporal lobes (e.g., F7, T1, T3, T5). If benign childhood epilepsy with centrotemporal spikes is a consideration, the montage should include the parasagittal region (e.g., C3, C4). The specific montage that is used depends on the number of channels available for electroencephalography. Sample montages for 8, 10, 12, 14, 16, 19, and 24 channels are shown in Table 163-1.

The following montage of 16 electrodes in anterior-to-posterior chains provides excellent coverage for suspected seizures:

- Left temporal: Fp1-F7, F7-T3, T3-T5, and T5-O1
- Left parasagittal: Fp1-F3, F3-C3, C3-P3, and P3-O1
- Right parasagittal: Fp2-F4, F4-C4, C4-P4, and P4-O2
- Right temporal: Fp2-F8, F8-T4, T4-T6, and T6-O2

This montage allows the evaluation of interictal and ictal activity during sleep. Two additional anterior temporal electrodes, T1 and T2, may be added because they are particularly useful for detecting anterior temporal IEDs. In one study comparing abbreviated EEG montages with a standard 18-channel bipolar montage, seizures were more readily distinguished from arousal patterns using 7- and 18-EEG channel montages as compared with 4-channel EEG records.[11]

Short-Term and Long-Term Monitoring

When the history suggests frequent daytime spells or spells occurring during daytime naps, STM, a video-EEG recording, typically obtained in an electroencephalography or sleep laboratory for 6 to 8 hours during the day, may be helpful. The value of such studies for assessing patients with strictly nocturnal spells, however, is limited. Occasionally, STM is useful if sleep attacks—spells of diminished responsiveness due to sleepiness—are included in the differential diagnosis of daytime spells.

Unfortunately, one or two nights of monitoring in the sleep laboratory is not always sufficient to capture and characterize spells. LTM, an extension of STM that allows continuous recordings for up to several weeks, is used mainly for patients with known or suspected seizures.[12] For patients who are taking antiepileptic medications, these drugs may be tapered and discontinued during LTM; intermittent sleep deprivation also is commonly used to facilitate spells. Because frequent seizures or status epilepticus can occur in epileptic patients discontinuing medication, LTM is generally performed in a hospital setting, usually a specialized epilepsy-monitoring laboratory.

The lack of ictal activity during a clinical event does not exclude a seizure, particularly if awareness is preserved during the clinical event (e.g., a simple partial seizure) or if the event originates in the frontal lobe. Ictal activity may be apparent only with intracranial electrodes, such as depth electrodes that penetrate the brain parenchyma, subdural strips, and subdural grids.[13] These invasive electrodes, which rarely are used in diagnosing nocturnal spells because of the risks of infection and hemorrhage, generally are reserved for patients who require localization of epileptic foci before surgical resection and in whom scalp recordings are inconclusive. If an ictal pattern is not recorded but clinical seizures are strongly suspected on the basis of the history and stereotypical behavioral spells recorded on videotape, an empiric trial of an antiepileptic medication may be appropriate.

Ambulatory Monitoring

Ambulatory monitoring combines the extended recording time of V-EEG-PSG with the convenience of recording in a patient's home. Several commercial products allow patients to go home with 12 or more channels of EEG electrodes and a recording device. The recording device sometimes includes video monitoring. Ambulatory recording systems may use analog or digital recorders. Patients and their bed partners are instructed to keep an activity log to document events.

The indications for ambulatory monitoring in the differential diagnosis of nocturnal events are not established. This monitoring technique appears promising for evaluating interictal epileptiform activity during sleep. Depending on the sophistication of the video recordings, ambulatory monitoring can identify some cases of epileptic seizures, NREM arousal disorders, RBD, rhythmic movement disorder, panic disorder, and dissociative disorder.

ARTIFACTS AND PITFALLS

Artifacts, which are common during sleep-related spells recorded with any of these modalities, must be distinguished from interictal and ictal epileptiform activity and from EEG activity associated with parasomnias. Although artifacts can obscure the EEG, making diagnosis more difficult, they are sometimes helpful; for example, head or body rocking artifact may be diagnostic of rhythmic movement disorder, and rhythmic myogenic artifact may support bruxism (Figure 163-4). Other examples of frequently encountered artifacts include those caused by head tremor, eye movements, and tongue movements (glossokinetic artifact). Associated rhythmic activities can resemble ictal EEG patterns.

Pitfalls in interpreting recorded events are common; the inexperienced interpreter can mistakenly identify an artifact as EEG activity characteristic of seizures or parasomnias. When the clinician is in doubt about an EEG PSG pattern, a trained electroencephalographer should be consulted. Apart from artifacts, many other normal physiologic variants may be mistaken for epileptic activity; these include positive occipital sharp transients of sleep (see Figure 163-4); frequent and sharply contoured vertex waves, particularly in young patients (Figure 163-5); sawtooth waves; benign epileptiform transients of sleep; wicket spikes; and rhythmic temporal theta of drowsiness.[14]

RELATIVE INDICATIONS, ADVANTAGES, DISADVANTAGES, AND LIMITATIONS

Although V-EEG-PSG and other neurologic monitoring techniques can be useful in diagnosing nocturnal events, the incremental cost of these techniques over standard PSG necessitates that specific indications be met. Unfortunately, no standards exist for determining when to choose a specific monitoring technique, and the reliability and validity of the monitoring techniques described here have not been formally studied. Consensus guidelines for interpretation have not been developed. In addition, the role of ambulatory electroencephalography in monitoring parasomnias is not well defined. The indications, advantages, disadvantages, and limitations outlined here are based on my own clinical experience and that described by others.

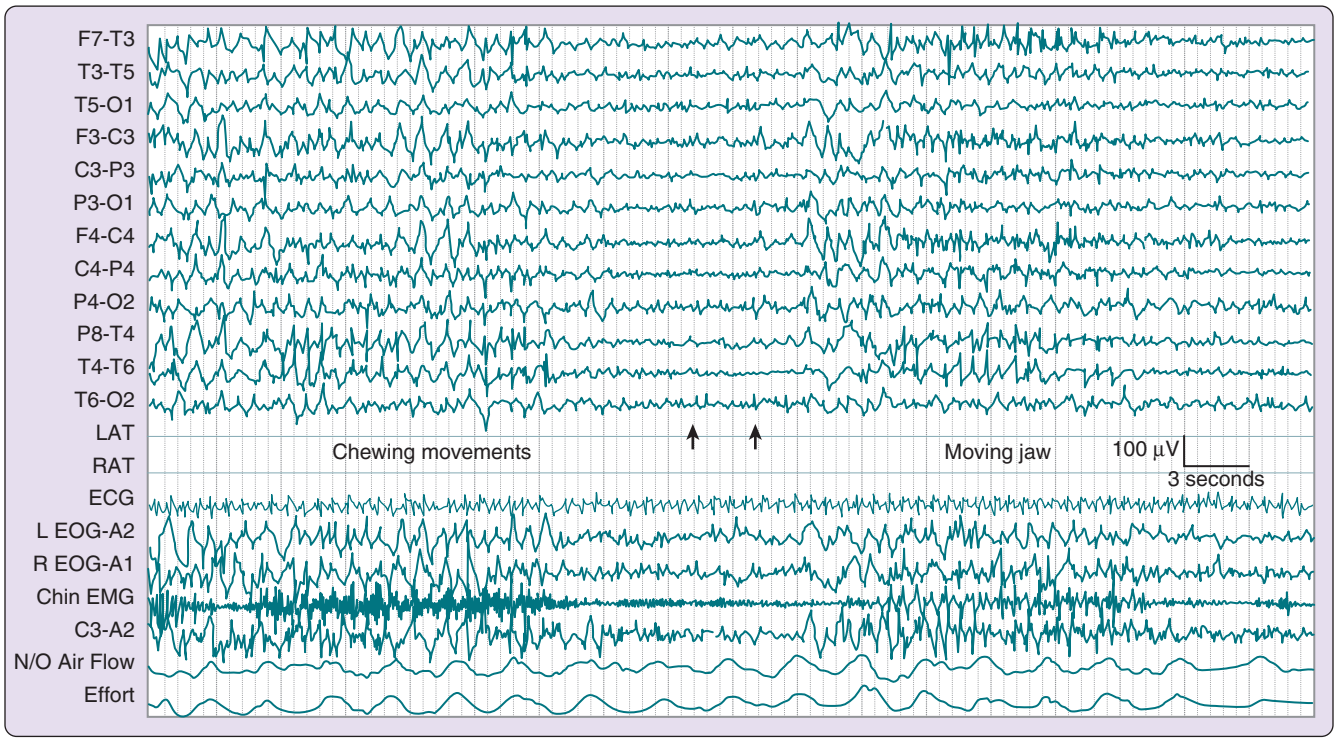


Figure 163-4 Artifacts resembling interictal epileptiform activity. Chewing movements produced by bruxism cause rhythmic activity with superimposed myogenic artifact in the electroencephalogram (EEG) channels, bearing a superficial resemblance to generalized spike-wave discharges. The *arrows* identify posterior occipital sharp transients of sleep, normal features of non–rapid eye movement (NREM) stage 1 sleep that appear sharply contoured and may be mistaken for pathologic occipital sharp waves at the standard polysomnograph paper speed of 10 mm/second. ECG, Electrocardiogram; EMG, electromyogram; EOG, electrooculogram; LAT, left anterior tibialis; N/O, nasal-oral; RAT, right anterior tibialis.

Video-Electroencephalography–Polysomnography

Indications for V-EEG-PSG include suspected sleep-related epileptic seizures, suspected NREM arousal disorders, RBD, and suspected dissociative disorder.

Suspected Sleep-Related Epileptic Seizures

Although some epileptic seizures can be diagnosed based on history (e.g., generalized tonic-clonic activity witnessed by a reliable observer), most events occurring frequently and suspected to be complex partial seizures are best confirmed with V-EEG-PSG monitoring. Monitoring is especially indicated for events that include the features of thrashing, kicking, hyperventilating, head rocking, screaming, or subtle arousals from sleep; such events may represent complex partial seizures.[15]

The advantage of V-EEG-PSG over conventional PSG or EEG without video in suspected epileptic seizures is the ability to analyze stereotypical activities characteristic of epileptic seizures, in association with ictal EEG activity. Figure 163-6 shows a recording with an extended EEG montage of an epileptic seizure, illustrating a clear evolution of activity. Apart from recording epileptic seizures, PSG with an expanded EEG montage allows sampling of IEDs throughout the night. The IEDs associated with partial epilepsy usually are more prevalent in sleep, especially delta NREM sleep, than in wakefulness.[16,17] Occasionally, IEDs missed on a routine daytime EEG are detected on an overnight recording.

CASE EXAMPLE 163-1

A 34-year-old woman with a remote history of daytime complex partial seizures presented with nightly nocturnal spells. Her husband reported that within 45 minutes of falling asleep she aroused, sat up, appeared frightened, breathed rapidly, looked around the room with a blank, wide-eyed stare, and then returned to sleep. These episodes were stereotypical and lasted less than 1 minute. She responded to his questions almost immediately and did not recall the episodes.

The differential diagnosis includes disorder of dissociative episodes as well as epileptic seizures. Nocturnal panic disorder is unlikely, because she does not recall the episodes. RBD is possible, but this disorder would be unusual in a young woman, usually does not cause stereotypical behavior, and rarely occurs soon after sleep onset.

Because the history was insufficient for diagnosis and because the spells were occurring nightly, V-EEG-PSG was performed, during which the patient exhibited several stereotypical spells arising from all stages of NREM sleep. These spells were associated with ictal discharges beginning over the right temporal lobe and consisting of rhythmic theta activity that increased in frequency and decreased in amplitude.

She was treated for complex partial seizures with carbamazepine, and the spells resolved. If this patient had been on antiepileptic medication chronically and had spells that were less frequent (e.g., once a week), an alternative approach would have been LTM with tapering of medications to promote occurrence of seizures.

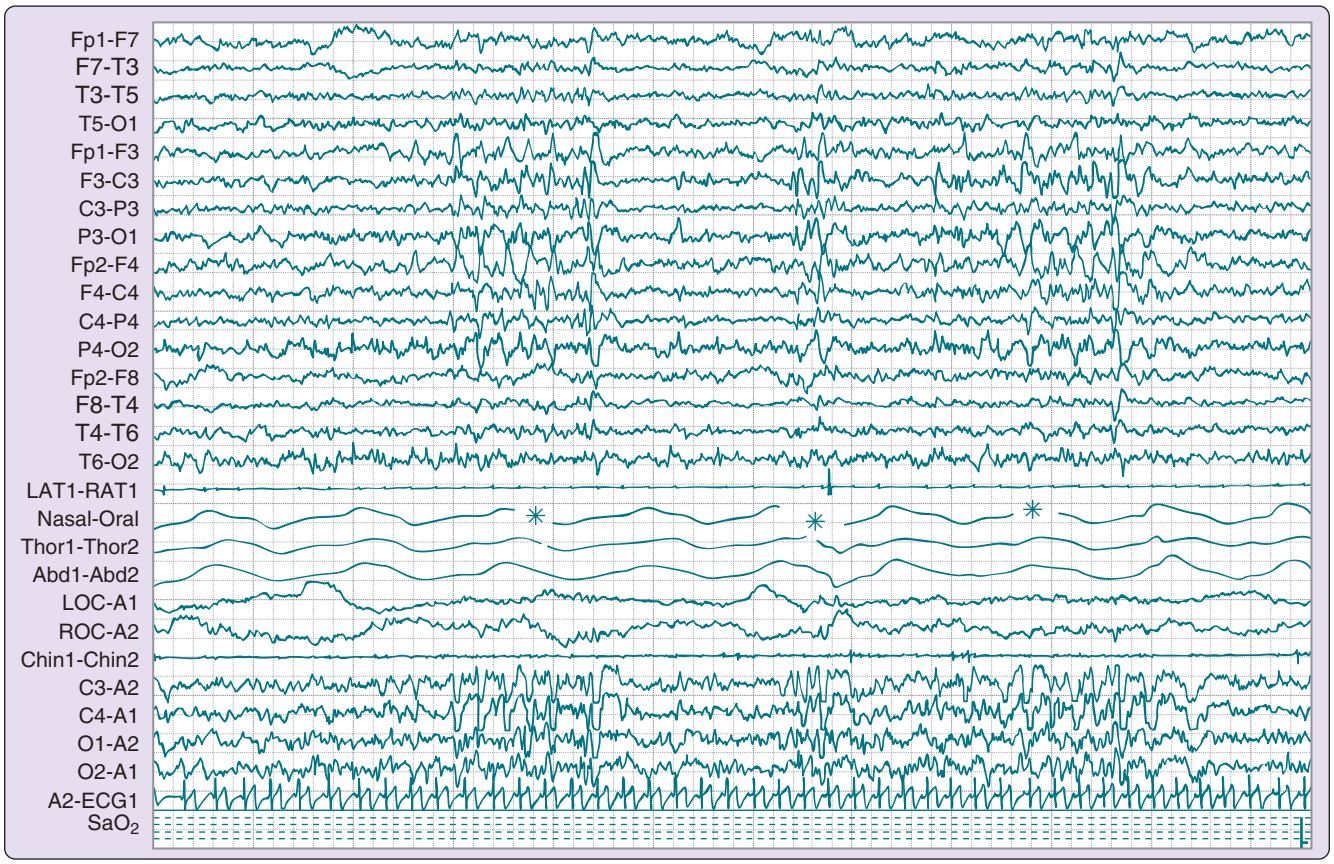


Figure 163-5 Sharply contoured vertex waves in a 6-year-old child. Although these physiologic waves resemble abnormal epileptiform activity, their morphology and distribution distinguish them from pathologic discharges. Compare with Figure 163-2. Thirty-second epoch. Calibration symbol (*bottom right*): 100 μV. ECG, Electrocardiogram; LAT, left anterior tibialis; LOC, left (electro)oculogram: RAT, right anterior tibialis; ROC, right (electro)oculogram; SaO_2, oxygen saturation.

Suspected disorders of arousal from NREM sleep (e.g., confusional arousals, sleepwalking, and sleep terrors) often can be diagnosed on the basis of history; V-EEG-PSG is indicated if behavioral features are atypical or stereotyped, multiple nightly episodes occur, onset is in adulthood, or spells do not respond to a trial of medications. An advantage of V-EEG-PSG in the diagnosis of NREM arousal disorders is the combination of video to characterize the event of interest, sleep scoring channels to determine the stage of sleep involved, and an extended EEG montage to exclude ictal EEG activity characteristic of epileptic seizures.

The V-EEG-PSG can capture a confusional arousal, night terror, or sleepwalking episode arising out of delta NREM (stage N3 sleep) accompanied by synchronous delta activity (Figure 163-7). Alternatively, the V-EEG-PSG recorded during an NREM arousal event might show asynchronous delta or theta activity, synchronous theta activity, a drowsy pattern, or nonreactive alpha activity.

Suspected REM Sleep Behavior Disorder

Although RBD may be suspected on the basis of the history, definitive diagnosis requires capturing a behavioral event on a video recording or demonstrating abnormal muscle tone or excessive limb movements during REM sleep (see Chapter 103). The advantage of V-EEG-PSG is that video is combined with sleep staging to identify REM, and an extended EEG montage is used to exclude ictal EEG activity.

CASE EXAMPLE 163-2

A 4-year-old girl had near-nightly episodes of screaming loudly with onset approximately 1 hour after falling asleep. During these episodes, her parents found her agitated and inconsolable. On rare occasions, she got out of bed and wandered out of her room. She was amnestic for these spells. An older sibling had had similar spells.

Because the history is compelling for sleep terrors, evaluation with V-EEG-PSG is not necessary. If any atypical features were present (e.g., automatisms or stereotypical behavior, multiple nightly episodes, or onset in adulthood) or if symptoms did not respond to treatment, a V-EEG-PSG would be warranted.

Suspected Dissociative Disorder

Dissociative episodes and other psychogenic spells occur during wakefulness, although the patient might appear to be asleep and might believe that he or she is asleep.[7] Because the manifestations of dissociative episodes can be quite bizarre and include thrashing, screaming, or bicycling movements, it often is impossible to distinguish these spells from epileptic

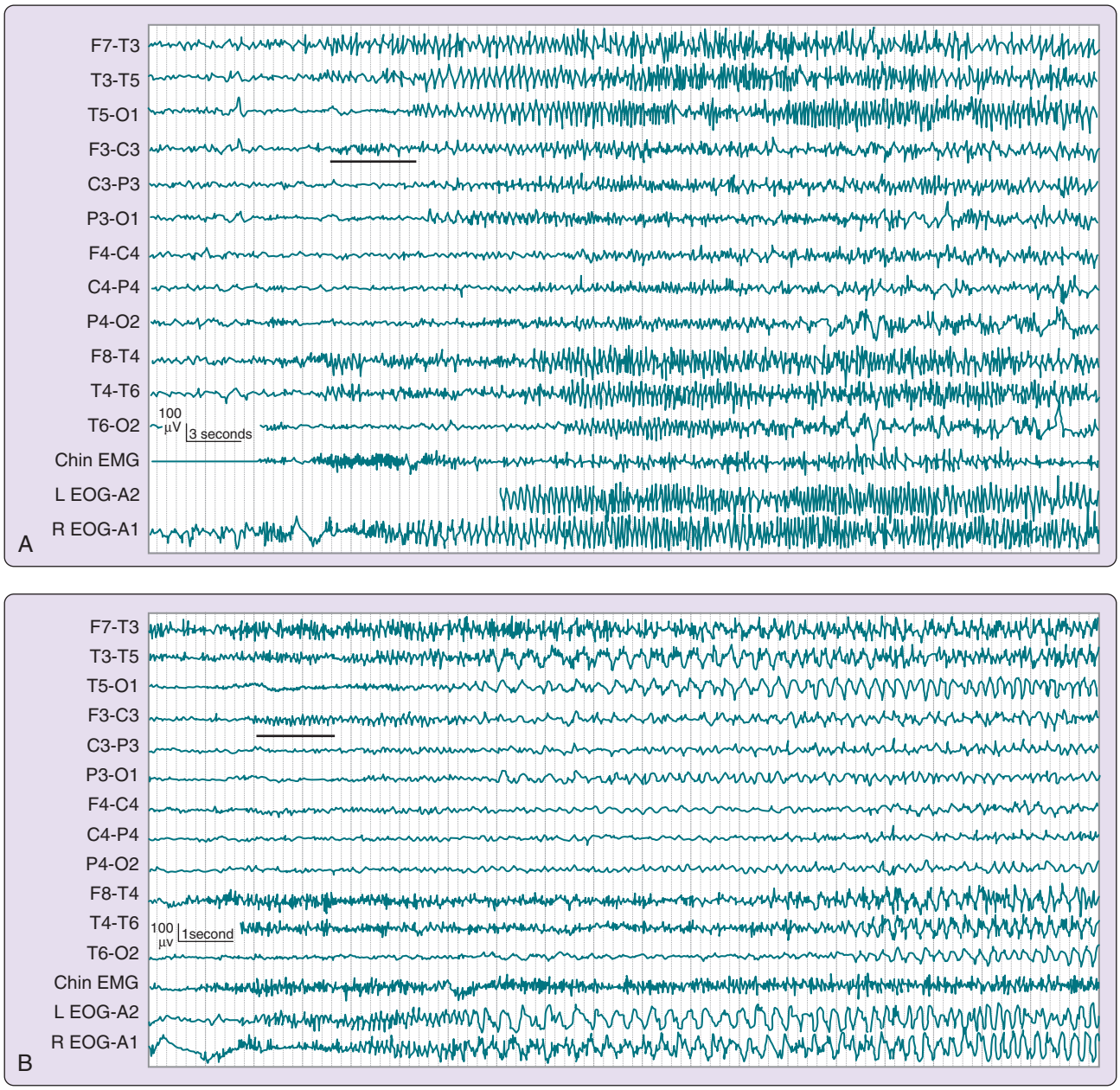


Figure 163-6 Partial seizure beginning during non–rapid eye movement (NREM) sleep. **A,** Polysomnogram recorded at 10 mm/second paper speed. Clinically, the seizure began with an abrupt arousal, followed by turning of head and eyes to the left and movements of the arms beneath the bedclothes. On the electroencephalogram (EEG), an initial reduction in voltage is followed by a progressive increase in the amplitude of the ictal discharge over the left hemisphere, with spread to the right hemisphere derivations. The *underlined activity* from the F3-C3 derivation appears to be muscle artifact; however, in **B,** at 30 mm/second paper speed, it is clear that the same *underlined segment* is the initial focal surface representation of the ictal discharge. Additional polysomnographic measures, recorded on channels 16 to 21, are not shown. EMG, Electromyogram; EOG, electrooculogram. (Modified from Aldrich MS, Jahnke B. Diagnostic value of video-EEG polysomnography. *Neurology* 1991;41:1060–6.)

seizures or parasomnias from the history alone. When nocturnal psychogenic episodes are suspected, V-EEG-PSG is advantageous in documenting the behavior of the patient, the presence of waking background EEG activity preceding onset of the spells, and the absence of ictal EEG activity.

The major disadvantage of V-EEG-PSG in the evaluation of suspected epileptic seizures, parasomnias, and dissociative disorders is the cost of the study. Additional technologist time is needed to place an extended EEG montage and to continuously observe patients throughout the study. In addition, physicians must review each spell to assess behavior and EEG patterns.

The V-EEG-PSG also has diagnostic limitations. The EEG recorded during a spell might not demonstrate an abnormality. Because epileptic seizures can lack surface EEG correlates, the absence of surface ictal EEG activity does not ensure that an epileptic seizure has not occurred. In addition, it can be difficult to differentiate an ictal EEG seizure pattern

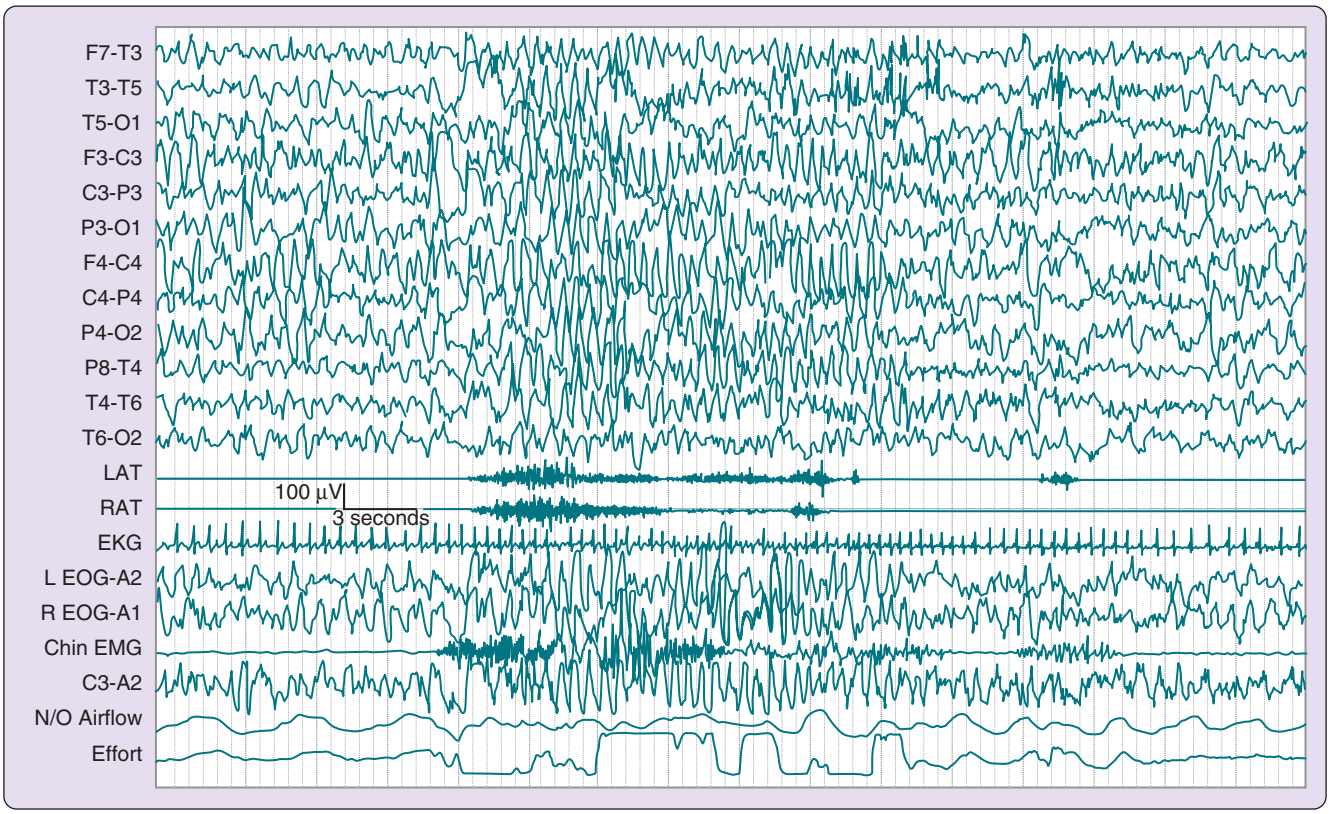


Figure 163-7 Arousal from delta non–rapid eye movement (NREM) sleep in a child with an NREM arousal disorder. Note synchronous delta activity during arousal from delta NREM sleep, associated with left anterior tibialis (LAT) and right anterior tibialis (RAT) EEG activity and a tonic increase in chin electromyogram (EMG). In contrast with the EEG of an epileptic seizure, the delta activity does not evolve in amplitude or frequency. EOG, Electrooculogram.

(which consists of rhythmic activity that evolves in frequency and amplitude) from the synchronous delta or theta activity or diffuse alpha activity occurring during an NREM arousal disorder. The best-developed portion of the ictal EEG pattern may be rhythmic delta or theta without a clear evolution, seizures can have bilateral onsets, seizures can arise from delta NREM sleep, and muscle or movement artifact can obscure the EEG. Two consecutive nights of V-EEG-PSG often are scheduled so that if no events are captured on the first night, the second night is available for study.

Daytime Electroencephalography

The advantage of daytime electroencephalography over V-EEG-PSG, standard PSG, or any of the other monitoring techniques is its brief recording time and low cost. The disadvantage is that spells, particularly sleep-related spells, are rarely captured. When the history is strongly suggestive of epileptic seizures, a routine EEG can demonstrate epileptic activity as supportive evidence of epilepsy. However, IEDs are not the equivalent of epileptic seizures and may be present in patients without epilepsy, such as those occurring in relatives of patients with benign childhood epilepsy with centrotemporal spikes. Conversely, patients with epilepsy might not have IEDs during EEG recordings. In addition, patients with epilepsy can have coexisting parasomnias. Therefore, in the absence of a compelling history, the occurrence of abnormal interictal epileptiform activity should not be used by itself to definitively diagnose nocturnal spells as epileptic seizures.

Short-Term and Long-Term Monitoring

The advantages of STM and LTM over routine electroencephalography are the acquisition of additional information over the longer recording time and simultaneous video monitoring. In patients with exclusively sleep-related spells, V-EEG-PSG is preferred over STM because the recording is performed during sleep. In patients with a mixture of daytime and sleep-related spells, STM is sometimes appropriate.

LTM is an alternative in patients in whom antiepileptic medication taper or discontinuation is planned. Medication taper or discontinuation is especially useful in facilitating seizure activity in epileptic patients with infrequent spells (e.g., once a week or less). The disadvantages of LTM are the cost of inpatient hospitalization and the need for a specialized epilepsy-monitoring laboratory. The limitations of LTM are similar to those of PSG in that spells may lack EEG correlates or might not occur even over many days of monitoring.

Ambulatory Monitoring

The advantage of ambulatory monitoring is the convenience of recording in the patient's home and the lack of need for continuous monitoring by a technologist. Cost varies, but it usually is lower than that of a recording in a sleep laboratory. A major disadvantage of ambulatory monitoring relates to the fidelity of the recording in the absence of a technologist. If electrodes become detached, ground wires break, or conductive media become dry during the study, adjustments cannot be made. In addition, systems may use a reduced number of

channels, thereby limiting the information provided, although some systems have the capability for expanded montages.

Furthermore, in contrast with the other monitoring techniques, the patient is not under constant observation and a technologist is not present. Consequently, interactions with the patient, critical for evaluating level of consciousness, are not possible. Also, interpreting rhythmic activities that resemble ictal discharges may be difficult in the absence of behavioral and consciousness assessment. The addition of synchronized video recordings to ambulatory monitoring has potential for facilitating correlation between EEG activity and clinical events.

CLINICAL PEARL

The clinician should consider performing video-electroencephalography in suspected cases of nocturnal seizures and parasomnias such as REM sleep behavior disorder or NREM arousal disorders. The video may be as helpful as the EEG in documenting stereotypical behavior in epileptic seizures.

SUMMARY

Patients with nocturnal spells present a unique diagnostic challenge to the sleep specialist and the sleep laboratory. Although standard PSG provides valuable information about the stage of sleep from which spells arise and the timing of the spell relative to sleep onset, the characterization of these spells is enhanced by video recording and an extended EEG (12 or more channels, and sometimes 21 or more). Addition of video and an extended EEG to the standard PSG is essential for precise definition of nocturnal spells, including epileptic seizures, REM sleep behavior disorder, and arousal disorders. The video component provides information on the behavioral and motor manifestations of the nocturnal spell. Depending on the clinical situation, a daytime EEG, an ambulatory EEG, daytime short-term monitoring, one or two nights of polysomnography, or long-term (over several days and nights) monitoring may be indicated.

Selected Readings

Berry RB, Brooks R, Gamaldo CE, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications,* version 2.2. Darien (Ill.): American Academy of Sleep Medicine; 2015 <www.aasmnet.org>.

Ebersole JS, Pedley TA. *Current practice of clinical electroencephalography*. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2014.

Foldvary N, Caruso AC, Mascha E, et al. Identifying montages that best detect electrographic seizure activity during polysomnography. *Sleep* 2000;**23**:211–29.

Malow BA. Sleep and epilepsy. *Neurol Clin* 2005;**23**(4):1127–47.

Schmitt B. Sleep and epilepsy syndromes. *Neuropediatrics* 2015;**46**(3): 171–80.

A complete reference list can be found online at ExpertConsult.com.

Chapter
164

Monitoring Techniques for Evaluating Suspected Sleep-Related Breathing Disorders

Max Hirshkowitz; Meir Kryger

Chapter Highlights

- Polysomnography evolved from a research tool into a medical diagnostic procedure largely applied for diagnosing and treating sleep-related breathing disorders.
- Principal components of an evaluation for sleep-related breathing disorders are airflow measurement, respiratory effort assessment, and oxyhemoglobin desaturation recording. Appreciating the relationship between sleep-disordered breathing events and sleep disruption also provides crucial information for patient care.
- Techniques are available to measure lung volume changes, blood pressure changes, and carbon dioxide; however, they are not routinely used for sleep-related breathing disorder evaluation in adults.

OVERVIEW

Soon after their discovery, sleep-related breathing disorders (SRBDs) became clinical polysomnography's dominant focus.[1] Polysomnography began as a laboratory research technique to study sleep, but with a few additional channels, it quickly found a new role as the "gold standard" modality for diagnosing obstructive sleep apnea. Penetration for assessing sleep-disordered breathing has been so complete that undoubtedly the vast majority of sleep studies performed tonight, or on any given night, will serve this function. Nonetheless, home sleep testing increasingly makes inroads as the technique to diagnose sleep-related breathing disorders.[2] With few exceptions, polysomnography and home testing devices use similar techniques. This chapter presents an overview of recording techniques and their application for evaluating breathing during sleep.

Abnormal breathing during sleep takes several different forms and arises from various underlying etiologies. It goes by many names, including sleep apnea, sleep apnea-hypopnea syndrome, sleep-disordered breathing (SBD), sleep-related breathing disorder (SRBD), periodic breathing, Cheyne-Stokes respiration (CSR), and hypoventilation (Box 164-1). However, obstructive forms of sleep apnea-hypopnea syndrome remain by far the most common SRBD. Detailed descriptions regarding SRBD-associated etiology, morbidity, and treatments appear elsewhere in this book.

The specific pathophysiologic events underlying SRBD include apnea episodes, hypopnea episodes, respiratory effort–related arousals, oxyhemoglobin desaturations, and snore arousals. As their names imply, *apnea* is cessation of breathing and *hypopnea* is shallow breathing. The face validity, that not breathing is undesirable for the health of the organism, seems obvious. However, the association between apnea and morbidity can be, but is not necessarily always, a matter of cause and effect. Other pathophysiologic processes, especially those related to metabolic and cardiac diseases, can produce downstream effects on breathing.

Polysomnographic and home sleep testing criteria for designating a breathing event as an apneic spell have been consistently defined for decades. Cessation of breathing for 10 seconds or longer constitutes an apnea. Ten seconds was chosen because it approximates missing two breaths. The definition for hypopnea, by contrast, remains controversial. *Hypopnea* essentially means "shallow breath." The difficulty in defining hypopnea stems from several sources. The first involves measurement technique, the second from the fact that hypopneas are not intrinsically pathophysiologic, and finally from decades of usage without any standard.

Polysomnographic and home sleep testing airflow monitoring techniques mostly rely on surrogate and uncalibrated measurements (as detailed further on). Furthermore, airflow signal magnitude from thermistors, thermocouples, capnographs, and nasal pressure transducers correlates poorly with tidal volume. Consequently, operational definitions for hypopnea based on a percentage decrease of flow signal use arbitrary cutoff points (regarding which considerable disagreement continues).

Brief hypopneic intervals routinely occur during wakefulness without causing harm. For example, hypopnea accompanies speaking, and except in extraordinary circumstances, talking does not produce adverse health consequences. During sleep, however, a hypopnea may provoke significant oxyhemoglobin desaturation or a central nervous system (CNS) arousal. Thus the consequence of the hypopnea (not the hypopnea itself) conceivably may be deemed pathophysiologic. Accordingly, prerequisites for designating sleep hypopnea as abnormal involve (1) accurately measuring physiologic consequences, (2) determining at what point these consequences reach significance, and (3) demonstrating their role in morbidity.

Box 164-1 SLEEP-RELATED BREATHING EVENTS, DISORDER CLASSIFICATIONS, AND PARAMETERS

- **Apnea** (A): The complete or near-complete cessation of breathing for 10 seconds or longer in an adult.
- **Oxyhemoglobin desaturation event** (ODE): A decrease in blood oxygen saturation. Usually, the magnitude must be 3% or greater to distinguish it from signal noise.
- **Hypopnea** (H): A reduction in (but not cessation of) ventilation. To be clinically significant, the hypopnea must be associated with an oxyhemoglobin desaturation event or a CNS arousal.
- **Desaturating hypopnea** (DH): A hypopnea with oxyhemoglobin desaturation of 4% or greater (also known as a "Medicare hypopnea").
- **Respiratory effort–related arousal** (RERA): A CNS arousal terminating obstructive breathing events that do not meet the criteria for apnea or hypopnea.
- **Obstructive apnea** or **hypopnea** episode (OA or OH): Breathing event caused by upper airway obstruction or reduced upper airway patency.
- **Central apnea** or **hypopnea** episode (CA or CH): Breathing event caused by absent or reduced respiratory effort (i.e., decreased output to inspiratory muscles from CNS respiratory control centers).
- **Mixed apnea** or **hypopnea** episode (MA or MH): Breathing event with both obstructive and central features. Many laboratories assign mixed events to the obstructive category.
- **Periodic breathing** (PB): A regularly repeating pattern in which normal or increased ventilation alternates with decreased or absent ventilation.
- **Central sleep apnea syndromes** (CSASs): These syndromes include primary central sleep apnea, Cheyne-Stokes breathing pattern, high-altitude periodic breathing, central sleep apnea due to medical conditions other than Cheyne-Stokes, and central sleep apnea due to drug or substance use/abuse.*
- **Obstructive sleep apnea syndromes** (OSASs): These syndromes include obstructive sleep apnea and obesity-hypoventilation syndrome (see Sleep-Related Hypoventilation Syndromes†). Primary snoring is a normal variant of an obstructive airway process.*

*Opiate and other sedatives are known to depress respiration by blunting respiratory drive.
†*ICSD-2* diagnostic classification.
CNS, Central nervous system.

Unfortunately, measurement and morbidity determination issues have remained unresolved for many decades. Many operational definitions for hypopnea* have appeared in published clinical and research literature. Emergence of differing criteria for percentage airflow decrease and oxyhemoglobin decrease and inclusion versus exclusion of CNS arousal as part of the definition seriously complicated matters. Definitional criteria set forth in a standardization attempt sponsored by the American Academy of Sleep Medicine, often referred to as the "Chicago Criteria," were adopted for research but not for clinical use.[3]

Subsequently on April 1, 2002, like it or not, the Centers for Medicare and Medicaid Services (CMS) defined a hypopnea as "an abnormal respiratory event lasting as least 10 s with at least a 30% reduction in thoracoabdominal movement or airflow as compared to baseline, and with at least 4% desaturation."[4] Eliminating CNS arousal from the hypopnea criteria and requiring 4% desaturation (rather than 3%) decreased the number of scorable hypopneic spells in many patients.[5,6]

On the face of it, respiratory-related and snore-related arousals certainly seem irrefutably abnormal (because they compromise sleep integrity); however, issues have arisen concerning detection reliability (and impossibility with use of many home sleep testing devices). Ignoring CNS arousal provoked by respiratory events seemed so ridiculously misguided to many clinicians that they began tabulating such events (even though they fell short of the 4% desaturation criteria). The designation *respiratory effort–related arousals* (RERAs) was commandeered from the Chicago Criteria even though it originally had quite different criteria.

In a major effort to standardize sleep medicine clinical practice, the AASM published *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* in 2007,[7] sometimes referred to as the *AASM Manual*. With regard to hypopnea, two definitional criteria were offered, neither of which agreed with the CMS criteria. Member centers were required to adopt and use the AASM criteria to maintain accreditation. One of the definitional criteria used a 3% oxyhemoglobin desaturation and the other included CNS arousal. These conflicting mandates and ambiguities placed American sleep specialists in an intellectual, clinical, and legal dilemma. Clinicians seeing even a single Medicare- or Medicaid-covered patient are required to use CMS criteria for all patients seen in their practice. Failure to comply with this regulation constitutes Medicare fraud. This CMS directive's intent is to prevent a two-tier medical system from developing. AASM issued clarifications of the rule in 2012, and ultimately, in 2013, AASM adopted CMS criteria[8] for scoring hypopnea.

Mitterling and colleagues examined scoring outcome using the different definitions.[9] Applying the 2012 scoring criteria produced higher apnea-plus-hypopnea indices compared with scoring based on the 2007 rules. These investigators used a sample of 100 healthy sleepers, ranging in age from 19 to 77 years.

Apnea and hypopnea episodes can be further categorized as obstructive, central, or mixed, depending on the presence or absence of respiratory effort during the entirety or some part of the breathing event. Beyond mere presence or absence, changes in the effort's magnitude can provide information concerning airway resistance. Increasing effort leading to a CNS arousal provides insight regarding pathophysiology. Therefore respiratory effort represents a crucial measure for evaluating patients with SRBD.

Sleep-disordered breathing severity can be based on a clinical dimension (e.g., sleepiness), event frequency (e.g., number of events per hour), or magnitude of the consequence (e.g., degree of oxyhemoglobin desaturation). Table 164-1 provides examples for dimensionally classifying severity of sleep-disordered breathing. General agreement is lacking on assignment of severity descriptors to indices of sleep-disordered breathing; however, two schemes are commonly used. In the first, "liberal" classification, an apnea-hypopnea index (AHI) between 5 and 15 is mild, between 15 and 30 is

*At last count, this author (MH) is aware of more than 15 different definitions for hypopnea.

Table 164-1 Clinical/Laboratory Features of Obstructive Sleep Apnea Syndrome by Severity

Dimension	Mild	Moderate	Severe
Sleepiness or unintended sleep episodes	During activities requiring little attention (e.g., watching television)	During activities requiring some attention (e.g., business meeting)	During activities requiring active attention (e.g., driving)
PSG SRBD events: number per night	5–15	15–30	>30
PSG/HST SRBD events and oxyhemoglobin (Sa_{O_2})	RDI or AHI 2–20 and/or Sa_{O_2} nadir >85%	RDI or AHI 20–40 and/or Sa_{O_2} Nadir 65%-85%	RDI or AHI >20 and/or Sa_{O_2} nadir <65%

AHI, Apnea-hypopnea index; HST, home sleep test; PSG, polysomnography; RDI, respiratory disturbance index; Sa_{O_2}, arterial oxygen saturation; SRBD, sleep-related breathing disorder.

moderate, and greater than 30 is severe. In the second, "conservative" classification, an AHI between 10 and 20 is mild, between 20 and 50 is moderate, and greater than 50 is severe.[10]

MEASURING AIRFLOW

General Considerations

Most clinical techniques use qualitative surrogate measures to estimate airflow changes. Fully quantitative airflow determination requires pneumotachography or having the patient sleep in a "body box"; however, such techniques are unsuitable for routine clinical sleep studies. Although semiquantitative measures are attainable using calibrated inductance plethysmography, most clinical evaluations rely on qualitative nasal-oral thermography and nasal pressure. This approach provides adequate data and minimizes patient discomfort, reduces costs, and simplifies data acquisition. The AASM Manual recommends using a thermal sensor to identify apnea and a nasal pressure sensor to detect hypopnea.[7] However, other methods also provide reliable assessment. Airflow also can be measured qualitatively by detecting chemical differences between ambient and expired air (e.g., capnography). However, respiratory activity requires careful classification. Sometimes the patient's airway can be completely occluded during inspiration but release small puffs upon expiration (detectable by thermistor or CO_2 analyzer). Such events are erroneously categorized as hypopneas or even normal (unobstructed breathing). Figure 164-1 illustrates the problem.

ERROR IN DETECTING APNEA

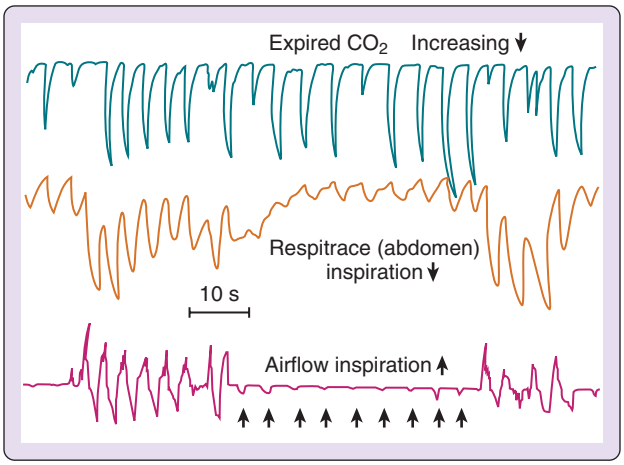


Figure 164-1 An example of the limitations of noninvasive airflow detection. Airflow is recorded simultaneously with a CO_2 analyzer and a pneumotachograph. During the obstructive apnea event, periods of expiratory airflow occur (recorded by the pneumotachograph and the CO_2 analyzer) in the absence of inspiratory flow (obvious in the pneumotachograph recording and unclear in the CO_2 recording). Without the information from the pneumotachograph, the recording from the CO_2 analyzer would be interpreted as evidence of uninterrupted inspiratory and expiratory airflow. *Top,* Airflow is detected with the CO_2 analyzer. *Middle,* Respiratory inductance plethysmograph (RIP) ("Respitrace"). *Bottom,* Airflow measured with a pneumotachograph. With each apnea-related expiratory deflection documented by the pneumotachograph (*arrows, bottom*), there is a sustained shift in the baseline of the RIP tracing. This correlation suggests an incremental decrease in functional residual capacity resulting from absence of inspirations with continued small expiratory puffs. If only the top two tracings were available, this pattern would have been mistakenly called hypoventilation or hypopnea, whereas it clearly reflects total occlusion on inspiration. (From West P, Kryger MH. Sleep and respiration: terminology and methodology. *Clin Chest Med* 1985;6:706.)

Measuring Temperature

Exhaled air usually is warmer than ambient temperature. Air in the lungs is warmed by core body heat, thereby creating a temperature difference between air entering and exiting the respiratory system. Consequently, measuring temperature fluctuation at the nares and in front of the mouth provides a simple surrogate measure of airflow. Measurement is possible using several different technologies.

Thermistors are thermally sensitive variable resistors that produce voltage alterations when connected in a low-current (but constant-current) circuit. Low current minimizes the tendency for the thermistor to heat itself. Thermistors maximize sensing area while minimizing sensor size and mass. Small temperature changes can produce large resistance changes that can in turn be transduced with a bridge amplifier. Care must be taken to ensure that the thermistor remains below body temperature (i.e., it must not rest on the skin); otherwise, expired air will not be warmed, and no resistance change will occur. In such a case, inspiratory activity and a respiratory pause will not be differentiable.

Thermocouples also sense temperature change but use a different approach. Different metals expand at different rates when heated. This difference can be transduced to voltage alterations displayable on polygraph systems. Like thermistors, thermocouples are placed in the airflow path in front of the nares and mouth, where expired air heats the sensor and increases its resistance. The transduced signal reflects oscillation between exhaled warm air and cooled inhaled air, thus providing a trace roughly corresponding to respiratory airflow.

Nasal Airway Pressure

During inspiration, airway pressure is negative relative to atmosphere. By contrast, expiration produces a relatively positive pressure in the airway. The resulting alteration in nasal airway pressure can provide a surrogate estimate of airflow and correlates favorably with pneumotachographically recorded signals.[11] The nasal pressure signal also offers greater sensitivity than that of nasal-oral thermography for detecting subtle flow limitations (Figure 164-2) when patients breathe through their noses.[12] Airflow limitation shows as pressure trace plateauing during inspiration. A direct current (DC) amplifier provides optimal interface; however, long time-constant alternating current (i.e., a very slowly coupled signal) can suffice. By contrast, rapid coupling can create artifact (Figure 164-3).

Expired Carbon Dioxide Sensors

CO_2 concentration in air leaving the lungs far exceeds that in ambient air. Thus measuring CO_2 in front of the nose and mouth can detect expiration. Infrared analyzers can determine the concentration. Because exhaled CO_2 reflects physiologic chemical change, it offers several advantages compared with physical changes detected by thermistor, thermocouple, and nasal pressure recordings.

In some patients, the end-of-breath CO_2 concentration provides evidence of elevated end-tidal Pco_2. The catheters sampling CO_2 typically entrain some room air, making the measured CO_2 lower than actual end-tidal Pco_2. Therefore an elevated CO_2 indicates that true Pco_2 is even higher, thereby providing a noninvasive technique (merely sampling the air stream) for detecting hypoventilation. The shape of the expired CO_2 curve can also offer useful information. When the patient's baseline expired CO_2 curve shows a clear-cut plateau, the loss of this plateau (or the curve's becoming smaller or dome-shaped) indicates a change in breathing pattern, usually a reduction in expiratory volume.

During central apnea, a low-volume catheter system set at its most rapid response time can show cardiogenic oscillations in the CO_2 signal. These oscillations result from small volume displacements caused by the beating heart.[13] These heartbeat-synchronized oscillations signify upper airway patency (Figure 164-4).

Capnography can be used in the sleep labratory to titrate noninvasive ventilation in patients with hypoventilation syndromes.[14] In infants and children with upper airway obstruction, severe hypoventilation may occur during sleep without observable apnea or hypopnea. Measuring expired CO_2 provides evidence for hypoventilation not detectable using thermistors or thermocouples.[13]

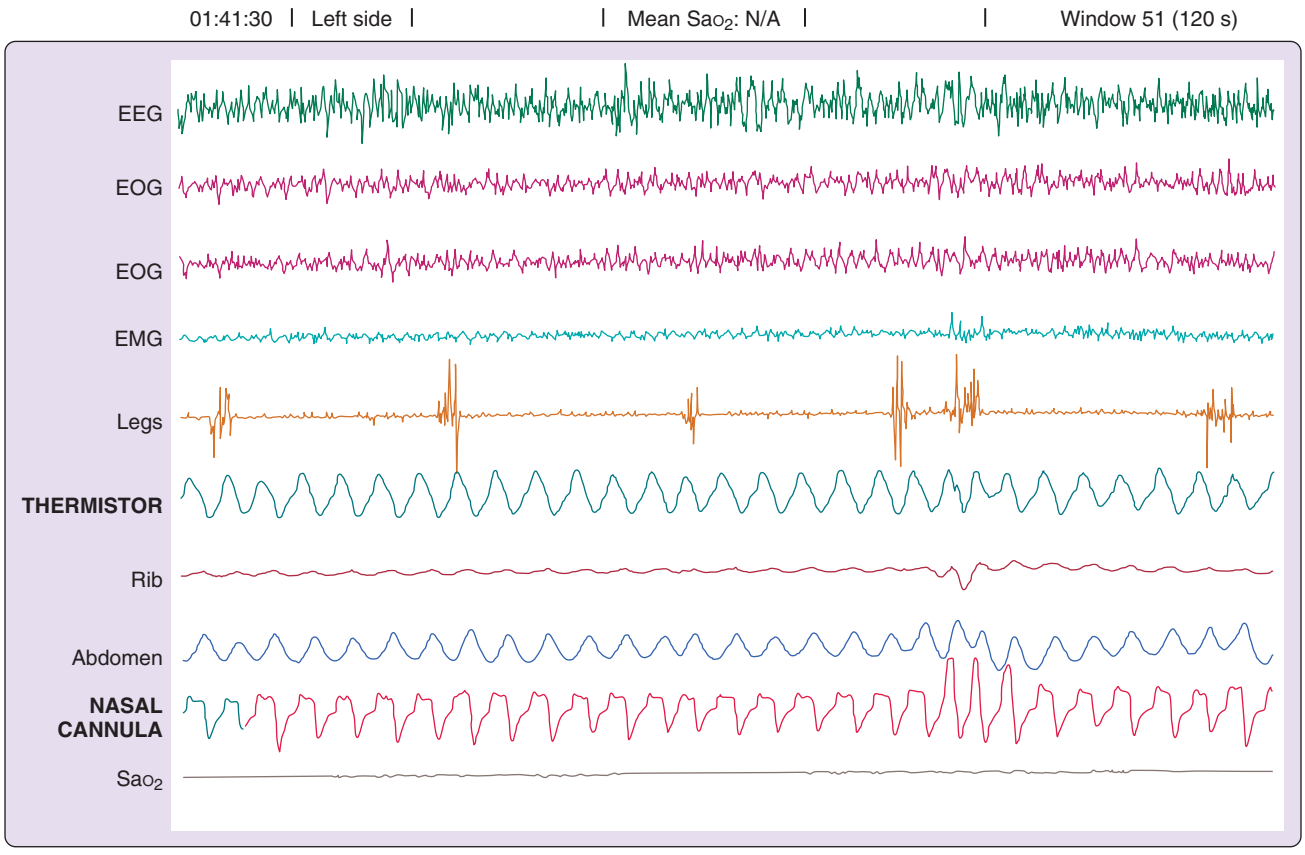


Figure 164-2 A 120-second section from a nocturnal polysomnogram in a subject undergoing simultaneous recording with a conventional thermistor and with a nasal cannula used for recording pressure. Nothing in the thermistor tracing suggests a respiratory event, and subtle movement barely registers in the rib and abdominal inductance plethysmographic tracings. In the nasal cannula tracing, however, the end of one flow limitation episode and the beginning of another are easily detected. Note the plateaus (*chopped-off tops*) of the pressure traces during flow limitation. EEG, Electroencephalogram; EMG, electromyogram; EOG, electrooculogram; Sao_2, oxygen saturation in arterial blood. (Courtesy Dr. David Rappaport, New York University, New York.)

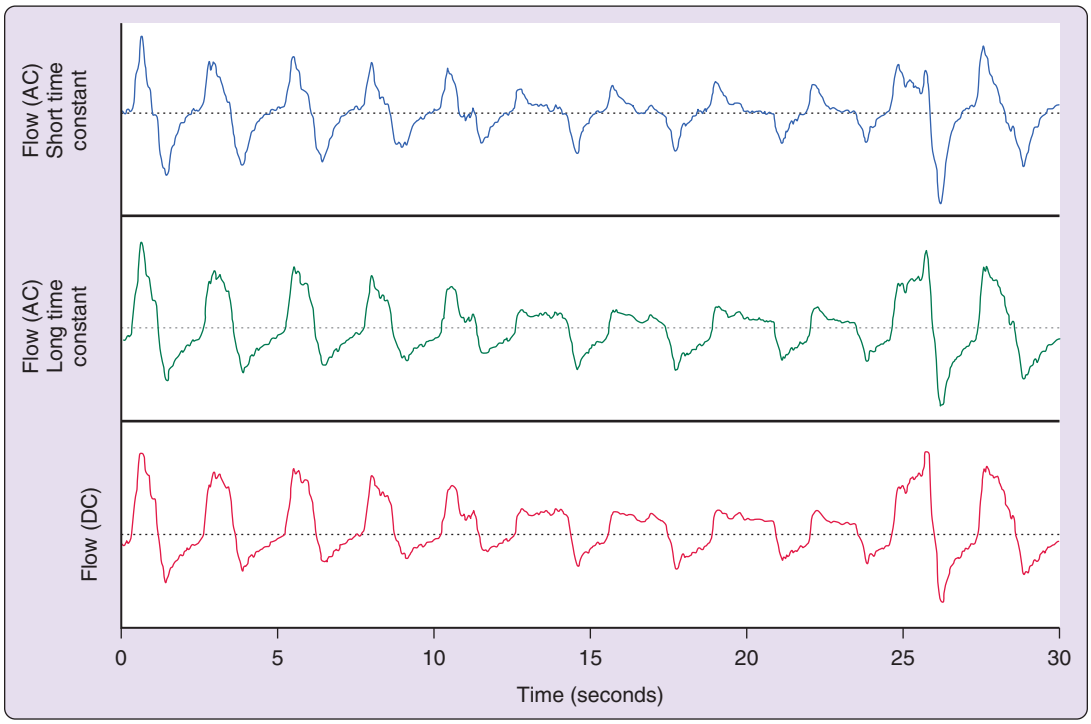


Figure 164-3 A flow limitation event recorded from the nasal cannula measuring pressure simultaneously amplified by three different amplifiers. The bottom signal is from a direct current (DC) amplifier with no filtering. The top two signals are from alternating current (AC) amplifiers with low-frequency filters, with time constants of 1.6 (*top*) and 5.3 (*middle*). The shorter time-constant filter (*top*) causes the flow signal to decay to baseline rapidly during a period of relatively constant flow (flow limitation plateau). The longer time-constant filter (*middle*) provides reasonably good reproduction of these constant flows.

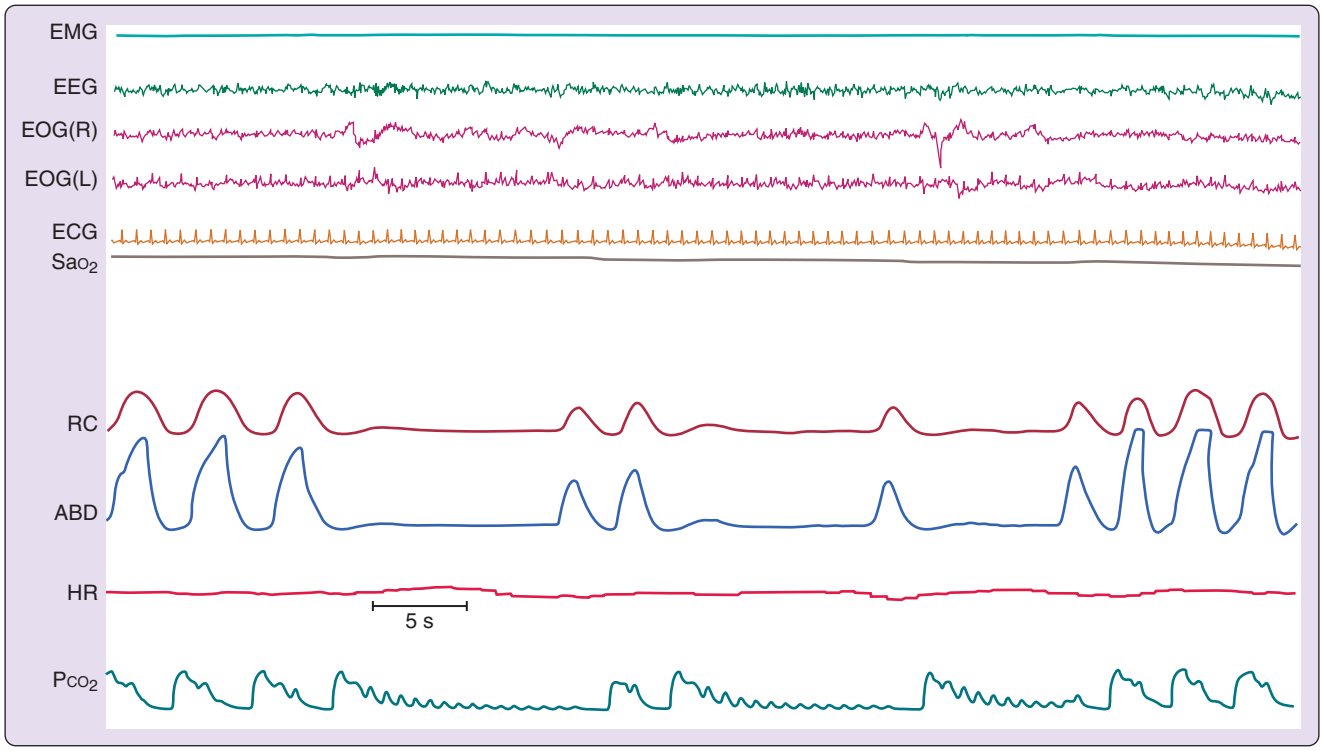


Figure 164-4 Cardiogenic oscillations in CO_2 are seen in the *bottom channel* of the recording in this example of central apnea. The presence of these oscillations, synchronous with the heartbeat, signifies that the upper airway is patent. ABD, Abdominal (movement); EEG, electroencephalogram; ECG, electrocardiogram; EMG, electromyogram; EOG, electrooculogram; HR, heart rate; RC, rib cage (movement); P_{CO_2}, partial pressure of carbon dioxide; Sa_{O_2}, oxygen saturation in arterial blood.

Pneumotachography

Pneumotachography accurately and quantitatively measures airflow volume. The patient usually wears a face mask, and the procedure can be uncomfortable. In awake subjects, pneumotachography is known to alter breathing; it increases tidal volume and reduces respiratory rate. Some positive airway pressure machines contain an integrated pneumotachograph usable to monitor airflow during laboratory titration. Pneumotachography is seldom used for routine SRBD diagnostics.

Several types of pneumotachographs are available. These devices differ with respect to measurement technique: They use either (1) differential pressure airflow transducers, (2) ultrasonic flow meters, or (3) hot-wire anemometers. Discussion is limited here to differential pressure flow transducers because they are the most widely used. In this technique, airflow directed through a cylinder exits through a small resistive field, usually composed of small parallel tubes or a grill promoting laminar flow. The pressure drop across this resistive field is measured using a differential manometer. When flow is laminar, the relationship between the pressure differences and flow is linear. Changes in gas density, viscosity, and temperature alter the pressure-flow relationship. To prevent condensation on the resistive element requires heating, so calibration should be conducted when the pneumotachograph is heated. After correction for errors introduced by alterations in these physical factors, the flow signal is integrated to determine volume.

MEASURING RESPIRATORY EFFORT

Measuring respiratory effort provides information needed to distinguish between respiratory events of obstructive versus nonobstructive (central) etiology. Categorizing an apnea or hypopnea as obstructive, mixed, or central in origin derives from differential respiratory effort and airflow patterns. As reviewed next, several techniques are available to detect and/or measure respiratory effort, including rib cage and abdominal motion, electromyography, pleural pressure changes, movement detected by static charge sensors in or on the bed surface, movement detected by standing wave patterns in the bedroom, and digital video recording.

Rib Cage and Abdominal Motion

Currently, the most common polysomnographic technique for measuring respiratory effort involves detection and quantification of rib cage and abdominal movements. During normal breathing, the major inspiratory muscles produce rib cage expansion and a downward movement of the diaphragm. These movements cause the pressure around and in the lung to become negative (relative to atmospheric pressure). The pressure gradient between ambient air and the lung draws air through the airways into the alveoli. Thus a change in lung volume is the sum of the volume changes of the structures surrounding the lungs, the rib cage, and the abdomen.[15] Other respiratory muscles (e.g., intercostal, sternocleidomastoid) also play a role in stabilizing the thoracic cage. Some clinicians erroneously interpret the abdominal and rib cage motion changes as implying separate activities of abdominal and thoracic respiratory muscles, but this is not the case. Virtually all of the changes in abdominal and rib cage volumes (including paradoxical motion) can be explained by changes in the status of the respiratory muscles directly inserting onto the thoracic cage. Paradoxical motion of the rib cage and abdomen can result from several changes, including loss of tone of the diaphragm, loss of tone of the other respiratory muscles, and upper airway obstruction (complete or partial). The mechanisms underlying this asynchronous motion of rib cage and abdomen are described in Box 164-2. Regardless of the pattern or its underlying mechanism, rib cage and abdominal movement reflect effort to breathe.

Box 164-2 MECHANISMS UNDERLYING PARADOXICAL MOTION OF THE RIB CAGE AND ABDOMEN

Loss of diaphragm tone. When the diaphragm ceases to contract and becomes flaccid, it merely reacts to pressure changes around it instead of generating pressure changes. In this situation, when the other respiratory muscles contract, the rib cage is enlarged, and pleural pressure becomes negative, sucking the diaphragm into the chest. This condition results in an increase in rib cage volume and a reduction in abdominal volume.

Loss of accessory respiratory muscle tone. When the accessory muscles lose tone, the rib cage, particularly the upper part of the rib cage, becomes unstable. When the diaphragm then contracts, the negative intrathoracic pressure causes the unstable part of the thorax to be sucked in during inspiration.

Partial upper airway obstruction. With partial upper airway obstruction, the diaphragm must generate very strong negative pressures for inspiration to occur. As the diaphragm contracts, it both pushes out the abdomen and creates great negative intrathoracic pressure. This highly negative intrathoracic pressure can overcome the mechanisms maintaining chest wall stability (accessory muscle tone and rigidity of the cage), so that the least stable portions of the rib cage will tend to move inward with inspiration. This potential for inward movement of the rib cage is a problem mainly in the very young, in whom the rib cage is quite pliable.

At a minimum, a single uncalibrated abdominal movement sensor can *detect* respiratory effort. Respiratory effort during airflow cessation usually signifies airway obstruction. Common approaches for *measuring* rib cage and abdominal movement use (1) strain gauges, (2) inductance plethysmography, and (3) piezoelectric transducers.

Strain gauges are sealed elastic tubes filled with conductive material through which an electric current is passed. When length is constant, current and resistance are constant. Stretching the strain gauge lengthens and narrows the cross-sectional area of the fixed-volume conductor. This deformation produces a proportional increase in electrical resistance. Current varies inversely in relation to the length of the gauge, thereby becoming an index of gauge length. A Whetstone bridge amplifier transduces this change to voltage for continuous display showing rib cage or abdominal expansion (depending on placement).

Inductance plethysmography electronically measures changes in the cross-sectional area of the rib cage and abdominal compartments by determining changes in inductance. Inductance is a property of electrical conductors characterized by the opposition to a change of current flow in the conductor.

Transducers are placed around the rib cage and abdomen—the physiologic equivalent of conductors. Each transducer consists of an insulated wire sewn into the shape of a horizontally oriented sinusoid and onto an elasticized band.

Piezoelectric transducers are used in yet another method of detecting movement. These sensors can be placed on the rib cage and abdomen and are sensitive to changes in length. When a piezoelectric crystal is squeezed, an electrical potential appears across its sides. The crystals can be arranged (usually as part of a belt) so that movement can be detected (see next).

Combined Sensors

Thermocouples, thermistors, and strain gauges are fairly old technologies. More recently, new materials such as polyvinylidene fluoride film have been introduced into the sleep laboratory. Such film has the interesting property of converting heat and mechanical energy into electrical energy that can be measured. Polyvinylidene fluoride films have both piezoelectric (responding to mechanical changes) and pyroelectric (responding to thermal changes) properties. The output from these films can be configured to measure airflow[16] (pressure and temperature), snoring[17] (pressure waveforms), and changes in length caused by abdomen and rib cage movement.[18-20]

Respiratory Muscle Electromyography

Recording intercostal muscle activity on the electromyogram (EMG) is one of the oldest polysomnographic techniques for detecting respiratory effort (Figure 164-5). These uncalibrated recordings are made using standard surface electrodes placed in pairs in the intercostal spaces on the right anterior chest. Obtaining an optimal signal requires practice, patience, and skill; recordings are prone to artifact, especially artifact from the electrocardiogram (ECG). Intercostal EMG activity, when recorded properly, can be extremely valuable for differentiating among central, obstructive, and mixed sleep-disordered breathing events. Furthermore, although signals are not calibrated, cascading increases in respiratory effort are readily apparent from recordings.

Pleural Pressure Changes

Some sleep centers use esophageal pressure to index inspiratory effort. In our experience, most patients undergoing all-night polysomnography find esophageal balloons unacceptable. However, the thin water-tip or catheter-tip piezoelectric transducers are better tolerated. Esophageal pressure measurements help verify central apnea or hypopnea episodes with a high degree of certainty. This technique also can detect very subtle respiratory events. However in subtle cases, a repeat

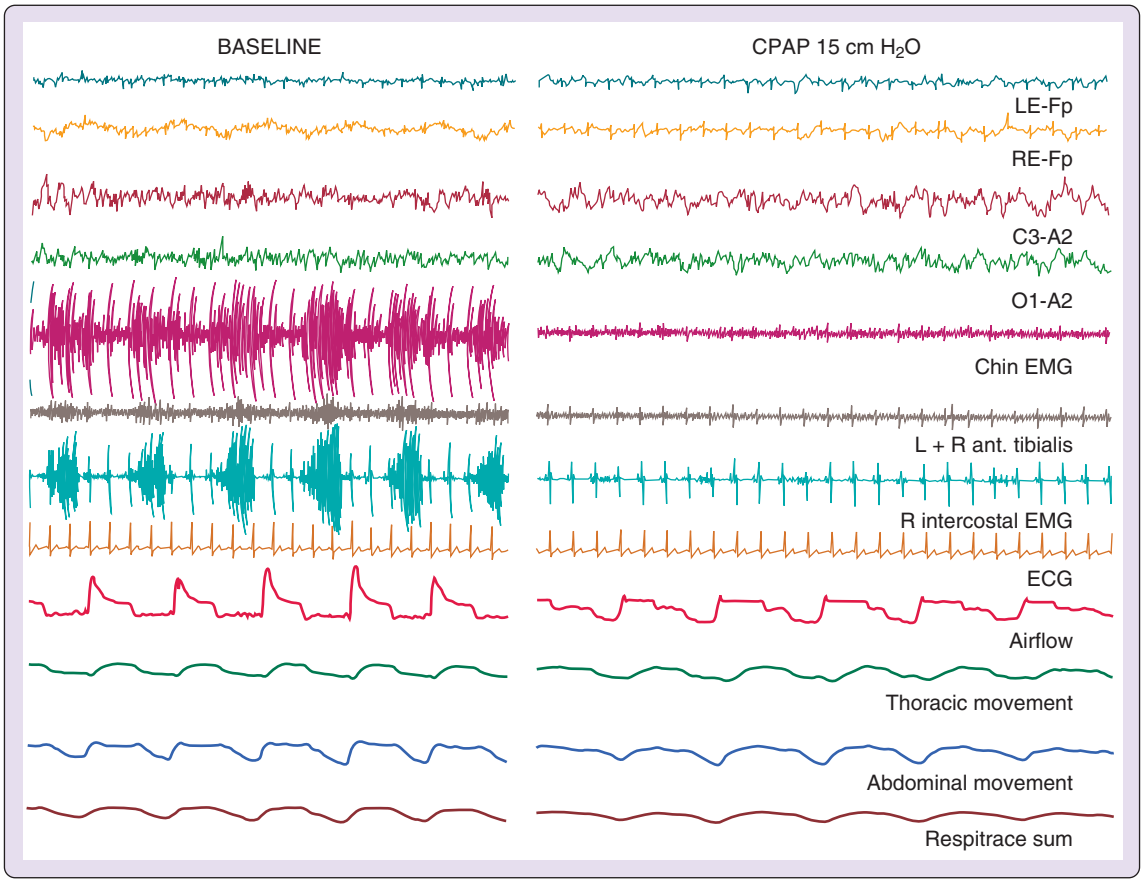


Figure 164-5 Surface respiratory (R intercostal) muscle EMG in sleep apnea. *Left,* The respiratory EMG signal is dramatically increased. *Right,* This signal is reduced on nasal continuous positive airway pressure (CPAP). A2, Right mastoid reference; C3, left central EEG; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; LE-fp, left eye referenced to frontal pole; L + R, linked left and right; O1, right occipital EEG; RE-fp, right eye referenced to frontal pole. (Courtesy Dr. J. Catesby Ware, Eastern Virginia Medical School, Norfolk, Virginia.)

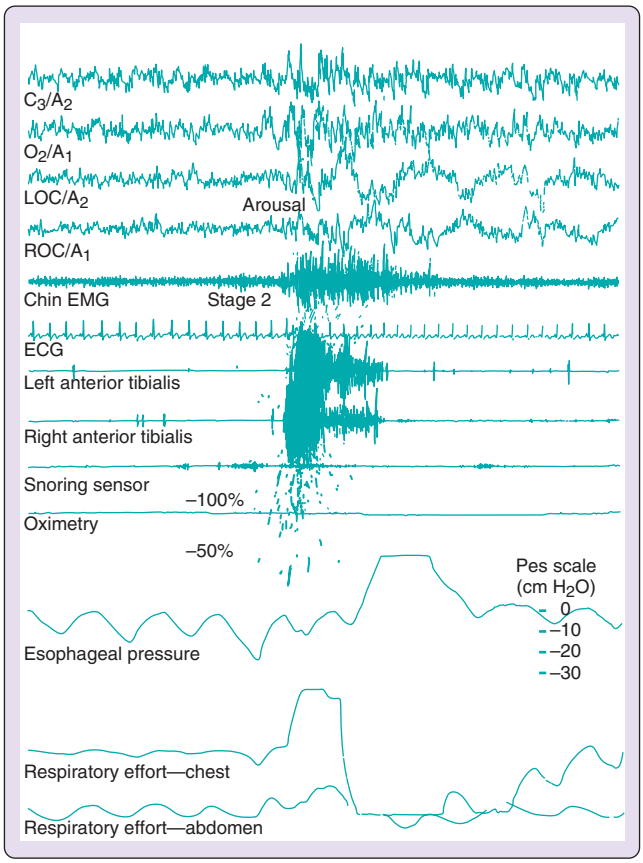


Figure 164-6 Esophageal pressure in the upper airway resistance syndrome. The esophageal pressure (Pes) swing was greatest just before the arousal. C3/A2, Left central EEG referenced to right mastoid; ECG, electrocardiogram; EMG, electromyogram; LOC, left outer canthus; O_2/A_1, right occipital lobe EEG referenced to left mastoid; ROC, right outer canthus. (From Butkov N. *Atlas of clinical polysomnography.* Ashland [Ore.]: Synapse Media; 1996, p. 224.)

polysomnogram may be preferable to using an esophageal catheter.[21]

In patients with *upper airway resistance syndrome*, the classic findings include progressively more negative pleural pressure until occurrence of a CNS arousal. The arousal is sometimes associated with an audible snort (Figure 164-6). After arousal, pleural pressure swings temporarily decrease until the next cycle begins, upon which pressure swings increase until another arousal occurs. Sometimes the CNS activation falls short of AASM duration criteria for arousal. However, the AASM's 3-second rule was established for determining reliability of spontaneous arousals using visual scoring.[22,23] On the electroencephalogram (EEG), signal changes of shorter duration are conceptually thought to represent CNS arousals, as are spectral analysis indices of alpha power loading, even when it is difficult to see on the raw data tracing.

Movement Detected by Static Charge and Pressure Sensors

The static charge–sensitive bed technology has been evaluated in sleep disorders.[24] The transducer is embedded within a thin mattress that responds to the slightest movement. Output from the bed is sensitive enough to detect heartbeat as a ballistocardiogram. Respiratory signal amplitude differs with body position changes; however, output is otherwise stable.

Figure 164-7 Synchronized digital video can be extremely helpful, as in this example in which a child with retrognathia slept with his neck arched and his mandible thrust forward (*arrow*). This sleep posture resulted in an unoccluded upper airway. The conventional polysomnography recording missed that a significant sleep breathing problem was present. (From Banno K, Kryger MH. Use of polysomnography with synchronized digital video recording to diagnose pediatric sleep breathing disorders. *CMAJ* 2005;173:28–30.)

A similar but newer technology uses piezoelectric sensors embedded in a strip placed perpendicular to the sleeper's body. Changes in pressure resulting from breathing movement and heartbeat can be analyzed.[25]

Movement Detected by Wave Technologies

Several systems approach movement detection using microwaves, radar, and/or changes in standing wave patterns in the bedroom. One approach directs a beam at the bed surface and analyzes the returning signal to evaluate the sleeper's upper body movement. In one such system, laser radiation can be used. Microwave and other radar-like technologies also can be applied. Additionally, a series of sensors could be used to monitor movement. Application of some of these emerging technologies to sleep medicine for evaluating SRBD is an important aspect of ongoing developments in the field.

Polysomnography-Synchronized Digital Video

Digital video is now a common feature of computerized polysomnography. Although recordings are widely used to evaluate parasomnias and seizures, they can be helpful in assessing SRBD. When the recording is properly synchronized with polysomnography, ambiguous and difficult-to-interpret tracings often become obvious. For example, it is easy to recognize a small dip in arterial oxygen saturation (Sao_2) as a significant sleep-disordered breathing event when it is followed by oxygen resaturation after an audible snort, a repetitive moving forward of the jaw, an arching of the neck, or a closing of a gaping mouth. Video is especially helpful in children, whose polysomnographic recordings may be difficult to interpret (Figure 164-7). An Sao_2 dip, without the other visual information, is quite likely to be missed, ignored, or dismissed as artifact. Video recordings are particularly useful in thin persons, who might not experience oxygen desaturation with their abnormal respiratory events. Furthermore, showing the sleep study video to patients can be very effective for promoting understanding of the problem and an appreciation of its severity.

MEASURING CHANGES IN LUNG VOLUME

Several methods can estimate tidal volume. These measures provide semiquantitative data concerning presumed airflow while also documenting respiratory effort. Calibration is

crucial in attempting to use these devices to gauge ventilation. Postcalibration movements, changes in body position, and shifting in placement of the recording device can introduce error. In sleep laboratories, strain gauges, inductance plethysmography, and impedance pneumography are sometimes used to measure volume changes. Other techniques include magnetometry, body plethysmography, canopy with neck seal, the barometric method, and pneumotachography; however, these are seldom used in a clinical setting.

Strain Gauges, Inductance Plethysmography, and Piezoelectric Transducers

In principle, length-sensitive devices can be used qualitatively to detect breathing abnormalities. If properly calibrated, these devices can be used quantitatively to measure dynamic volume changes.[26] Normally, the enlargement of the thorax and the outward movement of the abdominal wall occur together; that is, they are in phase. For a given change in lung volume, then, it is possible to quantify a change in rib cage volume and abdominal volume. For a given breath, the relative contributions of the rib cage and abdominal compartments also can be determined.

To quantify actual volume changes, the transducers must be calibrated against an independent volume-measuring system. In practice, two length-measuring devices are required for measuring changes in lung volume: one for the rib cage and one for the abdomen. The rib cage device is placed at the level of the axilla, and the abdominal device is placed just superior to the iliac crest. If it is assumed that the fractional contributions of the abdomen and the rib cage are constant, then changes in lung volume can be measured by calibrating transducers sensitive to rib cage and abdominal displacement. Once the transducers are calibrated, the sum of the rib cage and abdominal excursions will describe volume changes.

Unfortunately, the relative rib cage and abdominal contributions can change with posture and muscle tone occurring in sleep. Movement-related device migration from its original site and device deformability also must be considered. Such factors adversely affect calibration accuracy and stability. Nonetheless, calibrated inductance plethysmography appears to be sensitive enough to detect upper airway resistance syndrome events.[27]

Impedance Pneumography

Impedance defines the combined effects of two previously discussed properties of an electrical conductor: *resistance* and *inductance*. In physical terms, when impedance pneumography is used, the conductor is the thorax. Impedance is measured by applying a small current across the thorax using a pair of electrodes placed at the site of maximal thoracic excursion.

Transthoracic impedance changes reflect variations in the amount of conductive materials (liquids, including interstitial fluid, blood and lymph, and tissue) and nonconductive material (air) between the electrodes. The conductive and nonconductive materials affect the total impedance differently. Increased air in the lung increases impedance, and increased fluid in the thorax decreases impedance. A recording of the volume of air exchanged and the total of impedance changes can allow the differentiation between air-related and fluid-related changes in impedance. If total impedance is recorded in a single channel, changes related to air volume and fluid are measured.

Impedance alterations in obstructive apnea are complex. During apnea, lung volume decreases, whereas the negative intrathoracic pressure most likely temporarily pools blood in the pulmonary circulation. For these reasons, a precise measurement of respiratory volume and pattern may not be possible. Nonetheless, rate-adapting cardiac pacemakers using transthoracic impedance to drive ventilation have been used to screen for obstructive sleep apnea.[28]

MEASURING THE PHYSIOLOGIC CONSEQUENCES OF SLEEP-RELATED BREATHING DISORDERS

As previously mentioned, some SRBD events (e.g., hypopnea) are not intrinsically pathophysiologic; however, their consequences are. SRBD event consequences include oxyhemoglobin desaturations, carbon dioxide elevations, blood pressure changes, electrocardiographic abnormalities, and CNS arousals.

Oxygen and Carbon Dioxide Alterations

Routine clinical sleep evaluations require noninvasive, continuous, rapidly sampled measures to determine blood oxygen concentration. Thus directly measuring blood oxygen with an indwelling arterial catheter is ruled out on all counts. Pulse oximetry, however, serves the need well; consequently, it has become the standard technique for recording oxyhemoglobin desaturations during sleep. Pulse oximeters usually determine Sao_2 spectrophotoelectrically using a two-wavelength light transmitter and a receiver placed on either side of a pulsating arterial vascular bed (usually a finger, a toe, an ear, or the nose). Alternatively, in reflectance pulse oximetry, the light transmitter and the receiver are on the same surface. The light transmitted into the vascular bed is scattered, absorbed, and reflected. The amplitude of light detected in particular spectra by the receiver depends on the magnitude of the change in arterial pulse, the wavelengths transmitted through the arterial vascular bed, and the oxygen saturation of arterial hemoglobin (deoxygenated blood is bluer). These devices are sensitive only to pulsating tissues; consequently, venous blood, connective tissue, skin pigment, and bone theoretically do not hamper determination of Sao_2. Accurate measurement, however, requires a minimal pulse amplitude. Dyshemoglobinemias can cause problems.

Correct alignment of the light transmitter and receiver is critical to ensure measurement accuracy. If the sensor is applied to a digit, that digit must be immobilized. Significant bending of the digit can compromise detection of pulsatile flow and invalidate results. Although all pulse oximeters are based on similar technology, response characteristics differ both across manufacturers and even within a specific manufacturer's product line.[29] Sensor placement and device programming are crucial technical factors in obtaining optimal results. Reflectance oximeters also must contend with weaker pulse signals, because much less light is reflected back to the sensor. Differences in response characteristics and effect of sensor location cannot be overemphasized (Figure 164-8), because some oximeters appear to be completely insensitive to hypoxemia episodes clearly detected by other devices (Figure 164-9). Such devices can produce false-negative SRBD test outcomes.

A review of specific oximeters would be beyond the scope of this chapter; nonetheless, generalizations about (1) sensor location, (2) instrument filtering and sampling rates, and (3) potential pitfalls are worth noting here.

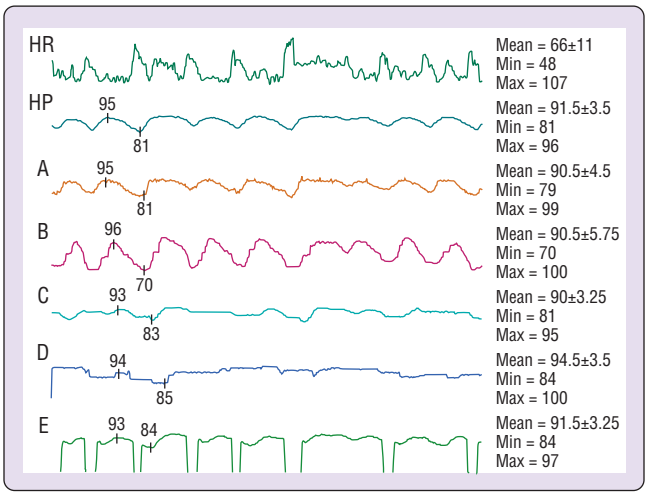


Figure 164-8 Heart rate (HR) and arterial oxygen saturation (Sa_{O_2}) during a 5-minute period in a patient with sleep apnea. The *top two traces* are for the HR monitor and a Hewlett Packard (HP) oximeter; A to E are traces for five different pulse oximeters. The scales for the six oximeters are identical. The *numbers* on the tracings represent the instantaneous Sa_{O_2} measured during the peak and trough of an apneic episode. The data listed to the *right* of the figure are the mean, standard deviation, and minimum and maximum values for HR and Sa_{O_2} for the six oximeters. Note that oximeters C and D do not track Sa_{O_2}, and that the recording for oximeter E has numerous artifacts.

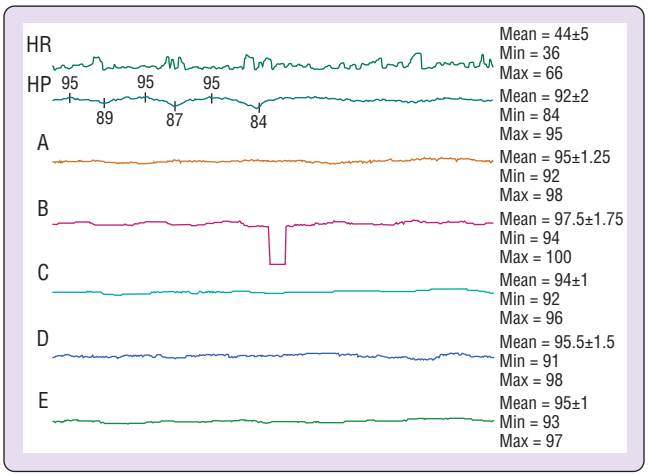


Figure 164-9 Heart rate (HR) and arterial oxygen saturation (Sa_{O_2}) during a 5-minute period in a patient with sleep apnea and bradycardia. Tracings are taken from six oximeters, as described in Figure 164-8. In this example, three apneic episodes are missed entirely by all of the pulse oximeters. This patient's problem would have gone completely undetected by pulse oximetry screening. HP, Hewlett Packard (oximeter).

Sensor Location

In our experience, the preferred sensor location in adults is the earlobe. Poor perfusion can be enhanced by applying a trace amount of vasodilator (e.g., nonylic acid vanillylamide plus nicotinic acid—Finalgon ointment [Boehringer Ingelheim, Ridgefield, Conn.]). Technicians must take care to avoid contact of the perfusion-enhancing agent with their own or the subject's eyes, because it is a powerful cutaneous vasodilator. When the ear site is not usable, reflectance pulse oximeter sensors placed on the forehead or another well-perfused surface will suffice. Recording from the ear also reduces circulator delay (compared with the finger). This advantage becomes especially important for associating a respiratory event with its subsequent oxyhemoglobin desaturation in patients with congestive heart failure.

Instrument Filtering and Sampling Rate

Most pulse oximeters filter the signal, and some filter algorithms use the heart rate. The degree of filtering is inversely related to heart rate; brief, mild hypoxemic episodes may therefore be missed during periods of very low heart rate, because they are filtered more strongly. During polysomnographic recording, filtering should be minimized (i.e., by setting the oximeter to the fastest response or the highest sampling rate, or both) to reduce missing transient oxyhemoglobin desaturation events.

Potential Problems

Because pulse oximeters use two wavelengths of light to estimate Sa_{O_2}, they cannot distinguish three or more hemoglobin species. In the presence of carboxyhemoglobin (as in heavy smokers, in whom carboxyhemoglobin levels can reach 10% to 20%), Sa_{O_2} is overestimated.[30] In the presence of a rising methemoglobin concentration, oximetry-determined Sa_{O_2} will plateau at approximately 85%, regardless of whether true saturation is much higher or lower.[31] Because light is transmitted through tissue, pigment in the skin can degrade oximeter performance, producing incorrect "probe-off" or "perfusion-low" error reports.[32] Some finger-clip devices can produce pressure-related injuries when worn for an entire night.

Partial pressure of arterial oxygen (Pa_{O_2}) estimated from the skin's surface depends on oxygen flux through the skin, local oxygen consumption, and the skin's diffusion barrier.[33] This measurement technique is most commonly used in neonates, whose skin is thin. Accurate measurement of transcutaneous P_{O_2} (P_{O_2}tc) requires maximal dilation of the local vasculature in the upper dermis. This is achieved by heating (to 43°C); however, heating shifts the oxyhemoglobin dissociation curve to the right, increases the resistance of the skin stratum corneum to oxygen permeation, increases the metabolic rate of the dermal tissue, and increases the rate of cutaneous blood flow. The shift in the oxyhemoglobin dissociation curve and the increase in metabolic rate effectively cancel each other out, leaving permeability and flow as the dependent factors in correlating P_{O_2}tc and Pa_{O_2}. An important advantage of heating is that the amount of blood present is maximal, and P_{O_2}tc is therefore unaffected by small changes in blood supply to the tissue.

The P_{O_2}tc may be misinterpreted when the state of blood flow is unknown. When flow and Pa_{O_2} are adequate, P_{O_2}tc reflects Pa_{O_2}. Under conditions of compromised flow and adequate Pa_{O_2}, P_{O_2}tc will change with flow. If Sa_{O_2} and flow are compromised, P_{O_2}tc tracks oxygen delivery. Transcutaneous measurement accuracy also depends on correct sensor application. To convert P_{O_2}tc measurements to Pa_{O_2} values precisely, a calibration curve for each subject is required, so this technique is too labor-intensive for use in routine clinical practice. In most laboratories, P_{O_2}tc serves to track, in relative terms, arterial oxygenation status. Device responsiveness is too slow for tracking rapid blood gas changes associated with brief SRBD events (less than 30 seconds), because oxygen diffuses slowly across the skin. The conditions governing transcutaneous Pc_{O_2} measurement are remarkably similar to those described for Pa_{O_2}. Although transcutaneous blood gas determinations are

of greatest value in neonates and young children, Pco_2tc is useful for assessing hypoventilation in adults.

Central Nervous System Arousals and Awakenings

Awakenings and CNS arousals provide critical information concerning sleep disturbance and fragmentations. Standardized scoring derives from electroencephalographic activity, preferably recorded from occipital sites. CNS arousals often take the form of bursts of alpha (7 to 13 Hz) activity, and the term *EEG speeding* is sometimes used to describe the event. When scored visually, a 3- to 15-second burst of alpha activity in non–rapid eye movement (NREM) sleep is considered an arousal. During rapid eye movement (REM) sleep, when alpha bursts can be an ongoing part of background activity, the alpha intrusion must be accompanied by increased muscle tone. In any sleep stage, a burst exceeding 15 seconds in duration is scored as an awakening.

The CNS arousal represents a clinically significant sleep parameter because it is associated with (in cross-sectional studies) and provokes (in intervention studies) tiredness, fatigue, and sleepiness. As a pathophysiologic consequence of obstructive sleep apnea, the arousal is thought to result from increased respiratory effort (and accompanying increase in autonomic nervous system sympathetic activation) triggered by increased airway resistance. Presumably, the CNS arousal terminates the SRBD event, because the arousal returns ventilation to voluntary control. As a result, the airway can be dilated and breathing resumes. However, some controversy exists about resumption of breathing after an obstructive event without CNS arousal and about the presence of arousals at the termination of central events.

Blood Pressure Changes

Clinical laboratories do not routinely record sleep-related blood pressure. Nonetheless, many pulse oximeters also can track pulse pressure (a magnitude index of pulsation occurring at the oximetry sensor site). In addition, pulse transit times (PTTs) have been used to indirectly estimate blood pressure.[34] PPT represents the time interval from heartbeat (ECG R wave) to its pulse recorded in the periphery. Negative pleural pressure provokes a blood pressure drop and in so doing lengthens PTT. The progressive increase in pleural pressure during obstructive apnea correlates with rising oscillations in PTT amplitude. During central apnea, this does not occur. Thus it has been suggested that PTT might provide an opportunistic estimate of inspiratory effort and thereby provide a means to differentiate obstructive from central apnea.[35]

Automatic self-inflating arm-cuff sphygmomanometers have long been available but are problematic for routine clinical use because they disturb sleep. By contrast, miniature finger cuff systems disturb sleep much less. Devices obtaining data in alternating fashion from adjacent fingers reduce finger injury risk, and some systems automatically perform hydrostatic correction when arm movement occurs. Finger flexion artifacts, however, remain a problem.

Cardiopulmonary Coupling

Cardiopulmonary coupling uses a single-channel ECG recording to extract signal features modulated by breathing.[36] Respiration induces small alterations in heart rate and amplitude variations in R-wave amplitude. Using computerized frequency domain analysis, a high-frequency component can be isolated that represents variations in breathing-induced vagal sinus pressure heart rate. A low-frequency component can be extracted that provides data about interbreath intervals. By examining a weighted composite of these data (increased variability and the envelope of amplitude change), time-domain periods (usually 2 to 10 minutes) with breathing pauses can be discerned by examining coherence and cross-power spectrums. In general, normal breathing loads high-frequency bands, and SRBD events mass in low-frequency troughs. For detailed information about cardiopulmonary coupling, see Chapter 166.

HOME SLEEP TESTING

In addition to standard attended polysomnography, guidelines for home sleep testing have been published.[37] Chapter 165 provides methodologic details. Because home sleep testing devices are less sensitive than laboratory polysomnography, the home test can rule in, but not rule out, SRBD.

To succeed diagnostically and economically, home sleep testing requires proper patient selection, appropriate portable recorder application, study interpretation by a qualified sleep specialist, readily available access to laboratory polysomnography when needed (as after a negative result on home sleep testing, or to investigate continuing problems despite treatment), and systematic follow-up assessments. Many health care coverage plans (including those administered by the Centers for Medicare and Medicaid Services) reimburse home sleep testing for diagnosing sleep apnea. Unattended studies are more prone to data loss as a consequence of uncorrected technical failures (e.g., electrode detachment) and patient tampering. Consequently, only patients with a high clinical suspicion for SRBD should be referred for HST.

CLINICAL PEARLS

- Detecting sleep-related breathing events and classifying each as obstructive, central, or mixed require measures of airflow, respiratory effort, and oxyhemoglobin saturation.
- Information about sleep-disturbing effects of SRBD events provides better sensitivity for the diagnosis of sleep-disordered breathing.
- Standardized guidelines developed by the AASM were developed to advance clinical practice and should not be construed as an impediment to advancing scientific inquiry.

SUMMARY

Evaluating patients with SRBDs is the most common application of laboratory and home sleep studies. Over the past four decades, some of the more frequently used procedures have evolved and become commonplace in clinical sleep laboratories. The American Academy of Sleep Medicine also published guidelines and a standards manual. Standard assessment for breathing disorders during sleep includes measures of airflow, respiratory effort, and oxyhemoglobin desaturation. Laboratory evaluation also includes sleep disturbance assessment. In this chapter we describe techniques to assess these and other, related physiologic activities. The underlying mechanism, advantages, and problems are also discussed.

Selected Readings

Berry RB, Brooks R, Gamaldo CE, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, Version 2.2. Darien (Ill.): American Academy of Sleep Medicine; 2015 <www.aasmnet.org>.

Iber C, Ancoli-Israel S, Chesson A, Quan SF. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Westchester (Ill.): American Academy of Sleep Medicine; 2007.

Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep* 2005;**28**(4):499–521.

Mitterling T, Högl B, Schönwald SV, et al. Sleep and respiration in 100 healthy Caucasian sleepers—a polysomnographic study according to American Academy of Sleep Medicine standards. *Sleep* 2015;**38**(6): 867–75.

A complete reference list can be found online at ExpertConsult.com.

Chapter
165

Home Sleep Testing

Thomas Penzel

Chapter Highlights

- Home sleep testing for diagnosing sleep-disordered breathing outperforms attended cardiorespiratory polysomnography. Diagnostic sensitivity and specificity are high enough for effective clinical application.
- Home sleep testing protocols have been established in terms of required signals. Monitoring usually includes respiratory effort and airflow, oxygen saturation, and body position/activity. Visual scoring is a necessary component. Clinical symptom assessment and home sleep testing should be combined to achieve high sensitivity and reliability.
- New developments are targeting use of fewer signals for diagnosing sleep-disordered breathing. Different systems achieve this with variable success. An economic benefit can be realized if sleep-disordered breathing can be diagnosed or managed using single-channel devices versus four- to six-channel devices, and such approaches currently are under investigation.

OVERVIEW AND BACKGROUND

Home sleep testing refers to portable monitoring for diagnosing sleep-disordered breathing. The term was introduced in the past decade. In 2014, a National Library of Medicine (NLM) PubMed database search for "home sleep testing" in all fields found 19 publications. The reference method for the diagnosis of sleep-disordered breathing is cardiorespiratory polysomnography. The recording technology and the scoring criteria used in home sleep testing are derived from this modality. Many studies on individual systems and a considerable number of evidence-based reviews indicate that home sleep testing for sleep-disordered breathing can be as specific and as reliable as sleep laboratory–based polysomnography recordings[1-3] in properly referred patients. Finally, adequate data are available to validate home sleep testing's use clinically, although it does have certain limitations.[4] A workshop consensus report presented the view of the participating medical societies (American Thoracic Society, American Academy of Sleep Medicine, American College of Chest Physicians, European Respiratory Society) on the use of this diagnostic procedure and provides directions for further research.[5] A clinical practice guideline from the American College of Physicians now recommends portable sleep monitors for diagnostic testing in patients suspected of having obstructive sleep apnea without serious comorbidity, and as an alternative to polysomnography when polysomnography is not available.[6] In Europe, home sleep testing has been widely used for several decades to diagnose sleep-disordered breathing.[7]

With respect to the guidelines and recommendations on home sleep testing, which are partially evidence-based, it is essential to consider the basis of the underlying studies used for evidence evaluation. The clinical studies evaluated for the guidelines usually are those performed by sleep centers on their patients, which correspond with clinical populations available in sleep centers.[8,9] Clinical populations differ from the general population in that the patients have been referred for evaluation for suspected sleep disorders. This selection leads to a high pretest probability for sleep disorders and for sleep apnea in particular.[1,10] Factors associated with this increased pretest probability include various physical examination measures and complaints reported by the patient or the bed partner, as follows:

- Loud and irregular snoring
- Observed or reported nocturnal cessation of breathing
- Excessive daytime sleepiness
- Nonspecific mental problems such as fatigue, low performance, or cognitive impairment
- Movements during sleep
- Morning dizziness, general headache, dry mouth
- Impaired sexual function
- Obesity
- Arterial hypertension and cardiac arrhythmias

A grading of the pretest probability factors could be useful to determine which patients may be more likely to suffer from sleep apnea or even from more severe sleep breathing disorder. This potential application has not been investigated in clinical practice.[11] Instead, clinicians commonly use a combination of validated questionnaires in conjunction with home sleep testing to confirm the suspected diagnosis.[12]

As revealed in a literature search, published reports are of several different types. Some papers describe new devices and comparative studies of such innovations with polysomnography. Reviews of data on existing devices are scarce. However, a good systematic review of data on existing systems that also provides categories for evaluation has been presented.[4] These categories are sleep, cardiovascular, oxygen saturation, position, effort of respiration, and respiratory flow—the SCOPER acronym. Sensors and systems are evaluated using these categories. A majority of published studies, however, focus on the role of home sleep testing in sleep apnea diagnosis. Some reports in the literature concentrate on general management

of patients with sleep apnea, whereas other reports focus on home sleep testing. Presented next is a short technical overview of available systems, followed by a discussion of home sleep testing with respect to requirements and special considerations.

HOME SLEEP TESTING WITH FOUR- TO SIX-CHANNEL SYSTEMS FOR DIAGNOSING SLEEP-DISORDERED BREATHING

Systems for diagnosing sleep-disordered breathing generally fall into one of four classifications defined in an American Sleep Disorders Association (ASDA) standard of practice guideline.[13]

- level I: attended cardiorespiratory polysomnography with at least 7 signals
- level II: cardiorespiratory polysomnography at home with at least 7 signals, unattended
- level III: unattended portable sleep apnea testing with at least 4 signals including airflow, respiratory effort, oxygen saturation, ECG, or heart rate or pulse rate
- level IV: unattended one or two signal recording, such as actigraphy or oximetry

Most diagnostic systems for home sleep testing attain level III device status and record four to six physiologic signals but do not record the electroencephalogram (EEG). Evidence-based home sleep testing reviews commissioned by health technology assessment agencies[3] revealed limited reliability. Up to 17% false-negative and between 2% and 31% false-positive findings of sleep apnea have been reported. These high error rates compared with those for polysomnography are considered unacceptable, and Ross and colleagues[3] concluded that portable monitoring of sleep apnea is not recommended. A few years later, studies with revised recording systems showed substantial improvement.[9,14] If systems incorporate a thoughtful selection of physiologic measures, have good signal acquisition, and use good signal processing technique, the number of false-positive diagnoses declines considerably.[1] When studies sampling from the general population are compared with studies using clinical populations, the importance of a high pretest probability becomes clear. A high pretest probability reduces the number of false-positive diagnoses. Altogether, the specificity increases enough to permit a conclusion that home sleep testing for sleep apnea can be recommended under certain conditions[4,6]:

1. Systems should be used only by certified sleep physicians based in certified sleep centers. This recommendation attempts to improve quality control and quality assurance. An interview of the patient and assessment of complaints should be conducted before home sleep testing is performed. This screening increases the pretest probability as explained earlier.
2. Home sleep testing for obstructive sleep apnea is recommended when no other comorbid pulmonary, cardiovascular, mental, neurologic, and neuromuscular disorder, or heart failure or another sleep disorder, is present. Other sleep disorders to rule out include central sleep apnea, periodic limb movement disorder, insomnia, circadian sleep-wake disorders, and narcolepsy.
3. In the earlier published studies, home sleep testing systems could not distinguish between central and obstructive sleep apnea events.
4. Home sleep testing for diagnosing sleep apnea needs to record oronasal airflow (using a thermistor or nasal pressure sensor); respiratory effort (using inductive plethysmography); oxygen saturation (with a short averaging period over few [3 to 6] pulses; pulse or heart rate; and body position.
5. Evaluation of the recordings should incorporate visual scoring of respiratory events using the same rules specified for polysomnography.[15,16] Editing of recorded events is necessary to remove artifacts occurring during the recording period. Furthermore, the visual scoring should be performed by trained personnel.
6. The technical specifications and sampling rates for the digital recording should be the same as specified in the evidence-based recommendation for cardiorespiratory polysomnography.[16]

Today, many devices fulfill the requirements necessary to meet level III device criteria. The home sleep testing devices all include pulse oximetry technology to record oxygen saturation and pulse rate. Many systems record oronasal airflow, which reflects intranasal pressure. Few systems still use thermistors for flow recording. Most devices record respiratory effort using either piezo sensors or respiratory inductive plethysmography. Some devices use one belt for rib cage movements, whereas others use two belts (for recording abdominal movements as well). Most systems record body position to identify positional apnea. Very few systems record the raw electrocardiogram (ECG), but many report heart rate derived by other means. A number of systems offer specific options to record signals in sleep apnea patients under therapy. The options are the recording of CPAP mask pressure in addition or as an alternative to other airflow signals. The signal will be split into a CPAP pressure reading, corresponding to set pressure level and to a respiratory flow reading which may be observed superposed to the CPAP pressure set. This may vary depending on pressure mode selection (e.g., bilevel or flex modes). This option is an important feature for using home sleep testing for treatment follow-up studies. Some systems allow an option for recording additional electromyogram (EMG) tibialis activity to detect leg movements; thus far, however, no systematic studies on this option's utility have been conducted.

It remains an open question whether home sleep testing can reliably diagnose periodic limb movement disorder. Similarly, a few systems can add electroencephalography channels for recording the sleep EEG, but no systematic studies have evaluated this option for its potential added diagnostic value or validated its sleep-related parameters. Nonetheless, many systems have been validated for clinical use with their basic signal setup together with their scoring and analysis software. Ingeneral, most systems show good performance, with some minor differences. No overall preference for one or another system emerges from published studies.

HOME SLEEP TESTING WITH ONE- TO THREE-CHANNEL SYSTEMS FOR DIAGNOSING SLEEP-DISORDERED BREATHING

Systematic reviews of home sleep testing for diagnosing sleep-disordered breathing have revealed that systems with one to three channels (pulse oximetry, long-term ECG, actigraphy, and oronasal airflow) are not suitable for routine diagnostic use. Specifically, these devices yield too many false-negative

(up to 17%) and too many false-positive (up to 31%) results.[3] This finding remained valid in the review by Collop and associates[1] and in a more recent systematic review using the new SCOPER criteria.[4] Therefore the application of these devices is not recommended for definitive diagnostic testing for obstructive sleep apnea, or to exclude the presence of obstructive sleep apnea.

Some of these devices, however, provide results in patients with severe sleep apnea that clearly suggest sleep-disordered breathing. Therefore high-quality recordings achieved with validated systems of this category can be used to increase the pretest probability before performance of cardiorespiratory polysomnography or even before four- to six-channel home sleep testing for sleep apnea.

Many technical innovations are currently emerging in this category of devices. A major challenge has been development of one- to three-channel devices that perform well and can diagnose sleep-disordered breathing. If reliable, such devices could facilitate diagnosis in new patient groups and provide tools for clinicians trained in other specialties with only basic knowledge of sleep medicine. Before initiation of therapy for sleep breathing disorders, however, a physician with a solid background in sleep-disordered breathing who is very familiar with the different treatment options should review the case.[17]

Described next are new technologies applicable with both one- to three-channel systems and four- to six-channel diagnostic devices.

NEW METHODS FOR HOME SLEEP TESTING

Different approaches are being explored for development of new technologies for diagnosing sleep-disordered breathing. Some approaches focus on developing new sensors to assess respiration for detecting breathing disturbances occurring during the night. Other technologies concentrate on assessing the patient's cardiovascular risk or sleep pathophysiology.

Assessment of Respiration

Several new sensors use surrogate signals to derive respiratory effort noninvasively. Some of these devices try to derive respiratory measures from direct respiration-related signals. These systems and concepts are discussed next.

A first-line approach entails recording respiratory airflow at the nose and the mouth. Usually these recordings are accompanied by pulse oximetry to determine oxyhemoglobin saturation. These simple screening devices provide a straightforward analysis for respiratory cessations. They even may distinguish obstructive from central respiratory events by analysis of flow limitation. Problems with obstructed nostrils, partial breathing through the mouth, blocked air tubes, and various artifacts pose logistical challenges to differentiating among the different types of apnea. Nevertheless, good validation studies are available,[18,19] with some limitations.

One approach tries to analyze respiratory sounds from the chest, with the goal of less obtrusive measures for detection of increased respiratory effort.[20] In other systems, respiratory sounds are recorded at the throat, and signal processing separates cardiac and movement signal first from breathing sounds and snoring. Together with oximetry, such recording quantifies respiratory measures, and snoring is tracked to detect respiratory cessations.[21,22]

Another approach involves recording midsagittal jaw movements based on magnetic distance determination.[23] A magnetic sensor is placed on the chin and another on the forehead to allow continuous determination of relative jaw movements. From this setup, it is possible to derive respiration and snoring. Analysis of this information is then used to detect respiratory events to diagnose sleep apnea.[24] By further analysis, a sleep wakefulness profile may potentially be estimated.[25,26] Combined with pulse oximetry and perhaps a cardiovascular parameter, this magnetic sensor–jaw movement detection feature is both simple and promising for clinical usefulness.

Pulse Wave Analysis

Many systems try to exploit the pulse wave on the finger or other peripheral sites. Such systems attempt to derive parameters from the pulse wave to assess cardiovascular event risk. The pressure wave may be detected with the photoplethysmograph already placed to measure oxygen saturation. In principle, this can be used to detect all forms of respiratory events[27] and cardiovascular event risk as associated with sleep apnea.[28,29]

Peripheral arterial tonometry[30] can be used to assess cardiovascular risk by measuring endothelial function during sleep-disordered breathing episodes. Arousals terminating sleep apnea events are accompanied by attenuated pulse amplitude. This decrease is due to peripheral vasoconstriction caused by sympathetic tone activation. If pulse rate also is analyzed, probability analysis can be used to distinguish between slow wave sleep and rapid eye movement (REM) sleep.[31] Several validation studies were published on use of the Watch-PAT, based on peripheral arterial tonometry, in patients with sleep apnea, with very good results.[29,32,33] A meta-analysis for this methodology is available and substantiates the diagnostic value of this device, even though it does not incorporate proximal sensors for effort and flow to record respiration, as recommended.[34]

Assessment of Electrocardiographic and Heart Rate Variability Parameters

ECG-derived respiratory parameters are very attractive for simple detection of sleep apnea owing to low costs and wide availability. To detect sleep apnea from the ECG alone does not require additional electrodes or additional hardware. The respiratory information is derived entirely from analytical software. This kind of analysis also could be performed retrospectively using previously recorded data. Sleep apnea is accompanied by a cyclic variation of heart rate, as already described many years ago.[35] Periodic changes in heart rate are related to the changes in sympathetic tone with apnea events.[36] Modern analysis of heart rate variability can satisfactorily derive cyclic variations of heart rate.[37,38] In addition, the morphology of the ECG wave itself is modulated by respiration. The derived respiratory curve—*ECG-derived respiration*[39]—correlates with respiratory effort and thus can be used to detect sleep-disordered breathing.[40,27] By combining ECG-derived respiration and sleep apnea–related heart rate variability, sleep apnea detection is possible.[41]

Electrocardiographic and Oximetry Assessments

A number of devices that use the ECG analysis techniques mentioned previously also try to link this approach to previous

techniques. Early on, pulse oximetry was applied (with limited success) for portable diagnosis of sleep apnea. Pulse oximetry alone has large diagnostic limitations in patients with arrhythmias or with additional lung diseases such as chronic obstructive pulomonary disease. Combining ECG-based sleep apnea analysis and oximetry is therefore a very promising approach.[42] An early study using pulse rate in addition to oximetry[43] could show that this improves the detection of sleep apnea. One retrospective study showed the advantage over pulse oximetry alone when combined with ECG analysis.[44] In that study, the ECG from a parallel polysomnography recording was evaluated. Based on these results, a combined long-term ECG recording system with oximetry was tested prospectively and provided very convincing results in terms of sleep apnea detection.[42]

MANAGEMENT OF HOME SLEEP TESTING IN A SLEEP CENTER SETTING

Many new studies show high reliability of home sleep testing in detecting sleep apnea.[19] A number of open research questions needing clarification concerning the conditions and restrictions for using portable monitoring are now being addressed in recent studies.[1,5] The important parameters are no longer technical limitations but often study limitations such as the preselection of patients. The inherent screening process corresponds with the characteristic high pretest probability of sleep apnea. This aspect prohibits the use of portable monitoring as a screening tool to exclude sleep apneas, such as in professional drivers and people with supervision tasks (in which symptoms and complaints have not been assessed and may conflict with employability or other issues).

Diagnostic and therapeutic approaches to management of sleep-disordered breathing differ among countries worldwide.[45] The development of sleep medicine in some countries is very advanced. In others, economically affordable strategies constitute the primary consideration.[46,47] In certain settings, sleep medicine may be very basic, with only home sleep testing available for diagnosing sleep apnea.[45] One potential reason for this restriction to home sleep testing alone is the limited availability of sleep medicine centers owing to unmet needs for qualified experts and funding for polysomnography beds. This is the case in countries in which sleep medicine is a young discipline. A second reason is that even in countries with a long history of sleep medicine, and with enough sleep centers and polysomnography beds, an economic decision may be made to limit access to such studies to patients with comorbid illnesses, and those with regular sleep apnea are diagnosed with home sleep testing alone. With improvements in the knowledge base for sleep disorders and sleep-disordered breathing among general physicians, the individual clinician can decide whether a particular patient should be evaluated for suspected sleep apnea alone or exhibits some comorbidity or has other risk factors. Then the patient can be referred for either home sleep testing or cardiorespiratory polysomnography. This approach would allow economic and thoughtful management of patients with respect to diagnosis and subsequent treatment, as appropriate.[46] In Germany a debate was initiated on different levels of sleep medicine service that would include different levels of medical expertise and correspondingly different levels of equipment complexity. Family physicians may have some basic knowledge about sleep-disordered breathing and already sometimes apply simple tests. A limited number of clinical centers would have clinical expertise, training, research, and other technical know-how as required in specialized sleep centers.[47] Many community-based centers may have basic sleep medicine knowledge and home sleep testing with four- to six-channel systems.

The other issue is health economy and patient care. A threshold regarding sleep apnea severity and assessing risk still remains to be established. How many apneas, how many hypopneas, what duration of apnea events, how much sleep fragmentation, or what degree of hypoxia represents substantial cardiovascular risk with increased mortality? How much increase in mortality justifies treatment with continous positive airway pressure (CPAP) or another lifelong therapy? In view of limited therapeutic adherence, how strict should researcher-clinicians be with respect to treatment follow-up studies? As the field of sleep medicine approaches the point at which diagnosis can be done easily with home sleep testing, a need is emerging for new clinical and economic decisions key for developing new strategies in managing sleep apnea.

Patients may be diagnosed and even treated at home. One home sleep testing study showed that the 4-week outcome in sleepiness and CPAP adherence was similar to that for sleep laboratory–based diagnosis and treatment.[48] An important limitation of the study was the short follow-up period.[49] Sleep-disordered breathing is a chronic condition, and long-term adherence with CPAP therapy may decline more at home. Thus more research is needed.

CONCLUSIONS

Attended cardiorespiratory polysomnography is the reference standard for diagnosing disordered breathing during sleep. Evidence-based literature, however, indicates that diagnosis of obstructive sleep apnea can be performed using home sleep testing under certain conditions in adults. The recording must include oxygen saturation, airflow, respiratory effort, heart or pulse rate, and body position. The SCOPER parameters summarize these requirements in a comprehensive and quantitative scheme.[4] Visual evaluation is needed to avoid misclassification of sleep apnea severity.[15] It is not possible to distinguish between central and obstructive respiratory events with certainty. Home sleep testing is reliable if it is performed under the supervision of personnel trained in sleep medicine and if screening has been adequate to achieve a high pretest probability among the study subjects of suffering from sleep-disordered breathing. In addition, the patients should not have other significant sleep or comorbid disorders (e.g., heart failure, stroke, diabetes mellitus, obstructive or restrictive lung diseases, or severe cardiac arrhythmias).

Home sleep testing systems with fewer channels can indicate the likelihood of sleep-disordered breathing but are not sufficiently validated for diagnostic purposes. Currently, these systems can point to the probable need for home sleep testing. Technological advances can be expected to improve these systems. Accordingly, it may well be that in the near future, systems with fewer channels may provide a sufficiently reliable diagnosis for disordered breathing during sleep. To prove the advanced usefulness of new technologies, good clinical studies, with sufficient sample size and testing of the new modality against a reference standard, are needed.

Technologic advances need to be accompanied by economy-driven strategies to diagnose and treat patients with sleep apnea. Recent approaches to diagnosing and even treating patients at home seem to provide effectiveness, in terms of outcome, similar to that for sleep laboratory–based studies. Economically proven home-based studies may be more feasible than sleep laboratory–based studies in light of the high prevalence of the disorders and the still-unmet clinical needs to recognize and treat patients with sleep-disordered breathing.

CLINICAL PEARLS

- Home sleep testing has been used for the portable diagnosis of sleep-disordered breathing for many years worldwide.
- Home sleep testing has been thoroughly compared against cardiorespiratory polysomnography.
- Sensitivity and specificity of home sleep testing are sufficient for a diagnosis if adult patients have a high pretest probability for sleep-disordered breathing.
- Home sleep testing requires monitoring of airflow, respiratory movement, oxygen saturation, pulse rate, and body position.
- Use of systems with fewer signals shows promising diagnostic results, but such streamlined systems require further validation.

SUMMARY

Home sleep testing is a well-validated technique to diagnose sleep-disordered breathing outside of the sleep laboratory and other clinical settings. Different devices are available. These devices usually record respiratory flow, respiratory effort, oxygen saturation, pulse or heart rate, and body position and/or activity. Some of these signals may be recorded indirectly and derived. The diagnostic sensitivity and specificity have been validated against those of polysomnography and generally show good agreement in patients with a high pretest probability for sleep-disordered breathing. Visual evaluation is needed to prevent misclassification of sleep apnea severity. Overall, the available evidence indicates that home sleep testing should be used in combination with a clinical assessment for factors and symptoms associated with sleep-disordered breathing. In view of the high prevalence of sleep-disordered breathing, methods to diagnose sleep apnea in a sufficiently accurate and economic way are urgently needed. Today, home sleep testing to assess sleep apnea is indispensable. In addition, home sleep testing offers a method for conducting therapy follow-up studies that may improve therapy compliance. Home sleep testing is developing an ever-more important role in the management of patients suffering from sleep-disordered breathing.

ACKNOWLEDGMENTS/DISCLOSURES

Several companies have furnished equipment for research or otherwise supported the work on which this chapter is based. Itamar Medical (Caesarea, Israel) provided sensors/consumables for research with the Watch-PAT device. Neuwirth Medical (Obernburg am Main, Germany) provided equipment to the Sleep Center for Research Studies. Cidelec (Sainte-Gemmes-sur-Loire, France) provided a grant for a validation study of its Pneavox system for laryngeal pressure and sound recording. Nomics (Liege, Belgium) provided a grant for a validation study of the Brizzy system. Somnomedics (Randersacker, Germany) provided Somnowatch equipment to the Sleep Center for Research Studies. Weinmann GmbH (Hamburg, Germany) provided Somnocheck equipment to the Sleep Center for Research Studies. Nox Medical (Reykjavik, Island) provided equipment and sensors for testing. Studies were done at the Interdisciplinary Sleep Medicine Center at Charite-Universitätsmedizin Berlin.

Selected Readings

Aurora RN, Swartz R, Punjabi NM. Misclassification of OSA severity with automated scoring of home sleep recordings. *Chest* 2015;**147**:719–27.

Berry RB, Gamaldo CE, Harding SM, et al. *The AASM Scoring Manual*, Version 2.2. Updates: new chapters for scoring infant sleep staging and home sleep apnea testing. *J Clin Sleep Med* 2015;**11**(11):1253–4.

Collop N. Home sleep testing: appropriate screening is the key. *Sleep* 2012;**35**:1445–6.

El Shayeb M, Topfer LA, Stafinski T, et al. Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: a systematic review and meta-analysis. *CMAJ* 2014;**186**:E25–51.

Gozal D, Kheirandish-Gozal L, Kaditis AG. Home sleep testing for the diagnosis of pediatric obstructive sleep apnea: the times they are a changing …! *Curr Opin Pulm Med* 2015;**21**(6):563–8.

Kuna ST, Badr MS, Kimoff RJ, et al. An official ATS/AASM/ACCP/ERS Workshop report: research priorities in ambulatory management of adults with obstructive sleep apnea. *Proc Am Thorac Soc* 2011;**8**:1–16.

Pack AI. Sleep medicine: strategies for change. *J Clin Sleep Med* 2011;**7**:577–9.

Pereira EJ, Driver HS, Stewart SC, Fitzpatrick MF. Comparing a combination of validated questionnaires and level III portable monitor with polysomnography to diagnose and exclude sleep apnea. *J Clin Sleep Med* 2013;**9**:1259–66.

Qaseem A, Dallas P, Owens DK, et al. Diagnosis of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2014;**161**:210–20.

Skomro RP, Gjevre J, Reid J, et al. Outcomes of home-based diagnosis and treatment of obstructive sleep apnea. *Chest* 2010;**138**:257–63.

A complete reference list can be found online at ExpertConsult.com.

Cardiopulmonary Coupling Sleep Spectrograms

Robert Joseph Thomas

Chapter 166

Chapter Highlights

- Cardiopulmonary coupling is a technique that uses heart rate variability as a measure of autonomic drive, in addition to respiration, to generate sleep spectrograms. The electrocardiogram contains a convenient set of signals for extracting both HRV and tidal volume fluctuations measured as R-wave amplitude fluctuations. Cardiopulmonary coupling analysis supports a bimodal (having only two distinct types) of non–rapid eye movement (NREM) sleep, rather than the conventional graded classification.
- High-frequency (0.1 to 0.4 Hz) coupling is a biomarker of integrated, stable NREM sleep, which is characterized by stable breathing, high vagal tone, generally a non–cyclic alternating pattern on the electroencephalogram, high relative delta power, and blood pressure dipping. This state may be considered to be "effective" NREM sleep. Effective sleep enables the normal functions of sleep, across multiple dimensions (e.g., neuronal network health, metabolic), such that spending periods in this state enables recovery and restoration processes.
- Low-frequency (0.1 to 0.01 Hz) coupling is a biomarker of integrated unstable NREM, with exactly opposite features: low-frequency tidal volume fluctuations, cyclic variation in heart rate, cyclic alternating pattern, electroencephalogram, low relative delta power, and stable (non-dipping) blood pressure. This state may be considered "ineffective" NREM sleep. Ineffective sleep fails to accomplish the normal functions of healthy sleep. Very-low-frequency (0.01 to 0.001) coupling is seen in the wake state and during healthy REM sleep; fragmented REM sleep has low-frequency coupling characteristics.
- A subset of low-frequency coupling, termed elevated low-frequency narrow-band coupling, identifies sustained periods of central apneas and periodic breathing.
- Sleep-fragmenting stimuli increase low-frequency coupling, and sleep-consolidating stimuli increase high-frequency coupling as a percentage of sleep, thereby allowing dynamic tracking of sleep physiology and pathology in health and disease.

In humans, the conventional scoring systems divide sleep into rapid eye movement (REM) and non-REM (NREM) types. NREM sleep is subdivided into grades of depth. These classifications are more than 50 years old and were developed at a time when the understanding of neurobiology was limited, modern brain imaging did not exist, and computational resources did not apply network concepts to biologic states. The primary features of REM sleep hold up well with modern neurobiologic circuit analysis. Stage N3 is known to have desirable biologic associations, such as blood pressure dipping.[1] By contrast, the current NREM sleep classification suffers from a major limitation in that deep and restorative "slow wave sleep" (once stage III and IV but now N3) makes up only a fraction of sleep time. Sleep restriction experiments clearly show that even minor reductions in total sleep time have adverse consequences, suggesting that other components of the sleep state must have restorative functions. It is equally plausible that the N3 stage does not capture all NREM sleep with restorative properties. A further limitation of the conventional staging system is that N3 decreases across the life span and may be "normally absent" after the age of 60 to 70 years even in persons who have excellent subjective and objective sleep quality (other than the absence of N3). The recognition and description of the network dynamics of slow (frequency of less than 1 Hz) oscillation as the basic underlying rhythm of NREM sleep[2] and the concept of coalescing multisite local sleep processes[3,4] remain largely unintegrated into clinical practice. Moreover, sleep is a multiphysiology state with specific changes in respiration, heart rate, and blood pressure, besides slow waves on the electroencephalogram (EEG).

The biologic role of NREM sleep associated with low delta power is unclear. Restricting such periods produces adverse consequences similar to those of total sleep deprivation, including sleepiness and metabolic dysregulation.[5-8] Electroencephalographic delta frequency activity (usually 0.5 to 4 Hz) is considered a biomarker of homeostatic sleep drive.[9] These slow waves are the scalp surface representation of a highly complex ensemble of oscillatory activity during NREM sleep, including the cortical "up" and "down" states that constitute the slow, less-than-1-Hz oscillation.[10-12] Delta power, as a proportion of total EEG power, is highest during the initial cycles of NREM sleep, decreases across the biologic night, and shows rebound effects after a period of sleep

deprivation.[13,14] The preponderance of slow wave (less than 1 Hz in frequency) oscillation is far greater in the first than in the final NREM period.[15] The "biologic worth" of stage N2, however, may not be fully accounted for by conventional scoring or by absolute delta power profiles. These disparities are especially apparent in persons older than 40 to 50 years of age, in whom stage N3 makes up less than 20% of the sleep period.[16]

Alternate characterizations for NREM sleep are available. The most extensively researched is cyclic alternating pattern (CAP). In essence, CAP proposes that NREM sleep has two fundamental modes of expression, one with phasic activity predominating and another relatively devoid of phasic activity.[17] Sleep-fragmenting stimuli increase CAP, whereas sleep-consolidating stimuli increase non-CAP. Heart rate variability analysis shows periods of NREM sleep with especially high vagal tone and strong sinus arrhythmia,[18-20] that is not restricted to stage N3 itself but is correlated with delta power. Sleep apnea researchers have long recognized periods of stable breathing that "come and go" during NREM sleep.[21,22] Figure 166-1 shows this "bimodality" of NREM sleep, in recordings from the same subject, within the same hour of sleep, in the same body position. Finding an appropriate term to describe these patterns of NREM sleep is challenging, but one proposal is for "effective" and "ineffective" NREM sleep, denoted by $NREM_E$ and $NREM_{IE}$, respectively.

Analysis of coupled sleep oscillations offers a complementary approach to sleep state estimations from a single signal. In theory, any two or more signals from physiologic subsystems during sleep may be coupled. Examples are blood pressure plus respiration, heart rate variability plus respiration, EEG plus heart rate variability, and so on. The cardiopulmonary coupling approach described here uses heart rate variability and electrocardiogram(ECG)-derived respiration, the latter being the R-wave amplitude fluctuations with tidal volume changes during respiration.

THE CARDIOPULMONARY COUPLING ELECTROCARDIOGRAPHIC SLEEP SPECTROGRAM

The cardiopulmonary coupling technique is based on a continuous ECG signal and uses the Fourier transform to analyze two features of the signal: (1) heart rate variability and (2) the

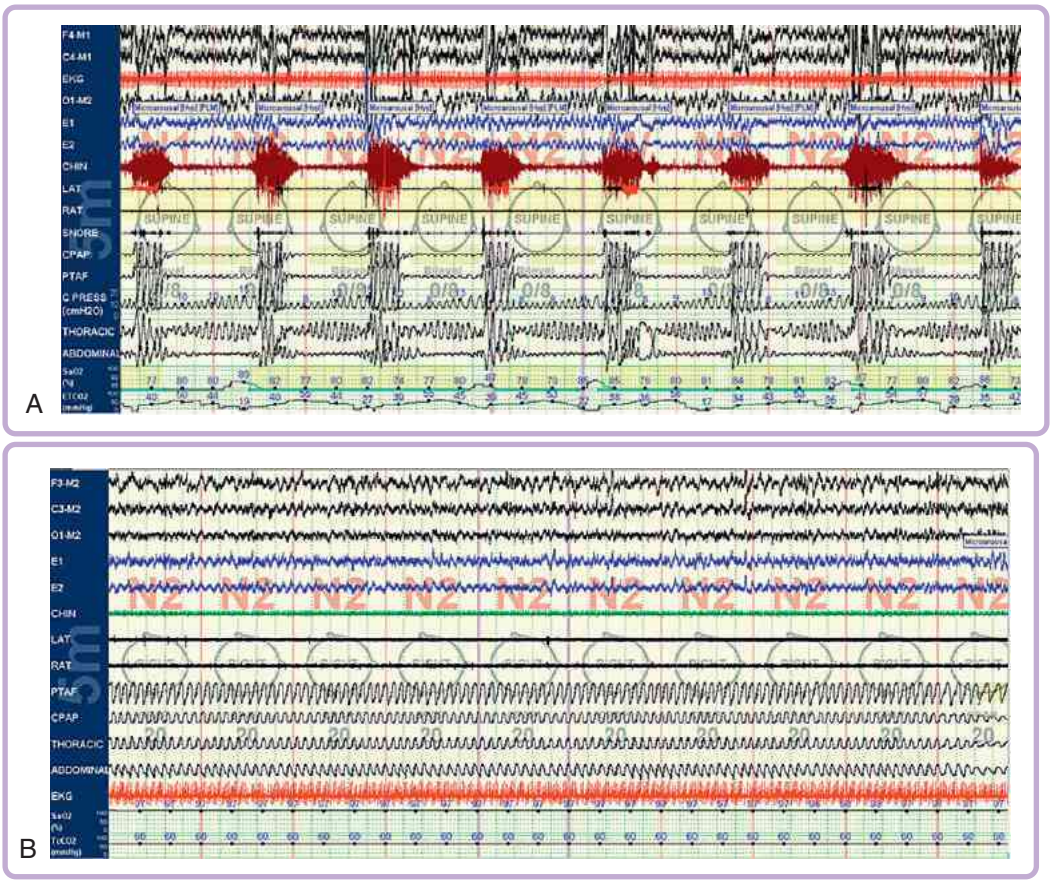


Figure 166-1 A, $NREM_{IE}$ in stage N2, 5-minute compression. Sleep apnea (in this case, poorly responsive to positive-pressure ventilation) readily brings out the features of $NREM_{IE}$. This snapshot shows temporally synchronized oscillations of EEG, EMG (chin), respiration (PTAF, C PRESS, THORACIC, ABDOMINAL), and oxygenation. **B, $NREM_E$ in stage N2, 5-minute compression.** This recording, from the same subject as in **A,** shows successful sleep apnea treatment, manifested by stable breathing, paucity of phasic EEG activation (non-CAP), and prominent sinus arrhythmia (not readily seen at this compression). Note prominent breathing amplitude modulation of ECG R-wave amplitude, which is one of the input signals for the cardiopulmonary analysis technique. C PRESS, Positive airway pressure device pressure trace; CAP, cyclic alternating pattern; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; $NREM_{IE}$, non–rapid eye movement sleep, ineffective; PTAF, pneumotachograph airflow.

fluctuations in R-wave amplitude induced by respiration.[23] These signals tend to have two basic patterns: a *high-frequency component* due to physiologic sinus arrhythmia reflecting breath-to-breath fluctuations and a *low-frequency component* reflecting cyclic variation across multiple breaths. Quantifying cardiac and respiratory interactions involves calculating the cross-power (measures the common power of the two signals at different frequencies) and coherence between these two signals.

The steps involved in the calculation of cardiopulmonary coupling are as follows:

1. An automated beat detection algorithm is used to detect beats, classify them as either normal or ectopic, and determine amplitude variations in the QRS complex. From these amplitude variations a surrogate ECG-derived respiratory signal (EDR) is obtained.
2. A time series of normal-to-normal sinus (N-N) intervals and the time series of the EDR associated with these N-N intervals are then extracted from the R-R interval time series.
3. Outliers due to false or missed R-wave detections are removed using a sliding window average filter with a window of 41 data points and rejection of central points lying outside 20% of the window average.
4. The resulting N-N interval series and its associated EDR are then cubic spline resampled at 2 Hz.
5. The cross-spectral power and coherence of these two signals are calculated over a 1024-sample (8.5-minute) window using the fast Fourier transform applied to the three overlapping 512-sample subwindows within the 1024-sample coherence window.
6. The 1024-sample coherence window is then advanced by 256 samples (totaling 2.1 minutes) and the calculation repeated until the entire N-N interval/EDR series is analyzed. For each 1024-sample window the product of the coherence and cross-spectral power is used to calculate the ratio of coherent cross power in the low-frequency (0.01 to 0.1 Hz) band to that in the high-frequency (0.1 to 0.4 Hz) band.

A preponderance of power in the low-frequency band tends to be associated with periodic sleep phenomena, whereas predominance of power in the high-frequency band is associated with respiratory sinus arrhythmia and sleep with stable respiration and EEG. High power in the very-low-frequency (0 to 0.01 Hz) band is associated with periods of wakefulness or REM sleep. This technique thus generates a moving average of the dominant oscillatory frequencies of autonomic drive coupled with respiration during sleep (Figure 166-2).

The discovery of this technique was largely serendipitous. Software developed to detect apnea fortuitously detected stable and unstable NREM periods. It was noted that these periods were clearly demarcated and did not associate strongly with conventional NREM stages but showed better correlation with CAP/non-CAP.[23] Sleep spectrogram analysis reveals that NREM sleep has a distinct bimodal-type structure marked by alternating and abruptly varying periods of strong high- and low-frequency cardiopulmonary coupling intensity, respectively. Much of high-frequency coupling occurs during stage N2, especially with the EEG non-CAP morphologic pattern. This pattern is associated with periods of stable breathing, a paucity of phasic EEG transients, and physiologic blood pressure dipping and is reduced in sleep apnea and fibromyalgia.[24,25] High-frequency coupling tags periods of stable breathing and repeated drops in blood pressure during sleep. The phenomenon of blood pressure dipping occurs exclusively during such periods (Figure 166-3).

Cortical slow wave kinetics can impact autonomic and respiratory function. Slow oscillation–like activity has been recorded in downstream neural elements (as reviewed by Vyazovskiy and Harris[26]), including the hippocampus, cerebellum, thalamus, basal ganglia, and even the locus ceruleus.[27] Increased slow wave activity after sleep deprivation in cats was reported in subcortical structures such as the hippocampus, amygdala, hypothalamus, nucleus centralis lateralis of the thalamus, septum, and caudate nucleus and the substantia nigra.[28] Cortical slow wave activity is likely to strongly modulate neuronal functions in "lower" brain centers and networks. This modulation of activity may in turn provide feedback to the cortex to enhance generation of sustained periods of high-density slow oscillations. Brainstem centers that induce slow wave sleep (e.g., subparafacial zone[29]) or arousals (e.g., parabrachial complex[30]) are potential targets for top-down modulation. Moreover, an increased probability of a stable breathing period enabled by the slow oscillation could, in turn, reduce arousing respiratory afferent stimuli, thus enhancing the likelihood of undisturbed and sustained slow-oscillation dense periods. These multicomponent and multisynaptic loops can stabilize the sleep state during a period of effective sleep.

The relationship between EEG delta power and cardiopulmonary coupling was tested using the Sleep Heart Health Study dataset. This analysis showed that slow wave power fluctuations positively correlate with high-frequency cardiopulmonary coupling. Furthermore, it identified stage N2 periods with potentially similar physiologic characteristics as N3 (Figure 166-4).[31] The key findings were as follows:

- Delta power measured from the surface EEG correlated with ECG-derived cardiopulmonary coupling high-frequency power, further supporting a link between cortical EEG electrical activity and brainstem-related cardiorespiratory functions.
- Normalized delta power provided improved correlation over that based on absolute delta power.
- A consistent lag (median of approximately 4 minutes) was observed between the start of the high-frequency power increase in relation to delta power increase.
- Correlations were reduced but still highly significant in the second half of the night compared with the first half.
- Age effects appeared to be small, with correlations being reduced only in the 80 years and older age group.
- Arousals tended to reduce the strength of the correlations.

This correlation between delta power and high-frequency cardiopulmonary coupling is consistent with strong "top-down" modulation of autonomic and cardiorespiratory activity. Delta power is positively correlated with high-frequency heart rate variability power—fluctuations of delta power across the night are associated with temporally linked heart rate variability power changes.[20,32-34] High-frequency heart rate variability, related to respiratory sinus arrhythmia, is a marker of a vagal-tone dominant state. The apparently paradoxical 4-minute lag between the onset of high-frequency power and delta power increases was unexpected but has several potential explanations. First, the lag could be an "artifact" of the filtering properties of the skull. Second, directly recorded thalamic

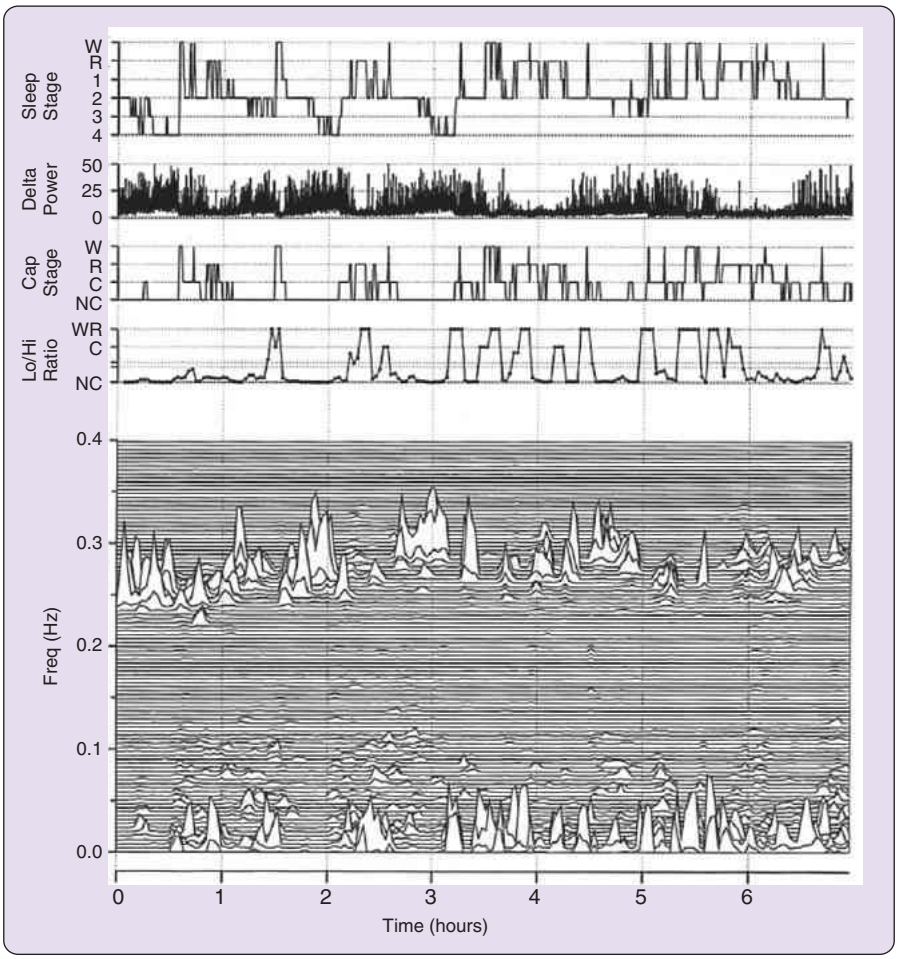


Figure 166-2 Cardiopulmonary coupling analysis in a 22-year-old healthy woman. The *four top panels* show (*top to bottom*) conventional sleep stage scored in 30-second epochs, second-by-second delta power from the C4-A1 electroencephalogram (EEG) montage (μV2/Hz), EEG-based manual cyclic alternating pattern (CAP) scoring, and the ratio of low-frequency (0.01 to 0.1 Hz) to high-frequency (0.1 to 0.4 Hz) coherent cross power (Lo/Hi ratio) used to detect sleep state. The *bottom panel* shows the cardiopulmonary coupling spectrogram across 7 hours of sleep in which the magnitude of the coherent cross power at each frequency is indicated by the height of the peak. The sleep spectrogram reveals spontaneous switching between high-frequency and low-frequency coupled states, represented by the two distinct bands of spectrographic peaks. Throughout the night, occurrence of cycles of increased delta power and high-frequency coupling that correlate with non-CAP sleep continues. Body position was supine throughout the study. C, CAP; NC, non-CAP; R, rapid eye movement sleep; W, wake.

spindles can appear before simultaneously cortical surface recordings, such that the lag could reflect a subcortical-thalamic sleep generation process that then entrains the cortex, followed in turn by increasing cortical influences on thalamic and downstream processes as sleep progresses.[35-37]

DETECTION OF STRONG RESPIRATORY CHEMOREFLEX ACTIVATION

Conventional scoring of central apneas and periodic breathing is impeded under certain conditions: The airway may be closed during "central" events, and flow limitation often is seen in otherwise typical periodic breathing. The two pathophysiologic mechanisms, obstructive (either upper airway collapsibility or a poor negative pressure response) and high loop gain, with or without a low arousal threshold, can coexist.[38-42] Alternative approaches to quantify excessive respiratory chemoreflex activation (high loop gain) include recognizing dominance in NREM sleep, measuring the CO_2 reserve with bilevel ventilation, pressure "dial-down," and use of proportional assist ventilation. These methods generally perturb stable NREM sleep and assess stability and instability characteristics.[40] Conventional scoring of central apneas and periodic breathing assesses unstable components of sleep respiration. An extension of respiratory instability analysis is possible by determining the spectral dispersion of single or coupled signals, the logic being that if the respiratory chemoreflex is driving the abnormality, metronomic self-similar oscillations will dominate. The cardiopulmonary coupling technique has been used to generate such spectral dispersion metrics and to quantify strong chemoreflex influences, without needing to consider an individual respiratory event's exact morphology (including flow limitation, described further on).

Cardiopulmonary Coupling Analysis and Estimation of Elevated Low-Frequency Coupling Subtypes

Analysis of the PhysioNet Sleep Apnea Database (http://www.physionet.org/physiobank/database/apnea-ecg/) using

the cardiopulmonary coupling technique showed that elevated power in the low-frequency coupling region coincided with periods of scored apnea/hypopnea. Optimal detection thresholds required that the minimum low-frequency power be greater than 0.05 normalized unit and that the low- to high-frequency ratio be greater than 30 to define periods of probable apnea-hypopnea, which has been termed *elevated low-frequency coupling* (e-LFC). Because the apneas and hypopneas in this database were scored in 60-second epochs and cardiopulmonary coupling measurements were made every 2.1 minutes, 60-second linear interpolation between consecutive 2.1-minute measurements was done. The 70 recordings in this database contained a total of 34,243 minutes, of which 13,062 (38%) were scored as containing episodes of apnea/hypopnea. Sensitivity and specificity rates for minute-by-minute apnea detection were calculated for a range of low-frequency coupling powers and low-frequency/high-frequency coupling ratios. Receiver-operator curves were then calculated, and the threshold giving the maximum combined sensitivity and specificity for apnea/hypopnea detection was selected as optimal. Thus e-LFC is defined here as a subset of low-frequency coupled cardiopulmonary oscillations, periods of which correlated significantly with periods of manually scored apneas and hypopneas in the PhysioNet Sleep Apnea Database.

Cardiopulmonary coupling analysis was modified to detect central apnea and periodic breathing. The PhysioNet Congestive Heart Failure Database (http://physionet.org/physiobank/database/chfdb/) was used, and a majority of samples were

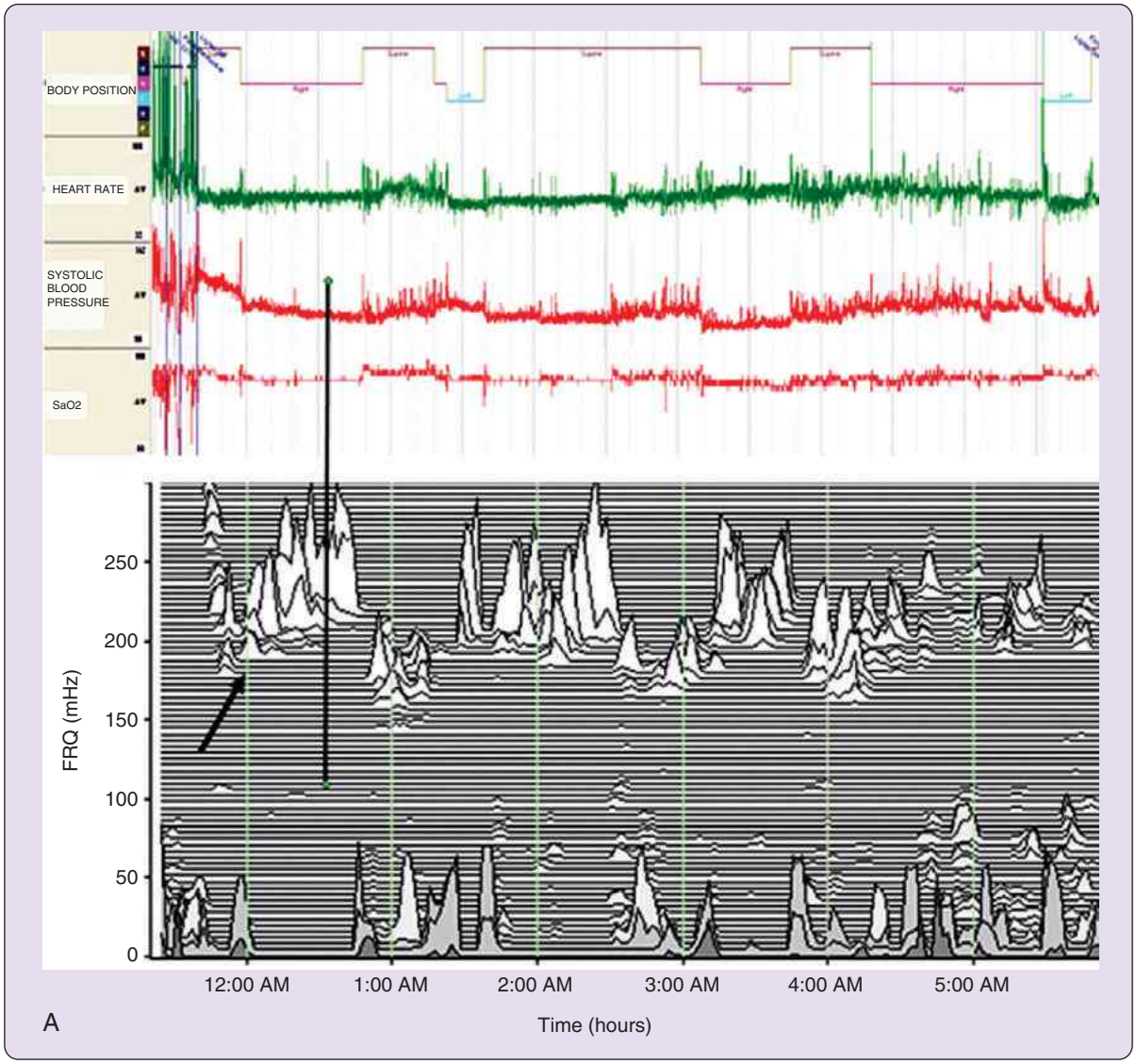


Figure 166-3 A, Blood pressure dipping and high-frequency coupling. Beat-to-beat systolic pressure, derived from pulse transit time, showing the simultaneous (*vertical line*) occurrence of blood pressure drop during sleep with a period of high-frequency coupling (*arrow*). This combined set of features occurs intermittently through the night. Standard ambulatory blood pressure, with measurements obtained on the half-hour or at 1-hour intervals, cannot capture this dynamic.

Continued

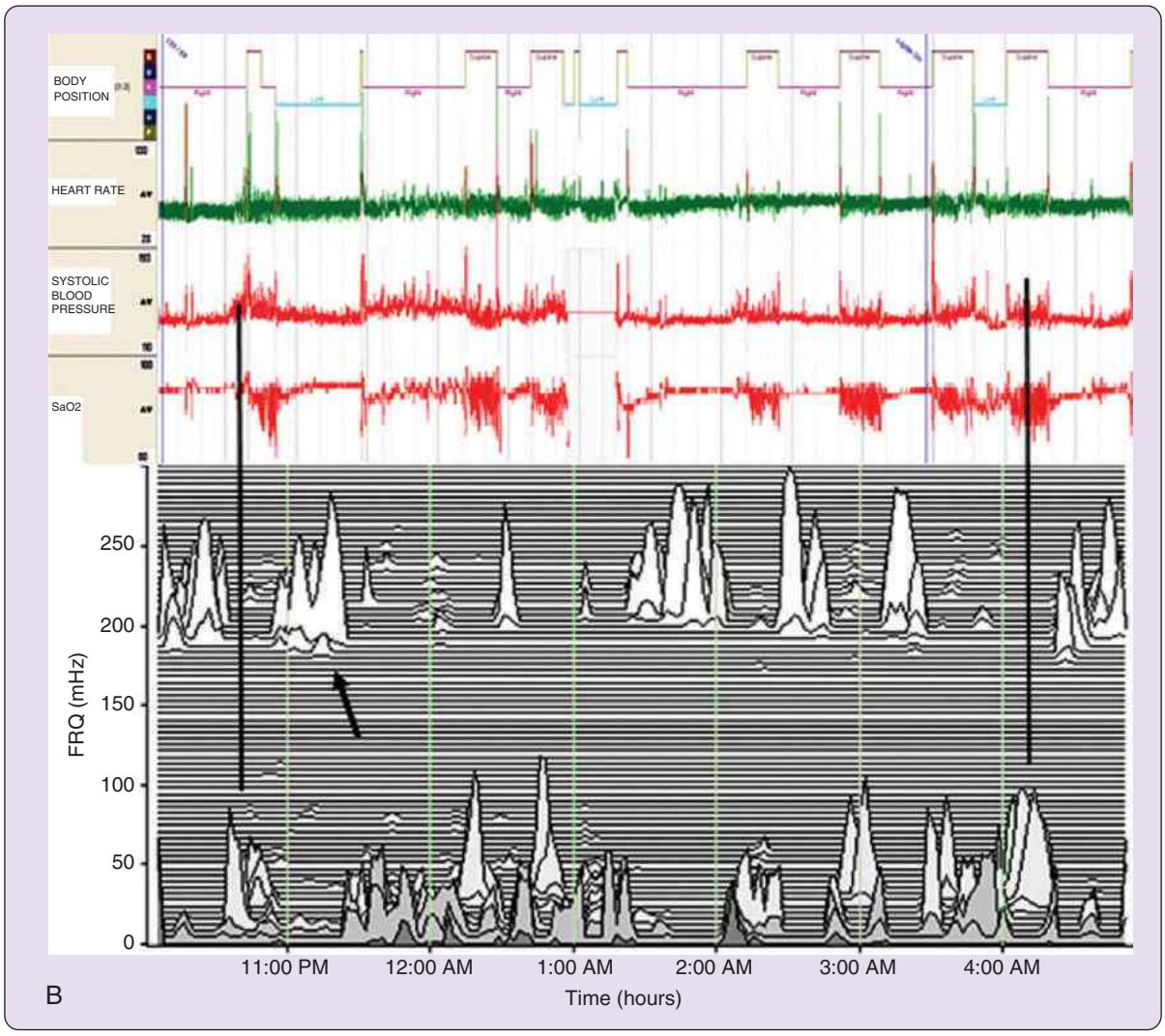


Figure 166-3, cont'd B, Blood pressure non-dipping or reverse dipping and low-frequency coupling, periods of stable breathing, and high-frequency coupling in a subject with sleep apnea. Note rise of blood pressure (*vertical lines* showing temporal concordance) during a period of low-frequency coupling; also note stable oxygenation, reflecting stable breathing, during periods of high-frequency coupling (*arrow*).

from subjects with Cheyne-Stokes respiration. To detect metronomic self-similar oscillations, which are characteristic of Cheyne-Stokes respiration, the researchers required that the coupling frequency of each pair of consecutive measurements remain within 0.0059 Hz of each other over five consecutive sampling windows, identifying narrow-band e-LFC.[43] The presence of e-LFC_{NB} increases the likelihood of emergent central sleep apnea[43] and is associated with increased hypertension and stroke risk.[25]; e-LFC_{NB} also has been found to have heritable characteristics.[44] Figure 166-5 shows the distinctive spectral dispersion characteristics of broadband and narrow-band coupling. Because the ECG spectrogram can be generated using ambulatory wearable devices, point-of-care tracking of sleep apnea and of the emergence or resolution of central sleep apnea and periodic breathing during therapy is possible (Figure 166-6).

SLEEP SPECTROGRAM APPLICATION TO DISEASE AND PATHOPHYSIOLOGY

ECG-derived sleep quality metrics show the following:

- The metrics are independent of absolute EEG amplitudes and thus are not constrained by the "loss" of slow wave sleep with age.[23]
- Specific spectrographic signatures of fragmented sleep are biomarkers of strong chemoreflex effects on sleep respiration.[43]
- Metrics are associated with EEG-defined stable and unstable sleep (CAP and non-CAP)[23] such that high-frequency coupling correlates with non-CAP/stable sleep. Blood pressure dipping occurs only during periods of EEG-defined non-CAP.[45]
- High-frequency coupling sleep is reduced in depression,[46] heart failure,[47] and fibromyalgia syndrome.[48]

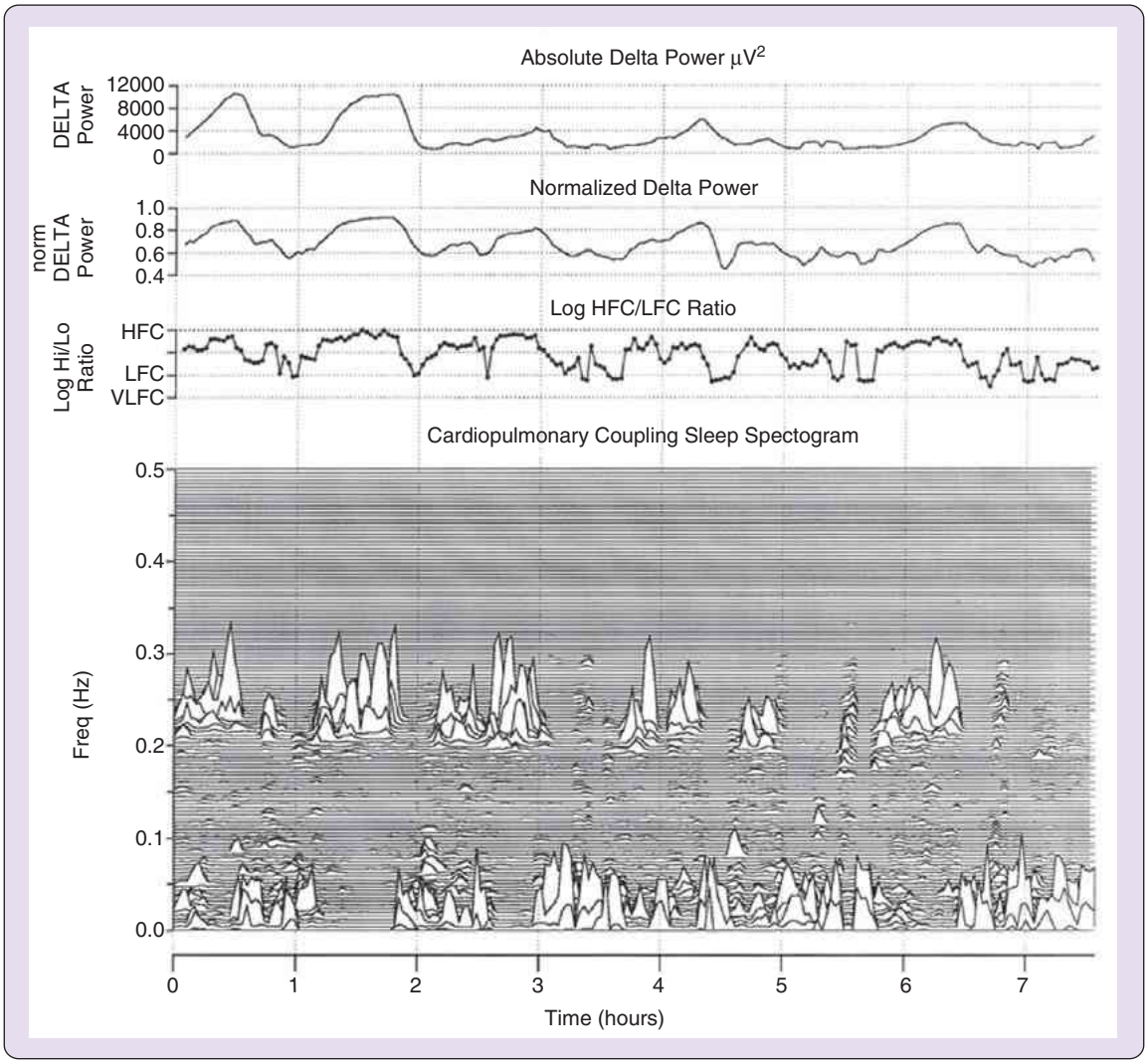


Figure 166-4 Correlations of EEG delta power (0 to 4 Hz) power and high-frequency coupling power. Delta power (0 to 4 Hz), cardiopulmonary coupling, and their cross-correlations for a representative subject in the Sleep Heart Health Study database. Absolute delta power of 0 to 4 Hz (μV^2). Note the higher absolute delta power in the first half of the night compared with the second half. Delta power (0 to 4 Hz) normalized to total EEG power. Note that the relative delta power in the first half and and that in the second half of the night are of relatively equal maximal magnitude. The logarithm of the ratio of high-frequency to low-frequency cardiopulmonary coupling (HFC/LFC). Note the correspondence between delta power fluctuations and the cardiopulmonary coupling ratios. The cardiopulmonary coupling sleep spectrogram. The cross-correlations between absolute and normalized delta power and high-frequency coupling in this subject were $r = 0.61$ and 0.75, respectively. EEG, Electroencephalogram; HFC, high-frequency coupling.

- High-frequency coupling is an independent determinant of the glucose disposition index.[49]
- Pre- and posttreatment effects in sleep apnea with use of an oral appliance[50] or upper airway surgery[51] are captured by changes in the ratio of high-frequency to low-frequency coupling.
- These metrics are improving the current understanding of subjective versus objective sleep characteristics in insomnia.[52,53]
- Certain metrics allow assessment of the impact of sleep fragmentation–related stress effects (e.g., leukocyte telomere lengths).[54]

AMBULATORY ASSESSMENT OF CARDIOPULMONARY COUPLING

The ease of acquisition of the ECG allows cardiopulmonary coupling to be assessed from any continuous ECG signal source, be it a formal polysomnogram or a continuous ECG monitor used for detection of cardiac arrhythmias. The software has been developed commercially through the SleepImage system (MyCardio LLC, Broomfield, Colo.). The first such device is called the M1 and records the ECG, body position, and trunk actigraphy (www.sleepimage.com). Five nights of recording is possible with one battery set, enabling

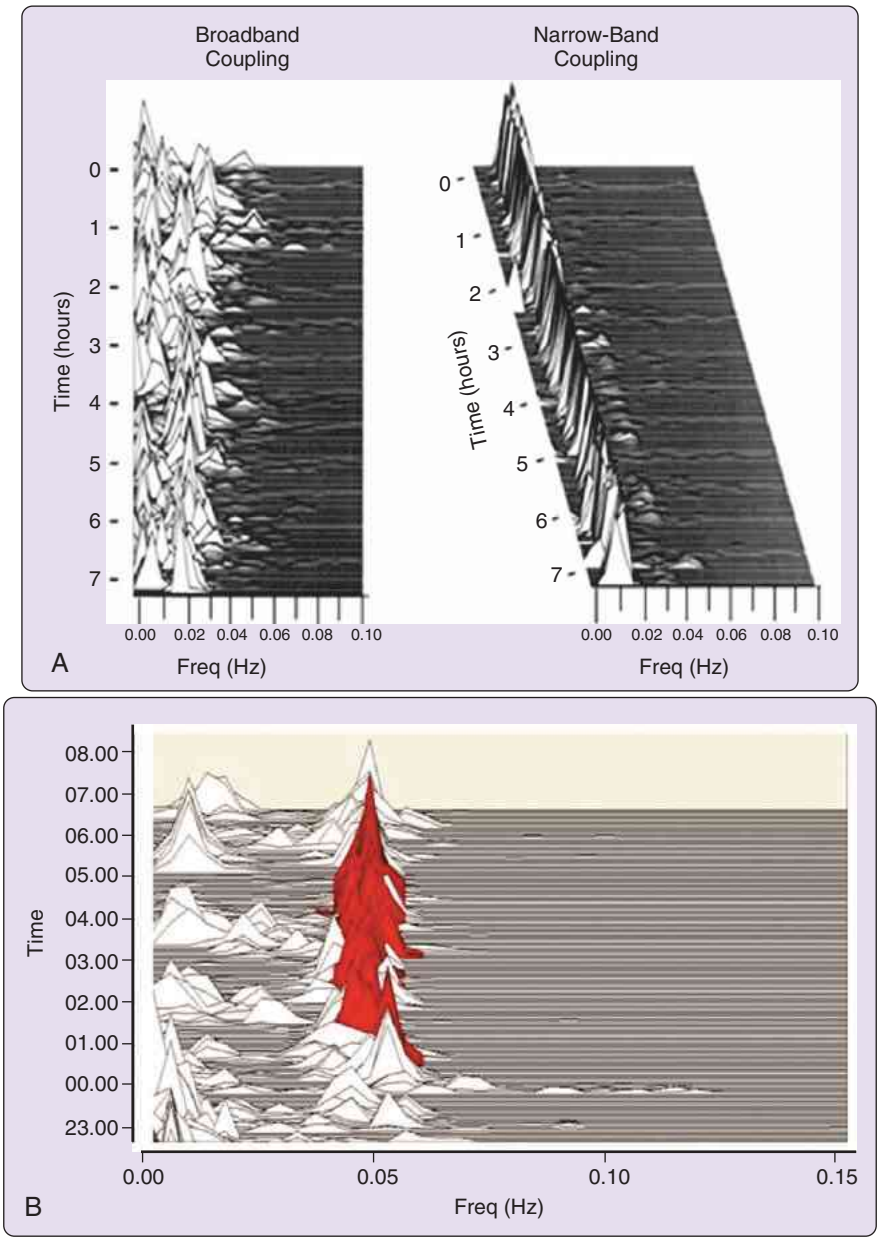


Figure 166-5 A, Broadband *(left)* and narrow-band *(right)* low-frequency coupling. Note the variable and tight dispersion of coupling frequencies across the night. The differences are visually recognizable and mathematically quantifiable, as percentage of analysis period with elevated narrow-band low-frequency coupling (e-LFC_{NB}). **B,** Ambulatory detection of e-LFC_{NB} on continuous positive airway pressure (CPAP) therapy.

assessment of night-to-night variability and minimizing the effects of such variability by averaging. The percent of the actigraphic sleep period in high-frequency, low-frequency, and very-low-frequency coupling and in narrow-band and broadband e-LFC is calculated through a cloud-based system. As the number of wearable devices proliferates, any continuous ECG may be used to generate the ECG spectrogram. Analysis can be readily accomplished using a smartphone application format. The night-to-night stability of the signals is high, with the intraclass coefficient of high-frequency coupling over 14 consecutive nights of recording in the range of 0.7 to 0.8 (unpublished data). In a given state or stage of sleep and body position, sleep physiology or pathology is relatively stable. The proportions of these "combinations," however, interact with time-of-night effects to provide mean values for a night.

CLINICAL PEARL

Cardiopulmonary sleep spectrograms suggest that NREM sleep is bimodal, with concordant and predictable sets of signal (and thus physiologic) characteristics from multiple sleep subsystems. Because sleep-fragmenting and sleep-consolidating stimuli move the proportion of high-frequency and low-frequency coupling in predictable directions, both absolute values and ratios can be tracked over time to measure sleep quality. The detection of strong respiratory chemoreflex activation using the narrow-band coupling biomarker allows tracking of the dynamics of sleep respiratory control in sleep apnea management.

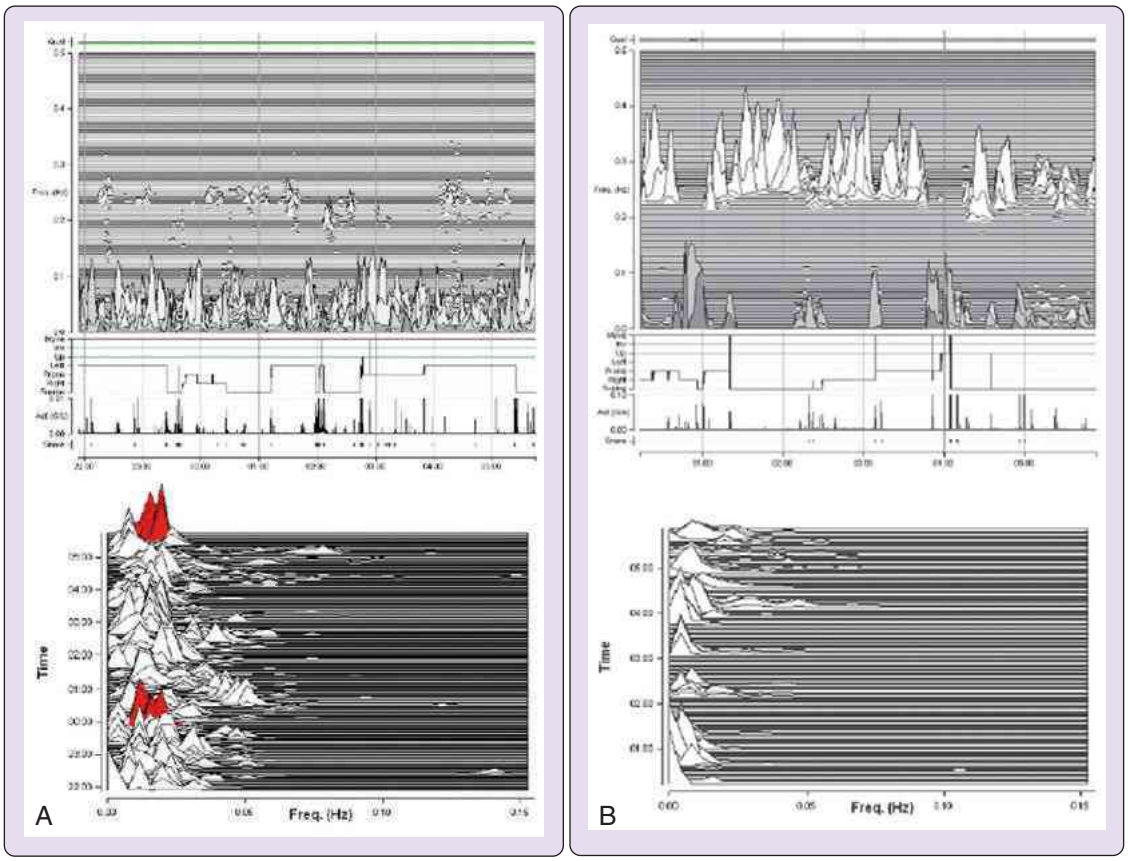


Figure 166-6 **Ambulatory tracking of cardiopulmonary coupling using the M1 device.** Both snapshots show body position and activity transients; actigraphy allows analysis to be restricted to the actigraphic sleep period. An actigraphic fragmentation index, sleep efficiency, and total sleep time also are calculated. **A, Failing CPAP.** Note low high-frequency coupling, predominance of low-frequency coupling, and two periods of elevated narrow-band low-frequency coupling (e-LFC_{NB}), suggesting treatment-persistent or emergent/complex sleep apnea. **B, Successful CPAP.** Note predominance of high-frequency coupling periods. CPAP, Continuous positive airway pressure.

SUMMARY

Mapping coupled oscillations during sleep provides new insights into sleep physiology and pathology. Both intrinsic brain oscillations and those driven by influences outside the brain (such as respiratory control) can sculpt these signals. The ECG is a readily available and repeatable study and is increasingly available through mobile technology, allowing tracking of long-term dynamic features of sleep. From a physiologic standpoint, the readout of the cardiopulmonary coupling ECG spectrogram strongly suggests that NREM sleep has spontaneously switching bimodal characteristics or modes. In one mode, desirable sleep features dominate, including high vagal tone/sinus arrhythmia, blood pressure dipping, high slow wave power, and stable breathing. In another mode, generally less desirable features dominate, such as cyclic variation in heart rate, absence of blood pressure dipping, tidal volume fluctuations (with sleep apnea of a degree exceeding clinical thresholds), and lower delta power. These features probably reflect the integrated output of network activities of multiple sleep subsystems.

Selected Readings

Choi JH, Thomas RJ, Suh SY, et al. Sleep quality change after upper airway surgery in obstructive sleep apnea: electrocardiogram-based cardiopulmonary coupling analysis. *Laryngoscope* 2015;**125**(7):1737–42.

Lee WH, Hong SN, Kim HJ, et al. A comparison of different success definitions in non-continuous positive airway pressure treatment for obstructive sleep apnea using cardiopulmonary coupling. *J Clin Sleep Med* 2016. [In press].

Thomas RJ, Mietus JE, Peng CK, et al. Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method. *Sleep* 2007;**30**:1756–69.

Thomas RJ, Mietus JE, Peng CK, et al. Relationship between delta power and the electrocardiogram-derived cardiopulmonary spectrogram: possible implications for assessing the effectiveness of sleep. *Sleep Med* 2014; **15**:125–31.

Thomas RJ, Mietus JE, Peng CK, Goldberger AL. An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep. *Sleep* 2005;**28**:1151–61.

Thomas RJ, Weiss MD, Mietus JE, et al. Prevalent hypertension and stroke in the Sleep Heart Health Study: association with an ECG-derived spectrographic marker of cardiopulmonary coupling. *Sleep* 2009;**32**: 897–904.

A complete reference list can be found online at ExpertConsult.com.

Chapter
167

Pulse Wave Analysis During Sleep

Ludger Grote; Ding Zou

Chapter Highlights

- The beat-to-beat digital pulse wave signal can be monitored by various recording techniques (e.g., photoplethysmography from pulse oximetry, peripheral arterial tone). Characteristics of the finger pulse wave form are modulated by skin sympathetic nerve activity and hemodynamic variables such as stroke volume, blood pressure, and central arterial stiffness.
- There has been increasing interest in using information embedded in the finger pulse wave signal alone or in combination with other physiologic signals to detect autonomic activation, sleep stage, and sleep-disordered breathing. Sleep diagnostic devices based on finger pulse wave signal analysis have been integrated in clinical routine procedures. They are recognized by the American Academy of Sleep Medicine for home sleep testing (e.g., peripheral arterial tone technology).
- Additional information on cardiovascular reactivity, beyond that obtained by classic diagnostic devices, can be assessed using finger pulse wave signal. Pulse rate variability, blood pressure, arterial stiffness measures, and signs of vascular aging can be detected from the pulse wave contour. One novel approach involves overall cardiovascular risk assessment using several parameters derived from the oximeter-based photoplethysmography signal. The clinical relevance of pulse wave derived cardiovascular parameters needs further validation in larger prospective outcome studies.

INTRODUCTION

We spend approximately one third of the 24-hour cycle during sleep, and sufficient and restorative sleep is essential for both physical and mental health. Consequently, the diagnosis of sleep-related dysfunction is of major clinical interest, and several recent developments are aimed to simplify and improve the methodology of sleep studies. Polysomnography (PSG) is the gold standard to quantify sleep time, to differentiate sleep stages, and to assess sleep fragmentation. In addition, respiratory and motor dysfunction can be assessed in the context of sleep. However, technical demands, required technical skills, and high costs may limit the widespread use of PSG in daily clinical practice. Consequently, efforts have been made to simplify and improve sleep diagnostic methods. On the other hand, additional dimensions of sleep, like autonomic activity and cardiovascular reactivity, derived from pulse wave analysis, may be of importance for health-related outcomes like quality of life or survival, when compared with the classical PSG outcome variables.[1]

Pulse wave signals can be obtained in different vascular compartments like the ear lobe, the limbs, or any other suitable vascular bed. However, photoplethysmography (PPG) in the finger is the state-of-the-art technology in sleep diagnostics, and we will mainly address finger as the measurement site for pulse wave analysis.

THE PHYSIOLOGY OF THE DIGITAL VASCULAR BED

Like in other extremities, the finger skin vascular beds are rich in arteriovenous anastomoses amounting to approximately 500/cm^2 in the fingernail beds. Arteriovenous anastomoses are coiled vessels with thick, muscular, and densely innervated walls connecting the arterioles and venules in the dermis. Blood from the digital artery bypasses the high-resistance arterioles and the capillaries of the papillary plexus, flows directly through the dermal arteriovenous anastomoses, and returns to the deep plexus of veins. This special feature enables large variations of digital skin blood flow, which ranges from 1 to 90 mL/minute/100 mL of tissue.[2] The skin vascular bed accounts for the majority of total digital blood flow.

The finger vascular bed is highly innervated. Despite local factors affecting finger blood flow, the microcirculation in the digital skin vascular bed is mainly controlled by the systemic vasoconstrictor tone. A high correlation between increased peripheral sympathetic nerve activity and reduction of finger pulse wave amplitude (PWA) has been shown during temperature changes.[3] Vasodilation is induced by elevated ambient temperature, sedation, and vasoactive drugs like nitroprusside. Vasoconstriction is induced by sympathetic excitation (e.g., stress, pain), cold, and vasoactive drugs such as noradrenalin and ephedrine.[4]

Autonomic and cardiovascular regulation relevant for pulse wave analysis during sleep is complex, sleep-stage dependent, and regionally differentiated. Non–rapid eye movement (NREM) sleep is associated with reduced sympathetic and increased parasympathetic activity when compared with wakefulness. Rapid eye movement (REM) sleep, in this aspect, is similar to wakefulness. Peripheral vascular smooth muscle sympathetic nerve activity measured by microneurography is reduced by approximately 50% during NREM sleep (stage 4) and doubled during REM sleep compared with wakefulness.[5,6] Regional differences in sympathetic output during sleep may also exist. For instance, sympathetic activity in vasoconstrictor fibers of limb skeletal muscle is increased in parallel with reduced output to the splanchnic, cardiac, lumbar, and renal vascular beds in a pharmacologic model of REM sleep.[7] Skin blood flow measured by laser Doppler showed a clear increase during sleep compared with wakefulness possibly due to thermolytic vasodilation.[8]

Quiet NREM sleep is characterized by general hemodynamic stability. However, arousal from sleep creates a significant change; the most pronounced changes can be noticed in the acceleration of heart rate combined with a peripheral vasoconstriction resulting in a marked elevation of systolic and diastolic blood pressure.[9] The autonomic response to an arousal from sleep, when started, appeared to follow a stable pattern over time. The amplitude of the cardiovascular reactivity is positively associated with the degree and the duration of the electroencephalographic (EEG) arousal. In some cases, a typical autonomic activation during sleep can be observed without an activation pattern in the cortical EEG, the so-called autonomic arousal. In this chapter, we are going to focus on different pulse wave analysis methods during sleep to assess autonomic arousal, sleep stages, sleep-disordered breathing, and cardiovascular function and risk.

METHODS FOR THE ASSESSMENT OF THE DIGITAL PULSE WAVE

Plethysmography is a well-established method for the quantitative assessment of volume changes over time and has been frequently used in humans for the quantification of respiratory and hemodynamic function. A number of different noninvasive plethysmographic techniques are available using water, air, strain gauge, impedance, or PPG to quantify volume changes of the peripheral pulse wave. For studies on finger microcirculation, additional methods like radioisotope clearance, capillaroscopy, and laser Doppler flowmetry have been used. In sleep medicine, the PPG and the peripheral arterial tone (PAT) are the most frequently used methods to assess the digital pulse wave. Recent developments in home sleep diagnostics have incorporated these technologies.

Water or Air-Based Plethysmography

The historic development of research related to the cardiovascular system was started with water or air-based plethysmography. The arm, finger, or leg is placed in a tightened compartment filled with water or air. Any pulsatile change in volume is transferred on paper or, more recently, to a digital signal in order to quantify blood flow and vascular resistance at rest and after certain type of provocations. For sleep-related assessments, this technology is less feasible due to obvious interference with sleep and artifacts during body movements.

Strain Gauge Plethysmography

The strain gauge technique for the in vivo assessment of blood flow has been used for decades. An elastic mercury in Silastic strain gauge is placed around the circumference at the area of interest (e.g., the forearm, the finger, or the leg). When using the venous occlusion technique, pulsatile volume increases of the compartment are recognized by changes in the circumference. The calibrated signal gives an exact measure of blood flow in milliliter per second. When blood pressure is monitored, even resistance of the vessels can be calculated. Characteristics of the pulse wave contour are similar when using strain gauge or digital PPG with a high correlation ($r = 0.9$).[10] The strain gauge technique has been widely used to quantify vascular reactivity after local arterial infusion or systemic application of vasoactive drugs. The concept of vascular and endothelial dysfunction in obstructive sleep apnea (OSA) patients was introduced by the findings of impaired responses to vasodilatory or augmented response to vasoconstrictor agents in normotensive and hypertensive OSA patients when compared with controls using venous occlusion plethysmography.[11,12]

Photoplethysmography

PPG measures the absorption/reflection of infrared light emission (wavelength around 940 nm) in the applied microvascular bed by a photosensor. The degree of light absorption/reflection correlates directly with the blood volume changes in the catchment, and a continuous pulse wave signal can be derived. This pulsatile component of the pulse wave is called the *AC signal component*. In addition, the baseline of the pulsatile wave may vary, and this component is described as the *DC component* and reflects to average blood volume and finger tissue.[13] This DC component may be affected by respiration, autonomic nerve activation and vasomotor activity, Traube-Hering-Mayer waves, hypovolemia, and thermoregulation.

The information embedded in the pulse wave can be divided in several aspects (Table 167-1; Figure 167-1). First, the systolic amplitude is correlated with the pulse volume increases in the finger compartment. The amplitude is modulated by both the cardiac stroke volume and the degree of skin sympathetic activity (high activity means reduced flow and low amplitude).[3] Using a daytime infusion protocol, PWA, derived from the finger plethysmography, was sensitive to the α–receptor agonist norepinephrine but not to the β_2–receptor agonist isoproterenol.[14] Second, the peak-to-peak interval of two consecutive pulse waves is considered as a surrogate measure of the electrocardiogram (ECG)-based RR interval and can be used for pulse rate calculation (see following sections). Third, the diastolic point of the pulse wave reflects the time taken for the pressure waves to pass from the heart to the sites of reflection at small arteries in the trunk/lower limbs and back to the upper limbs. This interval, also called *pulse propagation time* (PPT), has been proposed as a surrogate marker of pulse wave velocity and arterial stiffness.[15,16] A shortening of PPT indicates rigid and atherosclerotic vessels. The augmentation index (AI), calculated as the ratio between the systolic and the diastolic amplitude, is another way to mirror central arterial stiffness. Both PPT and the digital AI have been shown good agreement with the AI derived from radial artery or the pulse wave velocity calculated along the aorta and applied for assessment of atrial stiffness and

Table 167-1 Different Variables Derived from the Digital Pulse Wave Signal

Pulse Wave Parameter	Function Assessed	Modified by Dysfunction or Disease
Systolic pulse wave amplitude	Digital pulse volume, reduced by skin sympathetic activation (vasoconstriction), elevated by vasodilating drugs, correlated with stroke volume and blood pressure	Arousal from sleep, REM sleep, sleep-disordered breathing, hypertension, atherosclerosis, and metabolic disease
Pulse to pulse interval	Pulse rate and pulse rate variability	Bradycardia/tachycardia, arrhythmia (e.g., atrial fibrillation), ischemic heart disease, marker of central/autonomic arousal from sleep, baroreflex
Time between systolic and diastolic peak of the pulse wave (e.g., stiffness index, pulse propagation time)	Central and peripheral vascular wall compliance and stiffness	Atherosclerosis, vascular aging, hypertension, diabetes mellitus
Time between R wave of the ECG and systolic peak of the peripheral pulse wave (e.g., pulse transit time)	Marker of central arterial stiffness, correlation with blood pressure, sympathetic activation	Central or autonomic arousal, hypertension
Respiratory-related baseline shifts of the pulse wave DC component	Respiratory effort and negative intrathoracic pressure changes, repetitive Müller-/Valsalva maneuver	Obstructive vs. central sleep apnea, snoring
Proportion of pulse rate in the respiratory frequency band	Degree of respiratory sinus arrhythmia	Baroreflex sensitivity, autonomic neuropathy, cardiovascular aging, diabetes mellitus, disturbed breathing during sleep
Pulse amplitude ratio before and after Valsalva maneuver	Elevated left ventricular end-diastolic pressure	Heart failure with increased pre-load

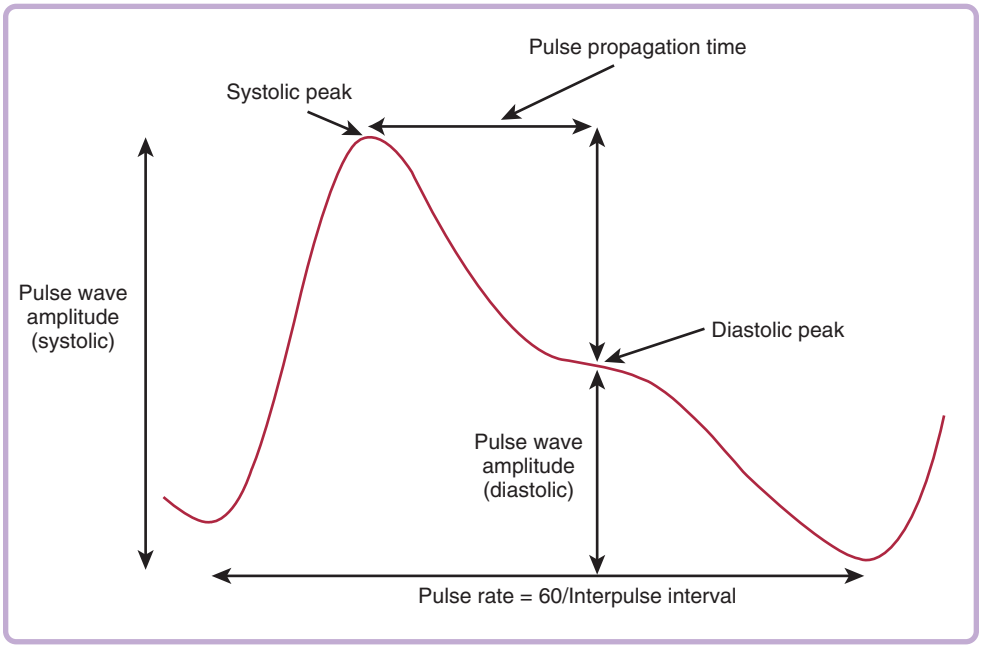


Figure 167-1 A typical PPG signal and its characteristic parameters.

cardiovascular risk.[15,17,18] During signal processing, first and second derivatives of the plethysmographic curve are used to more accurately calculate the pulse wave parameters explained above.[19]

The probe placement for assessment of the pulse wave form has been studied. It was suggested that the ear pulse wave may reflect better the passive effect of systemic hemodynamics, whereas the finger pulse wave may reflect local vasomotor fluctuations, including the neural mechanisms of skin sympathetic activation.[20] These studies were performed in spontaneously breathing wake subjects, and the site of pulse wave investigation may have quite different influences on the results

during sleep recordings. Indeed, one study looking on noninvasive markers of acoustically induced arousal from sleep showed that finger PPG was more sensitive for arousal response than ear PPG.[21] In addition, the finger site is most feasible for use during sleep recordings.

Peripheral Arterial Tone Technique

PAT is a novel technology for monitoring pulsatile arterial volume signals using a pressure-applied optical probe.[22] The technology is comparable to classic PPG with the difference of the specific probe features improving the signal quality. In detail, the PAT probe has a compliant elastic membrane surrounded by an outer rigid casing. Compared to conventional finger plethysmography, a pressurized region is used to cover the surface of the distal end of the finger. This region prevents the induction of venoarteriolar reflex vasoconstriction.[23] The balloon-like outer membrane creates a constant pressure leading to an unloading of arterial wall tension, thereby increasing the arterial wall motion and the size of the arterial volume change. The sensor region is also prevented from retrograded venous blood pooling commonly observed during finger movement. A transmission mode PPG is used to measure the optical density changes associated with pulsatile blood volume changes.

Pulse Transit Time

Pulse transit time (PTT) reflects the time interval for an arterial pulse wave to propagate from the aortic valve level to a given peripheral site. PTT is often measured as the time lapsed from the appearance of the R wave in the ECG to the start or the middle point of the systolic pulse wave appearing in the finger PPG recording. The PTT depends on the degree of stiffness of the arterial wall, and a shortened PTT is associated with vascular aging, atherosclerosis, and increase of blood pressure.

Finally, the usefulness of the three methods—PPG, PAT, and PTT—is limited in patients with cardiac arrhythmias, such as atrial fibrillation or frequent extra systolic activity. The high variability of cardiac stroke volume creates a beat-to-beat variation in the systolic PWA and the pulse contour, which sometimes invalidates the information embedded in the pulse wave signal. Indeed, existence of arrhythmia (but not type of arrhythmia) can be detected by PPG.

CLINICAL APPLICATIONS OF DIGITAL PULSE WAVE ANALYSIS

Periodic cardiovascular autonomic changes during sleep in patients with sleep-related disordered breathing have been documented since the 1970s. Heart rate and arterial blood pressure are the traditional physiologic parameters to study hemodynamic autonomic changes associated with sleep apnea events. More recently, the finger pulse wave signal in combination with other biosignals has been evaluated for the detection of autonomic arousal, sleep stages, sleep-disordered breathing, and the assessment of cardiovascular function[24] (Table 167-2). Improved and novel insights into the sleep state can be made with this rather simplified methodology. Despite the fact that even more outcome-based validation of such methods is needed, it is anticipated that they will be part of the state of the art in sleep diagnostics in the near future.

Identification of Autonomic Arousal/ Sleep Fragmentation

EEG arousal has been traditionally used for quantification of sleep fragmentation. However, it has been shown that transit sympathetic activations during sleep are not necessarily associated with visible EEG changes.[25] Hence, the term *autonomic arousal* is used to represent transit changes of autonomic activity associated with cardiac activation (e.g., heart rate and blood pressure increase). The finger vascular bed is densely innervated, and finger PWA drop has been shown to associate with increased EEG power density, suggesting that PWA attenuation is a useful surrogate marker for changes in cortical activity during sleep.[26] Other studies found digital vascular response (e.g., PWA) to variable degrees of arousal stimulus during NREM sleep to be more sensitive than other autonomic markers (e.g., heart rate, PTT, and pulse wave velocity).[21] Hence, digital vasoconstriction (PWA attenuation) may provide a useful tool to detect episodes of autonomic activation during sleep. By means of the PAT method, an autonomic arousal event was defined as ≥50% PAT attenuation or ≥30% PAT attenuation plus a 10% pulse rate increase, which correlated closely (r^2 = 0.67) with PSG scored EEG arousal.[27] Modified criteria for autonomic arousal detection in the ambulatory PAT device are based on either PAT attenuation plus increase of pulse rate or PAT attenuation ≥40% and short movement detected by actigraphic signal.[28] The correlation coefficient between PSG-derived arousal index and PAT autonomic arousal index (using actigraphy to detect sleep time) was 0.76. It should be noted that other conditions, such as periodic limb movements, can cause autonomic arousal without simultaneous cortical activation. It has been recently suggested that combined analysis of PWA amplitude and area under the pulse waveform may be helpful to reduce failure rate in respiratory arousal detection.[29]

The potential role of PAT for arousal detection was also studied in children. In general, EEG arousal in children is associated with sympathetic activation reflected by PAT attenuation.[30,31] However, PAT signal is a highly sensitive but less specific tool in this context. In fact, a substantial proportion of autonomic activations defined by PAT was found to occur in children without visually recognizable EEG changes. Whether these attenuations represent normal fluctuations of the autonomic sympathetic nervous system activity in children or reflect subtle sleep disruption that failed to be detected by traditional EEG scoring remains unclear. In children with upper airway obstructive disease undergoing noninvasive ventilation, a combination of 4% oxygen desaturation and microarousal detected by finger PWA was reported to be associated with movement and fragmentation index assessed by actigraphy.[32]

There is evidence for a gender-specific difference in the pulse wave response to arousal. Using conventional PPG, both spontaneous and acoustic arousals induced a strong pulse wave response in males but a less clear response pattern with a high inter-individual variability in females.[33] In addition, respiratory instability after arousal was also more pronounced in males. It is unclear whether those gender-specific findings in a small number of healthy subjects can be extrapolated to patients with cardiovascular disease or sleep disorders like sleep apnea or insomnia.

Table 167-2 Clinical Application of the Overnight Pulse Wave Analysis in Sleep Diagnostics

Method	Pulse Wave Parameter and Additional Signal	Results
Sleep Classification and Sleep Fragmentation		
PAT	Systolic PAT amplitude and PAT pulse rate	Autonomic arousal classification with high correlation to respiratory and nonrespiratory EEG-arousal in adults, overestimation in children
PPG	PPG systolic amplitude and pulse rate	Good correlation with EEG-arousal in adults
PPG and ECG	PTT and pulse rate	Good correlation with EEG-arousal in adults, overestimation in children
PAT and actigraphy	PAT amplitude and pulse variation (fractal signal analysis), actigraphy	Sleep/wake detection, REM/NREM classification, deep and light sleep classification
PPG	PPG signal (breathing related systolic amplitude variation)	Sleep wake analysis via PPG by breathing shape and breathing rhythmicity
Type and Amount of Sleep-Disordered Breathing		
PAT and oximetry	Pulse wave amplitude, pulse rate, and oxygen saturation	High correlation between AHI_{PAT} and AHI_{PSG} in multiple validation studies, AASM recommended method
PPG and nasal flow	PPG-derived respiratory effort	Improved differentiation between obstructive and central/mixed sleep apnea
PAT and oximetry	Pulse wave amplitude, pulse rate, and oxygen desaturation	Detection of Cheyne-Stokes respiration
PPG	Pulse wave amplitude, pulse rate, and oxygen desaturation	Detection of Cheyne-Stokes respiration
Cardiovascular Function and Risk		
PAT	Systolic PAT amplitude attenuations	Associations between overnight attenuations and office blood pressure
PTT and office blood pressure	A continuous measure of beat-to-beat blood pressure	Associations with intraarterial or oscillometric measure blood pressure at daytime and during sleep
PPG	Pulse wave derived augmentation index or pulse propagation time	Associated with vascular stiffness in atherosclerosis, vascular aging, and cardiovascular and metabolic disease; sensitive to vasoactive drugs
PPG and oximetry	PPG and oximetry-derived parameters reflecting cardiovascular, autonomic, and respiratory function during sleep	Associated with traditional cardiovascular risk prediction matrixes (e.g., ESH/ESC, Framingham, EU SCORE); identification of vascular stiffness in atherosclerosis, vascular aging, cardiovascular and metabolic disease; sensitive to cardiovascular medication

AASM, American Academy of Sleep Medicine; AHI, apnea hypopnea index; ECG, electrocardiogram; ESC, European Society of Cardiology; ESH, European Society of Hypertension; EU, European; NREM, non–rapid eye movement; PAT, peripheral arterial tone; PPG, photoplethysmography; PSG, polysomnography; PTT; pulse transit time; REM, rapid eye movement.

A transit dip in the PTT signal from baseline was reported to associate with EEG arousal.[34] Several studies have shown that PTT can improve microarousal detection in children and infants.[35,36] However, similar to PAT, this technique can sometimes be oversensitive for detection of arousal in children.[31]

Classification of Sleep Stages and Wakefulness

Wakefulness and the NREM/REM sleep stages differ in autonomic and hemodynamic regulation. This difference can be reflected by changes of the PWA and derived pulse rate signals.[37,38] For instance, a typical reduction of the PAT amplitude could be observed during REM compared with NREM sleep.[37] An automatic REM sleep detection algorithm based on a combination of the PAT signal and actigraphy was developed.[39,40] Subsequently, features from two time series of PAT amplitude and inter-pulse periods were used to further analyze NREM sleep and differentiate deep sleep from light sleep.[41] These sleep-staging algorithms were further validated and showed moderate agreement in a multicenter study.[42] In 38 normal subjects and 189 patients with suspected OSA, the overall agreement for detection of light/deep and REM sleep were 89 ± 6% and 89 ± 6% between PSG and PAT, respectively. OSA severity did not affect the sensitivity and specificity of the algorithm. However, in elderly patients with suspected OSA, deep sleep stage classification was underscored in PAT when compared with PSG.[43]

In a different approach, the PPG signal derived from single channel pulse oximetry has been used for sleep/wake detection.[44] The morphology and rhythm of breathing were derived from the PPG signal and analyzed. The similarity of consecutive breath and the "complexity of the breathing rhythm" were indicative for sleep and wake state, respectively.

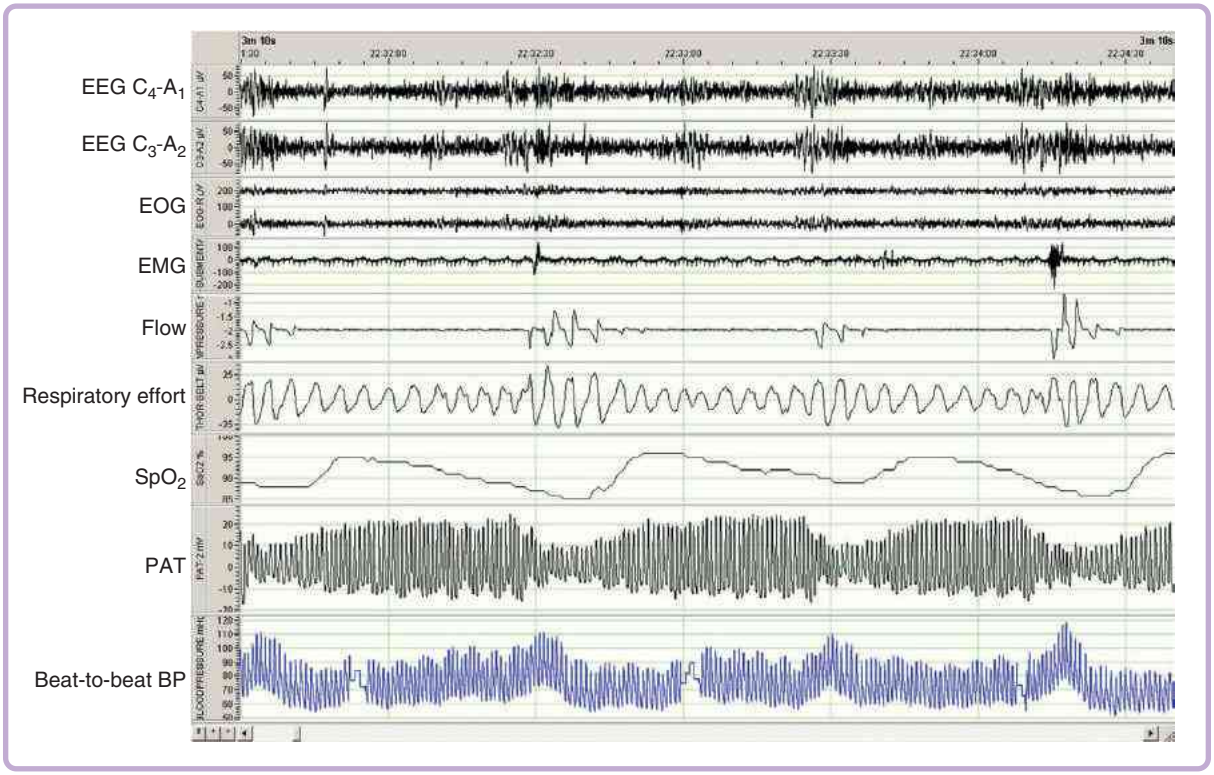


Figure 167-2 A 3-minute recording showing pulse wave amplitude changes from the peripheral arterial tone (PAT) signal associated with obstructive sleep apnea and arousal. The PAT signal fluctuates with increasing respiratory effort during apnea. The systolic PAT amplitude decreases significantly during central nervous arousal from sleep and subsequent resumption of breathing.

Sleep periods were identified when strong correlations between adjacent breaths, a constant rhythm, or periodic oxygen desaturation occurred. All noise and movement events were marked as wakefulness. In a study of 74 patients with sleep-disordered breathing, an agreement of 75% with PSG derived sleep classification was reached.[44] Two follow-up studies in a general population and in sleep apnea patients found a positive correlation/no correlation with PSG derived total sleep time, respectively.[45,46] The reliability of this interesting, but simplified, sleep/wake classification is therefore difficult to estimate, and further studies are meaningful.

Recognition of Obstructive Sleep Apnea

The use of plethysmography has demonstrated a typical attenuation pattern of PWA in the courses of repetitive obstructive apneas and hypopneas.[14,47] Similar changes could be observed by means of the PAT method[22,48] (Figure 167-2). Indeed, experimental data showed that the response is mainly mediated by the arousal response and, to a minor degree, by the response to significant upper airway obstruction.[49] Further, intra-arterial infusion of the alpha-receptor blocker phentolamine showed that this typical apnea-related response can be blocked, suggesting a strong association between PAT attenuation and alpha receptor sympathetic activity in the skin vasculature during apnea and subsequent arousal.[48]

The PAT technology uses the pattern recognition of pulse wave attenuations, oxygen desaturations, heart rate responses, and the above mentioned recognition of wake and sleep phases in a combined algorithm.[50] A recent meta-analysis of relevant validation studies performed in the general population and in several patient populations showed that this pulse wave based technology has a high agreement with PSG-derived indices of sleep apnea activity.[51] No specific effect of gender or age on the accuracy of the PAT technology for apnea detection has been found.

A similar approach was recently published with the use of single channel oximetry where saturation and the PPG signal were used together for sleep apnea recognition.[44] The PPG signal recognizes respiratory signals (effort and breathing frequency) and the sleep and wake rhythm (see previous description). A conventional oxygen desaturation pattern with the threshold of ≥3% is also used to detect breathing events. Although the algorithm is less well described in the publications, validation studies showed good agreement between PSG–apnea hypopnea index (AHI) and PPG–AHI, for example, the receiver operator characteristic (ROC) curve for an AHI threshold of 15 events/hour was 0.9.[46] At least, this technology can be seen as a significant improvement when compared with traditional single channel oximetry. In this context, an additional, so-called principal-component-analysis technique has been described, which allows the detection of respiratory efforts from the oximetry-derived PPG signal.[52]

The PAT and PPG signals have also been evaluated in the follow-up of patients with sleep-disordered breathing. The PAT technology was suggested to be specifically useful in the follow-up of treatment like positive airway pressure, oral device, and weight reduction.[53-55] The sleep structure, degree of remaining sleep fragmentation, and hypoxic load can be easily quantified in an outpatient setting. In particular, the capability of the PPG signal to detect arousal was validated

against PSG standard measures in patients using noninvasive ventilation. A good agreement between the methods was established at least for sleep fragmentation during NREM sleep.[56] The PPG technology was suggested to be used for simplified follow-up investigations in the growing number of patients with home ventilation.

During the past decade, the PPG signal has been made available in a number of PSG devices. One study postulated that the use of the PPG signal on top of all standard PSG channels was helpful to reduce interscorer variability.[57] The data have not, so far, been confirmed in a larger study, and the results may be highly dependent on the actual scoring rules used. However, it shows that the PPG signal as a marker of arousal and respiratory effort may be helpful in the quantification of sleep-disordered breathing in several settings.

Cheyne-Stokes Respiration and Central Sleep Apnea

Cheyne-Stokes respiration (CSR) is a condition that includes cyclic oscillation of breathing amplitude along with periodic fluctuation of sympathetic nerve activity. Interestingly, the PAT signal showed a sinusoid shape in accordance with the symmetric pattern of both respiratory effort and oxygen desaturation. The relative attenuation of the PAT amplitude was lower during wake CSR periods compared with CSR during sleep, and the maximum attenuation lagged the start of the crescendo breathing phase by 3 to 8 seconds. The sensitivity for CSR recognition varied between 73% (awake) and 91% (entire sleep period) with a specificity of 70% (REM sleep) and 97% (awake) when compared with PSG in a small study of 10 patients with CSR and heart failure.[58] These first positive results have not been replicated in a larger cohort, and it remains unclear whether the PAT technology in the absence of effort and flow signals can recognize CSR in a reliable way.

Another approach starts from the flow signal analysis and uses the PPG to indirectly assess respiratory efforts for improved apnea classification. The PPG-derived respiratory effort separates central/mixed apneas from obstructive ones. The correlation coefficients between flow/PPG signal derived and polygraphic scored central apnea index was 0.95 in a study of 66 sleep apnea patients.[59] This technology appeared to be useful for a better characterization of apnea type in limited channel devices.

PULSE WAVE ANALYSIS OF CARDIOVASCULAR FUNCTION DURING SLEEP

Blood Pressure

There are several methods using the PPG-based digital pulse volume curve for beat-to-beat analysis of systolic and diastolic blood pressure during sleep. The electropneumatic vascular unloading technique was introduced by Jan Peňáz to measure central blood pressure through a pressurized finger probe.[60] This method has been mainly used in research activities to assess continuous blood pressure change (Finapress or Portapres device). Another method uses the PTT to assess blood pressure continuously. Because central arterial stiffness is associated with blood pressure, shortening of the PTT is a marker of elevated blood pressure. This principle has been introduced recently in a sleep diagnostic device. The blood pressure values from the PTT curve need a calibration procedure against resting office blood pressure in order to enable continuous monitoring of the blood pressure surges associated with sleep apnea and/or arousal from sleep. Validation studies suggest a reasonable agreement for PPT-derived blood pressure when compared with values obtained by the conventional oscillometric method and finger arterial PPG (Finometer).[61-63] It is clear that the PTT method has advantages because it does not disturb sleep when compared with the oscillometric method or the Peňáz technique. However, accuracy of the diastolic pressure level and movement artifacts may limit the PTT methodology for continuous blood pressure assessment. Finally, the PWA alone using the PPG technology has also been used to calculate systolic blood pressure on a beat-to-beat basis during sleep.[64] The results showed also a reasonable agreement between PPG blood pressure and Portapres-derived pressure.

Pulse Rate Variability

Impaired heart rate variability is a marker of increased cardiovascular risk in patients with established ischemic heart disease and cardiac failure. The method is based on the recognition of RR interval by high resolution sampling of the ECG signal. The PPG signal can be used as a surrogate marker of heart rate variability. A large number of studies have been performed to analyze the accuracy between ECG-based heart rate variability and PPG-based pulse rate variability. Both methods are leading to quite similar results during resting conditions and undisturbed sleep.[65] However, frequency analysis during sleep apnea may differ between the two analysis methods.[66] In addition, respiratory sinus arrhythmia measured from pulse rate signal has been used for cardiovascular risk assessment.[67]

Vascular Stiffness

The peripheral pulse wave signal contains clinically important information about stiffness of the conduit arteries relevant for cardiovascular function and risk assessment.[16-18] The AI derived from finger pulse wave showed good intra-individual agreement during repeated measures. The PPG-AI demonstrated a dose–response relationship with mean arterial blood pressure and cardiovascular risk classification of the European Heart Score.[17,18] The mean PPG-AI increased in parallel with the classes of average to very high cardiovascular risk scores in 247 individuals and subjects with diabetes, hypercholesteremia, and hypertension had a significantly higher AI compared with healthy adults.[17] In hypertensive patients, it was shown that the measures indicating vascular stiffness in the PPG pulse wave were influenced by age, blood pressure, body mass index, and heart rate.[68] The vascular aging index was increased in the elderly and patients with uncontrolled hypertension. Finally, it has been shown that modification of vascular stiffness by vasoactive drugs can be mirrored by parallel changes in the finger pulse wave.[14-16,69,70]

ASSESSMENT OF OVERALL CARDIOVASCULAR RISK FROM A SLEEP RECORDING

Sleep-related respiratory and cardiovascular parameters are related to cardiovascular morbidity and mortality. For instance, failure to produce nocturnal dipping of blood pressure or heart rate has been independently associated with increased all-cause mortality.[71,72] The high frequency component of heart rate variability during sleep is blunted in patients with coronary artery disease, and the nocturnal arterial vascular tone

Table 167-3 Clinical Relevance of Physiologic Variables Derived from the Finger Photoplethysmography Signal for Cardiovascular Risk Assessment

Parameter	Function Assessed	Dysfunction Reflected by Parameter
Hypoxic variability	Recurrent hypoxia and reoxygenation	Hypoxic cardiovascular stress
Pulse wave attenuation	Frequency of pulse amplitude attenuations	Microvascular dysfunction, increased vascular sympathetic tone
Pulse rate acceleration	Sympatho-vagal balance at sinus node level, baroreflex	Coexisting cardiac/vascular/metabolic disease
Periodic pulse rate changes	Cardiorespiratory coupling, baroreflex sensitivity	Coexisting cardiac/vascular/metabolic disease
Pulse propagation time	Vascular wall compliance	Vascular aging, hypertension, atherosclerosis
Time of saturation <90%	Degree of nocturnal hypoxic load	Significant respiratory disease
Degree of symmetric nocturnal desaturation	Occurrence of central apneas and Cheyne-Stokes Respiration	Significant cardiac or CNS disease (e.g., heart failure, post stroke)
Cardiac response to nocturnal hypoxia	Heart rate response to intermittent hypoxia, chemoreflex	Cardiac or metabolic disease affecting autonomic response (e.g., diabetes)

CNS, central nervous system.

determined by finger PPG is elevated in patients with essential hypertension.[73,74] Sleep-related hypoxia, specifically intermittent hypoxia in sleep apnea, has been associated with increased cardiovascular mortality.[75] Therefore, there is a strong rationale for a systematic combined analysis of cardiac, vascular, and respiratory reactivity during the sleeping period as a measure of cardiovascular risk.

In a multicenter study, physiologic components of PPG signal were derived from the overnight finger oximetry recording using a Matching Pursuit algorithm for cardiovascular risk assessment (Table 167-3).[1,76] The parameters were selected based on their relevance in reflecting cardiovascular regulatory homeostasis and feasibility from the finger PPG signal. Variables reflecting cardiac rate variability were identified. A similar process was initiated for variables reflecting peripheral vascular reactivity and stiffness. In order to quantify autonomic events independent of respiratory and arousal events, PWA attenuations between 10% and 30% were included in the analysis. The PPT was used as a surrogate measure of pulse wave velocity and arterial stiffness. Finally, several measures of nocturnal oxygenation (constant, symmetric, and recurrent hypoxia) were also included in the analysis. A 2% oxygen desaturation was chosen for the hypoxia event threshold. The interaction between respiration and cardiovascular function using the respiratory frequency band of pulse rate and the heart rate response pattern to episodic hypoxia were calculated. The respiratory sinus arrhythmia is considered to reflect vagal-cardiac nerve activity, and a decrease in this pattern may reflect elevated cardiac sympathetic activity. Although these variables were modestly interrelated, the multivariate analysis suggested that all parameters contributed significantly to the cardiovascular risk assessment.[1] A composite cardiovascular risk score (range 0–1) was generated using a neuro-fuzzy system.[76]

In a cross-sectional validation study of this novel algorithm, all patients were classified according to the European Society of Cardiology/European Society of Hypertension (ESC/ESH) risk matrix.[1] The algorithm based on overnight PPG signal allowed the identification of high cardiovascular risk patients (high and very high added cardiovascular risk, according to the ESC/ESH matrix)[77] with sensitivity and specificity values of 74.5% and 76.4%, respectively, and area under the ROC curve was 0.80. Corresponding values for the AHI and the oxygen desaturation index were below 65%. The current data provide evidence that an algorithm based on a PPG signal offers a reasonable estimate of cardiovascular risk as expressed in standard matrix used in clinical medicine. However, it remains to be determined whether findings from this cross-sectional analysis can be extended to a prospective analysis. Such studies have been initiated.[78]

The PAT technology has also been evaluated regarding associations with cardiovascular risk markers. The magnitude of overnight PAT attenuations reflected daytime blood pressure in the general population independent of apneic measures from PSG.[79] This finding implies that vascular or autonomic phenomena recorded during the sleep period provide a marker for cardiovascular disease (e.g., hypertension). The AI calculated from PAT at rest was studied in 186 patients from a cardiology clinic.[80] Although PAT-AI significantly correlated with cardiac risk factors and coronary artery disease, the diagnostic capacity of this single pulse wave variable to differentiate patients with/without coronary artery disease was limited (area under the ROC curve 0.604), suggesting that multiple information processing of pulse wave signals for cardiovascular risk assessments may be needed.

SUMMARY

This chapter describes the hemodynamic and autonomic influence on the pulse wave signal during sleep. The information can be used for a better understanding of the associations among pulse wave change and autonomic arousal/sleep fragmentation, sleep stages, and sleep-related disordered breathing. A broad clinical application of digital pulse wave analysis in sleep medicine and sleep-related research is reviewed. The potential usefulness of using finger pulse wave signal for cardiovascular function assessment is discussed.

Given the simplicity, the cost-effectiveness, and clinical acceptance of the oximetry technology, the pulse wave signal obtained by modified oximeter technology opens new possibilities serving as an important additive or even alternative parameter in future sleep diagnostic device. Research is very active in the field and results are very promising.

CLINICAL PEARL

The finger pulse wave can be easily derived from PPG. Supportive information for sleep-stage classification, identification of autonomic arousal, and differentiation of obstructive versus central breathing event can be achieved, which improves the utility of ambulatory sleep diagnostic devices. Validation of the methodology has shown good agreement with conventional information gained from PSG. In addition, valuable information on cardiovascular dysfunction and risk can be obtained by advanced pulse wave analysis. Prospective outcome studies on cardiovascular risk prediction based on pulse wave analysis during sleep are necessary to establish the definite role of this novel methodology for sleep diagnostic procedures.

Selected Readings

Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas* 2007;**28**(3):R1–39.

Chan GS, Fazalbhoy A, Birznieks I, et al. Spontaneous fluctuations in the peripheral photoplethysmographic waveform: roles of arterial pressure and muscle sympathetic nerve activity. *Am J Physiol Heart Circ Physiol* 2012;**302**(3):H826–36.

Chen CY, Chen CL, Yu CC. Obstructive sleep apnea is independently associated with arterial stiffness in ischemic stroke patients. *J Neurol* 2015;**262**(5):1247–54.

Elgendi M. On the analysis of fingertip photoplethysmogram signals. *Curr Cardiol Rev* 2012;**8**(1):14–25.

Grote L, Zou D, Kraiczi H, Hedner J. Finger plethysmography–a method for monitoring finger blood flow during sleep disordered breathing. *Respir Physiol Neurobiol* 2003;**136**(2–3):141–52.

Gunarathne A, Patel JV, Hughes EA, Lip GY. Measurement of stiffness index by digital volume pulse analysis technique: clinical utility in cardiovascular disease risk stratification. *Am J Hypertens* 2008;**21**(8):866–72.

Hoyos CM, Yee BJ, Wong KK, et al. Treatment of sleep apnea with CPAP lowers central and peripheral blood pressure independent of the time-of-day: a randomized controlled study. *Am J Hypertens* 2015;**28**(10):1222–8.

Lavie P, Schnall RP, Sheffy J, Shlitner A. Peripheral vasoconstriction during REM sleep detected by a new plethysmographic method. *Nat Med* 2000;**6**(6):606.

Schnall RP, Shlitner A, Sheffy J, et al. Periodic, profound peripheral vasoconstriction–a new marker of obstructive sleep apnea. *Sleep* 1999;**22**(7):939–46.

Seetho IW, Asher R, Parker RJ, et al. Effect of CPAP on arterial stiffness in severely obese patients with obstructive sleep apnoea. *Sleep Breath* 2015;**19**(4):1155–65.

Sommermeyer D, Zou D, Eder DN, et al. The use of overnight pulse wave analysis for recognition of cardiovascular risk factors and risk: a multicentric evaluation. *J Hypertens* 2014;**32**(2):276–85.

Sommermeyer D, Zou D, Ficker JH, et al. Detection of cardiovascular risk from a photoplethysmographic signal using a matching pursuit algorithm. *Med Biol Eng Comput* 2015 Nov 4. [Epub ahead of print].

Sommermeyer D, Zou D, Grote L, Hedner J. Detection of sleep disordered breathing and its central/obstructive character using nasal cannula and finger pulse oximeter. *J Clin Sleep Med* 2012;**8**(5):527–33.

Yalamanchali S, Farajian V, Hamilton C, et al. Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis. *JAMA Otolaryngol Head Neck Surg* 2013;**139**(12):1343–50.

Zou D, Grote L, Eder DN, et al. Obstructive apneic events induce alpha-receptor mediated digital vasoconstriction. *Sleep* 2004;**27**(3):485–9.

A complete reference list can be found online at ExpertConsult.com.

Recording and Scoring Sleep-Related Movements

Chapter 168

Raffaele Ferri; Stephany Fulda

Chapter Highlights

- The development of scoring rules is a dynamic process with constant exchange between clinicians and researchers.
- The exchange between sleep medicine and research is vibrant, but the necessary time lag in incorporating new evidence into clinical scoring rules has led in many instances to the parallel existence of clinical rules and research rules for the same phenomenon.
- This chapter reviews both clinical and research recording techniques and scoring rules for sleep-related movements, with a focus on the most common ones, such as periodic limb movements during sleep, rapid eye movement sleep without atonia, and sleep-related bruxism. In addition, rhythmic movement disorder, propriospinal myoclonus of sleep onset, benign myoclonus of infancy, excessive fragmentary myoclonus, neck myoclonus, and leg movements other than periodic limb movements during sleep are briefly discussed.

The establishment and standardization of recording techniques and scoring criteria both reflect and afford progress in clinical sleep medicine and research. The development of scoring rules is a dynamic process, with constant interchange between clinical practice and research. This is illustrated by the fact that one of the standard references for clinical scoring rules, the manual published by the American Academy of Sleep Medicine (AASM),[1] is now continuously reviewed and updated based on new clinical evidence or advances in technology.

The exchange between sleep medicine and research is vibrant, but the necessary time lag in incorporating new evidence into clinical scoring rules has led in many instances to the parallel existence of clinical rules and research rules for the same topic. Presented in this chapter is an overview of both clinical and research recording techniques and scoring rules for sleep-related movements. The focus is on the most common such movements, such as periodic limb movements during sleep (PLMS), rapid eye movement (REM) sleep without atonia (RSWA)—manifested clinically as REM sleep behavior disorder (RBD), and sleep-related bruxism (SB). In addition, rhythmic movement disorder, propriospinal myoclonus of sleep onset, benign myoclonus of infancy, excessive fragmentary myoclonus, neck myoclonus, and leg movements other than PLMS also are briefly discussed. The finding of abnormal movements in sleep is important from both a diagnostic and a prognostic standpoint.

BASIC RECORDING METHODS

Electromyography

Surface electromyography is the "gold standard" for recording most muscle movements during sleep. The electromyogram (EMG) is acquired by means of silver chloride electrodes attached to the inputs of a differential amplifier, to obtain a bipolar derivation. It usually is recommended that interelectrode impedance be less than 10 KΩ (but better below 5 KΩ)[1,2]; the skin preparation procedure, before electrode placement, is very important (e.g., cleansing the skin with an alcohol pad, with light skin abrasion, and sometimes shaving excess hair). For better adhesion of the electrodes, the use of collodion is recommended because it is nonconductive, holds through hair (not only on the scalp), withstands oils and perspiration, and provides high performance for long-term recordings. Collodion is highly flammable, however, and produces fumes, so appropriate air purifier, fume extractor, or ventilation systems should be used. Finally, for long-term recordings, a conductive paste is needed to ensure good electrical contact between the skin and the electrodes.

EMG signals are produced by the muscle situated under the skin and by adipose tissue below the electrodes, which essentially record activity of superficial muscle. Muscle size and amount of adipose tissue significantly influence the amplitude of the surface-recorded EMG signal. This is the reason why surface EMG signals are considered semiquantitative and calibration of EMG activity is recommended.[2] The amplitude of surface EMG potentials depends also on the distance between the recording electrodes and can range between less than 20 μV and up to several millivolts, depending on the various factors noted previously.

EMG signals are constituted of superimposed motor unit action potentials produced by several motor units, each with a typical repetition rate of firing of approximately 7 to 20 Hz. Surface EMG records the sum of this activity and produces a signal with a wide spectrum. Figure 168-1 shows two examples of the EMG power spectrum from chin and tibialis anterior muscles (recorded with a sampling rate of 500 Hz); it is possible to note the wide extent of the power spectrum and the effects of notch filtering at 50 Hz.

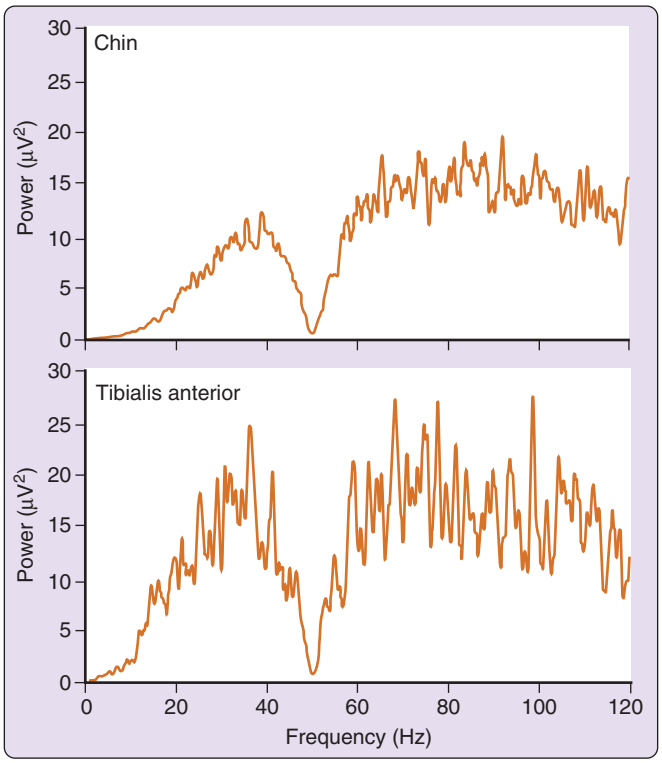


Figure 168-1 Average power spectrum of the electromyogram (EMG) signal sampled at 500 Hz and recorded from the chin and tibialis anterior muscles.

The spectral content of the EMG signal requires high sampling rates that should never be lower than 200 Hz; 400 to 500 Hz is usually recommended.[2,3] Band-pass filtering usually is applied, with typical settings at approximately 10 to 100 Hz and with a notch filter at 50 or 60 Hz, depending on the power line frequency.

Video Polysomnography

Video (and audio) recording, synchronized with the polysomnographic signals, is offered by most recording systems. This tool offers the possibility to examine simple or complex motor behaviors and vocalizations in conjunction with their accompanying electrophysiological correlates. Video polysomnography (PSG) is essential for the assessment of movements and behaviors that occur in several sleep disorders and allows their differentiation.[4-7] For example, this modality has a key role in differentiating parasomnias from nocturnal seizures and from behaviors arising during episodes of wakefulness, as sometimes occur in psychiatric conditions.

It should not be overlooked that video PSG provides objective documentation of all actions occurring in the laboratory, for both clinical and legal issues.

PSG equipment manufacturers can recommend fixed-focus video cameras, but devices that allow remote directional and zoom control (pan-tilt-zoom cameras) are preferable. Today's digital systems now allow "zooming" within the image also at the scoring or review stage.

To ensure darkness for the patient, an infrared light source is needed, with a corresponding appropriate camera.

A major limit of this method, however, is its high costs for both the recording and the reviewing/scoring processes. The evaluation of complex behaviors picked up by video PSG recordings is essentially a visual process that can be helped by the use of appropriate scales, developed for each sleep disorder.[7-15]

Time-lapse video recording, based on old time-lapse photography techniques,[16] has been proposed also to reduce the time (and costs) needed to review real-time video recordings, which are replayed in a much shorter time. The possibility to stop when a meaningful event occurs and to return to real-time replay allows finding the desired periods in a shorter time and analyzing them in detail.

In the future, automated video analysis can be expected to provide reliable help for the assessment of sleep-related behaviors, and some preliminary approaches have already been attempted.[17]

Actigraphy

Actigraphy is carried out using devices, usually worn on the wrist, containing motion sensors called accelerometers. These sensors integrate motion amplitude and speed, and their output is a signal with magnitude and duration depending on these motion features. This signal is appropriately amplified, filtered, and digitized to be stored in the device memory, most often in terms of movement counts per epoch. The length of the epoch is of crucial importance and can be fixed at 1 minute or can be chosen by the user (from a fixed list of epoch lengths).

In conjunction with the memory capacity of the actigraphic device, epoch length determines the maximum recording time. Because the most important clinical application of actigraphy is in the assessment of the circadian rhythm sleep-wake disorders,[3] it is necessary that these devices can reliably carry out recordings over a period of at least 7 days. Even if 5 nights have been indicated to be the minimum required to obtain reliable actigraph measures of sleep for children and adolescents,[18] it appears important to include all weekdays and the weekend for a complete report.

Data are stored in essentially three different ways by currently available actigraphs, some of which allow the user to select the preferred mode, whereas others do not. With the "Time Above Threshold" mode, the amount of time per epoch during which activity exceeds the set threshold is stored; the "Zero Crossing" mode assesses the number of times per epoch that the activity signal produced by the accelerometer crosses a threshold set around zero; finally, with the "Proportional (or Digital) Integration" mode, the area under the curve is stored for each epoch. Some devices allow the simultaneous use of multiple modes. Different studies have provided support for the various measurement modes. Conclusive statements regarding which is the most accurate among these modes are thus far lacking in the literature.

After the recording phase, data usually are downloaded on a computer for scoring. Each system has specific software tools but they typically are based on the algorithms proposed by Cole and associates[19] or Sadeh and coworkers.[20]

Actigraphy cannot be used for staging sleep; it can overestimate sleep, because when subjects are lying still while awake, they do not produce movements. Actigraphy shows good agreement with PSG in the measurement of total sleep time (TST) in healthy subjects, but in patients with sleep disorders (characterized by frequent arousals and reduced TST), it shows lower agreement. Actigraphy should always be

accompanied by a careful monitoring of sleep, along with a sleep diary kept by the patient, which allows a more reliable estimation of TST and wakefulness after sleep onset (WASO).

Actigraphic monitoring of foot or leg movements has been proposed,[21,22] and because it offers the possibility to record multiple nights in a home environment, it has been proposed as a tool to overcome the problems caused by the relatively large night-to-night within-subject variability reported to occur for periodic leg movements during sleep (PLMS).[23] Actigraphy alone, however, is unable to discriminate between PLMS and leg movements occurring during wakefulness; moreover, arousal or other related events (apnea) cannot be detected.

The technical aspects of actigraphic recordings are important and able to influence the results; earlier published criteria for reliable actigraphic recording should be carefully applied.[2] A recent review and meta-analysis[24] of the use of actigraphy for the measurement of PLMS has demonstrated significant heterogeneity among the few studies available, regarding the type of actigraph, position of the sensors on the legs, and methods for counting PLMS. In particular, an important limitation was noted in the possibility to reliably combine data from actigraphs placed on both limbs, in most devices.

Thus, at present, actigraphy cannot substitute for polysomnography within the process of diagnosing restless legs syndrome (RLS) but is useful in screening for RLS in large groups of subjects and may be more suitable for evaluation of the degree of night-to-night variability of PLMS.[23]

Other Methods

Several producers propose piezoelectric sensors for the recording of limb movements during sleep. These sensors, placed around the ankle or around the leg, transduce motion, vibration, and tension into an electrical signal. They are very sensitive, and there is no guarantee that the signal produced corresponds exclusively with events generated by the limb. No convincing validation studies are available for the use of these sensors, and they are not recommended by any guideline. The only advantage that they offer is that no skin preparation is required; however, this is counterbalanced by a cost higher than that for silver chloride electrodes.

A variety of other movements can be recorded during sleep; among them, of particular importance are those needed for the diagnosis of the different types of sleep apnea, such as thorax and abdomen movements. Strain gauges and piezoelectric sensors have been used, but the most recent and reliable belts are based on inductive sensors, which also allow estimation of volume changes in these anatomic structures.

PERIODIC LIMB MOVEMENTS

The first PSG EMG recordings of PLMS were carried out in Bologna by the group led by Lugaresi in 1965.[25,26] Some 20 years later, duration, amplitude, periodicity, and symmetry of PLMS were defined and manually measured in paper recordings by Coleman,[27] who created the bases of the scoring criteria established by the American Sleep Disorders Association (ASDA),[28] in 1993, which subsequently have been used for more than 20 years.

Currently, two sets of similar (but not identical) rules for scoring PLMS and periodic leg movements during wakefulness (PLMW) exist. These were partly informed by algorithms proposed for the automatic detection of leg movements in PSG[29] and include mathematically defined parameters such as thresholds, intervals, and amplitude.[29,30] First, in 2006 a task force of the International Restless Legs Syndrome Study Group endorsed by the World Association of Sleep Medicine (WASM/IRLSSG)[2] introduced a major revision of the scoring rules, which were then substantially (but not entirely) adopted by the AASM in 2007.[1,31,32] Table 168-1 lists the similarities and differences between the two sets of rules; in the following paragraphs, unless otherwise stated, recommendations, descriptions, and definitions refer to both sets.

Recording Methods

Surface EMG electrodes should be placed at 2 to 3 cm apart or a distance from each other that is one third of the length of the anterior tibialis muscle, whichever is shorter. Electrodes must be located longitudinally on the muscles, symmetrically around the middle. Impedance should be 10 KΩ or less for clinical studies, but a setting of 5 KΩ or less is recommended (AASM) or required for research studies (WASM/IRLSSG). EMG signals must be obtained from both the right and the left leg. Recording the two signals in one channel is strongly discouraged. For research studies, signals from each leg must be recorded separately (WASM/IRLSSG). Recording activity from other muscles besides the tibialis anterior is recommended only for research purposes or for special clinical conditions (e.g., arm restlessness). For the WASM/IRLSSG criteria,[2] sampling rate should be 200 Hz or greater (clinical settings) or 400 Hz or greater (research settings), whereas for the AASM criteria,[1,31,32] sampling rate should be 200 Hz or higher (minimal) or 500 Hz (desirable). Filter settings should be 10 to 100 Hz; for research purposes, settings of 10 to 200 Hz are recommended (WASM/IRLSSG); the AASM additionally recommends that use of the 60-Hz notch filter should be avoided.

Baseline resting EMG amplitude (i.e., in the relaxed muscle) should be ±2 to 3 μV (4 to 6 μV peak to peak; WASM/IRLSSG) or 10 μV or less (AASM). The WASM/IRLSSG rules recommend that before the recording, a calibration should be carried out to obtain from the relaxed anterior tibialis muscles a nonrectified signal of power no greater than ±5 μV (or 10 μV peak to peak; 5 μV for rectified signals) for clinical purposes and ±3 μV (or 6 μV peak to peak; 3 μV for rectified signals).

Scoring Rules

The scoring of periodic leg movements follows several general, sequential steps:

1. Leg movements are identified by the amplitude and duration of the EMG activation.
2. Bilateral leg movements are combined if a single PLMS index is to be calculated.
3. Leg movements occurring in the vicinity of sleep-disordered breathing events (e.g., respiratory-related leg movements [RRLMs]) are identified and excluded.
4. The remaining leg movements are classified as periodic or nonperiodic (isolated) on the basis of their occurrence within a series of such movements characterized by their number and the interval between movements.

The scoring process begins with the identification of candidate leg EMG events. Their onset is defined as an EMG increase of 8 μV or greater above the resting baseline, whereas

Table 168-1 Recording and Scoring of Periodic Leg Movements According to the Guidelines of the World Association of Sleep Medicine/International Restless Legs Syndrome Study Group (WASM/IRLSSG)[2] and the American Academy of Sleep Medicine (AASM)[1,32]

Feature/Component	WASM/IRLSSG	AASM
Recording of Leg Movements		
Electrodes	Surface electrodes	Surface electrodes
Electrode positioning	Tibialis anterior muscles Placed longitudinal, symmetrically around the middle, 2–3 cm apart or one third of length of anterior tibialis muscle, whichever is shorter	Tibialis anterior muscles Placed longitudinal, symmetrically around the middle, 2–3 cm apart or one third of length of anterior tibialis muscle, whichever is shorter
Combined recording of left and right legs	Bilateral recordings are required. Use of two channels, one for each leg, is strongly recommended for all studies and is required for research Clinical applications may, however, combine the electrodes from both legs into one recorded channel, although this practice is discouraged	Both legs should be monitored for the presence of leg movements. Use of separate channels for each leg is strongly preferred Combining electrodes from the two legs to give 1 recorded channel may suffice for some clinical settings, although this strategy may reduce the number of detected leg movements
Sampling rate	≥200 Hz in clinical studies ≥400 Hz in research studies	≥200 Hz 500 Hz desirable
Filter	10–100 Hz in clinical studies 10–200 Hz in research studies	10–100 Hz Use of 60-Hz (notch) filters should be avoided.
Impedance	≤10 KΩ in clinical studies ≤5 KΩ in research studies	<10 KΩ <5 KΩ preferred
Definition of Leg Movement		
Onset	EMG increase ≥8 μV above the resting baseline	EMG increase ≥8 μV above the resting baseline
Offset	EMG decrease to <2 μV above the resting level for ≥0.5 s	EMG decrease to ≤2 μV above the resting level for ≥0.5 s
Duration	Time between onset and offset, 0.5–10 s	Time between onset and offset, 0.5–10 s
Baseline	Resting EMG Relaxed muscle Absolute signal amplitude, 4–6 μV peak to peak *Calibration*: Relaxed anterior tibialis lasting: ≤10 μV in clinical setting ≤6 μV in research setting *Special criteria for events during wake time*: If EMG > 6–10 μV for ≥15 s, then new increased baseline is defined as average amplitude during this period	Stable resting EMG Relaxed muscle Absolute signal amplitude, ≤10 μV peak to peak
Scoring of Periodic Leg Movements (PLMs)		
Intermovement interval (IMI)	Onset-to-onset: 5–90 s	Onset-to-onset: 5–90 s
Number of leg movements	≥4 (Leg movements lasting <0.5 s or >10 s are disregarded)	≥4
IMI >90 s	PLM series ends	PLM series ends
IMI <5 s	PLM series goes on; leg movement with IMI <5 s is disregarded	*Not specified*
Sleep-wake	All leg movements form PLM series For PLMS, only those during sleep are counted	Only leg movements during sleep form PLM series
Bilateral leg movements	Offset-to-onset <0.5 s	Onset-to-onset <5 s
Respiratory-related leg movements (RRLMs)	Excluded from PLM series	Excluded from PLMS series
RRLM definition	Any leg movement occurring within: ±0.5 s around the ending of an apnea/hypopnea event	Any leg movement occurring within: 0.5 s before the start to 0.5 s after the end of an apnea or hypopnea, respiratory effort–related arousal, or sleep-disordered breathing event

EMG, Electromyogram; PLMS, periodic leg movements during sleep.

offset is marked at the point when the EMG amplitude decreases to less than 2 μV above the resting level and remains below that value for at least 0.5 second. An event can contain one or more periods with EMG amplitude below the offset level that each last less than 0.5 second. The duration of the event is the time between its onset and offset; it must be at least 0.5 second and no longer than 10 seconds; 15 seconds are allowed for research purposes (WASM/IRLSSG). All other events not meeting these criteria are discarded and not considered any further.

Next, if a single PLMS index is to be calculated, as opposed to a separate PLMS index for each leg, bilateral leg movements are combined. The WASM/IRLSSG criteria consider leg movements to be bilateral if the offset of the first event is less than 0.5 second before the onset of the subsequent event. By contrast, the AASM rules consider leg movements to be bilateral if the onset of the first event occurs less than 5 seconds before the onset of the next event. Because leg movements can be between 0.5 and 10 seconds in duration, the AASM rule also classifies some overlapping leg movements as bilateral.

Both sets of rules exclude leg movements that occur in the vicinity of sleep-disordered breathing events (RRLM), from the inclusion into PLMS series, and these leg movements have to be identified and excluded before calculating the PLMS index. For the WASM/IRLSSG,[2] a leg movement must be excluded from the PLMS analysis when it is associated with the resumption of respiration at the end of an apnea/hypopnea event, defined as any part of the leg movement in the interval of ±0.5 second around the end of the breathing event. By contrast, for the AASM,[1,31,32] a leg movement should not be counted when it occurs during a period starting 0.5 second before the beginning of a sleep-disordered breathing event to 0.5 second after the end of this event. Sleep-disordered breathing events include apneas and hypopneas and respiratory effort–related arousals.

After identifying leg movements, combining bilateral leg movements, and discarding RRLMs, among the remaining leg movements, those belonging to periodic leg movement sequences are identified. For this step, the crucial parameter is the interval between consecutive movements, measured from onset to onset—that is, between 5 and 90 seconds. If the interval between one leg movement and the next is greater than 90 seconds, any possible PLMS series ends with the previous leg movement. A crucial question is what happens when the interval between one leg movement and the next is less than 5 seconds. The WASM/IRLSSG rules specify that if the interval is less than 5 seconds, the second leg movement should be ignored, and the series goes on if the interval between the first leg movement and the next leg movement after the one that is being ignored is less than 90 seconds. The AASM rules, however, do not give any specification for this case.

Subsequently, a periodic sequence is defined as a series of four or more leg movement separated from each other by 5 to 90 seconds. A further crucial point at this step is the question of whether only leg movements during sleep or also those during wake are part of a periodic leg movement series and how intervening wakefulness affects a possible periodic leg movement series. Here, the WASM/IRLSSG rules recommend that all leg movements during sleep and wake can form part of a PLM series, and that for calculation of the PLMS index, only those during sleep are counted. Periodic leg movements during wakefulness are counted for the PLMW index. By contrast, in the AASM rules, this question is not addressed; the scoring of leg movements during wake is not foreseen, and it is unclear whether intervening wake categorically ends a PLMS series or whether in the case of short-lived (less than 90 seconds) wakefulness the series can go on, provided that the interval between the last leg movement during sleep before the wakefulness event and the first leg movement during sleep after the wakefulness event is less than 90 seconds.

Leg movements are considered to be associated with an arousal event when they are separated by less than 0.5 second—that is, between the end of one event and the onset of the other, regardless of which is first.

Once PLMS and PLMW are scored, several summary measures can be obtained. The following measures are recommended:[2]

- PLMS index: number of PLMS divided by the number of hours of sleep with leg movements recorded (PLMS/hour)
- PLMS with arousals index: number of PLMs associated with arousals divided by the number of hours of sleep with leg movements and arousal recorded (PLMA/hour)
- PLMW index: number of PLMW divided by the number of hours of wake with leg movements recorded (PLMW/hour)

Whenever possible, it also is recommended to report the PLMS index during non–rapid eye movement (NREM) sleep only; the PLMS index during REM sleep only; duration of PLMS and PLMW (separate for REM and NREM sleep), and intermovement interval of PLMS and PLMW (separate for REM and NREM sleep). Optional parameters are PLMS by sleep stages (including duration and intermovement intervals) and isolated leg movements.

Advanced Measurements of Periodic Leg Movements during Sleep and Periodic Leg Movements during Wakefulness

Sleep PSG recordings in patients affected by RLS or other sleep disorders usually also contain a significant amount of leg movements that cannot be classified as "periodic"[33-35]; moreover, it has been reported that leg activity during wakefulness in normal subjects is nonperiodic.[36] To analyze, in a more comprehensive way, this admixture of periodic and nonperiodic activities, new advanced measurements have been proposed as an important integration of the information provided by the previously reported scoring methods.[37]

In particular, an additional measure has been established, the *periodicity index,* indicating the degree of periodicity of the total leg movement activity.[33] This index quantifies the proportion, over the total, of intermovement intervals of $10<i\leq90$ seconds that are preceded and followed by another interval of the same length (this is equivalent to a series of 4 leg movements all separated by intervals of $10<i\leq90$ seconds). The numerical value of this index can range between 0 (absence of periodicity, with none of the intervals having a length between 10 and 90 seconds) and 1 (complete periodicity, with all intervals between 10 and 90 seconds long).[33] It is important to note that for the correct calculation of this index, leg movements occurring within 5 seconds from a previous movement should not be deleted from the analysis, taking into account intervals as short as 1 second (minimum movement duration

0.5 second + minimum interval duration 0.5 second). In a departure from the rules just presented, the range of the possibly periodic leg movement intervals also is considered to be 10 to 90 seconds by this measure.

The periodicity undergoes significant age-related changes, with a trajectory significantly different from that of the total leg movement count, indicating the need of age-adjusted normative reference values,[36,38] especially in elderly people, older than 75 years of age, and in children/adolescents, in whom a low periodicity can be expected even in the presence of clinically evident RLS.[39]

Another important feature characterizing PLMS is their time distribution during the night. In a majority of patients with RLS, PLMS count progressively decreases during the night.[38,40]

Unresolved Issues

Unresolved issues concern the discrepancies between the two set of rules. Among those, a major issue is the consideration of RRLMs. As detailed previously, the WASM/IRLSSG rules exclude RRLMs if they occur at the end of respiratory events, whereas the AASM rules exclude also LMs at the start of and during respiratory events. Neither of the sets of criteria to identify RRLMs were evidence-based. Recently, a study[41] on RRLMs as identified by the WASM/IRLSSG rules has argued that RRLM might represent part of a phenotypic spectrum that includes PLMS and has questioned the need to exclude these RRLMs from the analysis of PLMS. In addition, a further recent study,[42] and the first one to systematically analyze the distribution of LMs with respect to respiratory events, has reported that LMs in patients with sleep-related breathing disorders (SRBDs) were not increased at the beginning or middle of respiratory events but clustered around the end of events over a period significantly longer than specified by the AASM and WASM/IRLSSG rules. Of importance, both sets of rules underestimated the number of RRLMs and thus overestimated the number of PLMS in patients with SRBD. Together, these studies underline the need for continued evaluation and development of scoring rules for PLMS.

REM SLEEP WITHOUT ATONIA

RSWA is the polysomnographic hallmark of REM sleep behavior disorder (RBD), a parasomnia characterized by dream-enacting behavior in which the physiologic atonia during REM sleep is absent or greatly diminished (see Chapter 103). According to the current *International Classification of Sleep Disorders* (ICSD3),[3] among the criteria for the diagnosis of RBD is the demonstration of RSWA based on the definition provided by the most recent AASM guidelines.[1]

The AASM guidelines[1] distinguish between sustained, tonic muscle activity and excessive, transient, phasic muscle activity in REM sleep and provide scoring rules to identify REM epochs with atonia and/or excessive phasic activity based on chin and tibialis anterior EMG during REM sleep. Besides these guidelines, there is a multitude of different recording set-ups and scoring algorithms, both visual and automated, currently used in research protocols.

Recording Methods

It generally is accepted that scoring sustained, tonic muscle activity during REM sleep is based on the chin EMG.[1,43,44] Thus far, no consensus has emerged regarding practices or recommendations for assessing phasic muscle activity during REM sleep. Various approaches have been used, such as the recording of the chin muscles only,[45-50] or of chin muscles in different combinations with tibialis anterior,[51,52] brachioradialis,[53,54] biceps brachii,[55-57] extensor digitorum,[58,59] or carpi radialis.[60]

Comparing quantitative EMG analysis in 13 different muscles in patients with RBD, the Sleep Insbruck Barcelona (SINBAR) group showed that phasic EMG activity differed significantly depending on which muscles were evaluated.[61] Although the highest rate of phasic EMG activity was found in the mentalis muscle, reliance on only this muscle misses 45% of all phasic activity.[61] As indicated by analysis of different combination of muscles, the three-muscle combination that detected the highest number of mini-epochs with phasic activity was that composed of the mentalis muscle, the flexor digitorum superficialis in the upper limb, and the extensor digitorum brevis in the lower limb. Use of this combination identified 82% of all mini-epochs containing phasic EMG activity during REM sleep, in any recorded muscle. This montage was adopted as the SINBAR EMG montage. The SINBAR group subsequently reported that behavioral events and vocalizations during REM sleep in patients with RBD were almost exclusively (94.4%) linked to phasic EMG activity identified with this EMG montage.[62] Of importance, recording of the mentalis muscle alone did miss 35% of the behavioral events.[62] An important implication of this observation is that the quantification of RSWA and subsequently determined cutoff thresholds for the diagnosis of RBD probably will vary, depending on the specific montage and number of EMG signals used.[43,44] Several different protocols are currently in use for recording and scoring RSWA:

- *AASM recommendations*[1]: The AASM scoring system is based on the chin EMG for assessment of sustained, tonic EMG activity and on the chin and tibialis anterior EMG for assessment of excessive phasic activity.
- *SINBAR EMG montage*[61]: On the basis of the study in 13 different muscle groups just described, the SINBAR group has recommended the recording of the mentalis muscle, in association with the flexor digitorum superficialis in the upper limb and the extensor digitorum brevis in the lower limb, for research protocols. This group has recently proposed normative values for RSWA based on the SINBAR EMG montage.[43]
- *Montreal research group recommendations*[44,45]: The Montreal research group and many other researchers have assessed tonic and phasic EMG activity based on the chin EMG alone.[11,45,46,51,55-58,63-72] Cutoff values for the diagnosis of RBD have been recently evaluated.[44]

Scoring Rules

Visual Scoring of REM Sleep without Atonia

Since the first introduction of scoring rules for RSWA by Lapierre and Montplaisir in 1992,[45] there has been a wide variance in scoring approaches, many research and clinical groups defining their own, often with only slightly different rules (see Table 168-2 for a nonexhaustive list; a more complete list is available in the review by Fulda and colleagues[73]). The main differences concern the amplitude criterion to

Table 168-2 Scoring and Quantification of Tonic, Phasic, or Any Electromyogram (EMG) Activity during REM Sleep*

Definition No.	Study	Epoch Length (s)	Duration (% of Epoch)	Scoring: Amplitude of EMG Activation	Quantification: Name	Quantification: Measure	Reference(s) to Relevant Studies
Tonic Activity							
1a	Lapierre and Montplaisir, 1992[45]	20	>50%	Tonic EMG activation	Atonia %	% of 20-s epochs	11, 45, 46, 51, 55-58, 63-72
	Modification 1	30	>50%	Tonic EMG activation		% of 30-s epochs	47, 49, 80, 149-155
	Modification 2	30	>50%	>2 μV above background			156
	Modification 3	30	≥50%	>2 × amplitude during N3 sleep		% of 30-s epochs	157
	Modification 4	20	>50%	>2 × individual atonia amplitude and >10 μV		% of 20-s epochs	48, 158, 90, 159
	Modification 5	20	>50%	≥2 × background or >10 μV	Tonic EMG density	% of 20-s epochs	44†
	Modification 6	30	>50%	≥2 × background or >10 μV	Tonic EMG	% of 30-s epochs	74†
1b	AASM, 2007/2014[1,31]	30	≥50%	> minimum of amplitude during NREM sleep			160-162
1c	Khalil et al., 2013[96]	20	≥50%	>2 × background of adjacent REM epochs	Tonic activity %	% of 20-s epochs	96†
1d	SINBAR, 2012[43]	30	>50%	≥2 × background or >10 μ	Tonic EMG activity	% of 30-s epochs	43,† 163
1e	McCarter et al., 2014[94]	30	>50%	≥2 × background or ≥10 μV	Tonic EMG %	% of 30-s epochs	94†
Phasic Activity							
2a	Lapierre and Montplaisir, 1992[45]	2	0.1–5 s	>4 × background	Phasic EMG %	% of 2-s epochs	11, 45, 46, 51, 55-58, 63-68, 72, 90
	Modification 1	3	0.1–5 s	>4 × background		% of 3-s epochs	49, 69, 80, 149, 150, 152-156
	Modification 2	3	0.1–5 s	>2 × background		% of 3-s epochs	61, 62, 151
	Modification 3	2	0.1–5 s	>4 × lowest amplitude during present REM episode		% of 2-s epochs	164
	Modification 4	3	0.1–5 s	>4 × background, separated by ≥1 s		% of 3-s epochs	157
	Modification 5	2	0.3–5 s	>4 × background		% of 2-s epochs	158
	Modification 6	2	0.1–2 s	>50 μV		% of 2-s epochs	70
	Modification 7	2	0.1–10 s	>4 × background	Phasic density %	% of 2-s epochs	44,† 74,† 165, 166
2b	Bliwise et al., 2006[54]	2.5	> 0.1 s‡	≥4 × presleep baseline and "superimposed on a discernible background activity of not more than 25% of burst amplitude"	PEM	% of 2.5-s epochs	54,167
	Modification	2.5	> 0.1 s‡	≥4 × background activity		% of 2.5-s epochs	168

Continued

Table 168-2 Scoring and Quantification of Tonic, Phasic, or Any Electromyogram (EMG) Activity during REM Sleep—cont'd

Definition No.	Study	Epoch Length (s)	Duration (% of Epoch)	Scoring: Amplitude of EMG Activation	Quantification: Name	Quantification: Measure	Reference(s) to Relevant Studies
2c	AASM, 2007/2014[1,31]	30	0.1–5 s[§]	≥4 × background			160, 169
2d	SINBAR, 2012[43]	3	0.1–5 s	>2 × background until >0.25 s; then return to background In the presence of sustained tonic activity: ≥2 × tonic background in the same 3-s mini-epoch with waxing/waning morphology	Phasic EMG activity	% of 3-s epochs	43[†]
2e	Khalil et al., 2013[96]	2	0.1–10 s	>4 × background of adjacent REM epochs	Phasic activity %	% of 2-s epochs	96[†]
2f	McCarter et al., 2014[94]	3	0.1–14.9 s	>4 × background until 0.2 s; then return to background In the presence of tonic activity: ≥2 × tonic background in the same 3-s mini-epoch	Phasic EMG %	% of 3-s epochs	94[†]
"Any" EMG Activity							
3a	Gilman et al., 2003[150]	30		Tonic (def. 1a) or phasic activity (def. 2a)	RBD PSG measure	% of 30-s epochs	72, 150
3b	Arnulf et al., 2005[52]	1	Yes/no	≥ amplitude during quiet wakefulness	%RWA	% of 1-s epochs	52, 60, 170-174
3c	Consens et al., 2005[49]	30/3		Tonic (def. 1a, mod. 1)[‡] and phasic activity (def. 2a, mod. 1)	PSG score	Average of % tonic and phasic activity	49, 80[†]
3d	Zhang et al., 2008[153]	30/3		Tonic activity (def. 1a, mod. 1)[‡] or phasic activity (def. 2a, mod. 1)[‡]	REMREEA	% of 30-s epochs with tonic activity + % of 3-s epochs with phasic activity	153, 175
3e	Gagnon et al., 2006[176]			≥2 × background or >10 μV	% muscle activity	% of REM time	176, 177
3f	SINBAR, 2012[43]	3	≥0.1 s	>2 × background	Any EMG %	% of 3-s epochs	43,[†] 88, 163
3g	McCarter et al., 2014[94]	3		Tonic (def. 1e) or phasic (def. 2f) activity	Any EMG %	% of 3-s epochs	94 [†]

*Considered muscle groups are chin EMG for tonic activity and chin EMG and other muscles for phasic or any activity.
[†]Quantifications for which diagnostic thresholds for the identification of patients with RBD have been evaluated.
[‡]Return to baseline had to be clearly present within each 2.5-second epoch.
[§]At least 5 to 10 (or more) 3-second mini-epochs contain phasic activity.
AASM, American Academy of Sleep Medicine; PEM, phasic electromyographic metric; RBD, REM behavior disorder; REM, rapid eye movement (sleep); REMREEA, REM-related EMG activity; RWA, REM without atonia; SINBAR, Sleep Innsbruck Barcelona group.

identify EMG activations and the duration of phasic EMG bursts (see Table 168-2). In addition, besides scoring and quantifying tonic and phasic EMG activity during REM sleep, several groups of investigators defined and used summary measures that quantify "any"[43] EMG activity during REM sleep. Usually, tonic activity is scored as such only when it occupies more than half of an epoch. At the same time, phasic activity has a maximum duration that is between 5 and 10 seconds, depending on the definition used. Consequently, EMG activity that lasts longer than the maximum duration for phasic activity and shorter than half of the epoch seems to be neglected by most scoring approaches. The term *"any" EMG activity* was therefore introduced to take into consideration EMG activity of any length during REM sleep.

Lapierre and Montplaisir Scoring Rules[45] and Variations. The first formal scoring rules were introduced by Lapierre and Montplaisir in 1992.[45] They quantified:

- *Atonia* as the percentage of 20-second epochs that contained tonic EMG activation in the chin EMG for more than 50% of the epoch
- *Phasic EMG* as the percentage of 2-second mini-epochs that contained EMG bursts of between 0.1 to 5 seconds that had amplitude larger than 4 times the background

This influential definition continues to be used with some modifications (see Table 168-2). The penultimate modification (Table 168-2, definition 1a, modification 5) was reported in 2010, in the first study[44] that explored diagnostic thresholds for the diagnosis of RBD based on this scoring of RSWA (see further on). In this study, the following rules were applied:

- *Tonic EMG density* was defined as the percentage of 20-second epochs that contained tonic EMG activation with an amplitude at least twice the background or larger than 10 μV.
- *Phasic EMG density* was scored as the percentage of 2-second mini-epochs that contained EMG bursts with amplitude larger than four times the background and lasting 0.1 to 10 seconds.

Finally, in 2014[74] the scoring rules for tonic EMG were adopted to an epoch length of 30 seconds (Table 168-2, definition 1a, modification 6) and diagnostic thresholds for the diagnosis of RBD were evaluated.

AASM Scoring Rules.[1] The AASM scoring rules evaluate each 30-second epoch for the presence of sustained, tonic or excessive, phasic EMG activity:

- An epoch with *sustained, tonic activity* is characterized by at least 50% of the epoch with chin EMG activation with amplitude greater than the minimum amplitude during NREM sleep.
- An epoch with *excessive phasic activity* contains at least 5 mini-epochs, based on the division of the 30-second epoch into 10 sequential 3-second mini-epochs, which contain EMG bursts of 0.1 to 5 seconds each with an amplitude at least four times the amplitude of the background EMG activity.

The AASM rules[1] provide no recommendation on how many epochs scored as tonic and/or with excessive phasic activity would be considered abnormal or consistent with a diagnosis of RBD.

SINBAR Scoring Rules.[43] The scoring approach of the SINBAR group quantifies tonic, phasic, and "any" EMG activations during REM sleep, for which normative values have recently been proposed.[43]

- *Tonic EMG activity* is defined as the percentage of 30-second epochs with chin EMG activity with an amplitude at least twice the background or larger than 10 μV present for more than 50% of the epoch.
- *Phasic EMG activity* is scored as the percentage of 3-s mini-epochs containing EMG activity with duration 0.1 to 5 seconds and an amplitude that is at least twice the background EMG amplitude. The end of a phasic EMG burst is determined by a return of 0.25 second or longer to background EMG levels. To score phasic activity in the presence of tonic activity the amplitude has to be at least twice the amplitude of the tonic background as determined in the same 3-second mini-epoch and the phasic burst has to have a waxing and waning morphology.
- *"Any" EMG activation* is scored as the percentage of 2-second mini-epochs that contain EMG activity of any length with an amplitude that is larger than twice the background amplitude.

Automated Scoring of REM Sleep without Atonia

Currently, several automated scoring algorithms are available for scoring EMG activity during REM sleep[75-78] or sleep in general.[79] Of note, all of the available algorithms quantify different entities and measures of EMG activity that are not directly comparable with those derived from visual scoring. Among the algorithms evaluated in larger groups of patients with RBD and control subjects are the so-called STREAM algorithm,[80] discussed next, and the REM atonia index.[78,81]

The Supra-Threshold REM EMG Activity Metric (STREAM)[80] was proposed by Burns and coworkers. It quantifies the percentage of 3-second mini-epochs during REM sleep with increased muscle activity identified by the variance of the EMG signal, which has to be above the fifth percentile of variance values during NREM sleep. The correlation between STREAM scores and the average of percentages of epochs with tonic and phasic activity visually scored, according to Lapierre and Montplaisir, was 0.87.[45]

The *REM atonia index* (RAI) is by far the most widely used automated scoring algorithm for RSWA. Introduced by Ferri and coworkers in 2008,[78] it was improved in 2010[81] with the addition of a noise reduction technique. The RAI is based on the automated analysis of the rectified, bandpass (1 to 100 Hz) and notch (50/60 Hz) filtered submentalis muscle EMG signal. For each 1-second mini-epoch, the average amplitude is corrected for the local noise level by subtracting the minimum amplitude of the EMG signal in the ±30-second interval around it.[78] The resulting average amplitude for each 1-second mini-epoch is classified in 20 distinct categories as 1 μV or less, between 1 and 2 μV (i.e., greater than 1 μV up to 2 μV or less), between 2 and 3 μV, and so on, until the

category of between 18 and 19 μV and the final category of greater than 19 μV.

The RAI is calculated as the proportion of 1-second mini-epochs with average amplitudes of 1 μV or less, reflecting atonia, with respect to all other mini-epochs except those with average amplitudes between 1 and 2 μV, which are thought to reflect both atonia and EMG activation. The RAI can vary, ranging from 0 to 1—from complete loss of atonia and the absence of mini-epochs with average amplitudes of 1 μV or less (RAI = 0)—to complete atonia with the amplitudes of all mini-epochs 1 μV or less (RAI = 1). Aside from the RAI, the algorithm quantifies the number of movements during REM sleep, defined as the number of consecutive 1-second mini-epochs with average amplitudes greater than 2 μV, which are additionally classified by their duration in 20 distinct categories (from 1 second to more than 19 seconds).

Evaluation of the original algorithm[82] compared against visual scoring of loss of atonia and phasic density as proposed by Lapierre and Montplaisir[45] showed adequate agreement. The average correlations in four large groups of subjects—young control subjects, old control subjects, patients with iRBD, and patients with MSA—were between 0.745 and 0.963 for the REM sleep atonia index and percentage of visually scored atonia, and between 0.628 and 0.915 for the number of EMG activations and visually scored phasic density. This has recently been confirmed for the improved RAI with noise correction.[74] In approximately 80 patients with RBD and 80 healthy control subjects, the correlation between the RAI and the visually scored tonic density (definition 1a, modification 6 in Table 168-2) was 0.87.

The RAI has been used in several studies employing larger samples of healthy control subjects in different age groups[83] and of patients with RBD,[78,81,82,84-86] Parkinson disease,[87,88] or other neurologic disorders[81,82,86] and cutoff values for the diagnosis of RBD have been established.[81]

Classification of Movements during REM Sleep

The classification of movement events during REM sleep is a dynamically evolving field. Approaches include a simple listing of observed behaviors,[48] the categorization in a few broad categories,[55,56,62,89,90] classification of each event along several dimensions,[11] and the development of a standardized rating scale.[14] Examples of these approaches are given in Table 168-3.

The only standardized rating instrument currently in use is the REM Sleep Behavior Disorder Severity Scale (RBDSS).[14]

Table 168-3 Examples of Reported Descriptions/Classification of Movements during REM Sleep

REM Behavioral Manifestations:	Classification	Descriptive Categories	REM Sleep Behavior Disorder Severity Scale (RBDSS)
Gagnon et al., 2002[48] Grimaces; limb and body jerking Sitting; gesturing; reaching; kicking Grimaces; talking; limb and body jerking Talking; limb and body jerking Grimaces; limb and body jerking Grimaces; limb and body jerking Grimaces; talking; limb and body jerking; gesturing Limb and body jerking Grimaces; talking; gripping; reaching; turning; gesturing Grimaces; talking; punching; leaping	**Sforza et al., 1988**[89] Simple motor behaviors Complex motor behaviors **Kumru et al., 2004**[55] *Mild*: Excessive proximal or distal limb jerking with minimal separation from the body, head or body jerking, murmuring, whispering, groaning, smiling, repetitive mouth opening *Moderate*: Gesturing, raising the arms, turning the head abruptly >90 degrees, talking, laughing, crying, singing *Severe*: Waving the arms vigorously, kicking, punching, sitting up in bed, jumping out of bed, loud talking, shouting **Iranzo et al., 2011**[62] Movements: • Head • Right upper limb • Left upper limb • Right lower limb • Left lower limb Vocalization: • Present • Absent	**Frauscher et al., 2007**[11] 1. Type of motor event Elementary motor events • Myoclonic • Simple • Stereotypical Complex behaviors • Complex/scenic motor events • Violent motor events Orofacial events Vocalizations • With comprehensible speech • Without comprehensible speech • Without visibile motor events 2. Emotional state (for vocalizations and complex/scenic events) • Positive • Negative • Unchanged 3. Involved body parts • *Segmental*: one or more contiguous body parts • *Multifocal*: noncontiguous body parts • *Generalized*: multiple or all body parts 4. Laterality • Unilateral • Bilateral	**Sixel-Döring et al., 2011**[14] 1. Motor events 0—no visible movement but registration of RSWA 1—slight movements of the distal extremities 2—movements involving proximal extremities, complex and/or violent behaviors 3—any axial movements with the possibility of falling or observed falls 2. Vocalization 0—absent 1—present *RBDSS score*: the highest score in each of the two categories, separated by a full stop (from 0.0 to 3.1)

The scale was created by Sixel-Döring and colleagues[14] to provide an easy-to-use classification of motor events during REM sleep. As detailed in Table 168-3, based on video PSG, all motor events are classified on a scale from 0 (RSWA but no visible movements) to 3 (any axial movements with the possibility of falling or observed falls. The motor event score is the highest score observed for each patient and on each night. In addition, vocalization is classified as absent (0) or present (1). The overall RBDSS score for a single night and for an individual patient is the combination of the motor score with the vocalization score, usually separated by a full stop (period). The score ranges from 0.0—that is, RSWA but no visible movements and no vocalizations—to 3.1 observed axial movements with either an observed fall or the distinct possibility of a fall and vocalizations. This scale has since been used in an increasing number of studies.[9,85,91,92]

Diagnostic Thresholds for REM Sleep Behavior Disorder

Some motor activity during REM sleep commonly occurs in healthy sleepers, mostly in the form of short, transient muscles activity and twitches.[1] Because motor activity during REM sleep forms a continuum ranging from occasional transient muscle activity to complete loss of atonia, it is essential to establish cutoff values that distinguish RSWA from physiologic REM sleep movement activity and can serve as a diagnostic threshold for the diagnosis of RBD. Of importance, neither the ICSD3[3] nor the AASM manual[1] provides definite cutoff values for RBD diagnosis.

One important methodologic challenge in the establishment of such thresholds is the need to independently replicate and validate them in independent samples. Usually, thresholds are identified as a cutoff value in a discovery sample that is optimized with regard to sensitivity or specificity, or their combination. Accordingly, a diagnostic marker developed and evaluated using a single clinical dataset will tend to overestimate sensitivity and/or specificity. The importance of these replications has recently been nicely illustrated by the study by Frauscher and associates[93] that explored sleep-related movements in a large sample of well-screened, healthy adult sleepers of a large age range. The same group of investigators had previously reported[43] normative values for various measures of RSWA with cutoff values in their discovery sample optimized to a specificity of 1; that is, none of the non-RBD cases had a value above the cutoff values. In the subsequent study[93] in healthy subjects, the specificity was considerably lower with 8% to 25% of healthy sleepers showing RSWA values above the previously established cutoff values (see Table 168-4). The only RSWA cutoff that continued to show a specificity of 1 was a value of 9.6% for tonic chin EMG activity.

Table 168-4 lists studies that have proposed RSWA cutoff values for the diagnosis of RBD and have been independently replicated to some extent.[43,44,74,81,94,95] Other cutoff values for different RSWA measures have been proposed[49,94,96] but await independent confirmation. As is apparent from Table 168-4, there is a tradeoff between *sensitivity*—here, the ability to correctly identify patients with RBD—and *specificity*—the correct exclusion of non-RBD cases. In addition, the table also illustrates that sensitivity and specificity can vary considerably between samples, with elderly subjects who do not have RBD posing a specific challenge.

Unresolved Issues

The scoring of RSWA is complicated by the particularly large variety in applied scoring criteria. Currently unknown is what—if any—is the magnitude of the effect that these variations have on the respective RSWA measures and on their ability to distinguish between patients with RBD and healthy subjects. Variations seem to be minor for the assessment of tonic EMG activity during REM sleep, because a large consensus favors use of the chin EMG and an epoch-wise quantification. For phasic EMG, by contrast, a rather substantial variation in applied scoring rules is evident, and standardized international criteria would greatly benefit the field. In addition, it must be stressed again that currently neither the ICSD3[3] nor the AASM manual provides any quantitative cutoff values for the diagnosis of RBD.[1]

SLEEP-RELATED BRUXISM

Sleep-related bruxism refers to regular or frequent teeth grinding during sleep (see Chapter 144). According to the ICSD3,[3] the diagnostic criteria for SB are the presence of regular or frequent teeth grinding sounds occurring during sleep and either the presence of abnormal teeth wear consistent with this report or transient morning jaw muscle pain or fatigue or temporal headache or jaw locking on awakening.

Several approaches and protocols are available to record and score SB activity, including the following:

- The ambulatory recording of masticatory muscle EMG (mmEMG), in which the identification and scoring of nocturnal bruxism activity are based only on the mmEMG signal[97,97-100]
- The ambulatory recording of mmEMG and heart rate with identification and scoring of nocturnal bruxism activity based on criteria for mmEMG and heart rate[101-103]
- Ambulatory PSG with additional recording of mmEMG but without video-audio recording, in which scoring of SB activity is based on mmEMG[104] or mmEMG and heart rate[105] during the PSG-identified sleep period
- AASM recommendations[1]: standard PSG with video-audio recording including chin EMG and only optional recording of mmEMG, in which SB activity is identified by chin EMG or chin and mmEMG activity and teeth grinding episodes are identified by audio recording during the sleep period
- Research diagnostic criteria (RDC)[106,107]: standard PSG with video-audio recording and mmEMG, in which SB activity is identified and scored on the basis of mmEMG and video-audio recordings during the sleep period

The main difference between these approaches is in the availability and/or use of video-audio recording to distinguish between SB activity and other nonspecific orofacial movements.

Orofacial movements with visible increases in mmEMG are very common during a night of sleep both in healthy sleepers and subjects with SB. Many of these movements may be unspecific and unrelated to SB activity but are indistinguishable from the mmEMG activation patterns that are used to define SB activity. Some examples of nonspecific movements are given in Table 168-5. In general, and as detailed subsequently, SB activity is being identified by EMG activations above a specified threshold that are either series of phasic, short activations or sustained tonic activations or

Table 168-4 Evaluation of Diagnostic Cutoff Values for the Diagnosis of REM Sleep Behavior Disorder

Measure*	Muscle	Cutoff	Study	Sample	Sens.	Spec.	PPV	NPV	FP%	FN%
Automated Measures										
REM atonia index (RAI)	Chin	<0.9	Ferri et al., 2010[81]	39 patients with iRBD vs. 10 a-m control subjects	0.74	0.7	0.91	0.41	30	26
				39 patients with iRBD vs. 35 control subjects	0.74	0.91	0.91	0.76	9	26
			Ferri et al., 2012[95]	16 patients with PD+RBD vs. 11 PD-RBD	0.94	0.91	0.94	0.91	9	6
				16 patients with PD+RBD vs. 19 a-m control subjects	0.94	0.84	0.83	0.89	6	6
			Ferri et al., 2014[74]	74 patients with RBD vs. 75 a-m control subjects	0.96	0.51	0.66	0.93	49	4
		<0.88	McCarter et al., 2014[94]	20 patients with PD+RBD vs. 40 control subjects	0.95	0.92	0.86	0.95	8	5
Visual Measures										
Tonic EMG density	Chin	≥30%	Montplaisir et al., 2010[44]	80 patients with RBD vs. 80 a-m control subjects	0.74	0.90	0.88	0.77	10	26
			Frauscher et al., 2014[93]	100 healthy sleepers		1.0			0	
Phasic EMG density	Chin	≥15%	Montplaisir et al., 2010[44]	80 patients with RBD vs. 80 a-m control subjects	0.80	0.88	0.86	0.81	12	20
			Frauscher et al., 2014[93]	100 healthy sleepers		0.64			36	
Tonic EMG	Chin	>9.6%	Frauscher et al., 2012[43]	30 patients with RBD (i + PD) vs. 30 s-a-m control subjects		1.0	1.0		0	
			Frauscher et al., 2014[93]	100 healthy sleepers		1.0	1.0		0	
"Any" mentalis, 30 s	Chin	>15%	Frauscher et al., 2012[43]	30 patients with RBD (i + PD) vs. 30 s-a-m control subjects		1.0	1.0		0	
			Frauscher et al., 2014[93]	100 healthy sleepers		0.85			15	
"Any" mentalis, 3 s	Chin	> 18%	Frauscher et al., 2012[43]	30 patients with RBD (i + PD) vs. 30 s-a-m control subjects		1.0	1.0		0	
			Frauscher et al., 2014[93]	100 healthy sleepers		0.80			20	
SINBAR, 3 s	Chin + FDS	>32%	Frauscher et al., 2012[43]	30 patients with RBD (i + PD) vs. 30 s-a-m control subjects		1.0	1.0		0	
			Frauscher et al., 2014[93]	100 healthy sleepers		0.92			8	
SINBAR, 30 s	Chin + FDS	>27%	Frauscher et al., 2012[43]	30 patients with RBD (i + PD) vs. 30 s-a-m control subjects		1.0	1.0		0	
			Frauscher et al., 2014[93]	100 healthy sleepers		0.92			8	

*Definitions of visually scored measures refer to those listed in Table 168-2: Tonic EMG density–def. 1a, mod. 5; Phasic EMG density–def. 2a, mod. 7; Tonic EMG–def. 1d; "Any" mentalis, 3 s–def. 3f; "Any" mentalis, 30 s–according to def. 3f but with % of 30-s epochs with >50% "Any" EMG activity; SINBAR, 3 s, 30 s–"Any" mentalis + phasic activity in FDS (def. 2d).

a-m, Age-matched; FDS, flexor digitorum superficialis muscle; FN%:,false negative rate; FP%, false positive rate; iRBD, idiopathic REM sleep behavior disorder; Narc, narcolepsy; NPV, negative predictive value; PD, Parkinson disease; PPV, positive predictive value; RBD, REM sleep behavior disorder; s-a-m, sex- and age-matched; Sens., sensitivity; Spec., specificity.

both—that is, a mixed episode. The total set of all mmEMG activations during sleep, then, has the following components:

A: activations that do not fulfill EMG criteria for SB episodes

B: activations that fulfill EMG criteria for SB episodes but are nonspecific movements such as coughing or swallowing (see Table 168-5) as identified with audio-video recording

C: activations that fulfill EMG criteria and are indeed SB episodes

Table 168-5 Examples of Orofacial Movements and Sounds that May Be Confused with Sleep-Related Bruxism Activity

Movements	Sounds
Swallowing	Coughing
Coughing	Grunting
Yawning	Snoring
Lip and tongue movements	Sleeptalking
Eye blinking	Tooth tapping
Light head movements	Temporomandibular joint clicking
Head rubbing or scratching	Tongue clicking
Lip sucking	Throat clearing

Rhythmic masticatory muscle activity (RMMA) is a subset of SB episodes (as defined for component C) that contains only the phasic and mixed episodes (see later). Within the complete set of mmEMG activations (A + B + C, or B + C), "true" SB episodes account for only 15% to 30% in healthy subjects and for 55% to 70% in patients with SB.[106,108,109] Audio-recording can help identify SB episodes with audible teeth grinding sounds but less than 30% of SB episodes are accompanied by teeth grinding noise.[106,110,111] For this reason, mmEMG recording with audio-video recording is the gold standard for assessment of SB activity, and the nonavailability of audio-video signals is a major limitation for the ambulatory assessment of SB.

The magnitude of misclassification of mmEMG activations as SB activity when audio-video recording is not available can be estimated from two recent studies. In healthy sleepers, Yamaguchi and colleagues compared RMMA episodes identified by the gold standard (video-PSG with

Table 168-6 Recording and Scoring of Sleep-Related Bruxism (SB) Using Research Diagnostic Criteria (SB-RDC), American Academy of Sleep Medicine (AASM) Criteria, and Rhythmic Masticatory Muscle Activity (RMMA)*

Feature/Component	SB-RDC[106,107]	RMMA[110]	AASM Criteria[1]
Electrodes for SB recording	Masseter muscle Temporalis muscle optional	Masseter muscle Temporalis muscle optional	Chin muscle Masseter muscle optional
Additional signals	Standard PSG signals	Standard PSG signals	Standard PSG signals
Audio-video recording	Yes	Yes	Yes
Sampling rate	≥128 Hz	≥128 Hz	≥200 Hz
Filter	Not specified	Not specified	10–100 Hz
Main outcome measure	SB episodes	RMMA	Bruxism
Exclusion of nonspecific activity by audio-video recording from scoring of main outcome	Yes	Yes	No
Amplitude threshold to identify EMG burst	≥20% of MVC[106,178] or >10% of MVC[110]	>10% of MVC	>2 × background
Included episodes	*Phasic*: ≥3 bursts, 0.25–2 s, separated by <3 s *Tonic*: ≥1 burst, >2 s, separated by <3 s *Mixed*: phasic and tonic episodes separated by <3 s	*Phasic*: ≥3 bursts, 0.25–2 s, separated by <3 s *Mixed*: phasic episode separated by <3 s from one or more tonic burst (>2 s)	*Bruxism, brief elevations*: ≥3 bursts, 0.25–2 s, in a regular sequence *Bruxism, sustained elevations*: >2 s Bruxism episodes separated by <3 s of stable background are scored as 1
Reported outcomes	Number of SB episodes/h % of SB episodes with teeth grinding sound Number of bursts/h	Number of RMMA episodes/h % of RMMA episodes with teeth grinding sound	

*See text for details.
EMG, Electromyogram; MVC, maximal voluntary contraction (during wake); RCD, research diagnostic criteria; RMMA, rhythmic masticatory muscle activity; SB, sleep-related bruxism.

mmEMG) with events recorded with a telemetric masseter muscle EMG device.[112] Both the telemetric mmEMG signal and the mmEMG signal of the PSG study were scored manually using the same criteria. Video-audio recording was then used to exclude nonspecific events from the scoring of RMMA episodes. The telemetric device correctly identified close to 99% of RMMA episodes identified by video PSG. However, 77% of all episodes identified by the telemetric device were other oromotor events and not SB-related. The positive predictive value (PPV) was therefore only 0.23—that is, the probability that an event identified with the telemetric device was indeed an RRMA episode was only 23%. When only the sleep period was considered, with information provided from video PSG, the PPV increased to 0.52.

These results correspond with those in another study in patients with SB,[113] in which video PSG was scored twice, without and with access to the audio-video recording. Thus the study was designed to compare PSG with video PSG, and the scorer had access to the complete PSG recording, including electroencephalogram (EEG) and electromyogram (EMG) recordings for other muscle groups, and used this information to distinguish RRMA from other orofacial activities. With use of video PSG, 33% of all masseter muscle EMG activations did not satisfy EMG criteria for RRMA. Of EMG activations that satisfied RRMA EMG criteria, approximately 69% represented other nonbruxism activity. These would have been false-positive identifications if only the masseter EMG activity had been available for scoring (PPV, 0.31). With use of the complete information provided by the PSG recording (with the exception of audio-video), the number of RMMA events was overestimated by only 23.8% relative to that for video PSG–RDC (PPV, 0.81)—a considerable improvement. Misclassification of nonspecific EMG activity, however, was still pronounced for tonic EMG activations: Only 41% of episodes identified without audio-video were indeed "true" RMMA events (PPV, 0.41).

Recording Methods

Surface EMG electrodes placed on masticatory muscles are the key component in assessment of SB activity. The different protocols vary in the selection and number of masticatory muscles, but most include the masseter muscles. Masseter muscles are easily identified by the muscle bulges between the cheekbone and the angle of the jaw during teeth clenching. The registration of other masticatory muscles, such as the temporalis muscle, often is optional but may increase reliability of SB activity identification.[114] Electrode positioning for the temporalis muscle usually is 1 to 2 cm above the zygomatic arch (cheek bone) and 1 to 2 cm behind the outer canthus orbital border (the cavity of the skull in which the eye is situated). The following recommendations have been made concerning the recording of SB activity:

- *ICSD3 recommendations*:[3] Although PSG is not strictly necessary for the diagnosis of SB, when SB activity is to be recorded and scored, the minimal requirement is one masseter muscle recording. Audio-video recording is strongly recommended to exclude nonspecific movements from the scoring of SB activity. For optimal diagnostic specificity and sensitivity, bilateral masseter and temporalis muscle EMG recordings, referenced to ear, mastoid, or zygomatic bone, are advised.
- *AASM criteria*:[1] AASM recommendations refer to in-laboratory PSG with the recommended parameters (EEG, EOG, ECG, EMG, respiration, oxygen saturation, body position), which include EMG of the chin muscles. Scoring rules for SB have been formulated for chin EMG only; the addition of masseter muscle recording is left to the discretion of the clinician. In addition, the audio signal is used to identify SB episodes with teeth grinding sounds.
- *RDC*[106]: Standard PSG with video-audio recording and bilateral masseter EMG is the minimum requirement for the scoring of SB activity according to RDC. The addition of temporalis muscle recording is recommended but not mandatory.[115]
- *Ambulatory recording*: There are no generally accepted recommendations for the ambulatory recording and scoring of SB activity. Currently used recording setups range from a single, unilateral EMG channel[100,112] to ambulatory PSG with additional mmEMG.[104,105]

Scoring Rules

Sleep Bruxism Episodes Scored According to Research Diagnostic Criteria

The scoring of SB activity according to RDC (SB-RDC)[106] involves the following general steps:

1. Identify all masseter muscle activations with mean amplitude above a predefined threshold.
2. Classify the identified EMG activations on the basis of their duration as myoclonic (less than 0.25 second), phasic bursts (0.25 to 2 seconds), or tonic bursts (more than 2 seconds).
3. Use video and audio recording to exclude non–(bruxism)-specific oromandibular activity.
4. Classify the remaining phasic bursts as belonging to a phasic episode if 3 or more phasic bursts are separated by less than 3 seconds; the remaining tonic bursts are considered tonic episodes.
5. Combine phasic and/or tonic episodes to one episode if they are separated by less than 3 seconds.
6. Classify episodes as phasic, tonic (episode with one or more tonic bursts), or mixed (episode with tonic and phasic episodes).
7. Characterize SB episodes regarding the presence or absence of audible teeth grinding sound, as recorded by audio.

These scoring criteria have been widely adopted in research studies. The only major variation concerns mmEMG threshold: In the original publication in 1996,[106] the mmEMG threshold had been set at 20% of background signal during maximal voluntary clenching (MVC) in the wake state. Since then, and based on later studies by the same group of investigators, a modification of this threshold to 10% MVC has been proposed and is generally used.[107,115]

Descriptive summary measures for SB activity include the number of phasic, mixed, tonic, and all SB episodes either as a total number per night or as the number per hour. In addition, the number of bursts per hour and the proportion of SB episodes with audible teeth grinding sound usually are reported.

Rhythmic Masticatory Muscle Activity

The scoring of rhythmic masticatory muscle activity (RMMA)[110] follows the outline described for the scoring of

SB episodes according to RDC, with the exception of two aspects:

- The threshold to identify mmEMG activations is 10% MVC.
- RMMA includes only phasic and mixed SB episodes as defined previously.

Because only approximately 10% to 20% of SB-RDC episodes are classified as tonic episodes,[106,116,117] a large overlap between RMMA and SB-RDC is to be expected. Perhaps because of this overlap, the distinction between RMMA and SB-RDC has become blurred over time, as evidenced by the fact that more and more recent research studies now include the tonic episodes in the scoring of RMMA.[113,118,119] In addition, in research studies, the tendency is to adopt a mmEMG threshold that is twice the background EMG,[113,120] as has been proposed by the AASM rules[1] (see further on).

RMMA usually is reported as the number of RMMA episodes per hour and the percentage of RMMA episodes with audible teeth grinding sound.

Sleep Bruxism Episodes Scored According to AASM Criteria

The AASM has specified scoring rules for bruxism based on chin EMG activity.[1] Bruxism is scored in accordance with the following criteria:

1. Chin EMG elevation is at least twice the amplitude of the background EMG, *and*
2. a. *Either* at least three brief, phasic chin EMG activity elevations (lasting 0.25 to 2 seconds) occur in a regular sequence,
   b. *Or* a sustained, tonic chin EMG activity elevation (lasting longer than 2 seconds) is present.
3. Chin EMG elevations that fit these criteria are counted as one episode when they are separated by less than 3 seconds.

Of note, audio-video recording is not used to exclude nonspecific orofacial movements from the scoring of bruxism episodes. Nevertheless, some use is made of the audio recording, because the manual states that "bruxism can be scored reliably by audio in combination with polysomnography by a minimum of 2 audible teeth grinding episodes/night of polysomnography (section VII.E).[1] Currently, the AASM manual provides no information about the reporting of bruxism activity.

Sleep Bruxism Episodes Described in the Current International Classification of Sleep Disorders

As mentioned previously, according to the ICSD3,[3] PSG documentation of SB activity is not a necessary diagnostic feature. Nevertheless, in the ICSD3 "Objective Findings" paragraph, some features of PSG patient monitoring are described, including minimum standards for recording that include the registration of at least one masseter muscle EMG (see earlier). According to the ICSD3,[3] three EMG patterns of SB episodes are described:

1. a. Phasic activity at 1-Hz frequency with EMG bursts of duration 0.25 to 2 seconds,
   b. Sustained, tonic activity with duration 2 seconds or longer, and
   c. Mixed patterns
2. SB episodes begin after 3 seconds or longer with no muscle activity.

Audio-video recording is recommended to identify teeth grinding sound and to distinguish SB from other orofacial or masticatory movements that normally occur during sleep, and from specific movement disorders.[3]

Sleep Bruxism Activity in Ambulatory Recordings

Currently, no generally accepted or validated criteria are available for the scoring of SB activity in ambulatory recordings without video-audio recording. Consequently, a wide variation in scoring criteria are in current use. Many, however, seem to be inspired either by the RDC formulated for video PSG[106] (see earlier) or by the criteria proposed by Ikeda and colleagues in 1996[101] for the identification of bruxism events based on ambulatory masseter muscle EMG and heart rate. These investigators[101] proposed to score as bruxism those events with the following features:

1. Masseter muscle EMG activity increases greater than 10% MVC for 3 seconds or longer
2. Separated by more than 5 seconds from other events—otherwise, events are combined
3. Accompanied by a change in heart rate greater than 5% over a 5-second period, compared with preevent baseline

A point worthy of emphasis is that these criteria were based almost exclusively on theoretical considerations and have never been empirically validated.

Scoring criteria used in research studies and in commercially available devices differ in the following ways:

- The amplitude thresholds for the identification of mmEMG activations (e.g., greater than 10% MVC[99,102,103,121-123]; greater than 30% MVC[124])
- The duration of identified mmEMG activations (e.g., longer than 0.25 second[99,102,103,121,122]; 0.25 to 2 seconds[125]; longer than 3 seconds[123,124])
- The interval between identified events (e.g., greater than 2 seconds[99]; greater than 3 seconds[121,122]; greater than 5 seconds[123,124])
- The classification[99,102,103,121] or nonclassification[122-124] of mmEMG activations in phasic, tonic, or mixed events

Partly responsible for this variation is the lack of validation studies for ambulatory SB assessment. Because both the gold standard and most ambulatory devices record the same signal, (i.e., mmEMG), measurement reliability per se is not the issue. The mmEMG signal from ambulatory devices has been shown to be in good agreement with the mmEMG signal from ambulatory or in-laboratory mmEMG. The main concern in validating ambulatory SB assessments is the lack of video-audio recording, with the probable misclassification of many mmEMG activations as SB episodes (see earlier). Two basic approaches to this dilemma can be postulated. In one approach, a direct comparison between the events identified by ambulatory assessments with the events identified by the gold standard, which includes video observation and a demonstration of adequate specificity in the identification. Several attempts have been made to develop automated classifiers that recognize different oromandibular movements from the EMG signal alone and could therefore distinguish between SB and nonspecific movements. So far, however, these have been tested with only simulated movements during wakefulness.[100,126-131] The second approach would be the validation of criteria that are specific to ambulatory assessments and take the lack of specificity into account. Such a validation study would establish

the sensitivity and specificity of new diagnostic criteria in distinguishing between healthy subjects and patients with SB, previously diagnosed according to the gold standard.

Research Diagnostic Criteria

Currently, only research diagnostic criteria are available for the identification of patients with moderate to severe SB.[106] In the pivotal study by Lavigne and coworkers,[106] using the RDC scoring criteria detailed earlier, various parameters of SB activity were assessed in 18 healthy sleepers and in 18 patients with moderate to severe SB. By considering sensitivity and specificity of different combinations and cutoff values, the following criteria were identified:

A + (B1 or B2 or C or D)

with:

A indicating 2 or more bruxism episodes with teeth grinding sounds/night
B1, >30 episodes/night
B2, >4 episodes/hour
C, >6 bursts/episode
D, >25 bursts/hour

The combination of criterion A with any one of the other criteria—B1, B2, C, or D—had a positive predictive value (PPV) (i.e., the probability that a subject identified with these criteria does indeed have SB) between 93% and 100% and a negative predictive value (NPV) (i.e., the probability that in a subject identified as not meeting the criteria for-SB, tooth grinding activity does not in fact occur) between 76% and 93%.[106] These criteria have been widely adopted[107]; over time, however, the most used and recommended combination of criteria is A + (B2 or D).[107,115]

The ability of these criteria to discriminate between patients with SB and healthy sleepers has recently been evaluated in a larger sample of 100 patients diagnosed with SB and 43 healthy sleepers.[116] The recording and scoring of SB activity were identical in both studies, with two critical differences: First, although general inclusion and exclusion criteria were identical, the original study included patients with SB who reported the nocturnal teeth grinding sounds at least 5 times per week during the last 6 months; by contrast, the other study included patients with SB reporting nocturnal teeth grinding sounds at least 3 times per week during the last 6 months—that is, with a wider spectrum of SB frequency. In addition, the diagnostic criteria employed in the study were any two of the following three: (A) two or more bruxism episodes with teeth grinding sounds per night, (B2) more than 4 episodes/hour, and (D) more than 25 bursts/hour, that is, different from the original study criterion (A) was no longer mandatory. With these modifications, only 54 of 100 patients with SB fulfilled the RDC criteria for SB, as did 9 of 43 healthy sleepers. Consequently, the sensitivity of the RDC was only 54% (PPV, 86%) and the specificity was 79% (NPV, 42%). These results emphasize the continued need for large and systematic studies in this field.

Unresolved Issues

The major unresolved issue in the recording of SB is currently the scoring without the availability of video-audio recording. As detailed earlier, nonexclusion of unspecific orofacial movements will considerably overestimate SB activity. In addition, diagnostic criteria include the observation of episodes with teeth grinding sounds, which also are not confirmable without at least audio recording.

Other issues concern the diagnostic criteria, which may lack sensitivity or specificity when applied with variations, and the inconsistent scoring of RMMA.

RHYTHMIC MOVEMENT DISORDER

Rhythmic movement disorder (RMD) is characterized by stereotypical rhythmic body movements occurring predominantly during sleep or drowsiness before sleep, which may involve the head, neck, trunk, or limbs in isolation or in combination. The criteria for scoring RMD are based on their PSG features[1]: frequency of 0.5 to 2.0 Hz; presence of at least four single movements, as required to form a rhythmic cluster; and a minimum amplitude for a single rhythmic movement that is twice the background EMG activity. In healthy infants or young children, the movements and behaviors frequently are self-limited and are called "benign rhythmic movements of sleep," to distinguish them from RMD[132]; however, no strict quantitative threshold has been established. The duration of rhythmic movement episodes may vary, ranging from some seconds to several minutes, and can be observed as well during wakefulness preceding sleep or after sleep onset. The most common pattern of RMD involves the head (head banging, or *jactatio capitis nocturna,* and head rolling), but also the body (body rocking) or occasionally the legs (leg rolling or leg banging) can be involved. The clinical diagnosis can be based on anamnesis, sometimes supported by homemade video recording; however, video PSG is useful in doubtful cases and allows definition of the type and site of movements. Additional EEG and EMG leads, in particular on the limbs, may help to distinguish RMD from other sleep-related repetitive movements (e.g., PLMS, ALMA) or motor seizures.

PROPRIOSPINAL MYOCLONUS

Propriospinal myoclonus (PSM) at sleep onset[133,134] is characterized by generalized and symmetric jerks, occurring at the wake-sleep transition, starting from the axial muscles of the abdomen, thorax, or neck and then spreading rostrally and caudally to the other myotomes by means of slow propriospinal polysynaptic pathways.[135] PSM is facilitated by relaxed wakefulness or drowsiness and inhibited by mental activation and sleep deepening. In PSM, there is no urge to move; however, the frequently repeated (but not periodic) jerks prevent the patient from falling asleep and reaching a deeper sleep stage, when the jerks disappear. No accepted quantitative PSG features have been documented for PSM, and its description is basically qualitative.

BENIGN SLEEP MYOCLONUS OF INFANCY

Benign sleep myoclonus of infancy (BSMI) is characterized by repetitive myoclonic jerks that occur during sleep in neonates and infants.[3] Its diagnostic features include the observation of repetitive myoclonic jerks involving limbs, trunk, or whole body. This movements occur only during sleep and stop abruptly and consistently when the infant is aroused. BSMI occurs in early infancy, typically from birth to 6 months of age.

Standardized scoring criteria for BSMI are lacking, but the muscle jerks last between 0.04 to 0.3 second and occur usually in clusters of 4 to 5 jerks per second.[3,136] These clusters may repeat in irregular series for 1 to 15 minutes, and in rare cases up to 60 minutes. The jerks often are bilateral and typically involve large muscle groups. BSMI is benign and relatively rare, but it has been included in the ICSD3 as a sleep-related movement disorder because it often is confused with epilepsy.[3]

MISCELLANEA

Excessive Fragmentary Myoclonus

Excessive fragmentary myoclonus (EFM)[137-139] is characterized by small EMG activations, not always corresponding with movements (twitches), of the fingers, toes, or corners of the mouth, resembling either physiologic hypnic myoclonus or fasciculations, occurring at the sleep-wake transition or during sleep. PSG recordings show recurrent and persistent, very brief (75 to 150 milliseconds) EMG bursts in various muscles, occurring asynchronously and asymmetrically, in a sustained manner, without clustering. The PSG criteria for scoring EFM are as follows: EMG burst duration usually is 150 milliseconds or less; EFM must be present in at least 20 minutes of NREM sleep; and at least 5 EMG potentials per minute must be recorded.[1] Regarding quantification, it has been proposed to quantify EFM rate as the number of EMG potentials of 150 milliseconds or less divided by the minutes of sleep.[139] Alternatively, a *myoclonus index* has been proposed by Lin and coworkers[140]; this index is defined as the number of 3-second mini-epochs containing at least one fragmentary myoclonus potential fulfilling the criteria, counted for each 30-second epoch, with totals ranging between 0 and 10. A mean fragmentary myoclonus index of 39.5 events per hour of sleep has been reported in patients with sleep disorders, and this index increased with age, was higher in men than in women, and was influenced by the presence of sleep breathing disorders.[141] The male predominance was recently confirmed in a large, well-screened sample of healthy sleepers.[93] Some fragmentary myoclonus was observed in all participants, however, and 9% even fulfilled criteria for EFM, suggesting that the criteria for quantification of EFM may benefit from an appropriate statistical revision.

Neck Myoclonus (during Sleep)

Neck myoclonus (head jerks) during REM sleep has been described in recent years[142] and is recognized by presence of a characteristic high-amplitude, short ("stripe-shaped") movement-induced artifact over the EEG lead signals that is verified in the video recording, which shows sudden myoclonic twitches of the head of variable intensity. It seems to be a frequent finding on routine PSG, being present in more than 50% of patients, but with a low frequency during the night (1 ± 2.7 events per hour of REM sleep) and an age-related decline. Neck myoclonus has been observed in 35% of healthy subjects, with variable frequency,[93] suggesting that it constitutes a physiologic phenomenon.

Other Leg Movements during Sleep

Alternating leg muscle activation (ALMA) indicates a rapidly alternating (frequency, 0.5 to 3 Hz) pattern of anterior tibialis activation (duration, 100 to 500 milliseconds) occurring during sleep, organized in sequences of at least 4 alternating activations,[1] lasting up to 20 to 30 seconds. ALMA can occur in all sleep stages but is seen particularly during arousals.[143,144] It is not yet clear whether ALMA is a separate nosologic entity or whether it belongs to the wide spectrum of nocturnal motor activities of RLS.

Hypnagogic foot tremor (HFT)[145,146] is a clinical condition similar to ALMA, with foot movements occurring at the transition between wake and sleep or during light sleep. PSG recordings show repeated short EMG potentials, typically at 1 to 2 Hz (range, 0.3 to 4Hz), in one or both feet, in sequences of at least 4.[32] The EMG events appear to be of longer duration than myoclonus (more than 250 milliseconds) and usually last less than 1 second; moreover, they are organized in trains lasting 10 or more seconds.[3]

The term *high-frequency leg movements* (HFLMs) has been proposed more recently for a phenomenon similar to both ALMA and HFT.[147] It was defined as a sequence of 4 or more short leg movements occurring, unilaterally but sometimes bilaterally, at a frequency of 0.3 to 4 Hz. Most HFLMs are observed during wakefulness, with only approximately one third occcurring during sleep. Thus far, no clear criteria have been established to score this phenomenon.

ALMA, HFT, and HFLMs are very similar phenomena of small or short EMG activations of the feet and legs: They all occur at sleep onset and are correlated with arousals, and they usually occur in trains. It is not clear yet if these phenomena are really separate entities or if they constitute slightly different definitions of the same condition; moreover, it is not yet established to what extent they are correlated with PLMS and/or RLS. In addition, it is unknown whether these are simple, physiologic phenomena or have any pathophysiologic significance. A case in favor of the first possibility is the observation that HFLMs have been found in 33% of well-screened, healthy sleepers.[93]

CLINICAL PEARLS

- Video polysomnography is the gold standard for recording and scoring sleep-related movements.
- For many sleep-related movements, more than one system of criteria or definitions exist. It is therefore essential to specify which scoring rules are applied.
- Scoring rules may and do change over time. In particular, those of the AASM are continuously reviewed and updated on the basis of new clinical evidence or advances in technology.
- The current *International Classification of Sleep Disorders* (ICSD-3) refers to the manual of the American Academy of Sleep Medicine (AASM) for the definition of sleep-related movements.

SUMMARY AND FUTURE DIRECTIONS

As reviewed in this chapter, a variety of clinical and research recording and scoring techniques have been applied to define, quantify, or diagnose sleep-related movements. By now, the use of computers and digital recordings is standard in sleep medicine and research. This approach offers numerous advantages, among them the precise measurement of sleep-related movement events. Many criteria to visually detect and measure motor events during sleep were introduced during the "paper era" of sleep medicine and continue to be the basis of the current rules. With the easy availability of very powerful

computers nowadays, it could be expected that such criteria would have become obsolete and substituted by more modern approaches that take full advantage of the digital possibilities. Nevertheless, many old rules still pervade the field and often are used with a simple transposition from the paper to the screen. The adoption of new approaches based on more quantitative and data-driven methods has been slow. The necessary programs and independent validations are prerequisite to effect change. In view of the strengths (e.g., objectivity and precision) of the new approaches, however, they can be expected to gradually enhance and ultimately replace many of the scoring techniques presented in this chapter.

Selected Readings

Berry R, Brooks R, Gamaldo C, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, version 2.2*. Darien (Ill.): American Academy of Sleep Medicine. Available at: <www.aasmnet.org>; 2015.

Ferini-Strambi L, Rinaldi F, Giora E, et al. REM sleep behaviour disorder. *Parkinsonism Relat Disord* 2016;**22**(Suppl. 1):S69–72.

Ferri R. The time structure of leg movement activity during sleep: the theory behind the practice. *Sleep Med* 2012;**13**:433–41.

Ferri R, Rundo F, Zucconi M, et al. Putting the periodicity back into the periodic leg movement index: an alternative data-driven algorithm for the computation of this index during sleep and wakefulness. *Sleep Med* 2015;**16**(10):1229–35.

Ferri R, Zucconi M, Manconi M, et al. Computer-assisted detection of nocturnal leg motor activity in patients with restless legs syndrome and periodic leg movements during sleep. *Sleep* 2005;**28**:998–1004.

Frauscher B, Gabelia D, Mitterling T, et al. Motor events during healthy sleep: a quantitative polysomnographic study. *Sleep* 2014;**37**:763–73.

Frauscher B, Iranzo A, Högl B, et al. Quantification of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder. *Sleep* 2008;**31**:724–31.

Fulda S, Plazzi G, Ferri R. Scoring atonia during normal and pathological rapid eye movement sleep: visual and automatic quantification methods. *Sleep Biol Rhythms* 2013;**11**:40–51.

Howell MJ, Schenck CH. Rapid eye movement sleep behavior disorder and neurodegenerative disease. *JAMA Neurol*. 2015;**72**(6):707–12.

Lavigne GJ, Rompré PH, Montplaisir JY. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. *J Dent Res* 1996;**75**:546–52.

Plante DT. Leg actigraphy to quantify periodic limb movements of sleep: a systematic review and meta-analysis. *Sleep Med Rev* 2014;**18**(5):425–34.

Van Der Zaag J, Lobbezoo F, Visscher CM, et al. Time-variant nature of sleep bruxism outcome variables using ambulatory polysomnography: implications for recognition and therapy evaluation. *J Oral Rehabil* 2008;**35**:577–84.

Yamaguchi T, Abe S, Rompré PH, et al. Comparison of ambulatory and polysomnographic recording of jaw muscle activity during sleep in normal subjects. *J Oral Rehabil* 2012;**39**:2–10.

Zucconi M, Ferri R, Allen R, et al. The official World Association of Sleep Medicine (WASM) standards for recording and scoring periodic leg movements in sleep (PLMS) and wakefulness (PLMW) developed in collaboration with a task force from the International Restless Legs Syndrome Study Group (IRLSSG). *Sleep Med* 2006;**7**:175–83.

A complete reference list can be found online at ExpertConsult.com.

Chapter 169

Evaluating Sleepiness

Max Hirshkowitz; Amir Sharafkhaneh

Chapter Highlights

- Evaluating sleepiness presents a difficult task for clinicians. Although sleepiness generally is regarded as an introspectively judged sensation, sleepiness assessment also must consider physiologic sleep drive and behavioral manifestations.
- Common methods used to assess introspective, physiologic, and manifest sleepiness include self-administered questionnaires, the Multiple Sleep Latency Test, the Maintenance of Wakefulness Test, the Epworth Sleepiness Scale and vigilance testing.
- Practical issues emerge in evaluating sleepiness in the regulatory, legal, and adjudication arenas. These include establishing the goal of testing, recognizing the specific purpose of a test, interpreting results, and considering safety and on-the-job performance.

OVERVIEW

Sleepiness is a sensation. Much like hunger and thirst, sleepiness occurs naturally as a physiologic drive informing the behavioral system about a presumed biologic need. Sleepiness naturally occurs at night, after prolonged wakefulness (sleep deprivation), or in response to rapid translocation across time zones. However, sleepiness also can arise from medical, neurologic, or psychiatric disorders. Many drugs also can provoke sleepiness, either by stimulating sleep-inducing mechanisms or by inhibiting wake-promoting brain centers.

Sleepiness (and wakefulness, for that matter) represents the composite output of different neurologic systems producing wakefulness and those generating sleep. Thus sleepiness reflects a dynamic balance between systems with opposing functions. When an increased physiologic drive to sleep begins to overwhelm the alerting system's ability to stave off that drive, sleepiness becomes excessive. The affected person, however, may or may not sense being sleepy.

Nonetheless, when excessive, drowsiness poses a potential hazard. The danger affects not only the drowsy person but possibly also family members, coworkers, and society at large. Excessive sleepiness during driving and work activities can constitute a serious, potentially life-threatening condition. When sleepiness surpasses the affected person's ability to maintain vigilance, response slowing, response lapsing, and transition into sleep can occur. This level of drowsiness constitutes dangerous sleepiness.

The focus of this chapter is on clinical assessment for sleepiness. Accordingly, three measurement issues merit special attention: problems associated with self-report; the need for normative-adjusted measures; and whether a particular measurement is appropriate given the circumstance.

In the clinical environment, the most common tool for assessing sleepiness involves self-report. Self-reported sleepiness can be ascertained easily and inexpensively. Multiple factors, however, compromise the reliability and validity of self-report measures. Some people stoically minimize their symptoms, whereas others tend to exaggerate. Both primary and/or secondary gain may color symptom self-disclosure (e.g., a person may deny sleepiness to avoid losing a motor vehicle license necessary for employment). Finally, sleepiness can influence self-perception of sleepiness.

Useful clinical measures usually require standardization to facilitate comparison with normative values. Unfortunately, many sleepiness measures can reliably detect changes in before-and-after experimental designs but not the differences from one subject to another. Consequently, although a measure can provide a superb metric for assessing change, it may fail miserably for determining whether a specific patient presenting for evaluation is sleepy.

The third measurement issue addresses metric validation. The commonly used term *excessive daytime sleepiness* underscores this point. Normative values for most sleepiness metrics rely mainly on data collected during the day. When clinical necessity requires nighttime assessment (e.g., in shift workers), the choices of validated instruments are limited. In other words, how sleepy must the subject be for sleepiness to be considered *abnormal* at 4:00 AM? or to be considered *excessive* at 4:00 AM? or to be considered *dangerous* at 4:00 AM? Although answers to these three questions may not differ at noon, they are likely to differ at 4:00 AM!

Carskadon and Dement[1] proposed a useful conceptualization for characterizing sleepiness. They considered sleepiness along three dimensions: introspective, physiologic, and manifest. *Introspective* sleepiness (when unbiased by a person's motives) derives from an individual self-reported assessment of his or her internal state. *Physiologic* sleepiness refers to the underlying biological drive to sleep; indexed by the speed with which a person falls asleep. *Manifest* sleepiness includes behavioral signs of sleepiness, inability to volitionally remain awake, and performance deficit on psychomotor or cognitive tasks. In theory, manifest sleepiness occurs when sleep drive overwhelms the system that maintains wakefulness.

The Carskadon-Dement model provides an organizational device for understanding differences between sleepiness measures. If sleepiness were a single measurable core phenomenon, rather than a composite output of multiple sleep and alertness mechanisms, one might expect equivalence between measures. However, different tests for sleepiness often produce varying results because they index quite different (although related) phenomena. The best concordance usually emerges at the sleepiness-wakefulness spectrum's extremes. That is, when all sleepiness dimensions are at nadir (no sleepiness) or when sleepiness reaches its zenith (maximum sleepiness), all measures usually agree. Thus, most of the time it is illogical to use physiologic, manifest, and introspective sleepiness measures interchangeably; it also misses the important differences between them.

INTROSPECTIVE SLEEPINESS

Introspective sleepiness is measured using self-administered questionnaires. In this section we describe the most commonly used clinical and research instruments. Some of these questionnaires request subjects to predict their own behavior (e.g., "How likely are you to doze off when . . . ?") or to judge their perceived symptoms over some time interval in the recent past (e.g., in the past month). By contrast, other instruments involve *momentary assessment* and query the person about how he or she feels right now. In general, questionnaires evaluating longer time domains (e.g., in the past month) are more useful clinically than those using momentary assessment. By contrast, momentary assessment instruments are more sensitive to variations in the level of sleepiness, making them more useful for investigating circadian oscillation or drug-induced alterations in sleepiness.

Ultimately, introspective sleepiness relies on self-report. In one respect, self-report constitutes the only means by which it is possible to know how another person feels. Self-disclosed information, however, is modified by individual ability to recognize internal states, memory, tendency to minimize or exaggerate feelings, capacity for honesty, and any underlying agenda.

Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS) is the most widely used clinical instrument for evaluating sleepiness. This specialized, validated, eight-item, pencil-and-paper instrument was developed by Murray Johns at the Epworth Hospital in Melbourne, Australia.[2] The ESS questions subjects about their expectation of "dozing" in differing circumstances. Dozing probability is designated as none (0), slight (1), moderate (2), or high (3) for eight hypothetical situations, as shown in Table 169-1.

The popularity of the ESS stems in part from its simplicity, brevity, and validation. Johns established reliability and validity using 54 patients with sleep apnea (before and after continuous positive airway pressure therapy) and 104 medical students.[3,4] Student control subjects had a mean score of 7.6, compared with 14.3 at baseline for patients with sleep apnea. After treatment, the mean score for patients with sleep apnea declined to 7.4. In another study, normative values were gathered from 942 patients waiting at outpatient clinics (e.g., dermatology, audiology, and ophthalmology clinics) and 1120 healthy people attending health fairs or community health lectures. The mean ESS total scores for these two groups were 8.1 and 5.2, respectively.[5] On the basis of this study and subsequent work in our clinic, we have categorized ESS scores ranging from 0 to 8 as normal, 9 to 12 as mild, 13 to 16 as moderate, and greater than 16 (double that of high normal) as severe.

Table 169-1 Epworth Sleepiness Scale Items

Question	Hypothetical Situation to Be Rated
1	Sitting and reading
2	Watching television
3	Sitting inactive in a public place (e.g., a theater or a meeting)
4	As a passenger in a car for an hour without a break
5	Lying down to rest in the afternoon when circumstances permit
6	Sitting and talking to someone
7	Sitting quietly after a lunch without alcohol
8	In a car, while stopped for a few minutes in traffic

The ESS differs from other tests in that respondents are not asked about how they feel but rather to make a probability judgment about their own behavior. Thus the ESS asks subjects to rate their own sleep drive; this may explain why ESS results correlate (albeit weakly) with Multiple Sleep Latency Test (MSLT)-determined sleep latency (an objective index of sleep drive). The ESS's main disadvantage is its questionable utility when it is readministered within a brief time interval.

Stanford Sleepiness Scale and Karolinska Sleepiness Scale

For many years, the Stanford Sleepiness Scale (SSS) served as the standard measure of introspective sleepiness.[6] Persons taking the SSS test choose one of seven statements to describe their self-assessed current state (choices are shown in Table 169-2). The SSS is a momentary assessment scale and can detect sleepiness as it waxes and wanes over the course of a day. Advantages include its brevity, its ease of administration, and its ability to be administered repeatedly. Experimentally induced sleep deprivation increases SSS scores; however, normative data do not exist, making it difficult to use for clinical decision making or comparisons between persons.

The Karolinska Sleepiness Scale (KSS) is quite similar to the SSS and consists of a nine-point scale ranging from 1, "very alert," to 9, "very sleepy, great effort to stay awake or fighting sleep." Scores of 7 or above are considered pathologic. The KSS has become increasingly popular for evaluating sleepiness in drug trial participants, flight crews, oil-rig workers, train engineers, and professional drivers. Its brevity, validation against EEG and behavioral parameters,[7] and its now-proven sensitivity to sleepiness put KSS on equal footing with the SSS.

Sleepiness-Wakefulness Inability and Fatigue Test

The Sleepiness-Wakefulness Inability and Fatigue Test (SWIFT) is a 12-item self-administered questionnaire

Table 169-2 Stanford Sleepiness Scale Items

Code	Scale Description Statement
1	Feeling active and vital, alert, wide awake
2	Functioning at a high level, but not at peak, able to concentrate
3	Relaxed, awake, not at full alertness, responsive
4	A little foggy, not at peak, let down
5	Foggy, beginning to lose interest in remaining awake, slowed down
6	Sleepy, prefer to be lying down, fighting sleep, woozy
7	Almost in reverie, sleep onset soon, lost struggle to remain awake

validated against the Epworth scale in normal subjects, patients with sleep-disordered breathing, and persons with narcolepsy.[8] The test asks how much of a problem the person has staying awake or how much of a problem he or she has with fatigue, tiredness, or lack of energy (i.e., "not at all," "just a little," "pretty much," "very much"). Six items focus on sleepiness (generally during the day, while driving, while stopped at a traffic light, while at work or while doing other daily tasks, while reading, in social situations) and six on fatigue (generally during the day, while driving, while at work or doing tasks, while reading or studying, in social situations, and when doing tasks that are not urgent). A validation study found good internal consistency (0.87), retest reliability (0.82), and criterion group differentiation (control subjects versus patients with sleep disorder). The SWIFT shows great promise, but clinical use would benefit from normative values establishment and further validation.

Pictorial Sleepiness Scale

Maldonado and colleagues[9] developed a nonverbal sleepiness scale for testing young children and poorly educated adults. They had subjects rank-order seven cartoon faces depicting sleepiness. Rankings were transformed to approximate linearity, and two cartoons were eliminated. The resulting five-picture scale was re-ranked by a different group of subjects to verify order for a final scale. Finally, a validation test indicated significant correlation with KSS and SSS scoring using a mix of normal adults, patients with sleep apnea, shift workers, and school-aged children. Whether this pictorial sleepiness scale will gain popularity remains to be seen.

Profile of Mood States

Although principally designed to assess mood, the Profile of Mood States (POMS) has often been used in sleep research.[10] Originally, the POMS was to include a dimension for sleepiness; however, the "sleepiness" proved to be nonindependent and was therefore eliminated. The "sleepiness" items loaded negatively on the Vigor scale and positively on the Fatigue and Confusion scales. To a lesser extent, sleepiness also emerged on Depression and Anger scales. The Confusion scale elevates more in response to severe sleepiness and the Vigor scale appears sensitive to partial sleep deprivation.[11]

Thus, early on, psychometricians found sleepiness to be a composite measure—a measurement difficulty that persists, because some researchers heedlessly view sleepiness as a unitary factor.

PHYSIOLOGIC SLEEPINESS

Multiple Sleep Latency Test

The feeling people refer to as "sleepiness" can be conceptualized as arising from a physiologic drive. One could therefore use the rapidity with which a person falls asleep to represent the drive's intensity. This relationship between sleepiness and sleep onset provides the foundation for the MSLT.[12] The MSLT provides nap opportunities across the day and indexes physiologic sleep drive with mean sleep latencies. Sleep deprivation hastens sleep onset (decreases sleep latency) when opportunity to sleep occurs. In a series of elegant studies, increased sleep drive provoked by aging, total sleep deprivation, partial sleep deprivation, and disorders of excessive somnolence were well characterized by MSLT-determined sleep latency.[13-15] Data relating homeostatic influences and sleepiness derive directly from MSLT studies.[16] Circadian influence also manifests as shorter MSLT latencies on midafternoon test sessions.

Methodology

The MSLT provides a widely used technique to scientifically assess physiologic sleep drive. In its traditional form, the MSLT involves a series of nap opportunities (four to six) presented at 2-hour intervals beginning approximately 2 hours after initial (morning) awakening (for details, see the report by Carskadon and associates[17]). To establish the prior night's sleep quantity and quality, the patient undergoes attended laboratory polysomnography the night before testing. A careful history of sleep habits, schedule, and drug use in the past month is essential (a sleep diary should be obtained). A clinical MSLT should not be conducted during drug withdrawal (especially from stimulants or from medications that suppress REM sleep), while sedating medications are pharmacologically active, or after a night of profoundly disturbed sleep.

Persons undergoing an MSLT are instructed to allow themselves to fall asleep or to not to resist falling asleep. Subjects are tested under standardized conditions in their street clothes and are not permitted to remain in bed between nap test sessions. Similarly, subjects should not engage in vigorous activity before nap opportunities, because this can alter test results.[18]

Obtaining reliable results depends critically on using standardized test conditions and techniques.[19] Sleep rooms must be dark and quiet during testing. Electroencephalographic (EEG; central and occipital), electrooculographic (EOG; left and right eye), and electromyographic (EMG; submentalis) recordings are used to recognize sleep onset and distinguish between sleep stages. MSLT guidelines also call for monitoring respiratory flow and sounds in patients known to snore (Box 169-1).

Two protocols exist for conducting the MSLT—one designed for research and the other for clinical assessment (Figure 169-1). The research protocol minimizes accumulated sleep by awakening the person when unequivocal sleep onset occurs. To meet criteria for unequivocal sleep, one of the

Box 169-1 MULTIPLE SLEEP LATENCY TEST RECORDING MONTAGE: PHYSIOLOGIC ACTIVITY RECORDED

Left or right central EEG (C3 or C4)
Left or right occipital EEG (O1 or O2)
Left horizontal or oblique EOG
Right horizontal or oblique EOG
Vertical EOG
Submentalis (chin) EMG
Electrocardiogram
Respiratory flow, as needed
Respiratory sounds, as needed

EEG, Electroencephalogram; EMG, electromyogram; EOG, electrooculogram

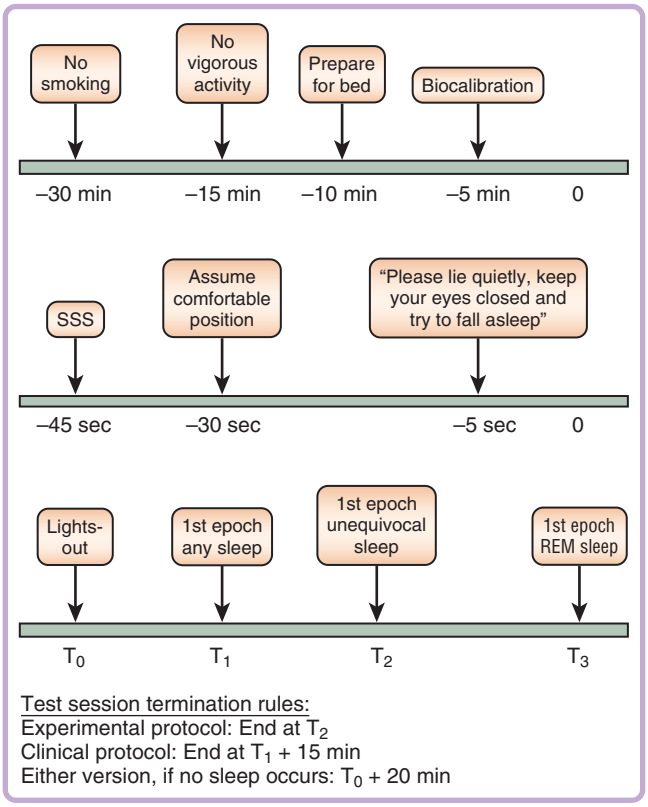


Figure 169-1 Specific procedures for multiple sleep latency test (MSLT) nap opportunities. REM, Rapid eye movement; SSS, Stanford Sleepiness Scale.

following must occur: (1) three consecutive 30-second epochs of stage N1 sleep or (2) a single 30-second epoch of stage N2, N3, or R (rapid eye movement [REM]). By contrast, the clinical MSLT protocol continues for 15 minutes after sleep onset occurs. The clinical MSLT attempts not only to index sleep drive but also to detect abnormally increased REM sleep tendency. Increased REM sleep tendency characterizes narcolepsy, and the added sleep time provides diagnostic information. Table 169-3 illustrates MSLT reliability in distinguishing patients with narcolepsy from control subjects on the basis of whether REM sleep occurred. Documentation of short sleep latency and REM sleep on two or more MSLT nap recordings confirms the diagnosis, especially in patients with cataplexy, sleep paralysis, or hypnagogia (or hypnapompia).

Table 169-3 Multiple Sleep Latency Test: Example Results

Parameter	Patients with Narcolepsy	Control Subjects
N (male/female)	57 (33/24)	17 (6/11)
Age (SD) in years	43.3 (12.3)	33.4 (9.9)
Percent who slept	99.0	63.5
Sleep latency mean (SD)	3.0 (2.7)	13.4 (4.0)
Minimum	0.6	4.8
Maximum	14.1	20
REM score	3.5	0

REM, Rapid eye movement (sleep).

In both the research and clinical versions of the MSLT, the test session is terminated after 20 minutes if no sleep onset occurs. Sleep latency is defined as the elapsed time from the start of the test to the first 30-second epoch scored as sleep. Sleep latency in normal adult control subjects ranges from 10 to 20 minutes. Traditionally, clinicians classified mean sleep latency on the MSLT of 5 minutes or less as highly pathologic sleepiness.[20] The *International Classification of Sleep Disorders* considers values of 8 or less to indicate sleepiness for diagnostic purposes.[21]

Utility

The MSLT can objectively document treatment response[22] and residual physiologic sleep drive[23] in patients independent of whether they self-reported sleepiness after treatment. The sensitivity of the MSLT to physiologic sleepiness makes it especially useful for detecting persistent sleepiness in patients with occult comorbid sleep disorders, ineffective treatment, poor adherence to a therapeutic regimen, or concomitant soporific medication.

As a technique to demonstrate a person's underlying sleepiness, the MSLT has the advantage of being a direct, objective, quantitative approach. It is generally thought that under normal circumstances, nonsleepy persons cannot make themselves fall asleep. By contrast, a sleepy person (if not overwhelmingly sleepy) can remain awake. Thus false-positive tests (MSLT-indicated sleepiness when the person is not sleepy) are theoretically minimal. Polysomnography documents the previous night's sleep quality and quantity, and if significant sleep disruption or disturbance has occurred, the MSLT should be rescheduled. Drug screening helps rule out pharmacologically induced sleepiness. The availability of a specific numerical criterion for characterizing pathologic sleepiness made the MSLT the standard modality for assessing sleepiness for many years.

Clinical Standards of Practice

In 2005, the American Academy of Sleep Medicine (AASM) published revised clinical practice parameters for the MSLT's clinical use.[24] Clinical standards, guidelines, and options were derived from a comprehensive evidence-based medicine review and a systematic protocol for expert consensus.[25] The conclusions can be summarized as follows:

- The MSLT is indicated as part of the clinical evaluation for patients with suspected narcolepsy.

- The MSLT may be helpful for clinical assessment of patients with suspected idiopathic hypersomnia.
- The MSLT is not indicated for routine evaluation of obstructive sleep apnea.
- The MSLT is not indicated for routine reevaluation of patients with sleep apnea treated with positive airway pressure therapy.
- The MSLT is not indicated for routine clinical assessment of insomnia, circadian rhythm disorders, or dyssomnia associated with medical, psychiatric, or neurologic disorders (other than narcolepsy and idiopathic hypersomnia).

Other Measures of Physiologic Sleepiness

Pupillography

Pupil stability and size are affected by exposure to light and by the level of nervous system arousal. In a darkened room, the pupil dilates to improve vision by widening the aperture and allowing more light to enter the eye. However, if the person begins to fall asleep, parasympathetic activation constricts pupillary diameter. Sleepiness also provokes pupil size instability and alters the magnitude and speed of pupillary constriction in response to a flash of light. These alterations reflect the autonomic nervous system balance changes associated with sleep and wakefulness.

Several researchers used pupillography to measure sleep tendency and evaluate narcolepsy.[26] Advanced mathematical techniques provide within-subject differentiation of sleepiness versus alertness using the spectral F-test.[27] Although pupillography would seem to be an attractive approach to objectively measure sleep drive, several barriers continue to impede its clinical utility. First, the procedure is not easily mastered. Second, it remains difficult to compare one patient with another and to designate a clinical numeric threshold for sleepiness. Finally, normative data are not currently available. Research continues on the development of this technique.[28]

Electroencephalography

Quantitative digital EEG analyses seems an obvious approach for assessing central nervous system arousal level. In essence, the MSLT uses the EEG (in conjunction with the EOG and the EMG) to quantify sleep onset from which sleep drive is deduced. It would therefore seem reasonable to expect subtle EEG waveform patterns (microarchitecture) to characterize physiologic sleepiness. Just before sleep onset, alpha frequency decreases and its amplitude increases. Additionally, it has long been thought that EEG delta activity might index sleepiness, because it increases in response to experimental sleep deprivation.[29] Fatigue-related differences, especially in alpha and theta EEG bandwidths are reported.[30] Some researchers believe the EEG cipher lies not in examining resting EEG spectral content but rather in assessing EEG responsiveness to sensory input. If sleepiness alters neurologic reactivity, ongoing task-related EEG changes or event-related potential alterations might better index physiologic sleepiness. Some recent work following this approach focuses on drowsy driving. The "B-Alert X10" system examines EEG power spectral densities recorded at frontal, central, parietal, and occipital scalp sites.[31] The study investigators replicated previous findings and found within-subject EEG indices associated with fatigue.

Although these approaches hold promise for within-subject comparison, the normative data limit their clinical application. The high degree of between-subject variation makes it difficult to compare results between individual subjects. Finally, techniques are not standardized, and many use proprietary algorithms that are declared as trade secrets.

MANIFEST SLEEPINESS

Manifest sleepiness encompasses observable signs and measurable behavior indicating that the person is either sleepy or about to fall asleep, is in the process of falling asleep, or has fallen asleep. Observable signs can include yawning, ptosis (upper eyelid drooping), and head bobbing. Of interest, continuous observation may actually be more sensitive than current EEG measures. Several investigators found EEG measures to be less predictive of task-related performance lapses than video recordings showing eyelid drooping and closure.[32] Manifest drowsiness signs, however, are not specific to sleepiness; for example, ptosis may be neurogenic (as in oculomotor nerve palsy) or myotonic (as in myasthenia gravis), and head bobbing can signify cysts in the third ventricle. By contrast, EEG-EOG-EMG polysomnographic recordings can objectively determine if the person is falling, or has fallen, asleep during a controlled test session. This determination provides the basis for the Maintenance of Wakefulness Test (MWT).

Maintenance of Wakefulness Test

The procedures for conducting the MWT are similar to those used for the MSLT.[24-25] The most significant difference is that rather than instructing the subject to not resist sleep, the examiner instructs the subject to attempt to remain awake. In this manner, the MWT is used to assess a person's capability to, as the name implies, maintain wakefulness. The subject's task is to resist being overwhelmed by sleepiness. To a large extent, the MWT evaluates the magnitude of sleepiness in relationship to the underlying wakefulness system's functioning. If the wakefulness system fails, sleepiness becomes manifest. This laboratory situation parallels circumstances in which sleep onset occurs inadvertently while the person remains passively sedentary in a nonstimulating environment. The MWT gauges the potential threat of inappropriately and nonvolitionally lapsing into sleep—that is, dangerous sleepiness.

Potentially identifying dangerous sleepiness has attracted the attention of regulatory agencies. With growing interest in sleepiness and public safety, the demand for tests to assess sleepiness has increased. Indeed, the Federal Aviation Administration recognizes the MWT as a means to determine whether noncommercial pilots may be licensed after treatment for sleep apnea.[33] Trucking companies and safety officers for companies with high-risk processes are beginning to follow suit, especially as the liability for and financial costs of accidents increase. MSLT and the MWT findings, however, do not correlate well with self-reported sleepiness.[34-35] Patients who fall asleep when not resisting the sleep drive (as on the MSLT) may be able to remain awake during the MWT.

Methodology

In the MWT, the subject's only task is to remain awake. The subject sits in a dimly lit but not totally darkened room. Dressed in street clothing and situated reclining on the bed with a bolster pillow, the subject is not permitted to read, watch television, or engage in other activities. During testing,

EEG, EOG, and EMG are recorded. As with the MSLT, test sessions are scheduled at 2-hour intervals, beginning approximately 2 hours after awakening from the previous night's sleep. Test sessions are terminated when unequivocal sleep occurs (i.e., either three consecutive 30-second epochs of N1 or a single 30-second epoch of N2, N2, or REM sleep).

Sleep latency for each test session, regardless of unequivocal sleep determination, is determined by the first epoch of sleep. The average sleep latency across the four test sessions provides the primary index. The recording also may be evaluated for microsleep (3- to 10-second) occurrences. As expected, studies comparing the MWT and the MSLT find longer mean sleep latencies when subjects are instructed to remain awake than when they are told not to resist sleep. Unlike with the MSLT, a prior night sleep study is not required, because if a subject successfully maintains wakefulness, how he or she slept the night before is moot. However, an important factor involves stimulant use. Consequently, caffeine consumption is restricted on the MWT (and MSLT) test day. Urinalysis and/or blood chemistry also may be required.

For many years, the MWT lacked standardization. Researchers and clinicians used various protocols. The MWT test session duration ranged from 20 to 60 minutes, with the longer tests attempting to avoid ceiling effects. An additional issue for clinical MWT interpretation stemmed from an absence of normative data. Fortunately, however, this situation has changed.[36] Normative data gathered by a consortium of sleep disorders centers established a range for expected values. The raw data from this project were provided for additional analysis by the AASM Standards of Practice Committee. The results enabled MWT to ascend to the status of preferred clinical tool for objectively indexing alertness.

Studies demonstrate MWT clinical utility for evaluating treatment outcomes in patients with narcolepsy and sleep-related breathing disorders. Also, MWT measures can detect improvement in treated patients with persistent MSLT-indexed sleep drive; thus MWT can extend sensitivity range for multiple sleep latency testing.[37] Nonetheless, MWT changes induced varying amounts of sleep deprivation, age, time of testing, and drugs remain grist for researchers' mills.

Standards of Practice Parameters

The AASM developed MWT clinical practice parameters and based standards on an evidence-based literature review[25-26] and expert consensus (when data were inadequate). MWT testing is indicated for clinical assessment of persons whose inability to remain alert constitutes a personal or public safety hazard. Another indicated use includes determining pharmacotherapeutic response in patients with narcolepsy or idiopathic hypersomnia. Clinicians are cautioned that although falling asleep rapidly during an MWT logically would seem to be a powerful indicator for dangerous sleepiness, direct evidence linking MWT sleep latency to real-world accidents is largely lacking. Thus clinical evaluation must integrate MWT findings with signs and symptoms, history, and treatment adherence.

Specific recommendations include conducting four 40-minute trials. Based on statistical analysis of normative data, a mean sleep latency less than 8 minutes is abnormal. Scores between 8 and 40 (the maximum value) are of uncertain significance. The mean sleep latency for presumed-normal volunteer subjects was 30.4 minutes. The ability to remain awake for the entire 40 minutes on all four test sessions (which is the upper limit of the 95% confidence interval) provides the strongest evidence for normal alertness. Nonetheless, clinical judgment is critical because even completely normal values do not guarantee safety.

Vigilance Tests

Response slowing and lapsing during vigilance tests also offer evidence for the consequence of sleepiness, inattention, or both.[38] Therefore a variety of tests requiring simple psychomotor responding (i.e., signal detection reaction time tasks) can index a sleep drive manifestation. Such assessments are called "vigilance tests" because they evaluate the person's ability to remain heedfully vigilant. Typically (but not always), these tests attempt to mimic the tedious, palling situation of watching for blips on a radar screen or for ships on the horizon.[39] Loss of vigilance, when faced with a nonstimulating task, is particularly relevant for patients with disorders of sleep and arousal. The task's monotony theoretically unmasks underlying sleepiness.

Vigilance tests measure arousal level, attention, or both. As with the MWT, performance cannot exceed ability or maximum effort. Although a person could intentionally perform poorly (i.e., faking "bad"), he or she cannot fake "good." However, teasing apart arousal and attention can make test interpretation difficult. Attention deficits in nonsleepy persons can confound test results. Fortunately, sleepiness and inattention often coexist; consequently, long, experimenter-paced, monotonous tasks are sensitive to sleep loss, sleep disruption, and circadian variation. The landmark studies of the 1970s, collectively referred to as the Walter Reed experiments (named for the institute where they were conducted), documented the effects of sleep deprivation on performance.[40] These pioneering studies established that increasing duration of prior wakefulness and time-on-task provoked response slowing and lapsing (for an excellent review, see reference [41]).

A variety of vigilance tests are available; however, the psychomotor vigilance test (PVT) developed by Dinges and colleagues is currently the best-validated and most widely used.[42-43a] The PVT is a visual signal detection test, approximately 10 minutes long, administered either by computer or by a hand-held display-and-response unit. Response latencies to visual target stimuli are recorded. Response slowing and lapsing correlate with findings on the SSS and the MSLT; these results provide convergent validity. Additionally, PVT results have been reported for a variety of subject groups, including normal control subjects, sleep-deprived volunteers, and patients with major sleep disorders. The test provides exquisitely sensitive within-subject measures for before-and-after experimental designs. PVT normative data are limited, however, making clinical between-subject comparisons difficult. Programs using portable devices are now available.[44]

The other vigilance test popular for sleep research is the Oxford Sleep Resistance (OSLER) test.[44a,45] The testing paradigm mimics that for the MWT but uses a visual signal detection task rather than EEG-EOG-EMG monitoring. The test uses four 40-minute test sessions, during which visual target signals are presented. Subjects are instructed to respond

Table 169-4 Comparison of Tests for Evaluating Sleepiness

Sleepiness Type Assessed	Test	Are Normative Data Available?	Is It Possible to Fake Sleepiness?	Is It Possible to Fake Alertness?
Introspective sleepiness	ESS	Yes	Yes	Yes
	SSS	No	Yes	Yes
	SWIFT	No	Yes	Yes
	Pictorial sleepiness scale	No	Yes	Yes
	POMS	Yes	Yes	Yes
Physiologic sleepiness	MSLT*	Yes	No	Yes[†]
	Pupillography	No	No	Unknown
	EEG	No	No	Unknown
Manifest sleepiness	MWT*	Yes	Yes[‡]	No
	Vigilance and performance tests	No	Yes[§]	No
	Postural balance test	No	Yes	No

*Standard protocol described in an American Academy of Sleep Medicine Practice Parameter: Test involves 4 to 6 test sessions per day, at 2-hour intervals. Test sessions are sometimes scheduled at shorter intervals (e.g., for children); however, this practice is not recommended by the authors.
[†]Assuming that the subject is not overwhelmingly sleepy, attempting to remain awake can undermine the test result.
[‡]Assuming that the subject is physiologically sleepy, not attempting to remain awake will make it appear that overwhelming sleepiness is present.
[§]Intentionally not attending or responding to the task can make the subject appear sleepy.
EEG, Electroencephalogram; ESS, Epworth Sleepiness Scale; MSLT, multiple sleep latency test; MWT, maintenance of wakefulness test; POMS, Profile of Mood States; SSS, Stanford Sleepiness Scale; SWIFT, sleepiness-wakefulness inability and fatigue test.

to each signal with a simple button press. A test session is terminated either after 40 minutes or after significant response lapsing (which is considered a failure to maintain wakefulness). OSLER has been validated against MWT test results, but specific normative data supporting clinical threshold scores are not currently available.

Postural Balance Testing

Sleep-deprived people experience difficulty maintaining balance. Extended periods of wakefulness also impair balance.[46] Time-of-day variation also occurs.[47] Research suggest posturographic indices may provide a technique for assessing manifest sleepiness. Balance can also be adversely affected by psychoactive drugs. One approach records pressure shifts on a force platform on which the subject stands, feet together, with crossed arms at the chest. The subject's direction of gaze is on a fixed point straight ahead. The body's *center of pressure* is sampled for 30 seconds at a rate of 1000 samples/second. Some studies conduct a trial at 2-hour intervals, but protocols are not yet standardized. Whether this approach can be refined for use as a clinical tool remains to be seen; however, it shows promise.[48,49] Standardized evaluation protocols, optimized data processing, normative data, and validation against other measures of sleepiness and sleep drive are needed.

PRACTICAL ISSUES AND CONCLUSIONS

Before initiation of an evaluation for sleepiness, several practical issues warrant consideration. An important distinction to be made is whether the goal is to (1) establish the presence of sleepiness, (2) confirm the absence of sleepiness, or (3) identify changes in sleepiness. Is testing being conducted for clinical assessment, research, or legal purposes? Does the subject have a self-interest in the outcome? (i.e., Is there any primary or secondary gain?) With increasing frequency, sleep specialists provide expert opinions in legal matters involving accidents and disability claims. Expert panels often render opinions concerning fitness for duty or disability adjudication. In such cases, objective testing is crucial. Furthermore, a normal test result does not guarantee fitness for duty. Table 169-4 summarizes characteristics of the tests described in this chapter. Ideally, physiologic, manifest, and introspective sleepiness should be assessed. In general, if the subject claims to be sleepy and the goal is to demonstrate sleepiness, the MSLT is likely to be the best confirmatory test. If the subject claims not to be sleepy and the goal is to demonstrate an ability to remain awake (as when concerns arise about the person's ability to operate a motor vehicle), the MWT has certain advantages.[50,51]

For clinical purposes, self-reported measures combined with MSLT have long been the sine qua non for establishing sleepiness. Sometimes, however, in cases involving severe sleepiness, the MWT can demonstrate improved alertness after treatment, whereas the MSLT shows little or no change. Such persons continue to be pathologically sleepy, but they are not overwhelmed by sleep during the brief testing interval. The relationship between this pattern of change and performance or behavior requires further study.

The dangers posed by excessive sleepiness are becoming increasingly apparent. The National Commission on Sleep Disorders Research catalogued a substantial list of sleep-related industrial and transportation accidents. Long ago, Kleitman proposed sleepiness as resulting from accumulation of bloodborne or cerebrospinal fluid hypnotoxins; however, clinical tests for such substances have not been developed or validated. Nonetheless, the search continues. Therefore the clinician can use one or a combination of the evaluation techniques described in this chapter to measure the underlying physiologic drive for sleep; the subjective, internalized consequence of that drive; and/or sleepiness's behavioral manifestations.

CLINICAL PEARLS

- Measuring sleepiness in a clinical setting is not a simple matter. The MSLT and the MWT represent standardized tests evaluating patients for physiologic and manifest sleepiness.
- Clinical practice standards recommend use of the MSLT for evaluating narcolepsy and idiopathic hypersomnia.
- The MWT is indicated for testing the person's ability to remain awake when safety is at stake.
- Sleepiness testing must always be viewed within a larger context of the clinical history and examination findings.

SUMMARY

Excessive sleepiness, although central to sleep medicine practice, remains a poorly defined concept. In clinical practice, assessment usually involves self-report; however, objective measures are available. The conceptual framework used in this chapter is based on the three faces of sleepiness: introspective, physiologic, and manifest sleepiness. The indications and techniques used to evaluate sleepiness in standard clinical practice include the ESS (for introspective sleepiness), the MSLT (for physiologic sleepiness), and the MWT (for manifest sleepiness). Other assessment procedures, each with its relative merits and limitations, include postural balance testing and vigilance testing. Practical considerations in choosing and evaluating sleepiness tests include the goal of testing, the purpose of the test, and interpretation of results.

Selected Readings

Carskadon MA, Dement WC. Daytime sleepiness: quantification of behavioral state. *Neurosci Biobehav Rev* 1987;**11**:307–17.

Cuthbertson DW, Bershad EM, Sangi-Haghpeykar H, Cohen HS. Balance as a measurement of fatigue in postcall residents. *Laryngoscope* 2015;**125**(2):337–41.

Franzen PL, Siegle GJ, Buysse DJ. Relationships between affect, vigilance, and sleepiness following sleep deprivation. *J Sleep Res* 2008;**17**(1):34–41.

Honn KA, Riedy SM, Grant DA. Validation of a portable, touch-screen psychomotor vigilance test. *Aerosp Med Hum Perform* 2015;**86**(5):428–34.

Kosmadopoulos A, Sargent C, Zhou X, et al. The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment. *Accid Anal Prev* 2015;pii: S0001-4575(15)30096-8.

Standards of Practice Committee of the American Academy of Sleep Medicine. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep* 2005;**28**:113–21.

Wilhelm B, Bittner E, Hofmann A, et al. Short-term reproducibility and variability of the pupillographic sleepiness test. *Am J Hum Biol* 2015. [Epub ahead of print].

A complete reference list can be found online at ExpertConsult.com.

Chapter 170

Chronobiologic Monitoring Techniques

John H. Herman

Chapter Highlights

- The suprachiasmatic nucleus is the master biologic clock of the mammalian brain and drives an organism's circadian rhythm. Suprachiasmatic nucleus properties can be estimated by measuring input variables such as light or food and output variables such as variations in temperature or the nightly onset of melatonin secretion.
- Under experimental conditions of continuous darkness for many days, the daily onset of wheel running by experimental animals (i.e., rodents) unmasks the period length of the circadian rhythm. The animal will begin running several minutes earlier or later each 24-hour period with remarkable consistency. Plotting the time of wheel running with consecutive days stacked upon each other reveals the period length of the animal's circadian rhythm, or tau.
- Humans are capable of entraining only to dark-light schedules close to 24 hours. In a forced desynchrony protocol, humans are exposed to dark-light schedules to which they cannot adapt. In these protocols subjects are permitted to sleep in the dark portion of the cycle only. The result is desynchrony—that is, sleep and wake lose their normal relationship to body temperature, endocrine secretion, alertness, and behavioral performance.
- Sleep onset normally occurs approximately 2 hours after the beginning of melatonin secretion. Melatonin secretion is emblematic of biologic night. The timing of dim-light melatonin onset can reveal the presence of phase advance or phase delay disorders. Recently, inexpensive salivary melatonin collection kits have been developed that enable sleep specialists to readily measure dim-light melatonin onset in patients.

BASIC CONCEPTS AND TERMINOLOGY

The purpose of chronobiologic monitoring techniques in humans is to measure biologic and behavioral properties related to circadian rhythms. These methods help promote understanding of normal and abnormal clocks controlling human alertness and sleepiness. Chronobiology research is based on the assumption that a given organism contains within it a core mechanism or clock for generating rhythmic expression of physiologic parameters. The fluctuations in any biologic or behavioral parameters that display an approximately 24-hour periodicity are assumed to be the direct or indirect consequence of an underlying pacemaker that possesses its own inherent and autonomous periodicity.

The internal clock has properties that are indirectly studied in circadian rhythm research by examining output variables. These include amplitude, period length (tau), phase, rate of change, and relative duration of the active period. *Amplitude* is the quantity of change in a parameter that occurs from apogee to nadir by measuring variables such as core body temperature, cortisol, or cognitive problem solving. *Period length* is the temporal duration of one complete circadian cycle under free-running conditions. *Phase* is the relation of the internal clock to the environment, such as sunrise and morning awakening. *Rate of change* is the rapidity with which a circadian parameter switches from its active to dormant or dormant to active modes. Finally, it is possible to measure the *duration of the active period* relative to the duration of the dormant period. Chronobiologic techniques can measure each of these variables and define how they may be manipulated and how they are interrelated.[1]

The process by which the internal clock stays in synchrony with the environment is called *entrainment.* Stimuli, such as light, that bring about entrainment are called *zeitgebers.* The circadian mechanism is always subject to the influences of masking. *Masking,* which is further described later in the chapter, is the ability of an external variable, such as light, sound, or physical activity, to alter a behavioral or physiologic parameter without affecting the core pacemaker. Much of the foregoing terminology, especially the melatonin secretory profile, is illustrated in Figure 170-1.

The definitions of entrainment and masking overlap to some extent. Exposure to daily sunlight, for example, entrains humans to their local time zone and masks their internal clock from expressing its endogenous rhythm. Thus daily exposure to normal sunlight both entrains mammals to and masks the true periodicity of the endogenous pacemaker. In animals, masking influences may be removed by placing the animal in continuous light or dark conditions.

As simple and straightforward as it would seem to be to measure the human circadian period length, the field of circadian rhythms research has undergone several modifications

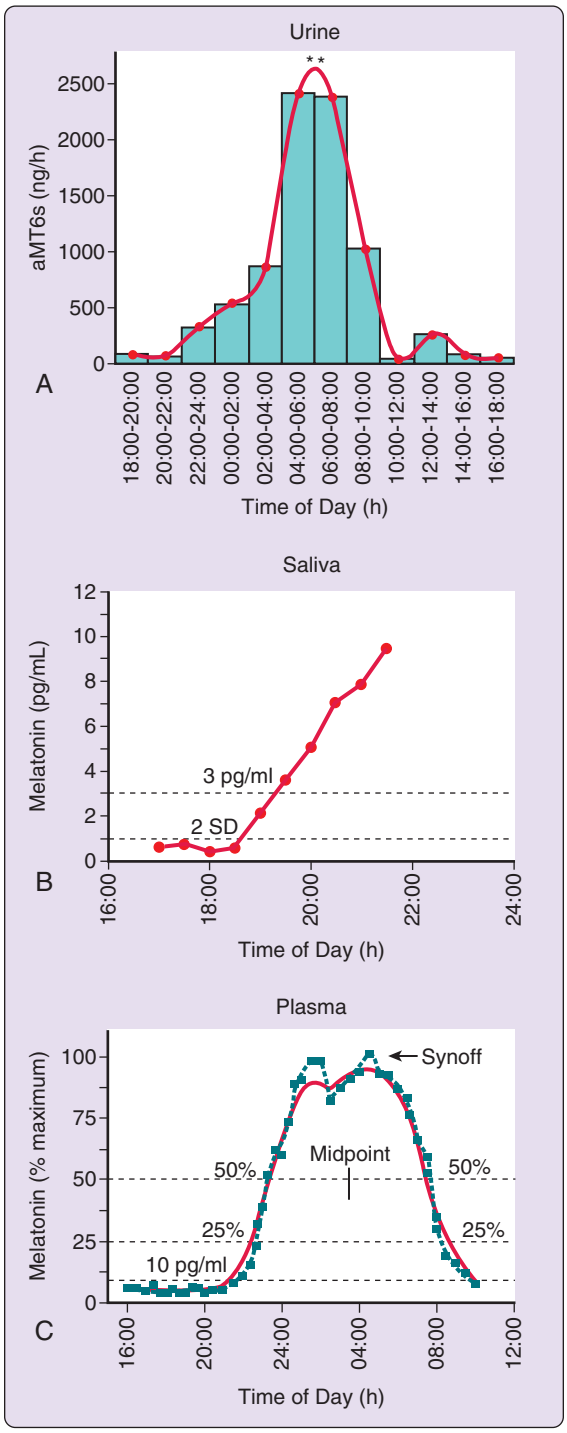


Figure 170-1 Three melatonin sample types and their associated phase estimates. **A,** 24-hour rhythm of the primary urinary melatonin metabolite 6-sulfatoxymelatonin (aMT6s) derived from urine samples collected in 2-hour bins under dim light. The fitted curve reveals a significant 24-hour rhythm with maximum levels observed between 04:00 and 08:00 (**$P < .01$). **B,** Salivary melatonin profile collected under dim-light conditions. The low-threshold dim-light melatonin onset (DLMO) was defined as either the first sample to exceed and remain above a threshold of 3 pg/mL or that was 2 SD above the mean of the first three baseline samples (2 SD). **C,** Overnight plasma melatonin profile, plotted as a percentage of maximum (*dashed line*) and smoothed with a Lowess curve fit to the raw data (*solid line*). Some frequently used phase markers are shown: DLMO at 10 pg/mL, DLMO or dim-light melatonin offset (DLMOff) at 25% or 50% of maximum levels, the midpoint, and the termination of melatonin synthesis (Synoff). (From Benloucif S, Burgess HJ, Klerman EB, et al. Measuring melatonin in humans. *J Clin Sleep Med* 2008;4:66–9.)

of estimated human cycle length. Estimates improve as methodologic errors are eliminated by new and more accurate paradigms. In humans, a number of corrections of the period length of the free-running circadian rhythm length, or period, have been introduced. Period length currently is estimated at slightly greater than 24 hours, 24.18 hours being the most recent value.[1]

In addition to a core pacemaker, each circadian system has input and output components. Input components are principally responsive to photic information but respond to a variety of stimuli. These stimuli are manipulated in circadian studies to examine the effect on output variables such as the sleep-wake cycle, temperature, or behaviors such as reaction time or problem-solving ability.

METHODOLOGIC CONSIDERATIONS IN THE INVESTIGATION OF CIRCADIAN RHYTHMS

During the preceding five decades, estimates of human circadian period length range from 13 to 65 hours.[2] Also, in humans, greater interindividual inconstancy in circadian variables had been observed than that seen in other mammals. In nonhuman animals, circadian rhythms are remarkably stable from day to day. In humans, there have been continued corrections (and narrowing) of the period length of the circadian rhythm of sleep-wake; most recent estimates indicate a period slightly longer than 24 hours, with a precision not that different from that in mammalian relatives.

Disagreements continue over what stimuli are capable of acting as zeitgebers to entrain or shift circadian rhythms. Initially, only bright light (7,000 to 12,000 lux) was considered to be capable of shifting the timing of circadian rhythms.[3] Subsequently, light of moderate intensity was described as being capable of shifting circadian rhythms,[4] and most recently, studies now maintain that ordinary room light is capable of inducing a circadian shift.[5,6] Some investigators claim that extraocular light is capable of influencing circadian rhythms. One study demonstrated that behind-the-knee (popliteal region) bright light exposure is capable of inducing shifts in circadian timing,[7] but multiple failures to replicate this finding have been documented.[8-11] As chronobiology research progressed, a series of modifications of experimental design successively eliminated some masking artifacts and external zeitgebers that previously introduced artifact.[1]

The field also has experienced some controversy regarding what constitutes a zeitgeber, or a stimulus capable of entraining a circadian rhythm. Some evidence suggests that nonphotic stimuli are capable of shifting the timing of such rhythms.[12] Exercise,[13,14] social activity,[15] feeding schedule,[16] ambient temperature,[17] napping in darkness,[18] administration of melatonin,[19] and knowledge of time of day or night[20] each have some capacity for circadian rhythm entrainment in human subjects. Some studies suggest a major role for such nonvisual zeitgebers,[12] whereas other studies imply a more modest role.[21] The extent to which exogenous melatonin administration is capable of shifting circadian rhythms was controversial until recently.[22]

Another area of research involves separating behavioral and biologic properties of circadian rhythms from other physiologic processes. Many behavioral and biologic phenomena are influenced by the sleep-wake cycle or the duration of previous wakefulness period. The homeostatic model describes an

increasing pressure to sleep with increased duration of wakefulness and a decreased pressure to sleep with increased duration of sleep.[23] Circadian rhythm studies must separate variance in a biologic or behavioral parameter into homeostatic and circadian components; that is, they must identify what portion of the parameter's variance is circadian and what portion is homeostatic.[24,25]

An additional area in which agreement is not universal is what biologic metric is best used to approximate the properties of the output of the underlying core pacemaker, such as period length. Core body temperature[24] and melatonin[26] secretory profiles have been used in most studies. Some research uses dim-light melatonin onset (DLMO) as the marker for the beginning of a new circadian cycle; other studies use dim-light melatonin offset (DLM offset),[26] but many use the core body temperature nadir. However, no stable point is identifiable as the minimum of temperature or the onset of melatonin secretion. Temperature fluctuates from moment to moment, and the minimum or maximum must be extracted mathematically. Melatonin rises gradually after sunset, and an arbitrary level must be selected to denote onset or offset. A modified RAND consensus panel consisting of melatonin investigators came to the conclusion that blood, saliva, or urinary melatonin concentration may each be optimal under different experimental circumstances.[27] Figure 170-1 shows (1) the urinary secretory profile of melatonin; (2) the salivary DLMO, and (3) the secretory profile of melatonin as determined from blood samples. It must be emphasized that extracting circadian rhythm data relies on mathematical criteria and data-analytic techniques selected in each experimental protocol. One study used nonorthogonal spectral analysis to identify core body temperature minimum in humans.[1]

The amount of information uncovered in the past few decades in the field of circadian rhythm research has been truly astounding. It includes the identification of basic anatomy, novel photoreceptors, and genetic mechanisms. However, researchers have not agreed on specific definitions in the areas of contention listed previously, nor has any chronobiologic monitoring technique been accepted as the "gold standard." Findings using different experimental paradigms contribute to the turbulence in this field.

This chapter describes various paradigms recently used in circadian rhythm research and reviews their strengths and weaknesses. For in-depth detailed descriptions of each model, the reader is directed to a few key references at the end of the chapter in which a fuller exposition is offered.

PARADIGMS

Fixed Light-Dark Schedules, Double-Plotted

The most common model for animal studies of circadian variables and activity schedule consists of placing the animal in a cage with a running wheel and subjecting it to a fixed schedule of light and dark. The running wheel's turns are counted continuously. Typically, the animal's running behavior on the wheel is plotted as a vertical or horizontal line per unit time, or as wheel turns per unit time. These counts are plotted successively for 24 hours, yielding a visual representation of when and how much the animal ran that day. Successive days are stacked, enabling the reader to visually appreciate changes in the timing of activity that occurred during the experiment. The entire plot is duplicated side by side, called *double plotting,* because seeing the image next to itself enables the reader to appreciate what happened in the experiment more readily. For example, if the animal's principal expression of wheel-running activity drifts from before to after midnight, double plotting allows the reader to more easily follow the continuity of changed wheel-running times. Drinking or feeding may be plotted in the same manner. Protocols are labeled as LL (constant light), LD (fixed light-dark schedule, most commonly 12 hours light and 12 hours dark), or DD (constant darkness). Constant dim light is sometimes used in such protocols, or light dim enough to prevent suppression of secretion of melatonin in the experimental animal (hamster or rat). In some studies, the LD period length may be longer than or shorter than 24 hours, and use of a period as short as 1 hour has been described.

The DD paradigm successfully demonstrates the remarkable consistency of a given species' free-running activity rhythm in constant darkness or dim light, elucidating the period length of its biologic clock.[1] Often, the timing of light and dark ise abruptly changed to measure the animal's reaction; adaptation can take several days. Such paradigms are at the core of much genetic research, in which normal (wild-type) animals (+/+) are compared with heterozygous (+/–, –/+) or homozygous (–/–) knockout animals. Studies reveal genetic phenotypes (behavioral expression), because knockout animals display altered periodicity, diminished rhythmicity, or absence of any circadian rhythm in running, feeding, or drinking behavior.

The strength of the fixed light-dark schedule with activity monitoring is its ability to accurately and economically measure an animal's rest-activity schedule. Alterations in rest-activity schedules after genetic knockout modifications constitute a standard genotype-phenotype model. One design weakness is an inability to examine complex interactions that would occur in more natural circumstances. If only light, dark, or light versus dark is studied, then all effects appear to be a function of the animal's illumination schedule. The paradigm is subject to overly simple interpretation, the implications of which are discussed later.

Entrained 24-Hour Protocol

In entrained 24-hour studies, subjects live in laboratory conditions with constant bedtime, wakeup time, meal time, and timed activities. Subjects are aware of time of day and are not isolated from normal zeitgebers such as light from windows. This protocol is used to study biologic or behavioral parameters (e.g., hormone secretory profiles or cognitive processes) under constant conditions. Often, subjects are fitted with indwelling catheters for recurring blood sampling.

Phase-Shifting Protocols

One of the most experimentally robust circadian rhythm research findings is the capacity of zeitgebers, such as bright light, to change the time of day at which the circadian system switches from its active or day mode to its dormant or night mode. This switching is referred to as a *phase shift.* It is similar to jet lag or shift work in the real world.

Many studies have examined the capacity of light, at various intensities and durations, to alter the timing of sleep, melatonin secretion, or the temperature nadir. These studies share a common methodologic feature: Environmental light

is controlled at baseline, but typically the timing is similar to that for natural environmental light. Often the baseline consists of dim light ranging in intensity from approximately 70 lux (equivalent to romantic restaurant lighting) to as low as 1.5 lux (similar to the light from a candle). In the experimental condition, light is then delivered to the subject in what normally would be the dark or night portion of the circadian rhythm.[28] Light is delivered for the same duration but at a different time during this experimental condition for one or several days. Some studies include a subsequent condition in which the subject is in constant dim light or in a constant routine (described later). One study examining the intensity of light required to accomplish a phase shift or melatonin suppression showed a "ceiling effect" at approximately 1000 to 3000 lux[29]

Circadian variables such as wheel running in rodents and hamsters, the timing of sleep, or the timing of melatonin secretion are monitored during each condition.[1-3] The magnitude of change in a circadian variable after the altered timing of light is measured. A phase response curve (see Figure 170-2) shows the magnitude of response to light exposure (typically 1 to 3 hours) at different times throughout the circadian cycle in changing the timing of a circadian output variable.[30] Most studies show circadian rhythm shifts in response to light occurring only if the light is delivered at specific times, typically at or near normal hours of darkness. This finding indicates that mammals are not responsive to bright light administered during normal hours of daylight.[31] Other studies claim that there is no dead zone and that bright light delivered at any hour, including normal daylight hours, has a phase-shifting effect.[32] It is possible that many subjects have partial desynchrony between sleep-wake and temperature or feeding rhythm.

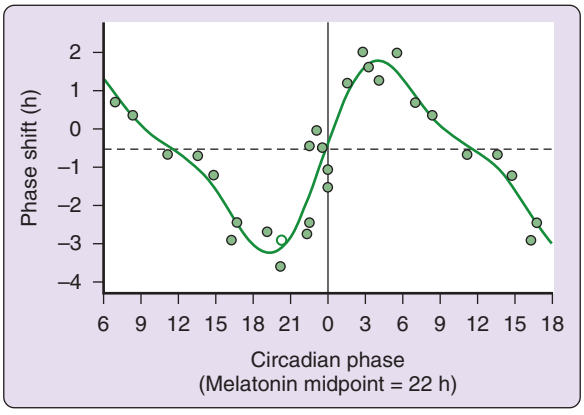


Figure 170-2 Accurately identifying the effect that light and melatonin have on advancing or delaying circadian timing is at the core of much circadian rhythm research. The magnitude of the phase advances (falling asleep and awakening earlier, represented by positive numbers) and delays (falling asleep and awakening later, represented by negative numbers) are plotted against the center of the period of light exposure relative to the core body temperature minimum at 0 hours. The *filled circles* represent data from plasma melatonin, and the *open circle* represents data from salivary melatonin in one subject from whom blood samples were not acquired. The *solid curve* is a harmonic function fitted through all of the data points. Light's phase-delaying effect increases as it approaches the temperature minimum; then it flip-flops in the 6-hour period surrounding the temperature minimum to a maximum phase-advancing effect 6 hours after the temperature minimum is reached. (From Khalsa SB, Jewett ME, Cajochen C, Czeisler CA. A phase response curve to single bright light pulses in human subjects. *J Physiol* 2003;549[Pt. 3]: 945–52.)

Time-Isolation Protocols

In time-isolation protocols, animals may be kept in total darkness or in light dim enough to presumably not entrain their circadian timing mechanism. Human subjects typically have been isolated in an environment free of time cues, such as an underground facility or an internal suite of rooms in an enclosed bunker. Great pains are taken not to provide any clues regarding time (including use of double-door entry chambers and randomized schedules for technicians interacting with subjects). In such studies, for many years, subjects were permitted to turn off their room lights and retire when they wished and awaken when they wished. Subjects were instructed not to nap.

One of the more unexpected findings in circadian rhythm research is that the subjective sense of sleepiness, or the wish to sleep, is at its minimum at the normal time of sleep initiation and that sleepiness is greatest at the normal time of awakening.[33-37] This finding of greatest alertness in the hours preceding sleep is called "the sleep-forbidden zone," and the occurrence of greatest sleepiness soon after awakening is called "sleep drunkenness." However, the tendency to be most alert in the hours preceding bedtime and to feel most sleepy at the hour of awakening is attributed to the respective maximum and minimum of the circadian systems effect on wakefulness. These phenomena demonstrate the capacity of circadian rhythms to influence subjective sleepiness and alertness.

Under conditions of time isolation, a variety of studies showed great variability in human sleep-wake cycle length. By contrast, the sleep-wake cycle in most other mammals is consistent and stable. Human studies showed sleep-wake cycles as long as 36 hours and great day-to-day variability in both sleep and wake duration.[38] Subjects who remained in time-isolation conditions for a sufficient interval developed desynchrony between their temperature rhythm and their sleep-wake cycle. Under normal conditions, the temperature minimum occurs somewhat after the middle of the sleep period. In subjects in temporal isolation protocols, the interval from one temperature minimum to the next has a cycle length of 24 to 25 hours, but the interval between sleep initiation on successive nights is longer. Consequently, the temperature minimum eventually strays elsewhere in the sleep-wake schedule.[39]

The time-isolation protocol is no longer used, because results from these studies are widely disparate from those for circadian rhythm studies in all other mammalian species. Of note, however, by disregarding this experimental approach, a basic attribute of human behavior is being ignored: Most persons appear to prefer a sleep-wake schedule longer than 24 hours, when allowed. This perhaps unique human preference for longer wake and sleep episodes has been overlooked in the circadian rhythm field's fervor to produce human data as similar as possible to those acquired in other mammals, invertebrates, and plants.

Forced Desynchrony Protocols

In forced desynchrony protocols, subjects are scheduled for rest-activity cycles anywhere from every 20 minutes to 28 hours. Forced desynchrony is obtained by subjecting the study participant to a sleep-wake schedule to which adaptation is impossible. The human circadian system is capable of adapting

to sleep-wake schedules within only a certain range surrounding 24 hours. If the forced sleep-wake schedule is outside of that zone, then the circadian rhythms of all other biologic and behavioral phenomena become desynchronized with the enforced sleep-wake schedule. In such studies, core body temperature, melatonin, or both are monitored continuously. Subjects live in dim light during the active part of their rest-activity cycle and in darkness during scheduled sleep episodes. Dim light is defined as a level from 1 to 70 lux in various studies. Under such conditions, the endogenous circadian pacemaker, as measured by DLMO or temperature rhythm, displays a stable period and is unable to follow the scheduled 28-hour (or 20-minute) sleep-wake cycle. The temperature rhythm then becomes desynchronized from the imposed sleep-wake schedule. The 20-minute sleep-wake cycle is referred to as the ultrashort sleep schedule.[40] The forced desynchrony protocol allows investigators to determine the extent to which various biologic rhythms are sleep-dependent or independent of sleep.

In forced desynchrony protocols, subjects are placed in a sound-attenuated and windowless room and have no access to any timepiece for the duration of the study. The contacts with staff members are limited, and they are trained not to transmit any information about the time of day. In such studies, technicians have only brief contact with the subjects to announce the moments for rising, meals, showering, testing, and going to bed. Subjects watch videos, listen to music, or engage in their own preference of leisure activities. In some studies, subjects are allowed caffeine-containing beverages in the scheduled morning.

In forced desynchrony protocols, subjects spend two thirds of their time in light or waking conditions and one third of their time in dark or sleep conditions—for example, 60 minutes of activity and 30 minutes of scheduled sleep in a 90-minute day, and 18.66 hours of activity and 9.33 hours of scheduled sleep in a 28-hour day paradigm. In all forced desynchrony protocols, from 20-minute to 28-hour cycles, the scheduled day cycle length is outside the range of entrainment of the human circadian clock (under dim-light conditions).[41] Under such conditions, the circadian clock becomes unmasked as it continues to oscillate at its near-24-hour intrinsic period despite sleep permitted only on the 28-hour schedule (Figure 170-3).

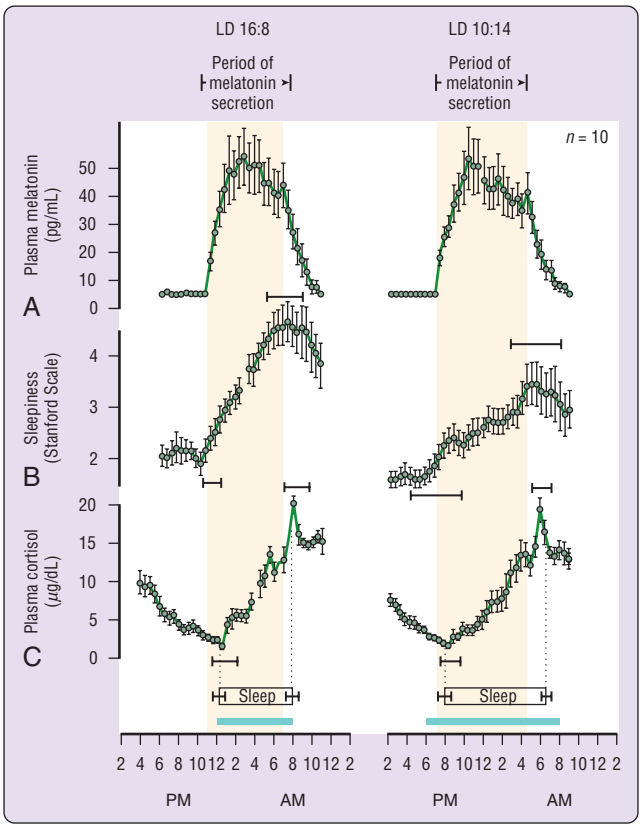


Figure 170-3 *Left plots,* Data for subjects in a 16-hour day and an 8-hour night (LD of 16 : 8), equivalent to midsummer at a latitude of 39 degrees (Washington, D.C.). *Right plots,* Data for subjects in a 10-hour day and 14-hour night, which occurs in midwinter at the same latitude. The circadian pacemaker imposes a pattern of neuroendocrine secretion influenced by the length of daylight. **A,** In summer and winter, the pattern of melatonin secretion is characterized by distinct diurnal and nocturnal periods with relatively discrete transitions between them. Melatonin is either at high levels (biologic night) or virtually absent (biologic day). **B,** The paradox described in the text: Subjects are sleepiest when they wake up and most wide awake when closest to bed time (as measured by the Stanford Sleepiness Scale). Of note, subjects are far sleepier during short nights than during long nights. This means that the introduction of artificial light has increased sleepiness of the human population as a whole. **C,** In humans, diurnal periods of absence of melatonin secretion, and falling levels of cortisol alternate with nocturnal periods of active melatonin secretion and high and rising levels of cortisol. The timing and duration of endocrine secretion are matched to the timing and duration of solar day and night. The pacemaker adjusts the duration of the biologic day and night to match seasonal variation in the duration of sunlight. After humans have been chronically exposed to long nights (*right*), the duration of melatonin secretion and rising levels of cortisol is longer than it is after they have been chronically exposed to short nights. (From Wehr TA. Effect of seasonal changes in daylength on human neuroendocrine function. *Horm Res* 1998;49:118–24.)

Constant Routine Protocols

The constant routine protocol requires the subjects to remain in a semirecumbent posture throughout the protocol's experimental phase, typically 24 to 48 hours, sometimes continuously awake. For this protocol, the subject remains in temporal isolation in dim light. Normal caloric and liquid intake is provided in hourly, equally divided portions. Food and beverages are at room temperature. In such studies, salivary melatonin samples may be acquired hourly, core body temperature may be monitored continuously, and psychometric testing may be administered at hourly intervals. Blood samples may be drawn at hourly intervals through an indwelling intravenous catheter. Constant routine protocols eliminate the effects of activity, light-dark cycles, sleep, and meals on temperature, enabling a more accurate estimate of the period length of the hypothetical "master clock" in the suprachiasmatic nucleus (SCN).[42] Other chapters in this book provide adequate evidence that the SCN is the body's master clock, synchronizing the clock cells throughout the body.

Long Nights Protocol

In the long nights protocol, subjects have an enforced dark period longer than the permitted light period. Such studies have been used in adults and adolescents. This protocol reveals an aspect of the human circadian rhythm that typically is ignored: the capacity of humans to adapt to an extended dark period with enforced immobility. Historically, humans in temperate zones of the northern and southern hemispheres obviously spent approximately half of the year with dark periods exceeding light periods. Society has used artificial light to impose permanent light conditions similar to the long

days of summer, to which humans are capable of entraining. The long nights protocol reverses the circumstances, testing the ability of the human circadian system to adapt to short days (10 hours) and long nights (14 hours) (Figure 170-3). Such studies have revealed increased total sleep time, increased duration of cortisol suppression, increased duration of melatonin secretion, increased duration of temperature suppression, and decreased daytime sleepiness under these conditions.[43] During long nights, subjects spend some of this time awake, typically after rapid eye movement (REM) sleep periods.[44] In this manner, the portion of the time spent with the body geared for sleep is increased and the time spent with the body geared to be awake is decreased (see Figure 170-3). This demonstrates an aspect of adaptability in the human circadian system that is not revealed in other circadian protocols.

Phase Shifting of Gene Expression Protocol

Many people work at times not in keeping with their biologic clocks. Many studies indicate that night shift or graveyard shift workers maintain their daytime peak and nocturnal drop in circadian activity, probably as a consequence of light exposure during and surrounding their hours of sleep and switching back to a normal day-awake schedule on days off. Night shift work is recognized as a health risk, increasing the likelihood of several serious, chronic morbidities.

The following study describes a protocol for flipping circadian rhythm activity to peak in the nocturnal work hours of night shift employees. One of the techniques used to measure phase shift was measurement of circadian genes in blood cells. This study examined clock gene expression in peripheral blood mononuclear cells in comparison with melatonin and cortisol secretion. In a laboratory study, five subjects awakened, worked, and slept 10 hours later than they normally would. Subjects were exposed to very bright white light of 6000 lux, equivalent to being outdoors in the shade on a sunny day, during their night shifts. They were restricted to dim light after night shifts and slept in total darkness for 8 hours, beginning two hours after the night shift. After 9 days on the night schedule, melatonin and cortisol secretory rhythms adapted to the shifted schedule. Gene expression of *HPER1* and *HPER2* in blood also shifted to the night work schedule. This study is significant for two reasons: First, it conclusively demonstrates that proper light exposure completely shifts circadian activity to a night shift schedule, and second, it shows that the underlying biology of these shifts can be measured with traditional hormonal sampling or with genes extracted from peripheral blood.[45]

Measuring Circadian Rhythms with Salivary Melatonin

A commonly used technique in human circadian research involves measuring melatonin in blood or saliva. Measuring melatonin levels in blood was first accomplished in 1978,[46] and suppression of melatonin secretion by the presence of bright light was soon demonstrated.[47] The evening rise in blood levels of melatonin, called *dim-light melatonin onset* (DLMO), and the decline in blood levels of melatonin, called *dim-light melatonin offset* (DLM offset), are markers used as the basis for standard techniques for measuring duration of the human circadian period in many experimental protocols as described earlier.

Melatonin Salivary Assays

More recently, melatonin salivary assays have replaced blood testing for measuring the timing of DLMO and DLM offset. This noninvasive procedure permits a clinician or investigator to indirectly measure circulating melatonin levels. Results plotted consecutively estimate time of melatonin onset and offset, from which a subject's circadian phase can be determined.[48] Salivary melatonin correlates directly with sleep propensity and inversely with core body temperature.[49]

Collecting Saliva

During the sampling period, subjects should avoid heavy exercise. Samples are taken under controlled light conditions to avoid light inhibition of melatonin: 250 to 300 lux for daytime and less than 50 lux for nighttime. Saliva is collected by placing a wool swab in the mouth and having the subject chew for 5 minutes. Saliva samples may be acquired once or multiple times over a 24-hour period.[50]

Salivary Melatonin and Cortisol Assay

Melatonin in saliva is measured by a direct radioimmunoassay. Various kits are manufactured in North America and Europe and are specially made for the quantitative determination of melatonin in saliva. The cross-reactivity of the assay for closely related chemical structures is less than 1%. The sensitivity of the assay is approximately 1 ng/L.[51]

Statistics in Salivary Melatonin Studies

To characterize the salivary melatonin profile for a given subject, the periodic function may be fitted to a set of data points using the baseline cosine function. The peak height (amplitude), acrophase, and duration of secretion are then calculated and may be subjected to further calculations or common statistical analysis.

Standardized procedures for measuring DLMO and DLM offset have been published, including guidelines for collecting and analyzing urinary, salivary, and plasma melatonin, to help clinicians and researchers to determine which technique of measuring melatonin is most appropriate for their particular needs. It is hoped that such standardization will facilitate comparison of data among laboratories.[51] One study showed that salivary estimation of DLMO is consistent with the time of sleep onset when samples are gathered from normal subjects and those who meet criteria for delayed sleep phase syndrome in the *International Classification of Sleep Disorders,* 2nd edition (ICSD2), occurring in each subject approximately 10.75 hours before wakeup time.[52]

Another study established a phase response curve for melatonin. This study found the greatest phase advance to be approximately 6 hours before the habitual hour of sleep onset and the greatest phase delay soon after habitual wakeup time. Also observed was a "dead zone," for melatonin to have a phase-shifting effect, lasting from shortly before sleep onset to midsleep, a period that encompasses the time when exogenous melatonin most typically is administered to patients with sleep onset difficulties.[53]

ACTIGRAPHY

Details concerning actigraphy methods appear in Chapter 171. Actigraphy is a noninvasive method for monitoring human rest-activity cycles. Actigraphs record activity using

accelerometers usually mounted in a wristwatch-like device. Actigraphy is used in circadian rhythm research and in sleep disorders centers. Clinically helpful for quantifying and diagnosing various sleep disorders, it is extremely valuable for providing objective data confirming a suspected circadian rhythm disorder. Typically, the patient wears the actigraph for 1 to 3 weeks and then returns to have the data uploaded. The plot of data for active versus quiescent periods can reveal circadian rhythm disorders, the degree of phase shift, and the consistency of the sleep-wake pattern. Specifically, the data also can be used for visual display of the normal timing and continuity of sleep in a normal adult (Figure 170-4). Advanced sleep phase disorder consistently shows an early bedtime and rising time (Figure 170-5). Delayed sleep phase disorder usually shows the patient's enforced wakeup time on weekdays and delayed wakeup and sleep onset on weekends (Figure 170-6). In delayed sleep phase disorder, the difference between the timing of sleep on work days and non–work days quantifies the magnitude of the disturbance. Free-running sleep disorder shows an incremental delay in the onset of sleep and wakeup time (Figure 170-7).

MEASURING CIRCADIAN RHYTHM PARAMETERS IN GENE EXPRESSION

A new method for measuring circadian rhythms has evolved recently. It is well known that most cells contain clock-like genes that express proteins in a manner similar to gene expression in the SCN. Studies reveal that different organs express gene products in a circadian manner, in vitro or in vivo, and each organ has its own rhythm of gene expression. It is believed that the SCN acts as the conductor or organizer of the expression of gene products throughout the body. What follows is a brief description of the methodology and some applications of the technique of studying circadian rhythms from blood, organs, or muscle fibroblasts.

The expression of clock genes in peripheral leukocytes is believed to reflect the circadian clock system. *Per1* and *Per2* demonstrate circadian oscillations in messenger RNA expression in mouse leukocytes, and these rhythms are virtually abolished in *Cry1-Cry2* double-knockout mice in which circadian rhythms are abolished.[54] In human leukocyte clock genes, only *PER1* exhibits a robust circadian rhythm, making this gene a candidate for further investigation.[54]

The suprachiasmatic nuclei are believed to synchronize biologic rhythms in peripheral tissues. In humans, blood is the most accessible tissue source. Various investigators have examined clock gene expression in human peripheral blood mononuclear cells. Clock genes in peripheral blood and in peripheral organs have been found to express circadian rhythms.[55] The liver, kidney, and spleen express the most clock genes—*Per1, Per2, Cry1, Cry2, Rev-erb-α, Clock,* and *Bmal1*—in a circadian rhythmic manner. Clock genes are examined simultaneously in peripheral tissues in rodents or humans every 4 hours for

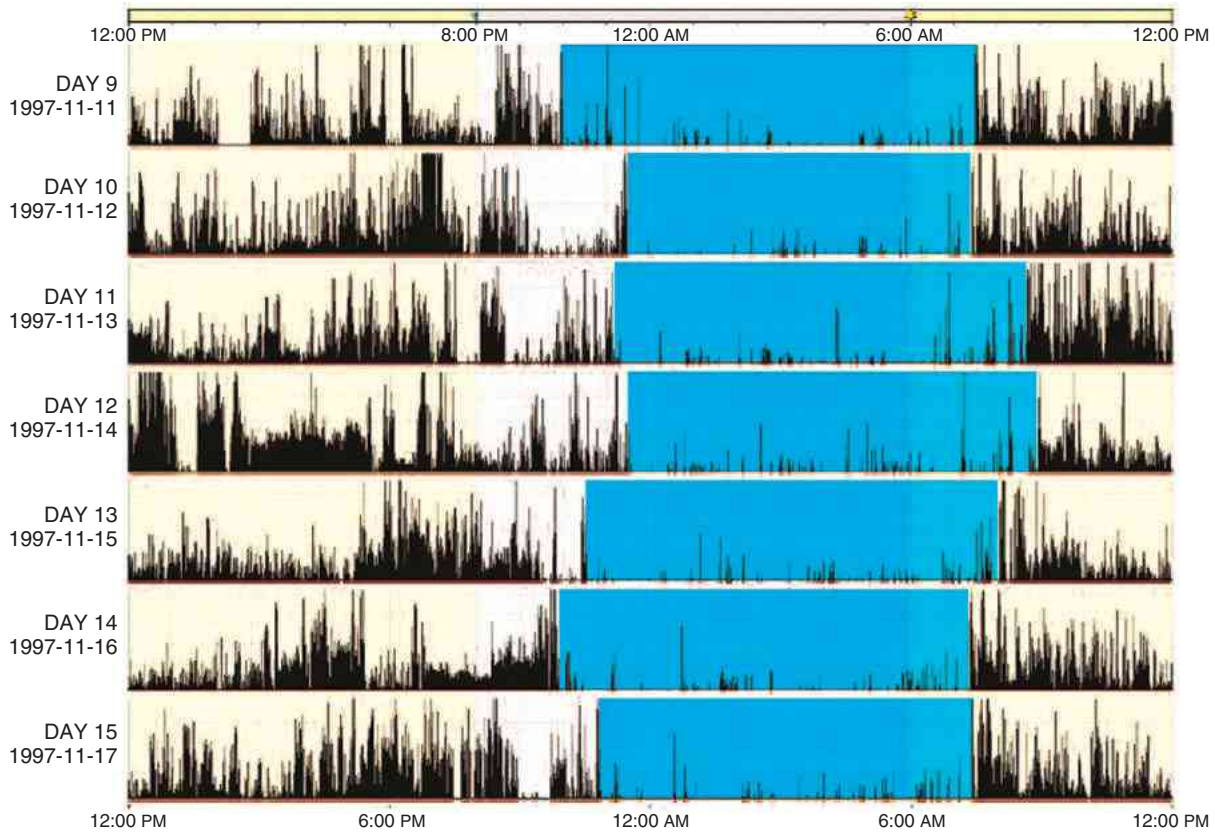


Figure 170-4 Traces from an actigraphy watch worn by a normal subject. The *blue shading* indicates the periods during which sleep most likely occurred. Data were generated by software developed by the actigraph manufacturer. Days 11 and 12 probably represent the weekend, because the subject sleeps later. (Permission granted by and copyright 2009 Koninklijke Philips Electronics NV.)

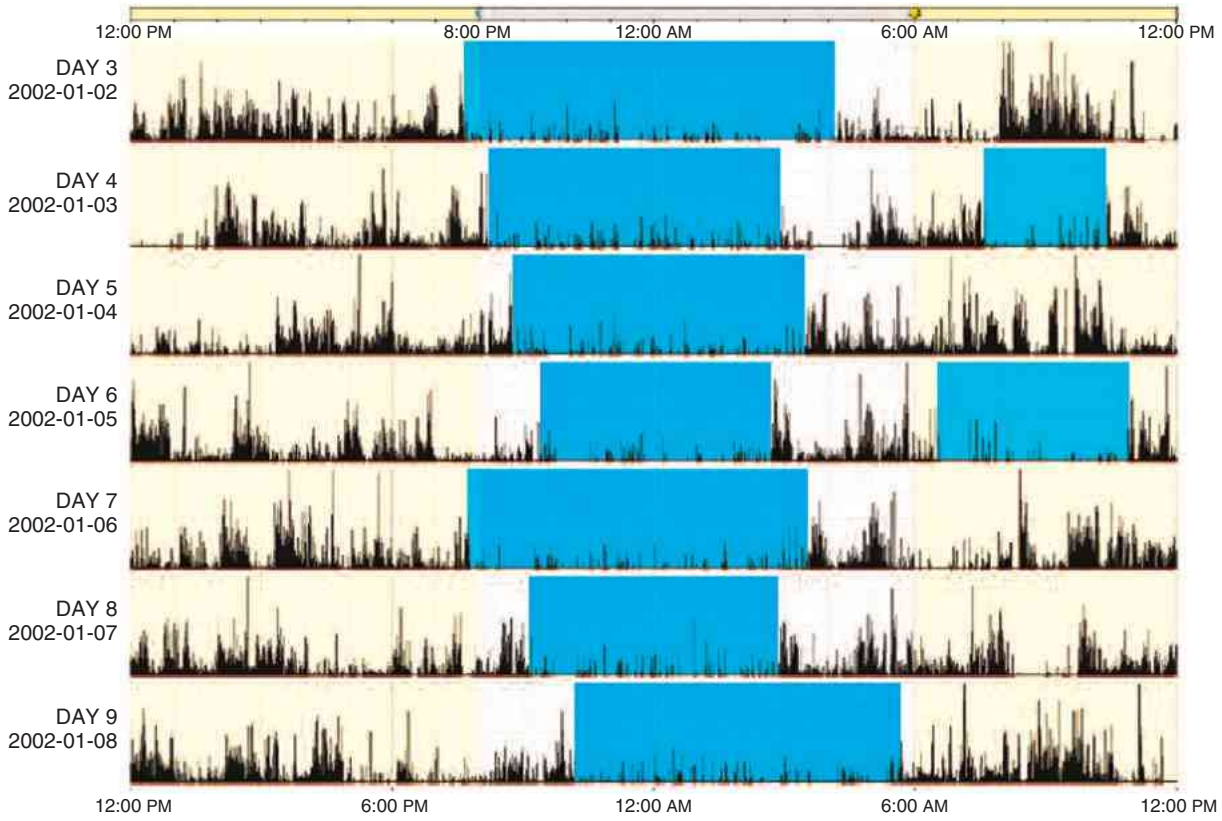


Figure 170-5 Actigraph traces showing typical features of advanced sleep phase disorder. The subject begins sleep by as early as 7:00 PM and as late as 10:00 PM. The subject awakens anywhere from 3:00 AM to 5:45 AM. The subject takes lengthy naps on two of the mornings, suggesting insufficient sleep. (Permission granted by and copyright 2009 Koninklijke Philips Electronics NV.)

24 hours in synchronized light-dark conditions using real-time polymerase chain reaction assays. One study using this methodology showed that different tissues expressed various genes to differing extents. For example, *Bmal1* and *Cry1* are more abundant in the thymus; *Per1, Cry1,* and *Cry2* are more abundant in the testis; and *Cry2* and *Per2* are expressed at a lower level in the kidney, spleen, and thymus. All genes tested are less abundant in peripheral blood. Of all peripheral tissues, the greatest circadian changes occur in the liver and kidney.[56] Of all circadian clock genes, *PER1, PER2,* and *PER3* are most rhythmically expressed in human blood samples.[56] One study claimed to be able to measure both circadian period and chronotype (owl versus lark) from genes from human blood.[57]

Perhaps the most interesting development in measuring circadian rhythms involves studying human and mouse fibroblasts. Most cells express circadian clock genes similar to those of the SCN. It is possible to engineer a lentiviral circadian reporter gene that permits characterization of circadian rhythms from forearm skin biopsy specimens, because the selected cell glows in proportion to gene expression as measured by a light meter. Several findings have emerged from these studies. First, the circadian period mean of gene expression from fibroblasts, 24.5 hours, is approximately the same as it is in behavioral (wheel running) and physiologic studies (melatonin secretion) in mice and humans.[58] However, the range of rhythms of human fibroblasts (from 23.5 hours to 26 hours) is greater than what is observed in human behavioral studies (23.9 to 24.5 hours).[1] In mice, the fibroblast period length correlates with the period length of wheel running.[58] A recent study in human volunteers examined the fibroblast period length of "larks" and "owls"—morning and evening chronotypes.[59] Morning types had significantly shorter fibroblast period lengths than evening types. Some of the morning types and evening types turned out to have normal period lengths, 24.5 hours ± 0.5 hour. In these subjects, a greater amplitude of the fibroblast gene signal was linked to more evening-type behavior. This novel finding suggests that the amplitude of circadian gene activity in addition to the cell's period length influences chronotype.[59]

As described previously, researchers can transfect cells with luciferase (glowing) to measure circadian period length of fibroblasts taken from mice and humans. Transfected cells glow in proportion to gene activity. With use of a photomultiplier tube, the gene's activity can be measured over time, revealing that it glows in a sine wave–like circadian pattern. Peak-to-peak or trough-to-trough luminance demonstrates the gene's period length. Trough-to-peak luminance demonstrates the gene's circadian amplitude. In one study of this phenomenon, the position of peak activity in the luminance cycle, early versus late, revealed the phase ("owl" versus "lark") of peak activity. Moreover, the transfected cells continued to

Figure 170-6 Actigraph traces for a subject with delayed sleep phase disorder. Sleep apparently begins at approximately 2:00 AM or later, and wakeup time is as late as noon. On days 3, 4, 5, and 8, there are periods of inactivity between noon and 2:00 PM suspicious for drowsiness. (Permission granted by and copyright 2009 Koninklijke Philips Electronics NV.)

glow in vivo for more than 20 cycles, showing little degradation. Fibroblasts were selected because they are easily obtained and their genetic period has been found to be similar to that of the mouse's SCN.[60] In both mice and humans, with different circadian behavior, fibroblast luminance characteristics correlated closely with the individual subject from which or whom the genes were taken. In mice, gene luminescence period length is similar to the period of wheel running in the same animals: If the mouse's clock genes are mutated to shorten, lengthen, or abolish the circadian period of wheel running, a similar effect is observed in the luminescence characteristics of the fibroblasts taken from these mice.[60]

Another study of dermal fibroblasts from normal humans and persons with bipolar disorder found that the amplitude of the clock gene's circadian rhythm was damped in the latter group.[61] An interesting application of the fibroblast technique was the extraction of fibroblasts from pregnant mouse dams in a forced-feeding protocol when food was available only during the dam's biologicl night. As expected, the dams developed a secondary feeding rhythm while maintaining their normal day-night circadian rhythm. The pups were sacrificed shortly before birth, and fibroblast analysis showed the pups' most active time, from a gene transcription perspective, was when the mother fed, and not during her biologic day, as measured by the mother's wheel-running activity and by her fibroblast gene activity.[62]

FUTURE DIRECTIONS

Monitoring techniques employed to elucidate various parameters of circadian rhythms continue to evolve. One novel protocol is based on the changing color of the sky observed at twilight. Previously, circadian rhythm studies have focused exclusively on brightness, or irradiance, as the light signal to entrain the circadian clock. Rodents are housed in cages exposed to light or dark but not to the normal color changes of the sky that occur in the field at twilight. As sky brightness changes at twilight, the illuminance stimulating retinal receptors decreases at dusk and increases at dawn (during sunset and sunrise). Another novel research technique monitors the changing color of the sky at twilight. The ratio of blue light to yellow light changes by a factor of 4 at twilight as blue wavelength light increases and yellow wavelength light decreases with increasing darkness at dusk.[63]

By measuring both sky irradiance and blue-yellow color ratios locally over a period of months (in Manchester, United Kingdom) at both dawn and dusk, it was found that the blue-yellow chromatic changes during twilight were more consistent on a day-to-day basis than the changes in irradiance at twilight. Irradiance varied more than the blue-yellow color ratio owing to the erratic presence of varying degrees of cloud cover.

This change in the ratio of blue to yellow light is detected by blue-yellow spectral opponent neurons in the SCN.

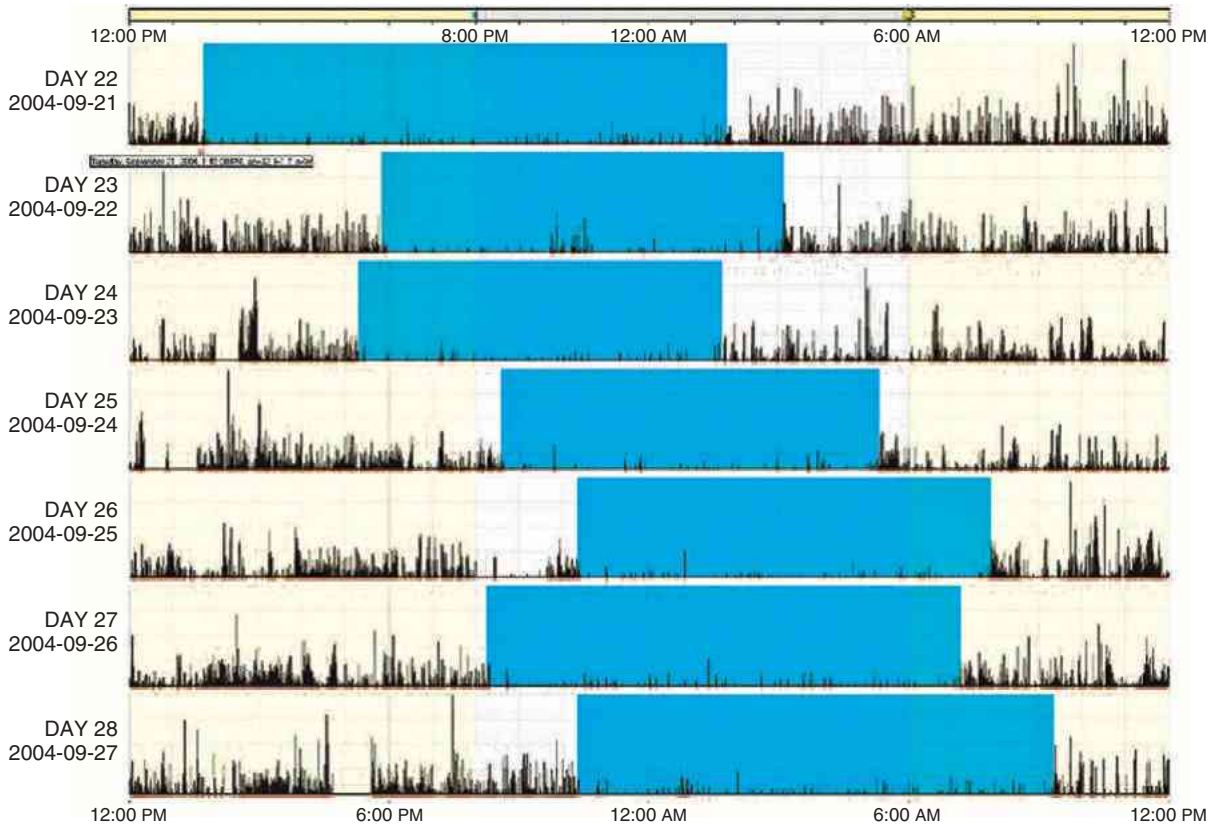


Figure 170-7 Actigraph traces showing the rest-activity pattern in a subject with free-running or nonentrained-type circadian rhythm disorder. This disorder most typically occurs in blinded persons whose biologic circadian rhythms are not entrained or masked by conventional time. The time of sleep onset does not follow an orderly progression, suggesting that social factors may contribute to the variance. (Permission granted by and copyright 2009 Koninklijke Philips Electronics NV.)

Electrophysiologic recordings of these neurons demonstrate changes in firing rates at twilight based on changes in blue-yellow light ratios independent of changes in illumination. These blue-yellow sensitive clock neurons are highly responsive to changes in spectral composition at twilight, which could be a basic mechanism for its detection and critical to detecting twilight and setting the body's clock.

Experimentally created simulated twilight transitions demonstrated that spectral changes that occur during twilight are sufficient to entrain circadian alignment to them. This novel technique for detecting twilight would be potentially of benefit to all mammalian species capable of blue-yellow light discrimination.[63]

Another novel experimental technique used to manipulate neural firing rate both in vivo and in vitro is optogenetics. With the technique, genes that express optically sensitive proteins are inserted into target cells make them respond to light. Then light is delivered by fiber optics to excite or inhibit the targeted cell. Widely used in neuroscience, this technique has recently been employed in the study of circadian rhythms.[64] Recent studies have combined optogenetically induced alterations of firing rate in the SCN with electrophysiologic recordings of SCN neurons in conjunction with observations of behavioral circadian rhythms. This is achieved with bioluminescence imaging and locomotor activity monitoring.[65]

Manipulating the firing rate optogenetically is capable of resetting the circadian rhythms in wild-type rodents, either phase advancing or phase delaying the sleep-wake cycle. This suggests that SCN firing rate is fundamental to circadian pacemaking as both an input to and an output of the molecular clockwork. The significance of the foregone study is in demonstrating that the firing rate of SCN neurons is not passively controlled by environmental illumination (dark versus light) but that the firing rate of neurons in the SCN is capable of resetting an organism's circadian rhythm. This study changed an animal's sleep-wake rhythms by artificially stimulating the neurons in the SCN by means of a laser light source and an optical fiber.

Until this study, neuroscientists had thought that the firing rate of SCN neurons was strictly an output of the biologic clock's activity. In accordance with this concept, experimentally altering the level of neuronal activity would not be expected to result in a phase shift of the circadian pacemaker. This study, however, demonstrated that stimulating (increasing) or suppressing (decreasing) the SCN's neurons in a fashion that emulates their day and night activity levels can force the clock to reset.

The project involved genetically engineering a strain of mice to manifest neurons in the SCN containing an optically sensitive protein that triggers neuronal activity when exposed to light and genetically engineered another strain of mice with

neurons expressing a similar protein that suppressed neuronal activity when exposed to light.

Current methods for assessing circadian pacemaker properties are indirect. In the future, current experimental techniques probably will be viewed as clever but rudimentary. Temperature is an indirect output of the core pacemaker, and melatonin also is indirectly related to it. Genetic advances in chronobiologic technique could allow measurement of gene transcription or translation products that are direct outputs of the circadian pacemaker. Studies using techniques such as functional magnetic resonance imaging could make it possible to directly monitor the functional activity of neuronal assemblies with purported pacemaker activity, such as the suprachiasmatic nuclei, which, after all, is the heart of the matter.

IMPLICATIONS

Methodologies in circadian rhythm research are rapidly evolving. Much has been learned about the circadian pacemaker in humans, its inputs and outputs, and factors that affect circadian rhythms. The protocols described in this chapter are continually being refined; additional controls are incorporated to eliminate unintended factors that could have masking effects.

Some of these protocols are complex and require sophisticated laboratories with extensive time-isolated floor space and equipment. Human data mostly derive from young male subjects, and little is yet known concerning sex differences. Each protocol has advantages and disadvantages.

Forced desynchrony studies can have a major psychological impact that may disrupt the emotional well-being of subjects. Subjects live continuously in a timeless environment; have no contact with the outside world; are awakened at various times throughout the night and day; are permitted to sleep only at times when natural sleep is unlikely; and have meals at times when the person would not be hungry. In countries where physical torture per se is not permitted, circumstances similar to forced desynchrony are used to interrogate prisoners of war. It is possible that the results from forced desynchrony protocols are affected by the unusual and unsustainable nature of the experimental conditions.

The long nights protocol reveals great flexibility in the human capacity to adapt to long periods of darkness, both biologically and behaviorally. It reveals biologic changes with such adaptation, such as longer cortisol suppression and decreased daytime sleepiness. These changes may prove o be advantageous in certain populations, such as children, people with narcolepsy, or patients with affective disorders. Little is yet known about this chronobiologic approach.

Chronobiology research in human subjects appears to be searching for experimental protocols that minimize differences between humans, other animals, and plants. Rodent studies have become the gold standard to which human circadian research aspires. Rodent and human studies maximize stability of period length and minimize variability. If gathering phenotypic information for genetic studies is the goal, it is understandable that maximal stability and minimal variance are sought. However, something about what makes us human is lost when sleep scientists abandon paradigms that show individual variance in preferred day length and design studies to exclude people who prefer to remain awake and asleep for extended periods.

CLINICAL PEARLS

- Circadian monitoring techniques have evolved rapidly and improved dramatically.
- Persons with suspected circadian rhythm disorders, such as phase advance, phase delay, and non–24-hour circadian rhythms, may be rapidly and accurately diagnosed using a wrist-worn actigraph or by measuring nightly onset of melatonin secretion using inexpensive saliva-gathering kits.
- The phase response curve to light and exogenous melatonin administration in advancing or delaying circadian rhythms is well established.
- In general, morning administration of bright light or melatonin advances a subject's circadian rhythm and evening administration of bright light or melatonin delays the subject's circadian rhythm.

SUMMARY

In adition to the fundamental methodologies in circadian rhythm research, among those reviewed in this chapter, exciting advances in recent years have occurred in techniques for measuring circadian rhythms. It is now possible to measure circadian rhythm period length from saliva gathered from a subject or patient, and to measure circadian rhythm period length, phase, and amplitude from genes extracted from a subject or patient's forearm. These are remarkable advances considering that the earliest studies of circadian rhythm properties required that the subject live in a cave for several weeks (and the results were flawed). Although methods can be described to measure the circadian rhythm properties of single-celled organisms, worms, flies, or mammalian organs, methodologies applicable to mammalian and, most specifically, human research probably are of greatest interest to sleep medicine professionals. As highlighted in this chapter, controversies and unresolved issues remain, and ongoing critical evaluation of the various chronobiologic protocols is indicated.

Recent studies using optogenetic techniques have demonstrated the ability to advance or delay circadian rhythms by manipulating suprachiasmatic nuclei firing rates indicating an unexpected ability of neural activity to influence circadian rhythms.

Selected Readings

Benloucif S, Burgess HJ, Klerman EB, et al. Measuring melatonin in humans. *J Clin Sleep Med* 2008;**4**(1):66–9.

Brown SA, Fleury-Olela F, Nagoshi E, et al. The period length of fibroblast circadian gene expression varies widely among human individuals. *PLoS Biol* 2005;**3**(10):e338.

Burgess HJ, Revell VL, Eastman CI. A three pulse phase response curve to three milligrams of melatonin in humans. *J Physiol* 2008;**586**(2):639–47.

Carskadon MA, Dement WC. Sleep studies on a 90-minute day. *Electroencephalogr Clin Neurophysiol* 1975;**39**:145–55.

Czeisler CA, Allan JS, Strogatz SH, et al. Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. *Science* 1986;**233**:667–71.

Czeisler CA, Duffy JF, Shanahan TL, et al. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science* 1999;**284**: 2177–81.

Fukuya H, Emoto N, Nonaka H, et al. Circadian expression of clock genes in human peripheral leukocytes. *Biochem Biophys Res Commun* 2007; **354**(4):924–8.

James FO, Cermakian N, Boivin DB. Circadian rhythms of melatonin, cortisol, and clock gene expression during simulated night shift work. *Sleep* 2007;**30**(11):1427–36.

Jones JR, Tackenberg MC, McMahon DG. Manipulating circadian clock neuron firing rate resets molecular circadian rhythms and behavior. *Nat Neurosci* 2015;**18**(3):373–5.

Liu S, Cai Y, Sothern RB, et al. Chronobiological analysis of circadian patterns in transcription of seven key clock genes in six peripheral tissues in mice. *Chronobiol Int* 2007;**24**(5):793–820.

Micic G, Lovato N, Gradisar M, et al. Nocturnal melatonin profiles in patients with delayed sleep-wake phase disorder and control sleepers. *J Biol Rhythms* 2015;**30**(5):437–48.

Walmsley L, Hanna L, Mouland J, et al. Colour as a signal for entraining the mammalian circadian clock. *PLoS Biol* 2015;**13**(4):e1002127.

A complete reference list can be found online at ExpertConsult.com.

Chapter 171

Actigraphy

Katie L. Stone; Sonia Ancoli-Israel

Chapter Highlights

- Wrist actigraphy offers many advantages over polysomnography for estimating sleep. It is more convenient and cost-effective, as well as less invasive, and allows data collection in a more natural sleep environment. It also is possible to collect data for a longer duration of time, over a period of days to weeks, for a more representative characterization of sleep patterns.
- Actigraphy is reliable and valid for detecting sleep in normal populations; however, it has limitations for detecting specific sleep disturbances.
- Actigraphy is a useful adjunct to the routine evaluation of insomnia, circadian rhythm disorders, and excessive sleepiness. It is helpful for diagnosing periodic limb movements in sleep and/or restless legs syndrome.
- Actigraphy is useful in special populations, such as children or demented elderly, who may not tolerate polysomnography.
- Actigraphy is increasingly used in clinical settings and research. Several cohort studies have used actigraphy to determine the association between sleep patterns and health outcomes. Actigraphy also is very useful for examining treatment effects in randomized trials.

INTRODUCTION

The "gold standard" modality for evaluating sleep is polysomnography (PSG) incorporating, at a minimum, electroencephalogram (EEG), electrooculogram (EOG), and submentalis electromyogram (EMG) recordings. Depending on the patient's sleep complaints, monitoring of other physiologic variables may be added (e.g., respirations, heart rate, tibialis muscle movement, oximetry). PSG, therefore, allows for the collection of detailed and comprehensive information about sleep. The EEG, EOG, and EMG records can be scored for sleep stages (non–rapid eye movement [NREM] stages N1 to N3, and rapid eye movement [REM]), total sleep time, total wake time, sleep onset latency, and percent time in REM versus NREM sleep.

This information is crucial for certain types of evaluations; however, the recording process may disturb the subject's sleep, and PSG studies are very costly both to record and to score. In addition, PSG typically provides data about sleep episodes during the time of recording, typically lasting 6 to 10 hours. Because recordings generally are made during the major sleep period, little information is available about daytime (waking) or napping behavior. In certain instances, the only information required may be on whether the person is awake or asleep, and knowledge of the specific stage of sleep and other physiologic variables is not necessary.

By contrast, actigraphy is much less expensive than PSG and provides 24-hour recordings of activity from which wake and sleep can be scored. Actigraphs (or actimeters) record limb movement. Traditionally, the actigraph is placed on the wrist, although sometimes activity from the leg is recorded. The data collected are displayed on a computer and are examined for activity-inactivity and analyzed for wake-sleep. This chapter reviews the major clinical applications for actigraphy, provides tips for successful performance of the actigraphy study, and identifies limitations with its use.

BACKGROUND

Wrist activity technology is based on the fact that during sleep, little movement occurs, whereas during wake, periodic increases in movement are seen. Activity monitors have been available for many years.[1] With the advent of microprocessors and miniaturization, most contemporary actigraphs include a movement detector (such as an accelerometer) and sufficient memory to record for long time periods. The only time the actigraph needs to be removed by the patient is during bathing or swimming, although many actigraphs now are water-resistant. Thus nearly continuous 24-hour recordings over several days or weeks are possible. Newer models are the size of a wristwatch and collect digitized data. Physical movement generally is sampled several times per second and stored in 1-minute epochs, although sampling and epoch rates can be set by the clinician or investigator. The three primary ways in which signals can be digitized are *time above threshold* (TAT), *zero crossing mode* (ZCM), and *digital integration mode* (DIM) (Figure 171-1).[2] The TAT method counts the amount of time per epoch that the motion signal is above a given threshold.

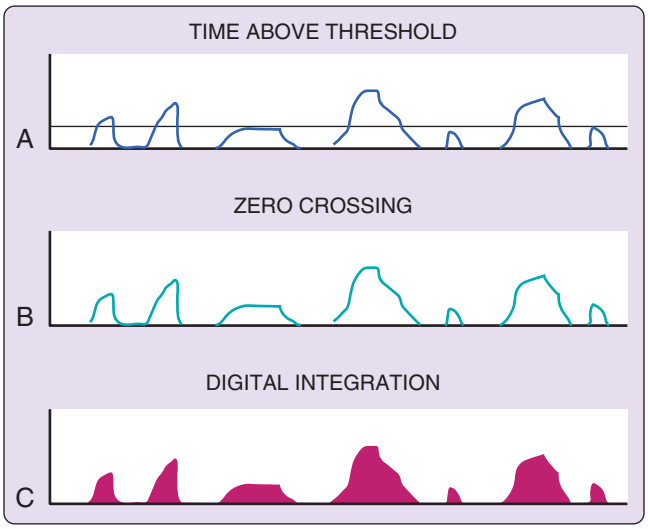


Figure 171-1 Three methods of deriving activity counts in actigraphy. The time above threshold method **(A)** derives the amount of time per epoch during which the activity is above some defined threshold (represented by the *thin horizontal line*). Zero crossing mode **(B)** counts the number of times the activity reaches zero (represented by the *solid baseline*). Digital integration mode **(C)** calculates the area under the curves represented by *dark pink shading*. (From Ancoli-Israel S, Cole R, Alessi CA, et al. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003;26:342–92.)

However, neither the amplitude of the signal nor the acceleration of the movement is reflected in this strategy. The ZCM method counts the number of times per epoch that the signal crosses zero. In this strategy, once again neither the amplitude nor the acceleration is taken into account, and high-frequency artifact could potentially be counted as movement. DIM, on the other hand, samples the accelerometry output signal at a high rate and then, for each epoch, calculates the area under the curve. This result reflects the amplitude and acceleration of the signal but not the duration or frequency of the signal. In studies comparing the three methodologies, DIM was best for identifying movement amplitude, followed by TAT and then ZCM.[3] Some actigraphs use more than one method, thereby decreasing the deficits of each method alone.

Once the data are digitized, computer algorithms automatically score wake and sleep and provide the user with summary statistics. These computer algorithms generally supply information on total sleep time, percent of time spent asleep, total wake time, percent of time spent awake, number of awakenings, time between awakenings and sleep onset latency.[2,4] The use of actigraphy in sleep disorders medicine has gained popularity. Research has examined the reliability of activity versus EEG for distinguishing wake from sleep. Results vary depending on the populations studied and on the actigraph and scoring software used. Most studies indicate actigraphy estimates of total sleep time correlate well with PSG data in normal sleep; reliability coefficients range from 0.89 to 0.98.[2,5] Correlations are somewhat lower for severely disturbed sleep. For example, in patients attending a sleep disorders clinic, reliability ranged from 0.78 to 0.88.[6,7] In infants and in children, the reliability ranges from 0.90 to 0.95.[8,9] Actigraphy has been compared to sleep and wake EEG in demented nursing home residents.[10] Correlations with total sleep time were 0.91 for averaged activity (i.e., the average activity recorded per minute) and 0.81 for maximum activity (i.e., the maximum activity recorded per minute). Given the problems with obtaining EEG recordings in this setting, actigraphy offers a relatively accurate and feasible alternative in these patients.

A multicenter study of 68 community-dwelling older women (mean age, 82 years) participating in the Study of Osteoporotic Fractures examined the reliability of estimates of total sleep time. All three modes of actigraphy data collection—ZCM, TAT, and DIM—were compared with unattended in-home EEG recording. Results showed that estimates based on DIM mode were most highly correlated with EEG activity; reliability coefficient (r) was 0.76.[11] A similar study compared actigraphy with electroencephalography as part of the Osteoporotic Fractures in Men (MrOS) sleep study (n = 889). In older men, total sleep time based on actigraphy scored in DIM mode was more highly correlated with EEG-based total sleep time as compared with ZCM and TAT modes. Even based on scoring in DIM, however, the correlations were more modest (Pearson correlation coefficient of 0.61).[12] Furthermore, actigraphic total sleep time systematically overestimated total sleep time by an average of about 13 minutes per night in this cohort of older men. Correlations were lower in certain subgroups, such as subjects taking antidepressants and those with severe sleep apnea (apnea-hypopnea index of 30 or higher). In another study involving similar methods in a sample of 181 adolescents, TAT mode performed best in terms of agreement with EEG activity, but overall agreement was modest (r = 0.41). Further analysis, however, revealed better concordance in girls than in boys (r = 0.66 versus 0.31), and among those without sleep-disordered breathing (0.55).[13]

Reliability of estimated sleep time from actigraphy varies considerably, depending on the device used, the setting, and the specific population being studied. In addition, although total sleep time generally correlates well with EEG signal, the actigraph data do not necessarily correlate well on a minute-by-minute basis, particularly if sleep is very disturbed. Most studies are in agreement that with correct use, the actigraph is reliable for certain populations. Actigraphs with different algorithms are now commercially available; accordingly, reliability has become a major question. Ideally, each actigraph needs its own reliability studies, and results from one population may not be generalizable to other populations. In general, validation studies for normal subjects show greater than 90% agreement with EEG recording.[14] To date, however, few head-to-head comparisons between different actigraphs have been performed, so no conclusion can be drawn about which collecting and scoring method agrees better with PSG results.[2] In one study, two common research models of actigraphs were compared with each other and with overnight PSG. The sample included recordings from 115 adolescent boys and girls.[15] Each device had fairly high sensitivity (for detecting sleep) but relatively poor specificity (for detecting wake) compared with PSG. The investigators also noted that reliability of the actigraphs versus PSG recording varied considerably, depending on scoring mode, specific age group studied, and sensitivity setting. Furthermore, the two devices compared poorly with each other, suggesting the need for caution in comparing results across different studies and populations in which various devices have been used.

Wrist actigraphy records continuously for long time periods, so behavior occurring both during the night and during the day can be studied. For example, collecting nighttime and daytime information for long time periods is particularly useful in evaluating the sleep of patients complaining of insomnia. It economically allows recognition of a pattern of difficulty sleeping. Among patients with complaints of insomnia, sleep can vary in quality and quantity from night to night and is easily disturbed by novel environments. The actigraph therefore has potential for examining sleep over several nights in the patient's home environment. Compared with a sleep diary, actigraphy yields data that are continuous and objective, rather than discrete and subjective.

Another advantage of collecting data over long periods is the ability to examine the circadian rhythms of the sleep or activity cycle, particularly in studies of patients with sleep-wake schedule disorders (e.g., jet lag, shift work, advanced or delayed sleep phase). Most commercial software will now perform circadian rhythms analysis, including parameters such as the mesor (mean of the rhythm), the amplitude (peak of the rhythm), and the acrophase (time of the peak of the rhythm). These analyses need 5 to 7 days of data collection to be most accurate. Several adaptations and alternative methods of circadian rhythm analyses in humans have been proposed.[16]

The wrist actigraph may be particularly valuable for studying patients who would have difficulty sleeping in a laboratory or sleeping with the wires used by traditional PSG biosensors—for example, people with insomnia, children, and demented elderly persons. With actigraphy, patients sleep in their natural environment. In demented elderly subjects, it has been shown that the traditional recording process disturbs sleep. In addition, the EEG in such patients often does not allow distinguishing wake from sleep.[17] Actigraphy avoids both of these problems and enables the recording of sleep-wake activity in an easy, unobtrusive manner.

Some actigraphs may record other parameters in addition to activity, such as ambient light, skin temperature, and sound. These additional features enhance the capability of actigraphy to allow for both clinical and investigational studies of circadian rhythms in the home environment. Most actigraphs also have optional event buttons that the subject can push when turning out the lights or for noting other events. Additionally, some devices include software to estimate both sleep-wake parameters and energy expenditure.

An actigraph's concurrent recording of light is particularly useful for determining "lights out," as well as observing the beginning of morning light when the sun begins to rise. Light measurements also are helpful in studies of advanced and delayed sleep phase because they help determine light exposure amount and duration. For example, several studies using activity monitors that also record light exposure found that normal elderly persons are exposed to only 58 minutes of bright light per day,[18] patients with Alzheimer disease living at home are exposed to 30 minutes of bright light a day,[19] and nursing home residents are exposed to only 1.7 minutes.[20] These data aid the current understanding of the changes occurring in sleep in these populations and can assist in developing treatment plans.

In recent years, numerous consumer-oriented devices and smartphones have begun to provide interested users with estimates of their sleep characteristics, such as nighttime sleep duration, awakenings, and even time spent in various sleep stages (see also Chapter 160). Many of these products are lightweight wearable devices (usually worn on the wrist or placed in a clothing pocket, but sometimes provided as a necklace or clip) that include accelerometers and sometimes include other sensors to monitor heart rate, skin temperature, and other physiologic parameters. Typically, these devices interface by way of Bluetooth or similar software to a smartphone application that the user may use to view summary reports including both daytime activity and nighttime sleep patterns. Although these devices are appealing because they generally are lightweight and inexpensive and provide user-friendly reports of sleep characteristics, some limitations to their use in research or clinical applications have become evident. In particular, device manufacturers often use proprietary algorithms for sleep-wake parameter estimation, and it often is unclear how well the estimated sleep parameters compare with PSG data, or even to similar estimates from research-grade actigraphs. Another limitation has been the lack of access to raw (i.e., epoch-by-epoch) data, which many sleep researchers use to perform more complex analyses. For some of the devices, problems also have emerged with continuous long-term collection of data, based on the need to remove the device regularly for recharging. In keeping with the widespread use and popularity of such devices, however, interest in their potential use for research is growing. A few recent studies have independently explored the reliability of consumer-level devices with research-grade accelerometers/actigraphs for sleep estimation. In one study of healthy young adults, estimates for total sleep time were compared among seven consumer-level monitors and two research-grade devices. Correlations generally were above 0.80.[21] In another study, sleep estimates were compared for 23 subjects with simultaneous collection using a consumer-level wearable monitor and a multisensor device including a piezoelectric sensor that is placed under the bed sheets, as well as sensors for temperature and light, and a microphone.[22] Overall agreement was poor, with R^2 values of 0.25 for total sleep time, 0.11 for sleep efficiency, 0.30 for sleep onset latency, and 0.09 for number of awakenings. However, agreement was found to be excellent for some subjects and poor for others. Technology is rapidly evolving in this area, and many new developments are likely over the coming years.

APPLICATIONS OF WRIST ACTIGRAPHY

Because actigraphy is unobtrusive and can record multiple days and nights, it is a useful tool for evaluating insomnia, particularly because insomnia can vary from night to night. In addition, actigraphy eliminates laboratory effects as patients can be recorded sleeping in their own beds in their own homes. However, the convenience of actigraphy must be weighed against its reliability and validity versus polysomnography.[2]

The American Academy of Sleep Medicine (AASM) published recommendations on the use of actigraphy in clinical assessment of sleep disorders.[23] The AASM investigators concluded that actigraphy is reliable and valid for detecting sleep in normal healthy adult populations and in patients suspected of certain sleep disorders. These sleep disorders include advanced sleep phase syndrome, delayed sleep phase syndrome, and shift work disorder. Additionally, when polysomnography is not available, actigraphy may be

useful to estimate total sleep time in patients with obstructive sleep apnea. In patients with insomnia, the use of actigraphy is recommended for characterizing circadian rhythms and sleep-wake patterns. The practice recommendations also specify that actigraphy is useful in special populations such as children or elderly populations, who may have reduced tolerance to PSG.

Insomnia

Several studies have used actigraphy to evaluate sleep in the patient with insomnia, yet few of the studies validated the actigraphy findings against those with PSG. In a large sample of subjects with insomnia, a high concordance was demonstrated for actigraphic estimates of total sleep time and measures of sleep fragmentation with corresponding data derived from PSG.[24] However, estimates of sleep onset latency were much less reliable. The study report noted the lack of standardization of actigraphy devices and scoring methods across studies. Consequently, the findings of this particular study are not sufficient to support a blanket recommendation regarding the use of actigraphy to assess sleep among persons with insomnia. Another study reported that actigraphy tended to underestimate total sleep time, sleep efficiency, and sleep onset latency and to overestimate sleep onset latency. However, actigraphy was more accurate than the sleep diary when both were compared against PSG.[25]

Although actigraphy cannot determine the etiology of insomnia, it can help evaluate the severity of the condition. In patients with insomnia, the most difficult distinctions for the actigraph to recognize are the transitions between sleep and wake and the ultra short sleep-wake cycles. In patients with insomnia and coexisting circadian rhythm disturbances, actigraphy can detect sleep phase alterations.

Sleep Apnea

Actigraphy cannot determine the presence or absence of breathing abnormalities. However, patients with sleep apnea have more disrupted sleep and thus more body movements. Several studies tested wrist activity for recognizing patients with sleep apnea. One study showed that compared with control subjects, persons with sleep apnea had significantly higher movement and fragmentation indices.[26] Once sleep apnea therapy began, activity recordings indicated decreased indices, indicating successful treatment. Activity measures of sleep also have been shown to successfully differentiate patients with sleep apnea from those with insomnia and from control subjects.[8]

In general, the conclusion from studies is that actigraphy may be capable of distinguishing normal patients from those with moderate or severe sleep apnea, particularly because actigraphy is more sensitive than sleep logs in identifying brief awakenings. It is unlikely, however, that a distinction could be made between the different types of sleep disorders that cause brief arousals.[2] The real value of actigraphy in sleep apnea is in combination with cardiopulmonary recorders or other apnea detection equipment that do not measures of wake or sleep. The addition of an actigraph to such a recording would allow determination of whether all respiratory events actually occurred during sleep. In a recent update, the Standards of Practice Committee of the American Academy of Sleep Medicine concluded that actigraphy is indicated for the assessment of total sleep time among patients with sleep apnea when PSG is not available.[23]

Periodic Limb Movements in Sleep

Evaluation for periodic limb movements in sleep (PLMS) generally is accomplished by measuring the EMG signal at the tibialis muscle. Because the leg movement is the primary characteristic of this disorder, the ability of actigraphy to measure movement is thought to be an advantage. One study found high reliability between tibialis EMG and actigraphy for the number of leg movements per hour of sleep.[27] Another study that compared alternative placements of the actigraph for assessment of PLMS in comparison with PSG found that locating the device at the base of the big toe provided good validity.[28] The intensity and severity of PLMS can greatly vary from night to night, so actigraphy may offer an advantages in that multiple nights can easily be recorded. Actigraphy also has been useful in the assessment of treatment efficacy.[29] More research is needed to determine optimal procedures for detecting PLMS using actigraphy.

Treatment Effects

Actigraphy is particularly appropriate for the study of treatment effects because it can identify changes over time. Because actigraphy is noninvasive and less expensive than PSG, it holds promise for assessing treatment effects. Furthermore, single-session measures of sleep (which frequently is all that is feasible with PSG) may not accurately reflect habitual behavior.[30] Actigraphy also has been used to monitor drug and behavioral treatment effects, suggesting that it also can be used for screening some patients before ordering the more expensive PSG. Similarly, it can be used for follow-up assessments once treatment has begun or to evaluate changes in sleep over the course of the treatment period.

Circadian Rhythms

Actigraphy allows the study of sleep-wake patterns occurring over many days; it is therefore well suited to the study of circadian rhythms. Activity is a valid marker of entrained PSG sleep phase and correlates strongly with entrained endogenous circadian phase.[2] It also is useful in identifying sleep altered by circadian rhythm changes.

Actigraphy has been used to study circadian rhythms in patients with cancer [31] and demented elderly persons.[32] Studies also have examined sleep schedules of adolescents,[33] shift workers,[34] in-flight crews,[35] and travelers with jet lag.[36]

More recently, actigraphy recordings collected in two large cohort studies of older adults—the Study of Osteoporotic Fractures and the MrOS sleep study—were used to study the relationship of circadian rest-activity rhythms and aging-related outcomes such as mild cognitive impairment and dementia,[37] depression,[38] cardiovascular disease,[39] and death.[40,41]

A variety of methodologies are available for analyzing circadian parameters of activity.[42-44] However, no studies have compared one methodology with another across different populations. Finally, no one standard methodology has been accepted.

Special Populations

Pediatric Patients

Using actigraphy in the pediatric population is becoming increasingly popular, particularly for evaluation of children

with behavioral, psychiatric, or neurologic problems. Actigraphy has been used successfully to characterize sleep of infants,[45] developmental differences,[46] and sleep in children with autism,[47] and to demonstrate sleep differences between children with depression and those who were abused and normal children.[48] One study used wrist activity to study treatment effects on the sleep patterns of 50 infants whose parents complained of sleep disturbances in their children.[49] Objective activity recordings were compared with parental reports. The activity records showed that percent sleep increased and the number of awakenings during the night decreased with behavioral treatments. By examining data from each successive night, it was determined that most changes occurred during the first night of intervention. In addition, parental subjective reports significantly differed from objective actigraphy on quality of sleep, with parents reporting fewer awakenings during the night. The presence of objective measures allows for the evaluation of activity during sleep that would otherwise be missed by observation alone.

Actigraphy was used to assess sleep and wake in children with sleep-disordered breathing. One study compared actigraphic estimates of sleep and wake with PSG data on an epoch-by-epoch basis, using various activity thresholds. Overall, very high predictive values and sensitivity were found for detecting sleep (all parameters were higher than 90%), but predictive values and sensitivity for detecting wake were much lower.[50] In a study of preschoolers with and without autism, poor agreement in detection of nocturnal awakenings was found between actigraphy and video observations.[51] Although actigraphy offers many advantages for quantifying sleep and wake in children, results may be less reliable for detecting awakenings. Validation studies are needed for the specific devices and for use in the specific population of children.

Elderly Patients

At the other end of the age spectrum, actigraphy has been used to study sleep-wake patterns in the elderly population. Older adults are particularly susceptible to sleep complaints secondary to circadian rhythm changes, sleep-disordered breathing, PLMS, medical illness, and medication use.[52] Although evaluation of and for some of these complaints may be accomplished in the laboratory, elderly persons often are more set in their ways, need to stay home to take care of a spouse, or just find it more comfortable to sleep in their own beds. In many large studies in older adults, actigraphy represents the only viable option for obtaining objective measures of sleep. Although many studies tend to rely on self-report of sleep duration, this may be quite inaccurate in older populations, particularly among elderly persons with poor cognition and functional disabilities.[53]

Using actigraphy, a number of studies show that sleep in nursing home residents is extremely fragmented, with most patients never sleeping for a full hour and never awake for a full hour throughout the 24-hour day.[54] In these settings, actigraphy was then used to show that a trial of light treatment consolidated sleep but did not lessen degree or prevalence of agitation in this population.[43,55]

Measures of sleep using actigraphy have been incorporated into several large epidemiologic studies of older adults. For such application, standardization of techniques across multiple clinic sites is essential. In one such study of older women participating in the multicenter Study of Osteoporotic Fractures, excellent agreement was reported between results of an expert scorer and of a second scorer when both followed standardized scoring procedures.[56] Further details are provided in the subsequent section Epidemiologic Studies of Sleep.

Other Medical Conditions and Populations

Actigraphy is increasingly used in clinical research in populations in which it would be difficult to conduct PSG studies. Actigraphy has been used to study sleep and circadian rhythms in menopausal women,[57] in older patients with schizophrenia,[58] and in patients with cirrhosis,[59] with coronary artery disease,[60] or with cancer.[61-63]

EPIDEMIOLOGIC STUDIES OF SLEEP

Actigraphy is useful for studying sleep in large populations, particularly when recording in the laboratory is potentially cost-prohibitive, leading to attrition of enrolled subjects.

Ancoli-Israel and Kripke, along with their colleagues,[64,65] used actigraphy to determine wake and sleep patterns in large samples of elderly (n = 426) and middle-aged adults (n = 355). Measurements were used to determine prevalence of sleep apnea and PLMS (that were recorded with additional sensors). Many of these volunteers would have been less willing to participate had they been required to sleep in the laboratory.

Over the past decade, several other large observational studies have incorporated measures of sleep using actigraphy. In the multicenter Study of Osteoporotic Fractures, actigraphy was performed in nearly 3000 older women, whereas in another multicenter study, the Outcomes of Sleep Disorders in Older Men (MrOS sleep study), actigraphy data were collected in more than 3000 subjects. Using actigraphy in these populations helped identify the relationship between poor sleep and increased risk of falls,[66] poor cognitive function,[67] and poor physical performance and functional limitations.[68,69] The Rotterdam Study, which collected actigraphy recordings in nearly 1000 older men and women, has been used to demonstrate a relationship between actigraphic sleep duration and fragmentation and risk of obesity[70] and serum cholesterol levels.[71] Actigraphy also has been used to obtain objective measures of sleep characteristics in early middle-aged (38 to 50 years of age) men and women in the Coronary Artery Risk Development in Young Adults study.[30]

TRICKS OF THE TRADE

Actigraphs traditionally are placed on the wrist of the nondominant hand. However, two groups of investigators have shown that either wrist can be used. Both groups found that although activity levels were different for the two hands, rates of agreement with PSG were essentially equivalent for data collected from both hands.[9] In studies in infants, actigraphs also have been placed on the baby's legs.[72]

Most actigraphs come with bands similar to plastic watch bands. For persons who might be sensitive to such bands, or who find it uncomfortable to wear them for long time periods, bands made of terry cloth and self-stick fabric (Velcro) can

be customized. To discourage patients from removing the bands, the two Velcro straps were reversed, with one opening from right to left and the other from left to right. Some investigators use locked hospital bands to keep actigraphs in place in "difficult" populations. These techniques have worked to keep even the most diligent patient from removing the devices; they also are effective in young children.

Patients wearing actigraphs should be asked to keep a sleep log or diary. They should note information about daily time to bed, time out of bed, and any unusual activity or times when the device is removed (such as for showers or swimming). This information is extremely helpful for editing and analyzing data.

When data are collected in 1-minute epochs and accumulated over a period of more than a few weeks, it is prudent to download data every week to minimize data loss. In those units with batteries that need to be replaced, battery levels should be checked on initializing the device and again during downloading. Batteries with levels below 90% of the original battery voltage should be discarded, because they are likely to fail. The battery life is approximately 30 days. In our laboratory, a battery log is kept that records the battery number, date of initialization of the activity monitor, date of the data download, total number of days of battery use, and starting and ending battery levels.

When actigraphy is used in large samples, with multiple examiners administering the instructions and downloading the data, it is strongly recommended that standardized protocols be developed, and that centralized training and certification be incorporated into the overall data quality assurance plan.

With devices that also record light exposure, it is extremely important that the light sensor not be covered by the person's sleeve. The sleeve can be tucked under the actigraph or pinned up to ensure that it does not occlude the light sensor. Another important consideration is that because the angle of the wrist differs from the angle of the eye, the lux reading from a light sensor may differ from that for ambient illumination. Some devices have external light sensors, in addition to the internal sensor, that can be clipped to a collar and conceivably may give more exact readings. All activity and light monitors should be checked on a regular basis to determine if calibration is needed.

In patients with sensitive skin, the bands can be removed for a few minutes each day to avoid pressure sores. The time the device is removed, the time it is replaced, and whether or not the person was awake for the few minutes it was removed all should be noted on the log. This information is needed during the data editing process.

EDITING ACTIGRAPHY DATA

Different software packages are available for scoring the rest-activity data and inferring sleep-wake status. Data are edited on a computer screen with the use of the daily sleep log. Time intervals during which the device is removed should be manually changed to wake status only if the investigator is sure that the person was awake (e.g., in the shower); otherwise, these data points should be marked as missing data. Lack of movement scored as sleep, such as when the device is removed for especially vigorous activity, also can be manually changed to the wake category. For example, in a study involving adolescents, some subjects removed the actigraph during football or volleyball practice. No activity was recorded during that time period; however, it was clear that the subjects were awake.[73] If no information is provided in the log about the activity during the time the device was removed, that time period should be scored as missing data.

Figure 171-2 shows examples of two actigraph outputs.

LIMITATIONS

Actigraph data scorers need to be aware of the limitations of these devices.[2] When compared with PSG, actigraphy is fairly valid and reliable in healthy normal subjects. It is best at estimating total sleep time. As sleep becomes more disturbed, the actigraphy recording becomes less accurate. In general, actigraphy may overestimate sleep and underestimate wake, particularly during the day.

CLINICAL PEARLS

- Actigraphy has good reliability and validity for the detection of sleep in normal populations but is less reliable for identifying disturbed sleep.
- Actigraphy is a useful adjunct to the routine evaluation for insomnia, circadian rhythm disorders, and excessive sleepiness and may help in the diagnosis of periodic limb movements in sleep and/or restless legs syndrome.
- Actigraphy is useful for sleep study in special populations such as children or demented elderly persons.
- Actigraphy is becoming a valuable tool for investigating sleep and sleep disturbances in large-scale epidemiologic studies.

SUMMARY

Wrist activity has several important advantages over PSG: It allows the recording of sleep in natural environments and of behavior that occurs both during the night and during the day, as well as for long time periods, and is cost-efficient. Although not a replacement for EEG or PSG recordings, in certain instances actigraphy provides clear advantages for data collection.

Actigraphy is particularly useful for evaluating patients who cannot tolerate sleeping in the laboratory—for example, patients with complaints of insomnia, small children, or elderly persons. It may provide a more accurate estimate of typical sleep durations by permitting patients to adhere more closely to their scheduled sleep and wake times. Actigraphy also is becoming an important tool in follow-up studies and for examining treatment efficacy in clinical outcome.

Actigraphy has some value in the assessment of sleep disorders, although it may not necessarily distinguish among different sleep disorders. The newer scoring algorithms have great accuracy in determining those variables that are most important in insomnia—specifically, improved ability to detect wake versus sleep, sleep latency, awakenings during the night, and total sleep time. Actigraphy is particularly superior to sleep logs for detecting brief arousals during the night. It also can be used for the evaluation and clinical diagnosis of circadian rhythm disorders. The ability to detect movements holds promise in the identification of sleep disorders characterized

Figure 171-2 A, Four days of unscored output from the Octagonal SleepWatch actigraph (Ambulatory Monitoring, Inc., Ardsley, New York). **B,** Four days of unscored output from the Actiwatch (Mini Mitter, a Respironics Company, Murrysville, Pennsylvania).

by frequent movements, such as PLMS, sleep apnea, or REM sleep behavior disorders. Actigraphy provides a particularly useful tool in situations requiring long-term monitoring.

Selected Readings

Ancoli-Israel S, Cole R, Alessi CA, et al. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003;**26**(3):342–92.

Bhat S, Ferraris A, Gupta D, et al. Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. *J Clin Sleep Med* 2015;**11**(7):709–15.

Blackwell T, Ancoli-Israel S, Redline S, Stone KL. Factors that may influence the classification of sleep-wake by wrist actigraphy: the MrOS Sleep Study. *J Clin Sleep Med* 2011;**7**(4):357–67.

Blackwell T, Redline SS, Ancoli-Israel S, et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep* 2008;**31**(2):283–91.

Cole RJ, Kripke DF, Gruen W, et al. Automatic sleep/wake identification from wrist activity. *Sleep* 1992;**15**(5):461–9.

Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act* 2015;**12**(1):42.

Gehrman PR, Marler M, Martin JL, et al. The timing of activity rhythms in patients with dementia is related to survival. *J Gerontol A Biol Sci Med Sci* 2004;**59**(10):1050–5.

Johnson NL, Kirchner HL, Rosen CL, et al. Sleep estimation using wrist actigraphy in adolescents with and without sleep disordered breathing: a comparison of three data modes. *Sleep* 2007;**30**(7):899–905.

Landry GJ, Best JR, Liu-Ambrose T. Measuring sleep quality in older adults: a comparison using subjective and objective methods. *Front Aging Neurosci* 2015;**7**:166.

Sadeh A, Acebo C. The role of actigraphy in sleep medicine. *Sleep Med Rev* 2002;**6**(2):113–24.

Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications. *J Ambul Monit* 1989;**2**(3):209–16.

Sadeh A, Hauri PJ, Kripke DF, Lavie P. The role of actigraphy in the evaluation of sleep disorders. *Sleep* 1995;**18**(4):288–302.

Stone KL, Ancoli-Israel S, Blackwell T, et al. Poor sleep is associated with increased risk of falls in older women. *Arch Intern Med* 2008;**168**(16):1768–75.

Tranah GJ, Blackwell T, Stone KL, et al. Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. *Ann Neurol* 2011;**70**(5):722–32.

A complete reference list can be found online at ExpertConsult.com.